Annotated Facial Landmarks in the Wild: A Large-scale, Real-world Database for Facial Landmark Localization *

Martin Köstinger, Paul Wohlhart, Peter M. Roth, Horst Bischof
Institute for Computer Graphics and Vision, Graz University of Technology
{koestinger,wohlhart,pmroth,bischof}@icg.tugraz.at

Abstract

Face alignment is a crucial step in face recognition tasks. Especially, using landmark localization for geometric face normalization has shown to be very effective, clearly improving the recognition results. However, no adequate databases exist that provide a sufficient number of annotated facial landmarks. The databases are either limited to frontal views, provide only a small number of annotated images or have been acquired under controlled conditions. Hence, we introduce a novel database overcoming these limitations: Annotated Facial Landmarks in the Wild *(*AFLW*).* AFLW *provides a large-scale collection of images gathered from Flickr, exhibiting a large variety in face appearance (*e.g.*, pose, expression, ethnicity, age, gender) as well as general imaging and environmental conditions. In total 25,993 faces in 21,997 real-world images are annotated with up to 21 landmarks per image. Due to the comprehensive set of annotations* AFLW *is well suited to train and test algorithms for multi-view face detection, facial landmark localization and face pose estimation. Further, we offer a rich set of tools that ease the integration of other face databases and associated annotations into our joint framework.*

1. Motivation

Face recognition is an intrinsic part of human visual perception. The significance of face recognition for humans is reflected by the variety of applications of computational face recognition. Thus, face recognition is also one of the core tasks in computer vision. For instance, it builds the basis for many applications in biometrics such as access controls or video face spotting. Similar methods could also be applied as aid for visually impaired people, *i.e.*, humans that suffer from prosopagnosia (also called face blindness). In addition, currently of broad interest, the rapid evolve of consumer digital photography leads to loose unlinked personal photo collections, where face recognition algorithms could help to automatically organize collections.

	Raw	HOG [6]	Felz. [10]	LBP [1]
not aligned	60,85%	63,22%	65,53%	66,13%
aligned	61,80%	65,68%	68,43%	70,13%
+	**0,95%**	**2,47%**	**2,90%**	**4,00%**

Table 1: Importance of face alignment: Face recognition accuracy on Labeled Faces in the Wild [13] for different feature types – a face alignment step clearly improves the recognition results, where the facial landmarks are automatically extracted by a Pictorial Structures [8] model.

For humans the recognition of a familiar face is straight forward, it has even been observed that humans are able to recognize familiar faces in very low resolution images [28]. Computational face recognition algorithms are able to match the performance of humans in controlled environments. However, in unconstrained real-world situations imaging conditions as diversity in viewpoint, lighting, clutter or occlusion severely lower the recognition performance. Therefore, the study of face recognition under real-world conditions is the way to go. Several large-scale benchmark databases have been proposed exhibiting large variability [4, 9, 12, 13, 17]. Hereby, the study of face recognition is typically divided into three succeeding steps: detection, alignment and recognition (DAR).

Face detection means estimating the coarse location of a face in an image. Face alignment is the process of registering two or more faces relative to each other. Face recognition means deciding if two faces match (face verification) or the identification of a certain person in an image (face identification). It is obvious that face detection and the recognition step are important for accurate face recognition performance. However, many authors [5, 24, 25, 29] observed that in a DAR pipeline an alignment step is very valuable. It is assumed that better aligned faces give better recognition

*The work was supported by the FFG projects MDL (818800) and SECRET (821690) under the Austrian Security Research Programme KIRAS.

results. One reason is that the description has not to cope with geometric invariance, thus enabling a more powerful description.

This, is also confirmed by some experiments we carried out on the *Labeled Faces in the Wild* (LFW) [13] dataset. The corresponding results are illustrated in Table 1, where it can be seen that a face alignment step for different feature such as LBPs [1], HOGs [6], Felzenszwalb HOG-like features [10] and raw patches clearly improves the recognition results. Moreover, the ROC curve for LBPs in Figure 1 further emphasizes this finding.

