

Review

Face Image Analysis Using Machine Learning: A Survey on Recent Trends and Applications

Muhammad Hameed Siddiqi ¹, Khalil Khan ^{2,3,*} , Rehan Ullah Khan ⁴  and Amjad Alsirhani ¹ 

¹ College of Computer and Information Sciences, Jouf University, Sakaka 72388, Saudi Arabia; mhsiddiqi@ju.edu.sa (M.H.S.); amjadalsirhani@ju.edu.sa (A.A.)

² Department of Information Technology and Computer Science, Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur 22620, Pakistan

³ Faculty of Computer Sciences and Information Technology, Superior University, Lahore 54660, Pakistan

⁴ Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia; re.khan@qu.edu.sa

* Correspondence: khalil.khan@fecid.paf-iast.edu.pk

Abstract: Human face image analysis using machine learning is an important element in computer vision. The human face image conveys information such as age, gender, identity, emotion, race, and attractiveness to both human and computer systems. Over the last ten years, face analysis methods using machine learning have received immense attention due to their diverse applications in various tasks. Although several methods have been reported in the last ten years, face image analysis still represents a complicated challenge, particularly for images obtained from ‘in the wild’ conditions. This survey paper presents a comprehensive review focusing on methods in both controlled and uncontrolled conditions. Our work illustrates both merits and demerits of each method previously proposed, starting from seminal works on face image analysis and ending with the latest ideas exploiting deep learning frameworks. We show a comparison of the performance of the previous methods on standard datasets and also present some promising future directions on the topic.

Keywords: face analysis; computer vision; deep learning



Citation: Siddiqi, M.H.; Khan, K.; Khan, R.U.; Alsirhani, A. Face Image Analysis Machine Learning: A Survey on Recent Resend Trends and Applications. *Electronics* **2022**, *11*, 1210. <https://doi.org/10.3390/electronics11081210>

Academic Editors: Jungong Han and Guiguang Ding

Received: 13 March 2022

Accepted: 6 April 2022

Published: 11 April 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human face image analysis is a fundamental and challenging problem in computer vision (CV). Face analysis plays a main role in different real-world applications, such as animation, surveillance, identity verification, forensic, medical diagnosis, human–computer interaction, and so on [1–3]. Some research work addresses various face analysis tasks (face, age, gender, race, head pose, etc.) as individual research problems [4–11]. CV researchers also report that all face analysis problems have a relationship with each other and can assist each other if addressed as individual tasks. Despite tremendous research work on the topic, face analysis is still an arduous task due to various factors such as changes in visual angle, facial appearance, facial expressions, background, and so on. In particular, it has more complications when tackled in uncontrolled and wild conditions. (Note that we are not considering facial recognition in our current study, since it is a much-explored area, and many good research and survey papers have already been reported [12–19]). Face image analysis has several applications; some of these applications are listed below.

- **Surveillance:** Face analysis and tracking is widely used for surveillance purposes by CV researchers. A surveillance system presented in [20] uses the attention of people in a particular scene as an indication of exciting events. Similarly, the work proposed in [21] captures the visual attention of people by using various fixed surveillance cameras.
- **Targeted advertisement:** One very interesting application of face analysis is targeted advertisement. Several interesting works are proposed by CV researchers regarding

targeted advertisement using face analysis. For instance, Smith et al. [22] present a work that tracks the focus of attention of people. The proposed system also counts the number of subjects looking at outdoor advertisements. This work also has some implications in human behavior analysis and cognitive science. Some recent work on targeted advertisement using CV and machine learning (ML) can be explored in [23–25].

- **Social Behaviour Analysis:** The human face is tracked and used in intelligent rooms to monitor and observe the participants' activities. The visual focus of participants' attention is particularly judged through head tracking [26–30]. This system follows the direction of speaking of individuals and also provides information about the gestures in a meeting. The semantic cues obtained are later transcribed with conversations and intentions of all participants, which further provides some searchable indexes to be used in the future. Some excellent works which use human face tracking in workplaces and other meetings can be explored in the references [31–35].
- **Driving Safety:** Face analysis also plays a key role in ensuring driver safety while driving. Some researchers designed a driver-monitoring system by installing a camera within the car and then tracking eyebrow and eyelid movements from fatigue [36–38]. It can also provide alert signals to drivers, for instance, if there is some danger of accident due to pedestrians [39]. Blind spots are detected in the method proposed in [40] while driving. This helps the driver to change their vehicle direction. Another method proposed in [41] combines head localization and head pose estimation (HPE) information to estimate pedestrian path. This helps drivers to make some vital decisions while driving.
- **Estimation of face, expression, gender, age, and race:** In the context of CV, human face analysis acquires high-level knowledge about a facial image. Face images convey several pieces of information, such as who the subject is, the gender and race of the person in the photograph, whether he/she is sad or happy, and what angle they are looking at. In all these tasks, a facial analysis infers knowledge from a face image. Some human face analysis problems are multidisciplinary, as well as intrinsically related to human science. However, very few research works combine various face image analysis tasks in a single unified framework. Our current research paper reports papers that combine at least two face image analysis tasks in one model. Estimating age, race, gender, expression, and head pose are the most important problems in face analysis, with further applications in forensics, the entertainment industry, cosmetology, security controls, and so on [42–45].

Despite immense research efforts, face analysis is still facing challenges due to several factors such as complex and changing facial expressions, occlusions, noise, illumination changes, and so on. To improve the recognition accuracy of a face analysis system, it is necessary to correlate some face analysis tasks with each other. For instance, it is most likely that males can have beards or mustaches, but females or children certainly cannot. Previous research work also reveals that most face analysis tasks are linked with each other; for example, research reported in the 19th century by Wollaston [46] suggests that face tracking is strongly and intrinsically linked with gaze estimation. The effect is also demonstrated by the authors with a picture, as can be seen in Figure 1. In Figure 1, two views of one person are taken at two angles. Although both the eyes and the head are the same, the gaze direction perceived by the human brain is dictated by two totally different head directions. The intrinsic relationship between various face image analysis tasks is later confirmed in [47]. The work in [47] claims that gaze estimation is a combination of HPE and eyes location estimation. The later research claims that head position provides a very strong indication of the eyes' gaze orientation, particularly when the eyes are not visible, for instance, in case of occlusion or when the resolution of the image is poor.



Figure 1. FIA and intrinsic relationship of face analysis tasks [46].

The intrinsic relation between various face parts is also confirmed in the latest research work reported in [48–52]. Research reported in [49–52] suggests that a mutual relationship between different face parts can be exploited to address several tasks in a single framework. The research work proposed in [53] segments a face image into six different face parts, which are later used for HPE in [49,50]. Similarly, gender recognition is combined with other tasks using the same strategy in [52]. A single framework is proposed in later research [51], combining gender recognition, race classification, HPE, and facial expression recognition. In short, CV researchers and psychology literature strongly suggest that face analysis tasks can be addressed in a better way if combined in a single unified framework. This paper aims to present the reader with a single review article where different methods are reported combining several face analysis tasks into one model.

2. Contributions and Paper Organization

To the best of our knowledge, the proposed paper is the first attempt that combines various face analysis tasks in a single comprehensive review article. Survey papers for each of these tasks (race, age, gender, expression, and HPE) are individually available [54–56]; however, there is an intense need for a single paper combining these demographic components. We specifically focus on the works that have been published in the last 10 years. Moreover, recently, a shift has been noticed in the state of the art (SOA) from conventional ML to deep learning (DL)-based methods, which justifies the need for an up-to-date review article. The main contributions of the proposed paper are summarized as follows:

- We present a detailed survey on different human face image analysis tasks. This provides researchers with a new and recent up-to-date overview of the SOA technology. We also introduce a taxonomy of all the existing methods for these tasks. We added a detailed discussion on the used technology's characteristics. Additionally, we explicitly elaborate on the open problems faced by the existing face analysis technology.
- We provide a list of all the publicly available datasets for each of the five face analysis tasks. We also describe the way these datasets are collected and their characteristics. We provide information such as the number of images contained, subjects involved, sex information, environmental diversity, and other details.
- We conducted a detailed study of all the methods for each of the five tasks, both in indoor and outdoor conditions. We summarize the results obtained using each method and present a detailed discussion. We present a concrete summary of the results for each task, provide a critical review, and point to some future directions. We also dedicate attention to various sensors used for data collection and discuss how ground truth data is collected and labeled.

The rest of this work is arranged as follows: in Section 3, we discuss already existing datasets (DBs) used to evaluate face analysis. Recent algorithms within face analysis are reviewed in detail in Section 4. How ground truth data is created for these face analysis tasks is discussed in Section 5. We present a comparison of various methods with their

reported results in Section 6. We also discuss the open challenges in face analysis with some recommendations in Section 6. Finally, we present the conclusion of the paper in Section 7.

3. Datasets

The performance of a face analysis system is measured with a DB available for experiments. A detailed list of the available DBs for all the tasks is presented in this section. A comprehensive list of the DBs is presented in Table 1. We include only those DBs that address at least two face analysis tasks. Following the progress of each face analysis task, the number of DBs has evolved in the last couple of years. Specifically, complexity in images and background are now considered.

Table 1. Publicly available DBs for gender, age, head pose, expression, and race classification.

| DB | No. of Subjects | No. of Images | Gender | Age | Head Pose | Expressions | Race |
|-------------------|-----------------|---------------|--------|-----|-----------|-------------|------|
| UNISA-Public [57] | 58 | 406 | ✓ | | ✓ | ✓ | |
| Adience [58] | 2284 | 25,580 | ✓ | ✓ | | | |
| IDMB [59] | 558 | 500 K | ✓ | ✓ | | | |
| VGGFACE [60] | – | 6000 | ✓ | | ✓ | | |
| VGGFace2 [61] | 9131 | 3.31 M | ✓ | ✓ | ✓ | | ✓ |
| SASE [62] | 50 | 30,000 | ✓ | | ✓ | ✓ | |
| CCNU [63] | 58 | 4350 | | | ✓ | ✓ | |
| Dali3DHP [64] | 33 | 60,000 | | | ✓ | ✓ | |
| McGill [65] | 60 | 60 | | | ✓ | ✓ | |
| AFLW [66] | 21,997 | 25,993 | ✓ | | ✓ | | |
| Multi-Pie [67] | 337 | 75,000 | ✓ | ✓ | | ✓ | ✓ |
| FacePix [61] | 30 | 5430 | ✓ | | ✓ | | |
| FEI [68] | 200 | – | ✓ | | ✓ | ✓ | |
| FERET [69] | 1199 | 14,126 | ✓ | | | ✓ | |
| CAS-PEAL [70] | 1040 | 100 K | ✓ | | ✓ | | ✓ |
| LFW [71] | 5749 | 13,233 | ✓ | | ✓ | | |
| CelebA [72] | 10,000 | 200,000 | ✓ | ✓ | | | |
| BJUT-3D [73] | 500 | – | ✓ | ✓ | | | |

- UNISA-Public [57]: The UNISA-Public is collected in real-time and unconstrained conditions. The DB consists of 406 face images taken from 58 individuals. The available DB does not provide cropped faces, but a face localization algorithm is needed for the images to be used for gender classification. The data is collected in a building with a camera fixed at the entrance point; therefore, lighting conditions do not change significantly. Different poses and facial expressions are taken, as the participants were not told about the data collection process beforehand. Some motion blur conditions can be seen due to the sudden movement of the individuals.
- Adience [58]: This is the latest DB released for face analysis tasks, including age and gender classification tasks. It is also a challenging DB collected in an outdoor environment. All the images are collected through a smartphone device. The data set is much more challenging, as different pose variations are included along with changing lighting conditions. The DB is large, as the number of face images is around 25,580. The total number of candidates who participated in data collection is 2284. The exact age of each participant is not given, but instead, each subject is assigned to a

- specific group. This clearly shows that the DB can be used for age classification but not estimation. The DB can be freely downloaded from the Open University of Israel.
- **IMDB [59]:** The IMDB contains images of celebrities collected from a website named IMDB. The DB is partitioned into two parts, namely IMDB and WIKI. The first part consists of 523, whereas WIKI consists of 51 face images. These images are manually labeled with both gender and age. There are some errors in the annotation of ground truth data. In fact, the authors assume that each face image belongs to the listed celebrity and then automatically annotate it with the gender declared in the profile. This automatic assumption results in errors in ground truth data annotation.
 - **VGGFace [60]:** This DB was explicitly built for face recognition but was later used for additional face analysis tasks. It is a large enough DB to train a DL framework. This DB was gathered in a very inexpensive way through google search. A huge quantity of weakly annotated face images are obtained in this way. These face images are filtered and annotated manually through a fast and inexact process. VGGFace2 is an extension of the VGGFACE DB.
 - **VGGFace2 [61]:** This DB was released in 2018, and it contains more than 3.31 million images. All the images are downloaded from the internet. Images in VGGFace2 have significant variations in illumination conditions, race, profession, and age. The DB is gender-balanced, as an almost equal number of male and female subjects are considered in the images collection. All rotation angles have been considered while collecting the DB.
 - **SASE [62]:** The SASE DB images are collected using a Kinect camera. They consist of RGB and depth images, taken from 32 males and 18 females. There are variations in subjects' hairstyle, race, and age. Three rotation angles are considered in the SASE DB. Unlike VGGFace2, facial expressions in the SASE are complex and variable. This DB is used both for HPE and expression recognition.
 - **AFLW [66]:** All images in AFLW are from a collection downloaded from the internet, with a very unconstrained background. Both profile and frontal images are considered in nine lighting conditions. All images have variations in face appearance, expression, and some other environmental factors. The total number of images is 13,000, with 5749 subjects. The DB can be used for face detection, landmarks localization, gender recognition, and HPE.
 - **Multi-Pie [67]:** The illumination conditions and facial expressions in Multi-Pie are variable. All data is collected in different sessions with 15 high-resolution cameras. The DB is large, as more than 75,000 images are collected. It is not a balanced DB, as 69% of the total participants are males and the rest are females. People of different races are considered, including European, Asian, and African communities.
 - **The FacePix [61]:** This is an imbalanced DB with 25 male and 5 female subjects. The total number of images in the DB is 5430. The yaw rotation is covered between $\pm 90^\circ$ with a step size of only 2° . Only one rotation angle yaw is considered in the FacePix DB. FacePix has been used for gender recognition and HPE.
 - **CelebA [72]:** This is a large DB with around 200,000 images. The total number of identities is 10,000. Each image in CelebA is labeled with 40 unique binary attributes. It is a very diverse DB, including entire body and cropped face images. Due to the large number of images, CelebA is adopted by many researchers for DL-based face image analysis tasks.
 - **FEI [68]:** This is also a multi-purpose DB that has been used for gender classification, expression recognition, and HPE. The DB is comparatively simple, as all images are collected in constrained conditions with a uniform and one-color background. The total number of participants is 200, half male and half female. The DB was collected over two sessions, called FEI A and B. Only two facial expressions are considered in FEI. Session A images are of a neutral facial expression, whereas session B images are smiling.

- FERET [69]: This is a DB that was introduced earlier for a number of face analysis tasks. The tasks include face recognition, gender classification, and HPE. It is also a simple DB that is collected in constrained laboratory conditions. It is a medium-sized DB, having 14,126 images. The number of subjects involved is 1199. The DB is somewhat challenging, as variations in facial expressions are present in images. The lighting conditions of the images are also not uniform.
- CAS-PEAL [70]: CAS-PEAL has more than 100,000 images. The number of participants is also sufficiently large (more than 1000). The DB is not balanced, having 595 male and 445 female candidates. The authors considered yaw and pitch angles in the range $\pm 45^\circ$ and $\pm 30^\circ$. The DB is simple to use in experiments, as less complexity is included in the data collection, and the number of poses is also less.
- LFW [71]: This is a comparatively challenging DB, as most of the images are collected from very unconstrained environmental conditions. The total number of participants in the DB is 5479, whereas the total number of images is 1323. All images are downloaded from the internet with very poor resolution. It is a highly imbalanced DB, having 2977 female and 102,566 male participants.
- BJUT-3D [73]: A total of 500 subjects participated in data collection, half male and half female. There are 93 poses in the DB. The pitch and yaw rotations have ranges of $\pm 40^\circ$ and $\pm 60^\circ$, respectively. The DB can be used for gender recognition and HPE.
- FairFace [74]: This is a large DB with more than 108,500 images. The DB is collected considering three tasks: age, race, and gender classification. It is a DB with seven race groups, which is the largest number of groups reported in the CV research community. The DB is balanced as far as the ethnicity of various races is concerned. Some images are very small; the minimum size of a face fixed is 50×50 . The DB is publically available for research purposes and downloading. It is challenging DB since the images are collected in real-time conditions with various difficulty levels due to occlusion, lighting variations, and so on.
- UTKFace [75]: This is a large DB with significant variations in the age range. The age range is from 0 to 116 years. Variations in the pose, illumination, occlusion, and so on are included in data collection, making the DB more challenging. The total number of images used for experiments is 23,688. The number of subjects is about 500+.

4. Methods

It is not easy to organize all research work on the topic into a single taxonomy. In the proposed work, we do not follow or claim a specific taxonomy, but we will try to label each system by the fundamental method that underlines its specific implementations. We discuss various types of methods and also present references where these approaches have been applied to a specific task. We will discuss both advantages and disadvantages of each method. A summary of all these methods is presented in Figure 2.

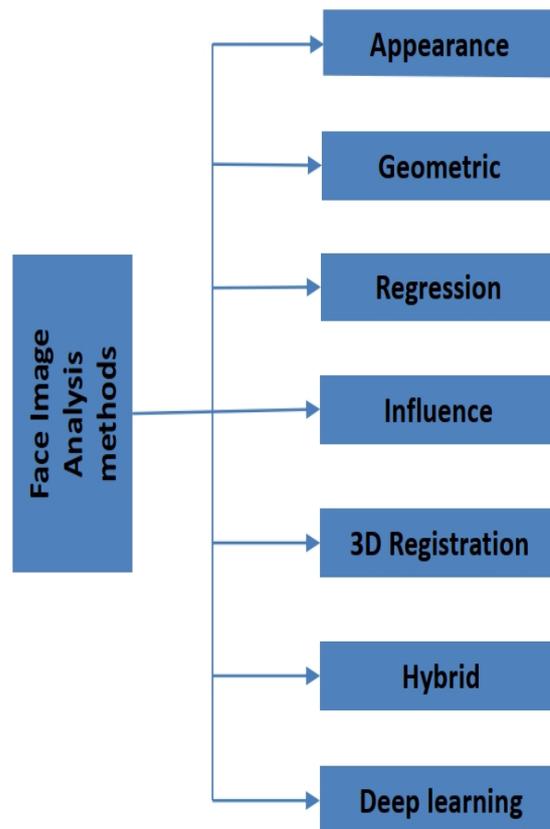


Figure 2. FIA methods and taxonomy.

4.1. Appearance Methods

These methods always assume that a specific correlation exists between a face analysis task and its 3D image properties. To know about this relationship, experiments are performed by training some classifiers. Each face image is treated as a one-dimensional vector, and certain visual features are extracted. This statistical information is used for human face analysis. Some methods that use appearance-based approaches are explored in [76–79].

Appearance-based methods are straightforward approaches suitable for both bad- and good-quality images. Along with positive data, negative training data is not needed. Expanding these methods is also very easy, allowing the framework to adapt to any changes when required.

Along with the above merits, these models also have some serious weaknesses. For these methods, the face part must already be detected by the system. If face localization is not correct, some drastically poor results are obtained. The amount of training data needed is large. When variations in local face appearance occur, no solution exists to tackle the algorithm (for example, problems of occlusion, subjects with facial hair, glasses, etc.). Lastly, pair-wise similarity is another significant issue faced by these methods. Although researchers have developed some algorithms to handle this problem, reported results are still not convincing.

4.2. Geometric Methods

These are also known as model-based approaches. These techniques need certain key points such as eyebrows, eyes, lips, nose, and so on. A single vector with local key points is extracted from these specific points. An algorithm is developed based on the relative position and mutual relationship between key points. This method is very similar to the appearance methods, as a human specifies the particular face task.

The literature reports the exploitation of different facial key points for human face analysis. Most of the methods use different key points in the eyes [1,77]. Other methods use the intra-ocular distance between more than one face part for face analysis. The tip of the nose is used as a discriminative cue for HPE in [80]. Due to the complex geometry of hair and its difficult extraction, hair was not previously used for face analysis. However, some recent methods report good results in HPE [81,82]. The processing time of the geometric methods is much less [83] since very few facial features are extracted, and the size of the feature vector is also small. The work reported in [84] uses nostril, eyes, cheek, and chin information. Another method proposed in [85] is used as a front-end face analysis framework. The last proposed method is used for face recognition and head pose prediction. The method proposed in [86] uses a combination of different features, which increases the processing time of the framework. In the last proposed method, the computational cost is increased as a result of improvement in the overall precision of the system.