Nevertheless, many face alignment methods require rather elaborate annotations. Only some of the many available face databases provide these, such as facial landmarks. However, in most cases these databases lack in several ways: First, they provide only a little number of annotated images or only sparse facial landmarks. Second, the databases are focused largely on frontal views of faces. Finally, the images are often captured under controlled conditions (uniform background, controlled lightning etc.) and therefore do not capture real-world problems.

Hence, the main motivation for the Annotated Facial Landmarks in the Wild (*AFLW*) database is the need for a multi-view, real-world face database for facial feature localization. The images of our database are collected on *Flickr*[1] exhibiting a large variety in pose, lightning, expression, ethnicity, age, gender, clothing, hairstyles, general imaging and environmental conditions. Further, the database offers various backgrounds and many other parameters. A wide range of images related to face relevant tags were gathered and manually processed. Therefore, as can be seen from Figure 6, the collection is not restricted to frontal faces.

The remainder of this paper is structured as follows. First, we provide an overview of related databases and discuss the main shared features as well as the main differences in Section 2. Next, we introduce the *AFLW* database in Section 3 and specify the intended uses following in Section 4. Finally, we give an overview on provided tools in Section 5 and summarize and conclude the paper in Section 6.

2. Related Databases

The huge interest in automatic face analysis can also be seen from the numerous face databases available publicly[2]. However, only a subset of these databases, which we summarized in Table 2, provides additional annotations such as facial landmarks. This number is even further reduced if multi-view faces or real-world imaging conditions would be required. For instance the popular benchmark dataset LFW [13] provides a huge set of real-world images that is gathered from news articles. Nevertheless, the faces are restricted to frontal poses and no annotations are provided. Other large-scale databases such as Caltech 10,000 Web Faces [2], CAS-PEAL Face Database [11] or the CMU / VASC [27] databases are designed for face detection. Therefore, they provide no or only a limited number of annotated landmarks. Databases with more annotated landmarks such as IMM [23] (58 landmarks, 240 images), MUG Facial Expression Database [22] (80 landmarks for a subset of 401 images) or AR Purdue [18] (22 point markup for 513 images, 130 for 897 images) provide only some hundreds of images.

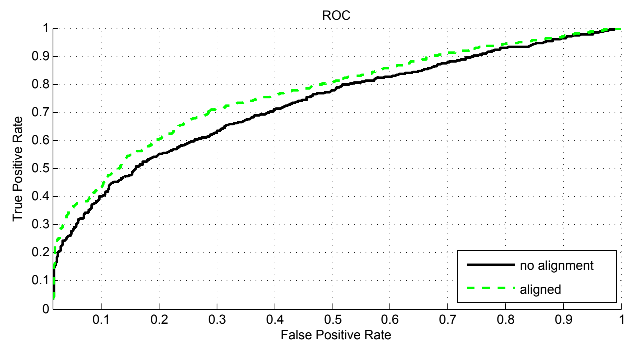


Figure 1: LBP receiver operating characteristic on Labeled Faces in the Wild [13]. An alignment step clearly improves the results. For the alignment facial landmarks are automatically extracted by a Pictorial Structures [8] model and are subject of a similarity transform least squares fit.

In the following, we discuss databases that are closely related ours in more detail:

The BioID Face Database [15] consists of 1521 gray level images at a resolution of 384×286 pixels. The images show frontal views of 23 subjects with slightly varying poses, expressions and some ad hoc modifications, *e.g.*, with and without glasses. The pictures were taken in an office environment with realistic background, although it stays constant for each subject. The initial eye position based markup scheme was extended by a 20 point markup scheme denoted in Figure 2 (a).

The XM2VTS data set [19] is intended to study the problem of multi-modal personal verification based on non-intrusive and user-friendly techniques such as speech recognition and face identification. The frontal image set, a subset of the audio-visual corpus, contains 2,360 color images at a resolution of 720×576 pixels. The images show frontal views of 295 individuals taken in 4 recording sessions. The markup scheme consists of 68 landmarks (Figure 2 (b)). The images were acquired with uniform background under constant imaging conditions. Subjects are not occluded and are mainly of Caucasian ethnicity.

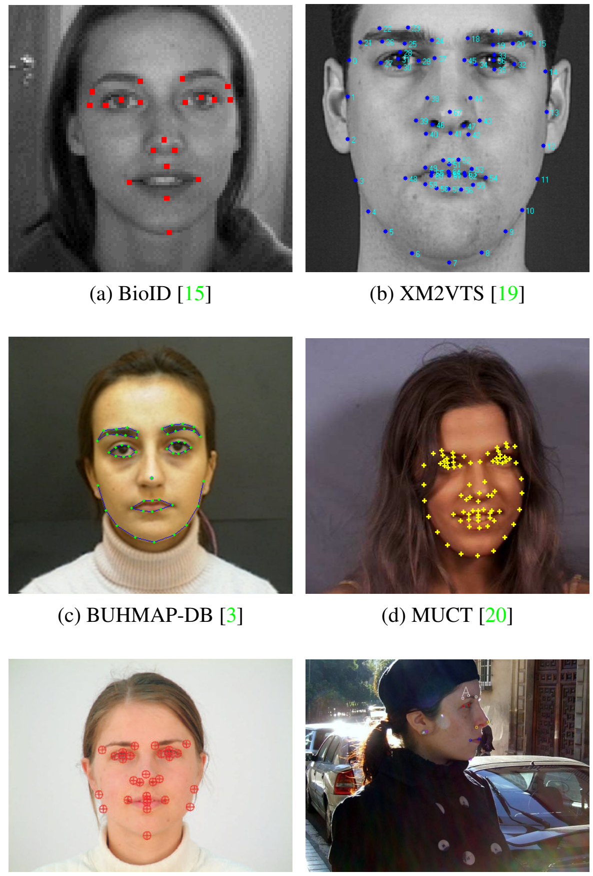
[1] http://www.flickr.com/
[2] For an overview see either [13] or the *http://www.face-rec.org/*

Boğaziçi University Head Motion Analysis Project Database (BUHMAP-DB) [3]. The database is intended to study Turkish Sign Language (TSL) and associated head/body motion and facial expressions. It involves 11 different subjects (6 female, 5 male) performing 5 repetitions of 8 different signs. In total the dataset consists of 440 videos. For a subset of 48 videos the dataset contains annotations of a 52 point markup (Figure 2 (c)). Roughly 2,880 images with a resolution of 640×480 are annotated. The images are taken under controlled conditions in a darkened room with constant, uniform background. Further, no subjects are occluded, have beards, mustaches or eyeglasses. The number of subjects is rather limited and also the ethnicity is restricted.

Milborrow / University of Cape Town (MUCT) Face Database [20]. The MUCT dataset provides 3,755 frontal faces with neutral expression or a smile at a resolution of 640×480 pixels, and the markup consists of 76 landmarks (Figure 2 (d)). One design goal was to provide more diversity of lighting, age and ethnicity compared to other datasets. In the image acquisition process controlled variation of lightning was introduced, up to three lightning sets per person. Further, the dataset contains a roughly equal number of males and females, with variation in age and ethnicity. Despite the introduced variation the dataset provides uniform background and no occlusions. The ethnic variation is predominately Caucasian and African.

Poznań University of Technology (PUT) Face Database [16]. The database contains 9,971 images of 100 subjects acquired at a resolution of 2048×1536 pixels. The intended use is the performance evaluation of face detection, facial landmark extraction and face recognition algorithms for the development of face verification methods. The authors argue that face pose is the main factor altering the face appearance in a verification system. Thus, the images were taken under controlled imaging conditions with uniform background showing various unconstrained face poses. The comprehensive set of annotations includes rectangles containing face and eyes and a set of 30 landmark points for all images (Figure 2 (e)). Further, for a subset of 2,193 near-frontal faces 194 control points are included. Despite the large-scale nature of the database and the comprehensive set of provided annotations, as a drawback, the images were acquired under controlled conditions with uniform background.