The main advantage of these methods is that the key points extracted are very robust to certain changes such as translation and rotation. However, extraction and detection of these key points is itself a big challenge in CV. In previous literature, active shape modeling (ASM) [87] was adapted for key points localization. However, ASM fails to locate these key points in some cases (for example, when variation in lighting condition is greater, the facial expression is complex, or if there is an occlusion problem). If the images' resolution is poor or there is a far-field imagery condition, then extraction of these key points is almost impossible.

4.3. Regression Methods

In non-linear methods, a face analysis task is estimated by learning a mapping from the face image space to a specific task. These methods have drawbacks (e.g., it cannot be assured that a regression tool will efficiently learn proper mapping). The high dimensionality in some cases also creates a challenge for which some dimensionality reduction tools can be used; for instance, support vector regressors (SVRs) [88] and localized gradient histograms [39]. Moreover, regression methods can be easily utilized on these low-dimensional features (e.g., [79]).

One of the main non-linear regression tools used is the neural network. Non-linear regression methods consist of multilayer perceptrons of many feed-forward cells. These feed-forward cells are defined in different layers [89,90]. In these methods, backpropagation is used for training perception. Such algorithms inversely propagate the classification error in the reverse direction. Such a strategy updates both weights and biases of the network. Some regression-based methods for head pose prediction are explored in [91,92].

Local linear mapping is another type of regression-based neural network method. These methods consist of multiple linear maps, such as [93]. In these methods, a weight matrix is learned by first giving the input training data and then comparing it to a centroid sample test for each map. The nearest neighbor search is performed in the last stage, followed by a regression process. For more details, see the work reported in [94].

One main advantage of these methods is their processing speed, which is less. These methods are straightforward to implement. Similarly, these methods can be regularized quickly, avoiding over-fitting problems. Stochastic gradient descent (SGD) is used to update regression methods by adding more data. The regression-based method's performance drops significantly if labeled images are not appropriately annotated. To improve the ground truth data, errors in classification can be used through distortion and applying invariance approaches, as in [95].

4.4. Influence Modeling

For the basic concept of influence-based modeling, see [96,97]. This model presents the idea that individuals influence each other in a system and predicts how an actor affects a partner. When the first state of one actor is already known relative to a location, the outcome is estimated more accurately. Since a single task may provide useful information for

different face tasks, all these tasks influence each other; therefore, several tasks can be handled in one framework.

Inspired by the influence-based modeling idea, some works are presented in [51,52]. These works address several face analysis tasks, including HPE, race, gender, and age estimation in a single model. The methods proposed in [51,52] use face segmentation information provided by a prior model developed in [53]. These methods do not extract landmarks information or high-dimensionality data, but instead perform face segmentation as a prior step.

Some excellent results are claimed by the authors through these methods, particularly for face analysis. However, one main drawback of these methods is the need for a manually labeled face DB. The creation of such a DB is a time-consuming and laborious task. Moreover, the cost of computation of these models is also more than that of competitive algorithms.

4.5. Model-Based 3D Registration

Measured data is registered with some specific reference models in these methods. Meyer et al. [98] proposed a method that predicts a head pose through registration of a morphable model with depth data. Similarly, 3D reconstruction and morphable model are used by Yu et al. [99]. Another work in [100] presents a model with a fitting process. A subject's face is modeled in 3D with depth data. Papazov et al. [101] propose a system for HPE, landmarks localization, and gender classification. The authors collected data using a depth sensor and triangular patch, which encodes the shape of the 3D surface. Similarly, Jeni et al. [102] perform registration by training a cascade regression framework with face scan images. Baltrusaitis et al. [103] combined depth and RGB data to regress an HPE framework using a random forest classifier. Compared to other methods, model-based face regression methods are less explored by computer vision researchers. Some recent methods for face image analysis using 3D registration methods can be seen in [104–106].

4.6. Hybrid Methods

Hybrid methods were introduced in [107] for the first time. The central idea is to combine several of the aforementioned methods into a single framework and perform some face analysis tasks. Appearance- and geometric-based methods are combined in [108]. Global and local features are exploited in the aforementioned method, and supplement each other in face analysis.

A method known as HyperFace [109] extracts features through CNNs and then performs face detection, gender recognition, landmarks localization, and head pose tracking and estimation. KEPLER [110] is another well-known hybrid method to address face analysis tasks. KEPLER learns both global and local features to explore some structural face dependencies. Some other recent hybrid methods can be seen in [97,109,111,112].

Since hybrid approaches obtain information from different cues through different methods and later fuse the estimates from different systems independently, we notice that this increases the prediction accuracy. Moreover, these architectures are modeled without initialization as some drift limitations. This helps overcome the limitation of one specific method with another one. To address face analysis tasks, data-mining methods are used in [113]. Similarly, gender recognition is explored in [114] through face segmentation. Extreme learning machine and CNNs are explored by Mingxing et al. [115] to address face analysis. The authors of the paper explored age and gender classification in their proposed work.

4.7. Deep Learning Methods

With the transition of ML from conventional ML to the recent DL, several limitations and drawbacks of conventional ML are now mitigated. Significantly improved results are reported with the introduction of DL methods in various visual recognition tasks. DL approaches, specifically convolutional neural networks (CNNs), outperform those

methods based on conventional features. The method proposed in [116] does not use landmarks information but extracts features through CNNs and then trains a classification tool. Similarly, CNNs are combined with a regression loss function in the method proposed in QuatNet [117].

The performance of conventional ML methods on face analysis tasks was satisfactory to some extent on images obtained in controlled lab conditions. However, when traditional ML models are exposed to images collected in wild conditions, a significant drop in performance is noticed. On the other hand, DL-based methods perform very well on these DBs that are collected in unconstrained conditions [118].

Gozde et al. [119] introduced a system based on deep learning for face analysis using customer interest. According to the authors, it is one of most exciting and innovative trends in customer interest. The system developed measures customer interest attention. In the proposed system, in the initial stage, customers whose heads and face are oriented towards the advertisement are detected by the system. Then, facial components are extracted for various tasks such as head pose estimation, expression recognition, and age estimation. Other papers on face analysis using deep learning include [120–123].

Although many improvements have been noticed in the performance of face analysis tasks with these DL-based methods, their use is still sporadic. Since these approaches are recent, there is still a need to establish their specific and complete potential in this domain.

5. Annotation Type and Processes

In the CV context, ground truth data consist of a set of images with labels on these images. The ground truth labels can be added by a human or in some automatic way, which totally depends on the complexity of the specific problem. These labels usually include interest points, histograms, corners, shapes, and feature descriptors from a model. For the training phase, ground truth data may be positive or negative. Negative images in training data are used to generate false matches, which helps the model building.

We argue that creating ground truth data may not be a cutting-edge research area; however, it is still as important as any proposed algorithm for CV tasks. No algorithm can be verified and assessed accurately unless the ground truth data has been prepared such that image contents and region of interest are selected and prepared effectively. Better analysis is only possible when ground truth data is better. Ground truth data and its preparation depend upon the task to be addressed. For instance, in 3D image reconstruction, attributes of ground truth data must be recognized accurately for each task. Creating ground truth data for some tasks is easy, for instance in gender, race, and expression classification. The labeling method can be either human-annotated or machine-automated annotation. Ground truth data is created using human annotation for the race, age, gender, and expression recognition. However, creating ground truth data for HPE is tricky; therefore, we present some details in the following paragraphs.

As compared to the above-mentioned face analysis tasks, creating ground truth data for HPE is not easy. The earlier method to create ground truth is through manual annotation by a human. A human assigns a specific label in this method. Creation of ground truth data through this method is easy for smaller DBs; however, when the size of the DBs increases, it is a time-consuming task. Moreover, a higher probability of human error exists in these methods. For example, Pointing'04 is collected by asking each participant to look at the points marked on a wall in the measurement room. Such methods assume that each participant accurately points their head to the point. Practically, such implementation is infeasible. Moreover, it is also assumed that each subject's head is always in an accurate physical position, which is, again, practically impossible.

Along with the above manual annotation, HPE DBs are also established through synthetic data, as in [70,124]. Typically, one model is placed on another model called the virtual ground, and then the camera is moved on a sphere's surface. The center of the sphere is the same as the head model's center. Creating ground truth data through this synthetic method also has drawbacks, for example, if a model represents a neutral

expression, or if the background and parts of the head are missing. Both of these drawbacks make the assessment of algorithms working in real-world conditions difficult. An image showing the creation of ground truth data for HPE is shown in Figure 3.

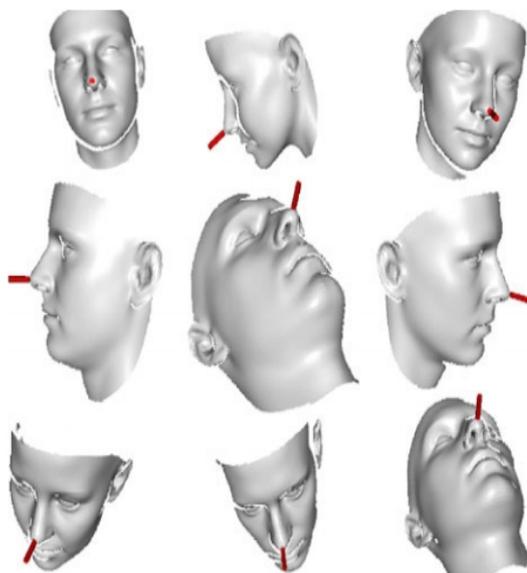


Figure 3. Ground truth data creation for HPE task using synthetic method.

In some other methods, such as [125], a laser pointer is attached to each subject's head, which helps to pinpoint every discrete physical location. A laser pointer is attached to participants' heads to annotate and produce ground truth data. For example, the BU [126,127] DB is created through this method. The magnetic sensor method produces nearly accurate ground truth data; however, the sensors used are very susceptible to small metals if present in the nearby environment.

6. Discussion: Open Areas of Research and Methods Comparison

This section includes a discussion on the results obtained using SOA DBs and some promising future directions for research on the topic.