If we recapitulate the characteristics and properties of the described databases it is obvious that each collection serves several interesting properties dependent on the intended purpose. Nevertheless, we notice that there is no large-scale, multi-view collection of face images in the wild, annotated with facial landmarks!



(a) BioID [15] (b) XM2VTS [19]

(c) BUHMAP-DB [3] (d) MUCT [20]

(e) PUT [16] (f) AFLW

Figure 2: Comparison of different databases and their landmark positions. *AFLW* provides less landmarks per image than other databases, however, it is the only database taken under real-world conditions.

3. The Annotated Facial Landmarks in the Wild Database

The motivation for the *AFLW* database[3] is the need for a large-scale, multi-view, real-world face database with annotated facial features. We gathered the images on *Flickr* using a wide range of face relevant tags (*e.g.*, *face*, *mugshot*, *profile face*) to collect the images. Due to the real-world nature of Flickr the images exhibit a large variety in pose, lightning, expression, ethnicity, age, gender, clothing, hairstyles, general imaging and environmental conditions. Further, the set of images was manually scanned for images containing faces. Thus, the collection, which is illustrated in Figure 6, captures typical real-world scenarios. The key data and most important properties of the database are:

[3] http://lrs.icg.tugraz.at/research/aflw/

Database	# landmarked imgs.	# landmarks	# subjects	image size	image color	Ref.
Caltech 10,000 Web Faces	10,524	4	-	-	color	[2]
CMU / VASC Frontal	734	6	-	-	grayscale	[26]
CMU / VASC Profile	590	6 to 9	-	-	grayscale	[27]
IMM	240	58	40	648x480	color/grayscale	[23]
MUG	401	80	26	896x896	color	[22]
AR Purdue	508	22	116	768x576	color	[18]
BioID	1,521	20	23	384x286	grayscale	[15]
XM2VTS	2,360	68	295	720x576	color	[19]
BUHMAP-DB	2,880	52	4	640480	color	[3]
MUCT	3,755	76	276	480x640	color	[20]
PUT	9,971	30	100	2048x1536	color	[16]
AFLW	**25,993**	**21**	-	-	**color**	

Table 2: Face databases with annotated facial landmarks.

- The database contains 25,993 faces in 21,997 real-world images, with realistic background. Of these faces 56% are tagged as female, 44% are tagged as male; some images contain multiple faces. No rescaling or cropping has been performed. Most of the images are color although some of them gray-scale.
- In total *AFLW* contains 389,473 manually annotated facial landmarks of a 21 point markup. The facial landmarks are annotated upon visibility. So no annotation is present if a facial landmark, *e.g.*, left ear lobe, is not visible.
- A wide range of natural face poses is captured The database is not limited to frontal or near frontal faces. To the best of our knowledge the ratio of non-frontal faces (66%) is higher than in any other database.
- Additional to the landmark annotation the database provides face rectangles and ellipses. Further, the face ellipses support the FDDB [14] evaluation protocol.
- A rich set of tools to work with the annotations is provided, *e.g.*, a database backend that enables to import other face collections and annotation types. For popular databases that provide facial landmarks or benchmark databases such as BioID [15], CMU / VASC profile [27] the importers are already included.

To recapitulate, *AFLW* contains more diversity and variation than any other face database with annotated facial landmarks. Further due to the nature of the database and the comprehensive annotation it is well suited to train and test algorithms for

- facial feature localization
- multi-view face detection
- coarse head pose estimation.

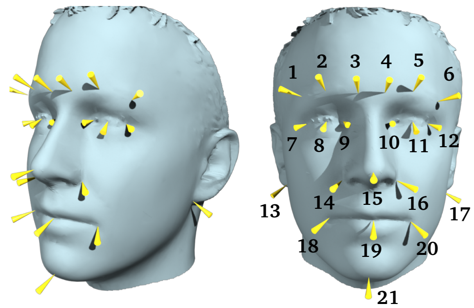
4. Intended Uses

The intended uses of our *AFLW* database are threefold. First, multi-view face detection under real-world conditions. Second, facial feature localization to support face recognition, face alignment or to train local detectors or descriptors. Third, face pose estimation to support, *e.g.*, face tracking. An important difference to many other databases is that our database not only suited for testing and evaluation, but also for training.

4.1. Facial Landmark Localization

Facial landmarks are standard reference points, such as the inner and outer corner of the eye fissure where the eyelids meet. In many cases the landmarks used in computational face analysis are very similar to the anatomical soft-tissue landmarks used by physicians. The task of automatically localizing these landmarks is beneficial for various reasons. For instance, an efficient estimate of the head pose can be obtained [21] with only some landmarks. Moreover, facial landmarks can be used to align faces to each other, which is valuable in a detection, alignment and recognition pipeline; better aligned faces give better recognition results. Further, we can extract properties that have a local nature such as face attributes (*e.g. bushy eyebrows*, *skin color*, *mustache*) [17], local descriptors [8] or to train flexible part-based detectors [31]. Nevertheless, facial landmark localization in unconstrained real-world scenarios is still a challenging task. Some face databases provide facial landmarks, however, they lack at least in some ways: For instance the images were acquired under controlled conditions, are limited to more or less frontal views or the variety in ethnicity is rather limited.

The landmark positions of *AFLW* are defined on a rigid 3D face model denoted in Figure 3. We use a markup of 21 reference landmarks mainly located in the area



(a) Side View (b) Frontal Face

Figure 3: The AFLW markup defines 21 facial landmarks that are located between eyebrows and chin.

ID	Description	Count
1	Left Brow Left Corner	16,545
2	Left Brow Center	20,624
3	Left Brow Right Corner	21,764
4	Right Brow Left Corner	21,979
5	Right Brow Center	20,790
6	Right Brow Right Corner	16,751
7	Left Eye Left Corner	19,461
8	Left Eye Center	21,439
9	Left Eye Right Corner	18,183
10	Right Eye Left Corner	17,877
11	Right Eye Center	21,873
12	Right Eye Right Corner	19,569
13	Left Ear	10,885
14	Nose Left	18,217
15	Nose Center	25,993
16	Nose Right	18,647
17	Right Ear	11,684
18	Mouth Left Corner	20,482
19	Mouth Center	25,448
20	Mouth Right Corner	21,262
21	Chin Center	24,641
		389,473

Table 3: Overview of landmark annotations in AFLW. The number of individual annotations ranges from 10,885 (left ear) to 25,993 (nose center).

between eyebrows and chin. Starting at the forehead three landmarks are located at each eyebrow, on the leftmost, rightmost and medial point. Each eye area is covered by further three landmarks. The inner and outer corner of the eye fissure where the eyelids meet (endocanthion, exocanthion) and the medial point. On the external nose we specified the left and right point of attachment of the nose cavity with the face (nasal alar crest) and the tip of the nose (pronasale). On the external ear we mark the lowest point of attachment to the head (otobasion inferius). On the mouth and lips the landmarks are placed on the left and right intersection point of the lips (cheilion) and the mouth center as medial point. Finally, on the chin the lowest point on the lower border (gnathion) is selected.

In the annotation process landmarks are marked upon visibility. So if a landmark is not visible it is simply not annotated. In total 230,189 landmarks have been annotated so far. For individual landmarks the number of annotations ranges from 6,203 (left ear) to 15,677 (nose center). Please see Table 3 for detailed statistics.

4.2. Face Pose Estimation

Head pose estimation in images captured under uncontrolled conditions in natural environments is still a challenging task. Some of the databases mentioned above include ground-truth pose information, some contain images taken under uncontrolled and wide variations of conditions, but none of them feature both properties and can thus be regarded as a valid benchmark for this task.