6.1. Comparative Assessment of Reported Results

Research papers and methods reported in this paper are selected from the last 10 years' (2012–2022) work. Table 2 summarizes results of all five face analysis tasks on SOA DBs. Similarly, Table 3 presents a summary of the latest research work on the topic and its yearly development. From both Tables and the reported methods, we summarize some conclusions, which are as follows:

- Previous methods reported their results using classification rate (CR) (age, gender, expression, and race classification). We also compared and present results for these four tasks with CR. CR is a single metric that is the ratio of correctly identified images to the total number of images. Mathematically, this can be written as follows:

$$CR = \frac{\text{correctly classified images}}{\text{total number of images}} \quad (1)$$

Along with CR, two other informative metrics are used for evaluating an HPE framework, that is, pose estimation accuracy (P_{ea}) and mean absolute error (M_{ae}). M_{ae} is comparatively more common than P_{ea} , as this metric provides a single value which provides an easy insight about the overall performance of an HPE system. On the other hand, P_{ea} purely depends on the poses used, and hence provides comparatively

less information regarding evaluation of the performance of a system. Sometimes, confusion matrices (C_{mat}) are also provided with results. C_{mat} are in a tabular form, where rows are indexed with original and column entries with estimated class. C_{mat} gives a deep visual insight into the classification errors of multiple classes. However, in this paper, we are not considering C_{mat} and P_{ea} , as very few papers reported these metrics, and a proper comparison cannot be made. Mathematically, M_{ae} can be represented as follows:

$$M_{ae} = \frac{1}{N} \sum_{i=1}^N |Y_i - Y'_i| \quad (2)$$

In Equation (2), N represents the test samples, Y_i the ground truth count, and Y'_i the count estimated for the i th sample.

- Face image analysis is a hot topic in CV. Table 2 shows more details on the performance reported for each method. Table 3 shows a summary of the face analysis work done from 2012 to 2022. It is clear from Table 2 that improvements are brought gradually in CR and M_{ae} values. A quick look at the results in Table 2 reveals that the performance for HPE, gender recognition, and age classification tasks on both traditional ML and newly introduced DL methods is not same. From [50–52], it is clear that influence-based modeling methods perform better than conventional ML-based methods. Moreover, in some cases, influence-based methods perform better than DL-based methods. Therefore, we believe that a much better understanding of the DL algorithm methods and their implementation on face analysis tasks is needed. The performance of influence-based modeling on simple DBs acquired in indoor conditions is better. However, when influence-based methods are applied to DBs acquired in open, uncontrolled conditions, a significant drop in performance can be seen. DL methods show much improved results for challenging DBs (for instance, AFLW [66]). Results of traditional ML methods on AFLW are very poor. However, DL has shown much better performance on the same set of images and the same DBs. As far as the performance of traditional ML algorithms is concerned, a mixed response with regard to face analysis tasks can be seen in Table 2. Much better results are shown by hybrid models, as is clear from Table 2.
- Existing methods for face analysis do not define any specific experimental protocol that can be used as a standard for experimental validation. Consequently, different authors use their own experimental setup for validation of their methods. Most of the researchers use 10-fold or 5-fold cross-validation experiments. The results and summary presented in Table 2 are with different experimental setups and validation methods. Therefore, results reported in Table 2 with the same DB may use different validation protocols and setups. For example, Geng et al. [128] use 5-fold cross-validation for Pointing'04, whereas Khan et al. [52] use a subset of the same DB. Therefore, the results presented in Table 2 can be used as a summary, but with a warning to avoid to draw concrete conclusions.
- From the results in Table 2, it can be seen that the performance of most of the methods on Adience DB is significantly lower than the other DBs. Adience is a large DB that is collected in unconstrained conditions. The results in Table 2 highlight the fact that the difficulty level in Adience DB is still high, and more research is needed to make the face analysis technology applicable to images collected in uncontrolled conditions. Moreover, the quality and type of available ground truth data are diverse. Gradual development in the methods of ground truth data can be seen from Section 5. The widely employed manual labeling methods introduce labeling errors, which are reduced with new methods. Synthetic DBs and the creation of manual data are comparatively simple. There is a lower chance of error creation with this method. Most of the images in the DBs mentioned in Table 1 are either centered or cropped faces. Very few methods are fully automated face analysis techniques. An automated face analysis method will detect a face in the initial step and then move to other

secondary face analysis applications or tasks. This could be an interesting area to be explored in the future.

Table 2. Performance comparison of various algorithms for age estimation (AE), gender recognition (GR), race classification (RC), facial expression recognition (FER), and head pose estimation (HPE).

| Paper | DB | Task | CR% | M_{ae}° |
|-------------------------|----------|------------------|----------|------------------|
| Xing et al. [129] | Morph-II | GR | 98.7 | – |
| | | RC | 99.2 | – |
| | | AE | – | 2.96 |
| Yoo et al. [130] | | AE | – | 2.89 |
| | | Wan et al. [131] | GR | 98.4 |
| AE | | | – | 2.93 |
| GR | | | 99.1 | – |
| Xie et al. [132] | | RC | 94.9 | – |
| | | AE | – | 2.81 |
| Sun et al. [133] | GR | 99.5 | – | |
| | RC | 99.6 | – | |
| Zhou et al. [134] | AE | 98.05 | – | |
| | GR | 93.22 | – | |
| Duan et al. [115] | AE | 52.3 | – | |
| Paper | DB | Task | A_{cc} | M_{ae} |
| Rodriguez et al. [135] | Adience | GR | 88.2 | – |
| | | AE | 61.8 | 2.68 |
| | | GR | 93.0 | – |
| Lapuschkin et al. [136] | | AE | 63.8 | – |
| | | GR | 94.0 | – |
| Tizita et al. [137] | | GR | 96 | – |
| | | AE | 65 | – |
| Levi et al. [138] | | GR | 86.8 | – |
| | | AE | 55 | – |
| Savchenko et al. [139] | | AE | 91.95 | 5.96 |
| | | GR | 99.86 | 94.01 |
| Eidinger et al. [58] | | GR | 77.8 | – |
| Liao et al. [140] | | GR | 78.63 | – |
| Hassner et al. [141] | | GR | 79.3 | – |
| Levi et al. [138] | | GR | 86.8 | – |
| Dehghan et al. [142] | | GR | 91.0 | – |
| Gurnani et al. [143] | | GR | 91.8 | – |
| Zhou et al. [134] | VGGFace | AE | 97.8 | – |
| | | GR | 98.3 | – |
| Smith et al. [22] | | GR | 98.2 | – |
| | | AE | – | 4.1 |
| Acien et al. [144] | | GR | 94.8 | – |
| | | RC | 90.1 | – |
| Sharma et al. [75] | | AE | 94.01 | 0.77 |
| | | GR | 99.86 | 2.9 |
| Das et al. [145] | | UTKFace | RC | 90.1 |
| | AE | | 70.1 | – |
| | GR | | 98.23 | – |
| Mane et al. [146] | IMDB | FER | 88.18 | – |
| | | GR | 96.0 | – |
| Mane et al. [146] | | FER | 88.18 | – |
| | | GR | 96.0 | – |

Table 2. *Cont.*

| Paper | DB | Task | A_{cc} | M_{ae} |
|---------------------------|----------|------|----------|----------|
| Thomaz et al. [147] | | GR | 99.0 | – |
| | | FER | 95.0 | – |
| Sergio et al. [50] | FEI | GR | 98.5 | – |
| | | FER | 96.25 | – |
| | | HPE | 81.0 | 2.79 |
| Khan et al. [52] | | GR | 97.3 | – |
| | | FER | 98.85 | – |
| | | HPE | 91.35 | 2.32 |
| Thomaz et al. [147] | FERET | GR | 84.0 | – |
| | | FER | 74.0 | – |
| Sergio et al. [50] | | GR | 96.5 | – |
| | | FER | 93.25 | – |
| Baltrusaitis et al. [103] | ICT-3DHP | HPE | – | 3.2 |
| | | FER | 59.4 | – |
| Ranjan et al. [109] | AFLW | HPE | 2.71 | – |
| | | GR | 94 | – |
| | CelebA | HPE | 52.3 | – |
| Ranjan et al. [109] | Dali3DHP | HPE | – | 6.07 |
| | | FER | 65 | – |
| Ranjan et al. [148] | AFW | HPE | 99.1 | 0.293 |
| | | GR | 99 | – |
| | | FER | 93 | – |
| | | AE | – | 0.20 |
| Shin et al. [149] | LFW | AE | 62.73 | – |
| | | GR | 92.2 | – |
| Karikkainen et al. [74] | FairFace | AE | 59.7 | – |
| | | GR | 94.2 | – |
| | | RC | 93.7 | – |

Table 3. Face image analysis yearly progress (2012–2022).

| Paper | Method | Face Task | DB |
|--------------------------|-----------|-----------------------------|------------------|
| 2022 | | | |
| Sharma et al. [75] | DL | age and gender | IDMB and UTKFace |
| Bhushan et al. [150] | DL | age and expression | YALE |
| Chethana et al. [151] | DL | age, gender, and expression | |
| Pandi et al. [152] | DL | gender and expression | IMDB |
| 2021 | | | |
| Karikkainen et al. [153] | DL | race, age, and gender | FairFace |
| Lahariya et al. [154] | DL | gender and expression | IMDB |
| Park et al. [155] | DL | age and gender | Mega Asian |
| Benkaddour et al. [156] | DL | age and gender | Adience |
| Micheala et al. [157] | DL | age and gender | FERET |
| Kale et al. [158] | DL | race, age, and gender | Face |
| 2020 | | | |
| Li et al. [159] | DL | age and gender | AFLW |
| Barra et al. [160] | geometric | age and gender | AFLW |
| Lim et al. [161] | DL | age and gender | IMDB |

Table 3. Cont.