Our database comes with approximate pose information for each face, derived from the annotated facial landmarks. To this end, we fit a mean 3D model [30] of the frontal part of a head (shown in Figure 3) to the annotated points in the image. The model is constructed by averaging over a set of 3D head scans. The pose parameters are adjusted, to minimize the distance between the projections of the corresponding points on the 3D model and the actual landmark locations in the image in a least squares manner. This is achieved by feeding the set of corresponding 3D and 2D points into our own implementation of the POSIT algorithm [7]. The resulting pose is stored in the database in terms of roll, pitch and yaw angles as depicted in Figure 4. Note that this estimated pose is coarse and not manually verified for all face instances, but nevertheless gives a valid ground-truth for approaches trying to find a rough approximation of the pose.

Further, the extracted pose estimate can be used to retrieve images from a limited range of poses only. This can be used to train sets of individual, pose dependent face detectors. Another possible application is the analysis of person independent relations between a given image representation and controlled variations in the pose.

4.3. Multi-View Face Detection

In the beginning, face detection methods focused on detecting frontal views of faces, looking straight into the camera. This task can essentially be regarded as solved, since there already exist satisfactory implementations in consumer products. However, multi-view face detection in uncontrolled environments is still a challenge.

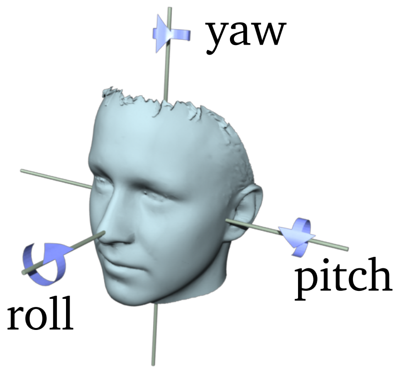


Figure 4: The pose of the head is described in form of the three rotation angles yaw, pitch and roll.

Most of todays face detectors represent their detection output by a rectangle around the detected face. While there is largely an agreement on how to define an anchor point and extents of the rectangle for frontal faces, it is not so obvious how to define them for profile and semi-profile views, considering the larger variation of face profile shapes. This makes it harder to get consistently annotated samples.

Some more recent detection methods indicate the in-plane rotation by rotating the rectangle or by an arrow that depicts the viewing direction. Others return the exact locations of facial landmarks in the image or even an object centered coordinate system. This multitude of possible detector outputs leads to a problem in comparative evaluations of the methods. For each of them, the test data would have to be annotated with its corresponding ground-truth data format.

Recently, a reasonable compromise was presented in [14]. All faces are annotated by an ellipse outlining the 3D ellipsoid capturing the front of the head. This gives a closer boundary of the region of interest and in-plane rotation information. While this representation captures most of the important information, it is generally not too hard to convert more complex detection outputs to this format. The major contribution of FDDB [14] is that it provides an evaluation protocol that specifies how the output of a detector should be evaluated against ground truth.

Hence, we propose a method to automatically generate the ellipses from facial landmark annotations. From the pose estimation procedure described in Section 4.2, we have an alignment of a mean 3D face model with the annotations. We define the center of the ellipse as the projection of a specified center point in the 3D model, into the image. The scale is estimated by the distance from the center to the projection of another fixed 3D point. The orientation can analogously be derived from the calculated pose.[4] The aspect ratio is kept fixed at 3:2. The result of this process is demonstrated in Figure 5.

Figure 5: Face ellipses automatically created from the annotated facial landmarks, following the specification in the FDDB evaluation framework.

Generally, also in cases when the pose estimation is not very accurate, because the mean 3D model cannot be well aligned with the depicted face, the resulting ellipse is very close to what the annotation guidelines in [14] specify and certainly within the range of variance of human annotators. Thus, our database is ready to be used in the FDDB evaluation protocol. Another advantage of this approach is that face ellipses can also be calculated easily for other databases with annotated facial landmarks.