| Paper | Method | Face Task | DB |
|---------------------------|-------------------|-------------------------------|----------------------|
| 2019 | | | |
| Yang et al. [162] | regression | age and expression | AFLW |
| HyperFace [109] | DL | appearance and DL | AFLW |
| Sergio et al. [50] | IBM | gender, age, and expression | FEI, FERET |
| Hsu et al. [117] | regression | HPE and expression | AFLW |
| Khan et al. [52] | IBM | HPE, age, and gender | FEI, FERET |
| Thomaz et al. [147] | IBM | gender and expression | FEI, FERET |
| Zhou et al. [134] | DL | gender and age | Adience, VGGFace2 |
| 2018 | | | |
| Gupta et al. [163] | regression | expressions and gender | AFLW |
| Ruiz et al. [116] | DL | age and gender | AFLW |
| Smith et al. [22] | DL | gender and age | VGGFace |
| Acien et al. [144] | DL | gender and race | VGGFace |
| Baltrusaitis et al. [103] | DL | HPE and expression | ICT-3DHP |
| Das et al. [145] | DL | gender, age, race | UTKFace |
| Mane et al. [146] | appearance | gender and expression | IMDB |
| 2017 | | | |
| Derkach et al. [164] | regression | appearance | SASE |
| Duan et al. [115] | hybrid | gender and age | Aidience |
| Dehghan et al. [142] | DL | age, gender, expression | Face+ |
| Ranjan et al. [109] | DL | age, race, gender, expression | AFW |
| Shin et al. [149] | DL | age and gender | AFW |
| 2016 | | | |
| Baltrusaitis et al. [109] | 3D morphable | HPE and gender | Multi-Pie+BU |
| Ranjan et al. [109] | DL | head pose and gender | AFLW and CelebA |
| Xia et al. [165] | geometric | race, gender, and age | FRGCv2 |
| Lapuschkin et al. [136] | appearance | age and gender | Adience |
| 2015 | | | |
| Afifa et al. [166] | appearance | age and gender | FacePix, CMU PIE, BU |
| Tulyakov et al. [64] | tracking | HPE and expr. | Dali3DHP |
| Peng et al. [167] | manifold embedded | gender and expr. | CMU-Multipie+BU+AFW |
| Sang-Heon et al. [168] | geometric | age and gender | FERET |
| Levi et al. [138] | DL | age and gender | Adience |
| liu et al. [169] | DL | race, age, gender | FERET |
| Chaudhari et al. [170] | DL | age and gender | FERET |
| Mery et al. [171] | hybrid | race, age, and gender | FERET and JAFFE |
| 2017 | | | |
| Tulyakov et al. [64] | tracking | HPE and expression | Dali3DHP |
| Laurentini et al. [172] | appearance | age, race, and gender | |
| Zhao et al. [173] | geometric | HPE and expression | LFW |
| Fazl-Ersi et al. [174] | appearance | age and gender | FERET |
| Eidinger et al. [58] | appearance | race, age, and gender | FERET |
| 2013 | | | |
| Jain et al. [78] | appearance | expression and gender | CMU-Pie |
| Zhu et al. [111] | 3D registration | age, gender, and expression | CUM-MultiPie+AFW |
| Tizita et al. [137] | appearance | age and gender | FERET |
| Guo et al. [175] | geometric | age, expression, and gender | FERET |
| Bolorizadeh et al. [176] | hybrid | age and gender | FERET |

Table 3. Cont.

| Paper | Method | Face Task | DB |
|------------------------|--------------|-----------------------|----------------------|
| 2012 | | | |
| zhu et al. [111] | appearance | gender and HPE | FERET |
| Hao et al. [177] | regression | HPE and gender | FacePix, Pointing'04 |
| Jang et al. [102] | appearance | age and gender | BU+CMU-MultiPie |
| Xiangyang et al. [178] | registration | expression and gender | FacePix |
| Huang et al. [48] | appearance | age and gender | CAS-PEAL |

6.2. Benchmark DB Development

The development and evaluation of an efficient face analysis framework require a sophisticated DB for experiments. CV researchers report various DBs for all face analysis tasks. The DBs range from a very simple one-color flat background (Pointing'04) to complex scenarios (LFW). Although the number of DBs is increasing gradually, as can be seen from Table 1 and Table 3, none of these address all face analysis tasks in a single DB. It is also clear from the discussion of the reported DBs that none can be used on its own for all face analysis tasks, as no DB covers all the modalities of a human face image. Secondly, it is also clear from Table 1 that the number of DBs addressing facial expressions is comparatively less. Very few DBs provide data about facial expression and race classification. Compared to these two, the remaining three areas (i.e., HPE, gender, and age classification) are more explored.

The number of images or videos is comparatively lower for DL-based experiments (except for a few, e.g., Adience). Another challenge these DBs face is not having multiple sessions per participant or information about specific time lapses between sessions. Diversity in lighting conditions, terms of the recording environment, ethnicity, gender, and viewing angle information are other aspects not addressed by researchers. Another challenge is not having multiple sessions per subject. Therefore, a preliminary step should be to contribute a DB that caters to all the modalities, including session-spanning information, ethnicity, age, gender, and facial expression.

6.3. Research on 3D Face is Required

As can be observed, most of the research conducted on all face analysis tasks is on 2D images. Few papers are reported using 3D face images. We believe exploring face analysis through 3D images will be an exciting topic. In most of the 3D face images, synthetic faces are used. In the case of 2D images, most of the time, valuable information is lost; for example, the exact position of the nose and chin in 3D images can provide more valuable information. The identity of a person is recognized using face image analysis methods in [179]. A pose-invariant learnt multiview 3D recognition approach is used to address the problem. Authors of the paper used four different datasets to address the task. These datasets include UMB-DB, GavabDB, Bosphorus, and FRGC database. Generative 3D face models are applied for the pose-invariant illumination method in [179]. According to the authors, the proposed 3D face model also fit the 2D along with 3D images under various conditions and sensors. Three-dimensional face analysis could be used in more research and application domains, such as face beautification, plastic surgery, and so on.

6.4. Knowledge Transfer and Data Augmentation: Possible Areas of Exploration

Looking for the new trends for other tasks in CV, we expect that face analysis will move more and more towards new DL methods. Since DL methods are facing some training problems due to less ground truth data, accurate knowledge transfer (KT) [180] and supervision learning [181] are possible options to be researched. Another possible area to be improved is data augmentation [182] and foveated architecture methods [183]. Data augmentation can overcome the limited data scenario of DL architectures. We also add that a comparatively less researched domain for KT in DL is heterogeneous domain adoption.

KT is extremely helpful in transferring knowledge from the training to testing phase when attributes are somehow different. All of this minimizes the labeling efforts for training data. As data from some other domain is utilized, the labeling task is reduced. Considering new developments in DL methods, the next probable keywords are temporal pooling, LSTMs, optical flow frames, and 3D convolution for human face image analysis. Even if some of the aforementioned methods are already being explored by researchers, more research is needed to improve the performance of these tasks.

7. Summary and Concluding Remarks

Human face image analysis is an essential step in various face-related applications since rich information is provided about intent, motivation, subjects' attention, and so on. Despite the extensive research on the topic, particularly in the last 10 years, face image analysis is still very challenging, particularly for data collected in unconstrained conditions. We also investigate aspects of some existing solutions: first, we review SOA methods based on hand-crafted representation, and then we move to the recently introduced DL architectures. We present an analysis of the SOA results obtained so far on the topic. Finally, we identify several promising open problems and present possible future work directions; for example, we expect to see more evaluations of the DL techniques on the most challenging DBs (i.e., those collected in uncontrolled environmental conditions). Another more interesting direction is a combination of DL and influence modeling. One additional possibility is the extension of influence modeling with geometric modeling. Classification and feature extraction can be exploited with DL architectures. We believe that the survey paper, DBs, and all the methods and algorithms mentioned here could help the research community working on the topic to improve current SOA performance and inspire some new research directions.

Author Contributions: Conceptualization, K.K. and R.U.K.; methodology, K.K.; software, K.K.; validation, M.H.S. and A.A.; formal analysis, M.H.S.; investigation, M.H.S.; resources, R.U.K.; writing, original draft preparation, K.K.; writing, review and editing, K.K.; visualization, A.A.; supervision, K.K.; project administration, R.U.K.; funding acquisition, M.H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Deanship of Scientific Research at Jouf University under grant No. DSR-2021-02-0350.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Valenti, R.; Sebe, N.; Gevers, T. Combining head pose and eye location information for gaze estimation. *IEEE Trans. Image Process.* **2011**, *21*, 802–815. [[CrossRef](#)] [[PubMed](#)]
2. Murphy-Chutorian, E.; Trivedi, M.M. Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness. *IEEE Trans. Intell. Transp. Syst.* **2010**, *11*, 300–311. [[CrossRef](#)]
3. Wang, K.; Zhao, R.; Ji, Q. Human computer interaction with head pose, eye gaze and body gestures. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; p. 789.
4. Asthana, A.; Zafeiriou, S.; Cheng, S.; Pantic, M. Robust discriminative response map fitting with constrained local models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 3444–3451.
5. Belhumeur, P.N.; Jacobs, D.W.; Kriegman, D.J.; Kumar, N. Localizing parts of faces using a consensus of exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2930–2940. [[CrossRef](#)] [[PubMed](#)]
6. Cao, X.; Wei, Y.; Wen, F.; Sun, J. Face alignment by explicit shape regression. *Int. J. Comput. Vis.* **2014**, *107*, 177–190. [[CrossRef](#)]
7. Dantone, M.; Gall, J.; Fanelli, G.; Van Gool, L. Real-time facial feature detection using conditional regression forests. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.
8. Murphy-Chutorian, E.; Trivedi, M.M. Head pose estimation in computer vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 607–626. [[CrossRef](#)] [[PubMed](#)]
9. Saragih, J.M.; Lucey, S.; Cohn, J.F. Deformable model fitting by regularized landmark mean-shift. *Int. J. Comput. Vis.* **2011**, *91*, 200–215. [[CrossRef](#)]
10. Shan, C.; Gong, S.; McOwan, P.W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vis. Comput.* **2009**, *27*, 803–816. [[CrossRef](#)]

11. Xiong, X.; De la Torre, F. Supervised descent method and its applications to face alignment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 532–539.
12. Zhao, W.; Chellappa, R.; Phillips, P.J.; Rosenfeld, A. Face recognition: A literature survey. *ACM Comput. Surv. (CSUR)* **2003**, *35*, 399–458. [[CrossRef](#)]
13. Masi, I.; Wu, Y.; Hassner, T.; Natarajan, P. Deep face recognition: A survey. In Proceedings of the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Parana, Brazil, 29 October–1 November 2018; pp. 471–478.
14. Kortli, Y.; Jridi, M.; Al Falou, A.; Atri, M. Face recognition systems: A survey. *Sensors* **2020**, *20*, 342. [[CrossRef](#)]
15. Sharif, M.; Naz, F.; Yasmin, M.; Shahid, M.A.; Rehman, A. Face Recognition: A Survey. *J. Eng. Sci. Technol. Rev.* **2017**, *10*, 471–478. [[CrossRef](#)]
16. Pandya, J.M.; Rathod, D.; Jadav, J.J. A survey of face recognition approach. *Int. J. Eng. Res. Appl. (IJERA)* **2013**, *3*, 632–635.
17. Wang, M.; Deng, W. Deep face recognition: A survey. *Neurocomputing* **2021**, *429*, 215–244. [[CrossRef](#)]
18. Lal, M.; Kumar, K.; Arain, R.H.; Maitlo, A.; Ruk, S.A.; Shaikh, H. Study of face recognition techniques: A survey. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 42–49. [[CrossRef](#)]
19. Alghamdi, J.; Alharthi, R.; Alghamdi, R.; Alsubaie, W.; Alsubaie, R.; Alqahtani, D.; Alqarni, L.; Alshammari, R. A survey on face recognition algorithms. In Proceedings of the 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 19–21 March 2020; pp. 1–5.
20. Benfold, B.; Reid, I.D. Guiding visual surveillance by tracking human attention. *BMVC* **2009**, *2*, 7.
21. Sankaranarayanan, K.; Chang, M.-C.; Krahnstoeber, N. Tracking gaze direction from far-field surveillance cameras. In Proceedings of the 2011 IEEE Workshop on Applications of Computer Vision (WACV), Kona, HI, USA, 5–7 January 2011; pp. 519–526.
22. Smith, K.; Ba, S.O.; Odobez, J.-M.; Gatica-Perez, D. Tracking the visual focus of attention for a varying number of wandering people. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1212–1229. [[CrossRef](#)]
23. Khan, Z.; Fu, Y. One label, one billion faces: Usage and consistency of racial categories in computer vision. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Toronto, Canada, 3–10 March 2021; pp. 587–597.
24. Zhang, L.; Shen, J.J.; Zhang, J.; Xu, J.; Li, Z.; Yao, Y.; Yu, L. Multimodal marketing intent analysis for effective targeted advertising. *IEEE Trans. Multimed.* **2022**, *24*, 1830–1843. [[CrossRef](#)]
25. Xiao, Y.; Wang, C. You see what I want you to see: Exploring targeted black-box transferability attack for hash-based image retrieval systems. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 1934–1943.
26. Zhang, S.; Zhang, S.; Huang, T.; Gao, W. Multimodal deep convolutional neural network for audio-visual emotion recognition. In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, ACM, New York, NY, USA, 6–9 June 2016; pp. 281–284.
27. Nihei, F.; Nakano, Y.I. Exploring methods for predicting important utterances contributing to meeting summarization. *Multimodal Technol. Interact.* **2019**, *3*, 50. [[CrossRef](#)]
28. Basu, S.; Choudhury, T.; Clarkson, B.; Pentland, A. Towards measuring human interactions in conversational settings. In Proceedings of the IEEE CVPR Workshop on Cues in Communication, Kauai, HI, USA, 9 December 2001.
29. McCowan, L.; Gatica-Perez, D.; Bengio, S.; Lathoud, G.; Barnard, M.; Zhang, D. Automatic analysis of multimodal group actions in meetings. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 305–317. [[CrossRef](#)]
30. Odobez, J.-M.; Ba, S. A cognitive and unsupervised map adaptation approach to the recognition of the focus of attention from head pose. In Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, Beijing, China, 2–5 July 2007; pp. 1379–1382.
31. Stiefelhagen, R.; Yang, J.; Waibel, A. A model-based gaze tracking system. *Int. J. Artif. Intell. Tools* **1997**, *6*, 193–209. [[CrossRef](#)]
32. Zhang, Z.; Hu, Y.; Liu, M.; Huang, T. Head pose estimation in seminar room using multi view face detectors. In *International Evaluation Workshop on Classification of Events, Activities and Relationships*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 299–304.
33. Zen, G.; Lepri, B.; Ricci, E.; Lanz, O. Space speaks: Towards socially and personality aware visual surveillance. In Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis, ACM, Firenze, Italy, 29 October 2010; pp. 37–42.
34. Reid, I.; Benfold, B.; Patron, A.; Sommerlade, E. Understanding interactions and guiding visual surveillance by tracking attention. In *Asian Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 380–389.
35. Chen, C.-W.; Aghajan, H. Multiview social behavior analysis in work environments. In Proceedings of the 2011 Fifth ACM/IEEE International Conference on Distributed Smart Cameras, Ghent, Belgium, 22–25 August 2011; pp. 1–6.
36. Zdzisław, K.; Czubenko, M.; Merta, T. Emotion monitoring system for drivers. *IFAC-PapersOnLine* **2019**, *52*, 200–205.
37. Braun, M.; Schubert, J.; Pfleging, B.; Alt, F. Improving driver emotions with affective strategies. *Multimodal Technol. Interact.* **2019**, *3*, 21. [[CrossRef](#)]
38. Ihme, K.; Dömel, C.; Freese, M.; Jipp, M. Frustration in the face of the driver: A simulator study on facial muscle activity during frustrated driving. *Interact. Stud.* **2018**, *19*, 487–498. [[CrossRef](#)]
39. Murphy-Chutorian, E.; Doshi, A.; Trivedi, M.M. Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. In Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Bellevue, WA, USA, 30 September–3 October 2007; pp. 709–714.
40. Ray, S.J.; Teizer, J. Coarse head pose estimation of construction equipment operators to formulate dynamic blind spots. *Adv. Eng. Inform.* **2012**, *26*, 117–130. [[CrossRef](#)]

41. Schulz, A.; Damer, N.; Fischer, M.; Stiefelhagen, R. Combined head localization and head pose estimation for video-based advanced driver assistance systems. In *Joint Pattern Recognition Symposium*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 51–60.
42. Fu, Y.; Guo, G.; Huang, T.S. Age synthesis and estimation via faces: A survey. *IEEE Trans. PAMI* **2010**, *32*, 1955–1976.
43. Ekman, P. Facial expression and emotion. *Am. Psychol.* **1993**, *48*, 384–392. [[CrossRef](#)]
44. Picard, R.W.; Vyzas, E.; Healey, J. Toward machine emotional intelligence: Analysis of affective physiological states. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1175–1191. [[CrossRef](#)]
45. Zeng, Z.; Pantic, M.; Roisman, G.I.; Huang, T.S. A survey of affect recognition methods: Audio, visual and spontaneous expressions. *IEEE Trans. PAMI* **2009**, *31*, 39–58. [[CrossRef](#)]
46. Wollaston, W.H. Xiii. on the apparent direction of eyes in a portrait. *Philos. Trans. R. Soc. Lond.* **1824**, *114*, 247–256.
47. Langton, S.R.; Honeyman, H.; Tessler, E. The influence of head contour and nose angle on the perception of eye-gaze direction. *Percept. Psychophys.* **2004**, *66*, 752–771. [[CrossRef](#)]
48. Huang, C.; Ding, X.; Fang, C. Head pose estimation based on random forests for multiclass classification. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 934–937.
49. Khan, K.; Mauro, M.; Migliorati, P.; Leonardi, R. Head pose estimation through multi-class face segmentation. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017; pp. 175–180.
50. Khan, K.; Ahmad, N.; Khan, F.; Syed, I. A framework for head pose estimation and face segmentation through conditional random fields. *Signal Image Video Process.* **2019**, *14*, 159–166. [[CrossRef](#)]
51. Benini, S.; Khan, K.; Leonardi, R.; Mauro, M.; Migliorati, P. Face analysis through semantic face segmentation. *Signal Process. Image Commun.* **2019**, *74*, 21–31. [[CrossRef](#)]
52. Khan, K.; Attique, M.; Syed, I.; Sarwar, G.; Irfan, M.A.; Khan, R.U. A unified framework for head pose, age and gender classification through end-to-end face segmentation. *Entropy* **2019**, *21*, 647. [[CrossRef](#)] [[PubMed](#)]
53. Khan, K.; Mauro, M.; Leonardi, R. Multi-class semantic segmentation of faces. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 827–831.
54. Raji, I.D.; Fried, G. About face: A survey of facial recognition evaluation. *arXiv* **2021**, arXiv:2102.00813.
55. Mascio, T.D.; Fantozzi, P.; Laura, L.; Rughetti, V. Age and Gender (Face) Recognition: A Brief Survey. In *International Conference in Methodologies and intelligent Systems for Technology Enhanced Learning*; Springer: Cham, Switzerland, 2021; pp. 105–113
56. Khan, K.; Khan, R.U.; Leonardi, R.; Migliorati, P.; Benini, S. Head pose estimation: A survey of the last ten years. *Signal Process. Image Commun.* **2021**, *99*, 116479. [[CrossRef](#)]
57. Azzopardi, G.; Greco, A.; Saggese, A.; Vento, M. Fast gender recognition in videos using a novel descriptor based on the gradient magnitudes of facial landmarks. In Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 29 August–1 September 2017; pp. 1–6.
58. Eidinger, E.; Enbar, R.; Hassner, T. Age and gender estimation of unfiltered faces. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2170–2179. [[CrossRef](#)]
59. Rothe, R.; Timofte, R.; Gool, L.V. Deep expectation of real and apparent age from a single image without facial landmarks. *Int. J. Comput. Vis.* **2018**, *126*, 144–157. [[CrossRef](#)]
60. Breitenstein, M.D.; Kuettel, D.; Weise, T.; Gool, L.V.; Pfister, H. Real-time face pose estimation from single range images. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 24–26 June 2008; pp. 1–8.
61. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. Vggface2: A DB for recognising faces across pose and age. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 67–74.
62. Usi, I.L.; Junior, J.C.J.; Gorbova, J.; Baro, X.; Escalera, S.; Demirel, H.; Allik, J.; Ozcinar, C.; Anbarjafari, G. Joint challenge on dominant and complementary emotion recognition using micro emotion features and head-pose estimation: DBs. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 809–813.
63. Liu, Y.; Chen, J.; Su, Z.; Luo, Z.; Luo, N.; Liu, L.; Zhang, K. Robust head pose estimation using dirichlet-tree distribution enhanced random forests. *Neurocomputing* **2016**, *173*, 42–53. [[CrossRef](#)]
64. Tulyakov, S.; Vieriu, R.-L.; Semeniuta, S.; Sebe, N. Robust real-time extreme head pose estimation. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 2263–2268.
65. Demirkus, M.; Clark, J.J.; Arbel, T. Robust semi-automatic head pose labeling for real-world face video sequences. *Multimed. Tools Appl.* **2014**, *70*, 495–523. [[CrossRef](#)]
66. Koestinger, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Annotated facial landmarks in the wild: A large-scale, real-world DB for facial landmark localization. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 2144–2151.
67. Gross, R.; Matthews, I.; Cohn, J.; Kanade, T.; Baker, S. Multi-pie. *Image Vis. Comput.* **2010**, *28*, 807–813. [[CrossRef](#)]
68. Centro Universitario Da Fei, Fei Face DB. Available online: <http://www.fei.edu.br/cet/faceDB.html> (accessed on 13 March 2022).
69. Phillips, P.J.; Wechsler, H.; Huang, J.; Rauss, P.J. The feret DB and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* **1998**, *16*, 295–306. [[CrossRef](#)]