5. Data and Tools

One of the main goals of this paper is to support the use of data coming from different face data collections in a joint framework. The collections are often associated with valuable annotations, though these come in different formats. Thus, along with our own collection of images and annotation data we provide a set of tools to view and manipulate them and import data from other collections.

5.1. SQLite Database

For individual data collections plain text file based annotations are suitable. Nevertheless, if different datasets are combined and used in a joint framework a common place to store the different annotations is beneficial. Thus, we provide a simple SQLite database to collect the annotations. Due to the features of a relational database this enables easy management of the annotations. Writing an SQL query needs by far less effort than writing traditional code that parses several text files and selects some options with never ending `if then else` statements. For instance to train a face detector it is easily possible to query for faces in certain poses, sizes, with specific visible landmarks and of course constrained to one or more databases.

In our SQLite database the central tables are `Faces` and `FaceImages`. These store the mapping to the faces asso-

[4] Unfortunately, the angle θ in FDDB is specified as the angle between the major axis of the ellipse and the x-axis, ranging from $-\pi$ to π, making it impossible to specify that a face is upside down. Since we have this information in our representation we store it in our database in an "upside-down" flag.

ciated database and the actual path to the image with some meta data. The facial landmark annotation resides in the table table `FeatureCoords`. Since every database comes with its own list of landmark definitions, a mapping to this shared list of landmarks has to be defined manually for each face database (table `FeatureCoordTypes`). For instance, to query for faces with annotated *ChinCenter* one has to select the records in `FeatureCoords` where the ID is 21.

The table `FacePose` holds the calculated yaw, pitch and roll angles of the head pose estimation in Sec. 4.2. Thus, it enables to query for a subset of the available face poses or to analyze the face pose distribution of a database. To query for some meta data of the face such as gender, or *e.g.* if the person wears eyeglasses, one can use the table `FaceMetaData`.

Annotation data in form of an ellipse (*e.g.*, to follow the FDDB protocol) as explained in Section 4.3, either manually annotated or calculated from facial landmark annotations, is stored in `FaceEllipse`. Rectangle annotations from databases such as CMU/VASC Profile are stored in `FaceRect`.

5.2. Label GUI

The cross-platform Label GUI provides an easy way to view and edit facial landmark annotations, as well as a set of meta informations about the annotated face. It is directly connected to the SQLite database.

5.3. Programming Tools

We provide a set of tools for easy access to the database. The `facedblib` defines the C++ interface. A generic SQL-statement class and an SQL-connection wrapper can be used to define and extend specific database queries and store the resulting data in a set of predefined face data structures, for further use in an application. On top of these core components, we built two exemplary implementations for importing meta data and annotations for our own, as well as for the CMU/VASC Profile [27] database.

The `facedbtool` application is meant as a central tool to inspect and manipulate the database. Up to now, it features displaying of annotation data for a selected face, pose estimation, face ellipse calculation and identification of possible duplicate annotations in the database.

Access to the database from MATLAB is possible through `mksqlite`[5]. A set of scripts demonstrating how to access the database to retrieve and display image data and annotations are included in our code. Also the script to export the ground-truth file for the FDDB evaluation protocol was implemented this way.

[5]http://mksqlite.berlios.de/

Figure 6: Impressions of the Annotated Facial Landmarks in the Wild (AFLW) database. The database provides variation in pose, ethnicity, realistic background and natural uncontrolled imaging conditions.

6. Conclusion

In this paper, we introduced the *Annotated Facial Landmarks in the Wild (*AFLW*)* database. *AFLW* provides a large-scale, real-world collection of face images, gathered from Flickr. Compared to other databases *AFLW* is well suited to train and test algorithms for multi-view face detection, facial landmark localization and face pose estimation. Since attribute-based face recognition approaches showed reasonable performance recently, further work will include to extend the database by facial attributes. Especially, because there are no databases available publicly for that task by date.