70. Fanelli, G.; Dantone, M.; Gall, J.; Fossati, A.; Gool, L.V. Random forests for real time 3d face analysis. *Int. J. Comput.* **2013**, *101*, 437–458. [[CrossRef](#)]
71. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled Faces in the Wild: A DB Forstudying Face Recognition in Unconstrained Environments. 2008. Available online: <https://hal.inria.fr/inria-00321923> (accessed on 13 March 2022).
72. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3730–3738.
73. Baocai, Y.; Yanfeng, S.; Chengzhang, W.; Yun, G. Bjut-3d large scale 3d face DB and information processing. *J. Comput. Res. Dev.* **2009**, *6*, 020.
74. Kärkkäinen, K.; Joo, J. Fairface: Face attribute DB for balanced race, gender, and age. *arXiv* **2019**, arXiv:1908.04913.
75. Sharma, N.; Sharma, R.; Jindal, N. Face-Based Age and Gender Estimation Using Improved Convolutional Neural Network Approach. *Wirel. Pers. Commun.* **2022**, 1–20. [[CrossRef](#)]
76. Burl, M.C.; Perona, P. Recognition of planar object classes. In Proceedings of the CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 18–20 June 1996; pp. 223–230.
77. Asteriadis, S.; Tzouveli, P.; Karpouzis, K.; Kollias, S. Estimation of behavioral user state based on eye gaze and head pose application in an e-learning environment. *Multimed. Tools Appl.* **2009**, *41*, 469–493. [[CrossRef](#)]
78. Jain, V.; Crowley, J.L. Head pose estimation using multi-scale gaussian derivatives. In *Image Analysis*; Kämäräinen, J.-K., Koskela, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 319–328.
79. Ma, B.; Huang, R.; Qin, L. Vod: A novel image representation for head yaw estimation. *Neurocomputing* **2015**, *148*, 455–466. [[CrossRef](#)]
80. Zavan, F.H.; Nascimento, A.C.; Bellon, O.R.; Silva, L. Nosepose: A competitive, landmark-free methodology for head pose estimation in the wild. In Proceedings of the Conference on Graphics, Patterns and Images-W. Face Processing 2016, Sao Paulo, Brazil, 4–7 October 2016.
81. Svanera, M.; Muhammad, U.R.; Leonardi, R.; Benini, S. Figaro, hair detection and segmentation in the wild. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 933–937. [[CrossRef](#)]
82. Muhammad, U.R.; Svanera, M.; Leonardi, R.; Benini, S. Hair detection, segmentation, and hairstyle classification in the wild. *Image Vis. Comput.* **2018**, *71*, 25–37. [[CrossRef](#)]
83. Sherrah, J.; Gong, S. Fusion of perceptual cues for robust tracking of head pose and position. *Pattern Recognit.* **2001**, *34*, 1565–1572. [[CrossRef](#)]
84. Nikolaidis, A.; Pitas, I. Facial feature extraction and determination of pose. In *Noblesse Workshop on Non-Linear Model Based Image Analysis*; Springer: London, UK, 1998.
85. Wu, J.; Trivedi, M.M. A two-stage head pose estimation framework and evaluation. *Pattern Recognit.* **2008**, *41*, 1138–1158. [[CrossRef](#)]
86. Cootes, T.F.; Taylor, C.J.; Cooper, D.H.; Graham, J. Active shape models-their training and application. *Comput. Vis. Image Underst.* **1995**, *61*, 38–59. [[CrossRef](#)]
87. Fleuret, F.; Geman, D. Fast face detection with precise pose estimation. In *Object Recognition Supported by User Interaction for Service Robots*; IEEE: Piscataway, NJ, USA, 2002; Volume 1.
88. Li, Y.; Gong, S.; Sherrah, J.; Liddell, H. Support vector machine based multi-view face detection and recognition. *Image Vis. Comput.* **2004**, *22*, 413–427. [[CrossRef](#)]
89. Bishop, C.M. *Neural Networks for Pattern Recognition*; Oxforduniversity Press: Oxford, UK, 1995.
90. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 2001.
91. Brown, L.M.; Tian, Y.-L. Comparative study of coarse head pose estimation. In Proceedings of the Workshop on Motion and Video Computing, Orlando, FL, USA, 5–6 December 2002; pp. 125–130.
92. Schiele, B.; Waibel, A. Gaze tracking based on face-color. In *International Workshop on Automatic Face-and Gesture-Recognition*; University of Zurich Department of Computer Science Multimedia Laboratory: Zürich, Switzerland, 1995; Volume 476.
93. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
94. Ma, B.; Zhang, W.; Shan, S.; Chen, X.; Gao, W. Robust head pose estimation using lgbp. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Washington, DC, USA, 20–24 August 2006; Volume 2, pp. 512–515.
95. Ma, Y.; Konishi, Y.; Kinoshita, K.; Lao, S.; Kawade, M. Sparse bayesianregression for head pose estimation. In Proceedings of the 18th International Conferenceon Pattern Recognition (ICPR'06), Washington, DC, USA, 20–24 August 2006; Volume 3, pp. 507–510.
96. Pan, W.; Dong, W.; Cebrian, M.; Kim, T.; Fowler, J.H.; Pentland, A.S. Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems. *IEEE Signal Process. Mag.* **2012**, *29*, 77–86. [[CrossRef](#)]
97. Dong, W.; Lepri, B.; Pianesi, F.; Pentland, A. Modeling functional rolesdynamics in small group interactions. *IEEE Trans. Multimed.* **2012**, *15*, 83–95. [[CrossRef](#)]
98. Malciu, M.; Preteux, F. A robust model-based approach for 3d headtracking in video sequences. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat.No. PR00580), Grenoble, France, 28–30 March 2000; pp. 169–174.

99. Yu, Y.; Mora, K.A.F.; Odobez, J.-M. Robust and accurate 3d head-pose estimation through 3dmm and online head model reconstruction. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 711–718.
100. Ghiass, R.S.; Arandjelović, O.; Laurendeau, D. Highly accurate and fully automatic head pose estimation from a low quality consumer-level rgb-d sensor. In Proceedings of the 2nd Workshop on Computational Models of Social Interactions: Human-Computer-Media Communication, ACM, Brisbane, Australia, 4–9 May 2015; pp. 25–34.
101. Papazov, C.; Marks, T.K.; Jones, M. Real-time 3d head pose and facial landmark estimation from depth images using triangular surface patch features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4722–4730.
102. Jang, J.-S.; Kanade, T. Robust 3d Head Tracking by View-Based Featurepoint Registration. 2010. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?jsessionid=8FE1D32FB3577413DD9E5D3200E0C734?doi=10.1.1.180.5065&rep=rep1&type=pdf> (accessed on 13 March 2022).
103. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.-P. Openface 2.0: Facial behavior analysis toolkit. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 59–66.
104. Hong, C.; Yu, J.; Zhang, J.; Jin, X.; Lee, K. Multimodal face-pose estimation with multitask manifold deep learning. *IEEE Trans. Ind. Inform.* **2018**, *15*, 3952–3961. [[CrossRef](#)]
105. Stéphane, L.; Mesejo, P.; Alameda-Pineda, X.; Horaud, R. A comprehensive analysis of deep regression. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2065–2081.
106. 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, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3359–3368.
107. Thrun, S.; Pratt, L. *Learning to Learn*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
108. Gee, A.; Cipolla, R. Determining the gaze of faces in images. *Image Vis. Comput.* **1994**, *12*, 639–647. [[CrossRef](#)]
109. 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.* **2017**, *41*, 121–135. [[CrossRef](#)]
110. Kumar, A.; Alavi, A.; Chellappa, R. Kepler: Keypoint and pose estimation of unconstrained faces by learning efficient h-cnn regressors. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, DC, USA, 30 May–3 June 2017; pp. 258–265.
111. Zhu, X.; Ramanan, D. Face detection, pose estimation, and landmark localization in the wild. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2879–2886.
112. Jebara, T.S.; Pentland, A. Parametrized structure from motion for 3d adaptive feedback tracking of faces. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 17–19 June 1997; pp. 144–150.
113. Thamilselvan, P.; Sathiaselan, J.G.R. Image classification using hybrid data mining algorithms—a review. In Proceedings of the 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 19–20 March 2015; pp. 1–6.
114. Khalil, K.; Attique, M.; Syed, I.; Gul, A. Automatic gender classification through face segmentation. *Symmetry* **2019**, *11*, 770.
115. Duan, M.; Li, K.; Yang, C.; Li, K. A hybrid deep learning CNN-ELM for age and gender classification. *Neurocomputing* **2018**, *275*, 448–461. [[CrossRef](#)]
116. Ruiz, N.; Chong, E.; Rehg, J.M. Fine-grained head pose estimation without key points. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2074–2083.
117. Hsu, H.-W.; Wu, T.-Y.; Wan, S.; Wong, W.H.; Lee, C.-Y. Quatnet: Quaternion-based head pose estimation with multi regression loss. *IEEE Trans. Multimed.* **2018**, *21*, 1035–1046. [[CrossRef](#)]
118. Patacchiola, M.; Cangelosi, A. Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. *Pattern Recognit.* **2017**, *71*, 132–143.
119. Gozde, Y.; Oztel, I.; Kazan, S.; Oz, C.; Bunyak, F. Deep learning-based face analysis system for monitoring customer interest. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 237–248.
120. Andrea, G.; Ceccacci, S.; Mengoni, M. A deep learning-based system to track and analyze customer behavior in retail store. In Proceedings of the 2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), Berlin, Germany, 2–5 September 2018; pp. 1–6.
121. Lopes, R.R.; Schwartz, G.M.; Ruggiero, W.V.; Rodríguez, D.Z. A knowledge-based recommendation system that includes sentiment analysis and deep learning. *IEEE Trans. Ind. Inform.* **2018**, *15*, 2124–2135.
122. Hou, J. Deep Learning-Based Human Emotion Detection Framework Using Facial Expressions. *J. Interconnect. Netw.* **2022**, 2141018. [[CrossRef](#)]
123. Apoorva, S.; Saha, S.; Hasanuzzaman, M.; Dey, K. Multitask learning for complaint identification and sentiment analysis. *Cogn. Comput.* **2022**, *14*, 212–227.
124. Liu, X.; Liang, W.; Wang, Y.; Li, S.; Pei, M. 3d head pose estimation with convolutional neural network trained on synthetic images. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 1289–1293.

125. Rae, R.; Ritter, H.J. Recognition of human head orientation based on artificial neural networks. *IEEE Trans. Neural Netw.* **1998**, *9*, 257–265. [[CrossRef](#)]
126. Cascia, M.L.; Sclaroff, S.; Athitsos, V. Fast, reliable head tracking under varying illumination: An approach based on registration of texturemapped 3d models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 322–336. [[CrossRef](#)]
127. Morency, L.-P.; Rahimi, A.; Darrell, T. Adaptive view-based appearance models. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003; I-803.
128. Geng, X.; Zhou, Z.-H.; Smith-Miles, K. Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2234–2240. [[CrossRef](#)]
129. Xing, J.; Li, K.; Hu, W.; Yuan, C.; Ling, H. Diagnosing deep learning models for high accuracy age estimation from a single image. *Pattern Recognit.* **2017**, *66*, 106–116. [[CrossRef](#)]
130. Yoo, B.; Kwak, Y.; Kim, Y.; Choi, C.; Kim, J. Deep facial age estimation using conditional multitask learning with weak label expansion. *IEEE Signal Process. Lett.* **2018**, *25*, 808–812. [[CrossRef](#)]
131. Wan, J.; Tan, Z.; Guo, G.; Li, S.Z.; Lei, Z. Auxiliary demographic information assisted age estimation with cascaded structure. *IEEE Trans. Cybern.* **2018**, *48*, 2531–2541. [[CrossRef](#)] [[PubMed](#)]
132. Xie, J.-C.; Pun, C.-M. Chronological age estimation under the guidance of age-related facial attributes. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 2500–2511. [[CrossRef](#)]
133. Sun, H.; Pan, H.; Han, H.; Shan, S. Deep Conditional Distribution Learning for Age Estimation. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 4679–4690. [[CrossRef](#)]
134. Zhou, Y.; Ni, H.; Ren, F.; Kang, X. Face and gender recognition system based on convolutional neural networks. In Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2019; pp. 1091–1095.
135. Rodriguez, P.; Cucurull, G.; Gonfau, J.M.; Roca, F.X.; Gonzalez, J. Age and gender recognition in the wild with deep attention. *Pattern Recognit.* **2017**, *72*, 563–571. [[CrossRef](#)]
136. Lapuschkin, S.; Binder, A.; Muller, K.R.; Samek, W. Understanding and comparing deep neural networks for age and gender classification. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 1629–1638.
137. Shewaye, T.N. Age group and gender recognition from human facial images. *arXiv* **2013**, arXiv:1304.0019.
138. Levi, G.; Hassner, T. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 34–42.
139. Savchenko, A.V. Efficient facial representation for age, gender and identity recognition in organizing photo albums using multi-output CNN. *PeerJ Comput. Sci.* **2019**, *5*, e197 [[CrossRef](#)]
140. Liao, Z.; Petridis, S.; Pantic, M. Local Deep Neural networks for Age and Gender Classification. *arXiv* **2017**, arXiv:1703.08497.
141. Hassner, T.; Harel, S.; Paz, E.; Enbar, R. Effective face frontalization in unconstrained images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
142. Dehghan, A.; Ortiz, E.G.; Shu, G.; Masood, S.Z. Dager: Deep age, gender and emotion recognition using convolutional neural networks. *arXiv* **2017**, arXiv:1702.04280.
143. Gurnani, A.; Shah, K.; Gajjar, V.; Mavani, V.; Khandhediya, Y. SAFBAGE: Salient approach for face soft-biometric classification—age, gender, and face expression. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2018.
144. Acién, A.; Morales, A.; Vera-Rodriguez, R.; Bartolome, I.; Fierrez, J. Measuring the gender and ethnicity bias in deep models for face recognition. In *Iberoamerican Congress on Pattern Recognition*; Springer: Cham, Switzerland, 2018; pp. 584–593.
145. Das, A.; Dantcheva, A.; Bremond, F. Mitigating bias in gender, age and ethnicity classification: A multi-task convolution neural network approach. In Proceedings of the European Conference on Computer Vision (Eccv) Workshops, Munich, Germany, 8–14 September 2018.
146. Mane, S.; Shah, G. Facial recognition, expression recognition, and gender identification. In *Data Management, Analytics and Innovation*; Springer: Singapore, 2019; pp. 275–290.
147. Thomaz, C.; Giraldo, G.; Costa, J.; Gillies, D. A priori-driven PCA. In *Computer Vision ACCV 2012 Workshops, Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 236–247.
148. Ranjan, R.; Sankaranarayanan, S.; Castillo, C.D.; Chellappa, R. An all-in-one convolutional neural network for face analysis. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 17–24.
149. Shin, M.; Seo, J.H.; Kwon, D.S. Face image-based age and gender estimation with consideration of ethnic difference. In Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Lisbon, Portugal, 28–31 August 2017; pp. 567–572.
150. Bhushan, S.; Alshehri, M.; Agarwal, N.; Keshta, I.; Rajpurohit, J.; Abugabah, A. A Novel Approach to Face Pattern Analysis. *Electronics* **2022**, *11*, 444. [[CrossRef](#)]
151. Chethana, H.T.; Nagavi, T.C. A Review of Face Analysis Techniques for Conventional and Forensic Applications. *Cyber Secur. Digit. Forensics* **2022**, 223–240. [[CrossRef](#)]
152. Pandi, C.; Adi Narayana Reddy, K.; Alladi, R.; Chandra Sekhar Reddy, V.; Sumithabhashini, P. Emotion and Gender Classification Using Convolution Neural Networks. In *ICT Systems and Sustainability*; Springer: Singapore, 2022; pp. 563–573.

153. Karkkainen, K.; Joo, J. Fairface: Face attribute DB for balanced race, gender, and age for bias measurement and mitigation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2021; pp. 1548–1558.
154. Lahariya, A.; Singh, V.; Tiwary, U.S. Real-time Emotion and Gender Classification using Ensemble CNN. *arXiv* **2021**, arXiv:2111.07746.
155. Park, G.; Jung, S. Facial Information Analysis Technology for Gender and Age Estimation. *arXiv* **2021**, arXiv:2111.09303.
156. Benkaddour, M.K. CNN based features extraction for age estimation and gender classification. *Informatika* **2021**, *45*. [CrossRef]
157. Micheala, A.A.; Shankar, R. Automatic Age and Gender Estimation using Deep Learning and Extreme Learning Machine. *Turk. J. Comput. Math. Educ.* **2021**, *12*, 63–73.
158. Kale, A.; Altun, O. Age, Gender and Ethnicity Classification from Face Images with CNN-Based Features. In Proceedings of the 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), Elazig, Turkey, 6–8 October 2021; pp. 1–6.
159. Li, J.; Wang, J.; Ullah, F. An end-to-end task-simplified and anchor-guided deep learning framework for imagebased head pose estimation. *IEEE Access* **2020**, *8*, 42458–42468. [CrossRef]
160. Barra, P.; Barra, S.; Bisogni, C.; Marsico, M.D.; Nappi, M. Web-shaped model for head pose estimation: An approach for best exemplar selection. *IEEE Trans. Image Process.* **2020**, *29*, 5457–5468. [CrossRef]
161. Abu Nada, A.M.; Alajrami, E.; Al-Saqqa, A.A.; Abu-Naser, S.S. Age and Gender Prediction and Validation through Single User Images Using CNN. 2020. Available online: <http://dspace.alazhar.edu.ps/xmlui/handle/123456789/632> (accessed on 13 March 2022).
162. 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, Long Beach, CA, USA, 16–17 June 2019; pp. 1087–1096.
163. Gupta, A.; Thakkar, K.; Gandhi, V.; Narayanan, P. Nose, eyes and ears: Head pose estimation by locating facial key points. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1977–1981.
164. Derkach, D.; Ruiz, A.; Sukno, F.M. Head pose estimation based on 3-d facial landmarks localization and regression. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 820–827.
165. Xia, B.; Amor, B.B.; Daoudi, M. Joint gender, ethnicity and age estimation from 3D faces: An experimental illustration of their correlations. *Image Vis. Comput.* **2017**, *64*, 90–102. [CrossRef]
166. Dahmane, A.; Larabi, S.; Bilasco, I.M.; Djeraba, C. Head pose estimation based on face symmetry analysis. *Signal, Image Video Process.* **2015**, *9*, 1871–1880. [CrossRef]
167. Yang, X.; Huang, D.; Wang, Y.; Chen, L. Automatic 3d facial expression recognition using geometric scattering representation. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 1, pp. 1–6.
168. Lee, S.; Sohn, M.; Kim, H. Implementation of age and gender recognition system for intelligent digital signage. In Proceedings of the Eighth International Conference on Machine Vision (ICMV 2015), Barcelona, Spain, 19–20 November 2015.
169. Liu, H.; Shen, X.; Ren, H. FDAR-Net: Joint convolutional neural networks for face detection and attribute recognition. In Proceedings of the 2016 9th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 10–11 December 2016; Volume 2, pp. 184–187.
170. Chaudhari, S.J.; Kagalkar, R.M. Methodology for Gender Identification, Classification and Recognition of Human Age. *Int. J. Comput. Appl.* **2015**, *975*, 8887.
171. Mery, D.; Bowyer, K. Automatic facial attribute analysis via adaptive sparse representation of random patches. *Pattern Recognit. Lett.* **2015**, *68*, 260–269. [CrossRef]
172. Laurentini, A.; Bottino, A. Computer analysis of face beauty: A survey. *Comput. Vis. Image Underst.* **2014**, *125*, 184–199. [CrossRef]
173. Zhao, X.; Kim, T.K.; Luo, W. Unified face analysis by iterative multi-output random forests. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1765–1772.
174. Fazl-Ersi, E.; Mousa-Pasandi, M.E.; Laganieri, R.; Awad, M. Age and gender recognition using informative features of various types. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 5891–5895.
175. Guo, G.; Mu, G. Joint estimation of age, gender and ethnicity: CCA vs. PLS. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–6.
176. Bolorizadeh, P.; Tojari, F. Facial expression recognition: Age, gender and exposure duration impact. *Procedia-Soc. Behav. Sci.* **2013**, *84*, 1369–1375. [CrossRef]
177. Ji, H.; Liu, R.; Su, F.; Su, Z.; Tian, Y. Robust head pose estimation via convex regularized sparse regression. In Proceedings of the 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 3617–3620.
178. Liu, X.; Lu, H.; Li, W. Multi-manifold modeling for head pose estimation. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 3277–3280.
179. Afzal, H.R.; Luo, S.; Afzal, M.K.; Chaudhary, G.; Khari, M.; Kumar, S.A.P. 3D face reconstruction from single 2D image using distinctive features. *IEEE Access* **2020**, *8*, 180681–180689. [CrossRef]

180. Hoffman, J.; Rodner, E.; Donahue, J.; Kulis, B.; Saenko, K. Asymmetric and category invariant feature transformations for domain adaptation. *Int. J. Comput. Vis.* **2014**, *109*, 28–41. [[CrossRef](#)]
181. Zhou, Z.-H. A brief introduction to weakly supervised learning. *Natl. Sci. Rev.* **2018**, *5*, 44–53. [[CrossRef](#)]
182. Wang, L.; Xiong, Y.; Wang, Z.; Qiao, Y. Towards good practices for very deep two-stream convnets. *arXiv* **2015**, arXiv:1507.02159.
183. Karpathy, A.; Toderici, G.; Shetty, S.; Leung, T.; Sukthankar, R.; FeiFei, L. Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1725–1732.