7. Acknowledgements

We want to thank all people who have been involved in the annotation process, especially, the interns at the institute and the colleagues from the Documentation Center of the National Defense Academy of Austria.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face Recognition with Local Binary Patterns. In *Proc. ECCV*, 2004. 1, 2

[2] A. Angelova, Y. Abu-Mostafam, and P. Perona. Pruning training sets for learning of object categories. In *Proc. CVPR*, 2005. 2, 4

[3] O. Aran, I. Ari, M. A. Guvensan, H. Haberdar, Z. Kurt, H. I. Turkmen, A. Uyar, and L. Akarun. A database of non-manual signs in turkish sign language. In *Proc. Signal Processing and Communications Applications*, 2007. 3, 4

[4] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D. A. Forsyth. Names and faces in the news. In *Proc. CVPR*, 2004. 1

[5] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In *Proc. CVPR*, 2010. 1

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. CVPR*, 2005. 1, 2

[7] D. F. Dementhon and L. S. Davis. Model-based object pose in 25 lines of code. *Int. Journal of Computer Vision*, 15:123–141, 1995. 5

[8] M. Everingham, J. Sivic, and A. Zisserman. "Hello! My name is... Buffy" – automatic naming of characters in TV video. In *Proc. BMVC*, 2006. 1, 2, 4

[9] L. Fei-Fei, R. Fergus, and P. Perona. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. In *Proc. CVPR Workshop on Generative-Model Based Vision*, 2004. 1

[10] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. on PAMI*, 32(9):1627 –1645, 2010. 1, 2

[11] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The CAS-PEAL large-scale chinese face database and baseline evaluations. *IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans*, 38(1):149–161, 2008. 2

[12] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007. 1

[13] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007. 1, 2

[14] V. Jain and E. Learned-Miller. FDDB: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010. 4, 6

[15] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz. Robust face detection using the Hausdorff distance. In *Proc. Audio and Video-based Biometric Person Authentication*, 2001. 2, 3, 4

[16] S. A. Kasiński A., Florek A. The PUT face database. *Image Processing & Communications*, pages 59–64, 2008. 3, 4

[17] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In *Proc. ICCV*, 2009. 1, 4

[18] A. Martínez and R. Benavente. The AR face database. Technical Report 24, Computer Vision Center, University of Barcelona, 1998. 2, 4

[19] K. Messer, J. Matas, J. Kittler, and K. Jonsson. XM2VTSDB: The extended M2VTS database. In *Proc. Audio and Video-based Biometric Person Authentication*, 1999. 2, 3, 4

[20] S. Milborrow, J. Morkel, and F. Nicolls. The MUCT Landmarked Face Database. In *Proc. Pattern Recognition Association of South Africa*, 2010. 3, 4

[21] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Trans. on PAMI*, 31:607–626, 2009. 4

[22] C. P. N. Aifanti and A. Delopoulos. The MUG facial expression database. In *Proc. Workshop on Image Analysis for Multimedia Interactive Services*, 2005. 2, 4

[23] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann. The IMM face database - an annotated dataset of 240 face images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2004. 2, 4

[24] E. Nowak and F. Jurie. Learning visual similarity measures for comparing never seen objects. In *Proc. CVPR*, 2007. 1

[25] S. Prince, P. Li, Y. Fu, U. Mohammed, and J. Elder. Probabilistic models for inference about identity. *IEEE Trans. on PAMI*, 2011. 1

[26] H. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. Technical Report CMU-CS-97-201, Computer Science Department, Carnegie Mellon University (CMU), 1997. 4

[27] H. Schneiderman and T. Kanade. A statistical model for 3D object detection applied to faces and cars. In *Proc. CVPR*, 2000. 2, 4, 7

[28] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 2006. 1

[29] J. Sivic, M. Everingham, and A. Zisserman. "Who are you?" Learning person specific classifiers from video. In *Proc. CVPR*, 2009. 1

[30] M. Storer, M. Urschler, and H. Bischof. R3D-MAM: 3D morphable appearance model for efficient fine head pose estimation from still images. In *Proc. ICCV Workshop on Subspace Methods*, 2009. 5

[31] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. on PAMI*, 1997. 4