IET Biometrics

*Research Article*

IET Journals
The Institution of Engineering and Technology

# ForenFace: a unique annotated forensic facial image dataset and toolset

ISSN 2047-4938
Received on 2nd November 2016
Revised 31st March 2017
Accepted on 16th May 2017
E-First on 18th July 2017
doi: 10.1049/iet-bmt.2016.0160
www.ietdl.org

*Chris G. Zeinstra[1] ✉, Raymond N.J. Veldhuis[1], Luuk J. Spreeuwers[1], Arnout C.C. Ruifrok[2], Didier Meuwly[1,2]*

[1]*University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*
[2]*Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands*
✉ *E-mail: c.g.zeinstra@utwente.nl*

**Abstract:** Few facial image datasets are suitable for forensic research. In this study, the authors present ForenFace, a facial image and video dataset. It contains video sequences and extracted images of 97 subjects recorded with six different surveillance camera of various types. Moreover, it also contains high-resolution images and 3D scans. The novelty of this dataset lies in two aspects: (i) a subset of 435 images (87 subjects, five images per subject) has been manually annotated, yielding a very rich forensically relevant annotation of almost 19.000 facial parts, and (ii) making available a toolset to create, view, and extract the annotation. The authors present protocols and the result of a baseline experiment in which two commercial software packages and an annotated facial feature contained in this dataset are compared. The dataset, the annotation and tools are available under a usage license.

## 1 Introduction

In forensic evaluation, trace material may for example consist of facial images extracted from CCTV footage taken at a crime scene and reference material may be (high quality) mugshots or 3D scans. During the comparison process, the forensic facial practitioner particularly pays attention to anthropomorphical features [1, 2]. The Facial Identification Scientific Working Group (FISWG) [3] has published recommendations for this process [4]. In particular, their Image Comparison Feature List for Morphological Analysis [5] contains characteristic descriptors (facial features) that can be taken into account. The nature of these descriptors ranges from broad and qualitative to narrow and quantitative. As an example, we show in Fig. 1 three vertical differences. In this example, A refers to the difference between inner and outer eyebrow tips, B refers to the difference between the outer eyebrow tip and outer eye corner, and C refers to the difference between the inner eye corner and the lowest point on the eyebrow outline in the vicinity of the inner eye corner.

The comparison process is largely manual, making it worthwhile to investigate whether biometric classifiers can assist the practitioner. However, quality of trace material is typically limited by technological and subject factors. Technological factors include image compression artefacts, perspective effects, low resolution, and interlacing. Subject factors include pose, illumination, expression, and partial occlusion of the face by hoodies or balaclavas. Therefore, it is not always possible to use 'off-the-shelf' classifiers that have been developed for a specific biometric non-forensic application. A different approach is to use classifiers (or rather the evidential value derived from their comparison score) that are specialised in a particular facial part or a set of characteristic descriptors. For example, such classifiers using the FISWG characteristic descriptors of the eyebrow and eye region have been studied by Zeinstra *et al.* [6, 7]. In Tome *et al.* [8, 9] results on automatic classifiers on forensic regions and shapes are presented.

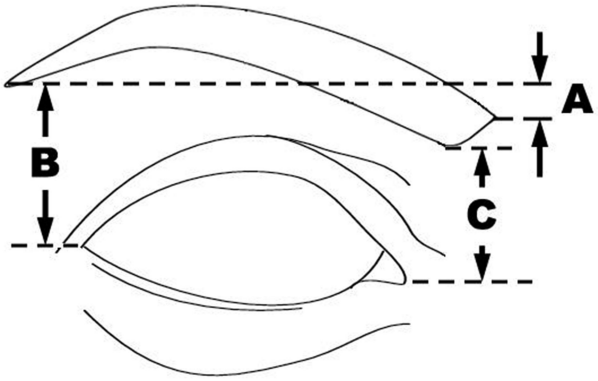


**Fig. 1** *Some FISWG eyebrow characteristic descriptors that capture the vertical position of the eyebrow, from [5]*

For the development and testing of such classifiers the availability of datasets that are representative of forensic trace and reference material is of paramount importance. This can be observed from the related field of automatic face recognition. This field has grown from initial work by Kanade [10] and Turk and Pentland [11] into a well-established, mature, and wide field of research. Many face recognition systems have been developed and successfully deployed in real world use case scenarios. A key success factor has been the availability of public facial image datasets (e.g. FRGC [12]) and vendor challenges using those datasets [13–15]. Initially, those datasets mainly consisted of images acquired under controlled conditions, but gradually there has been a shift towards sets acquired under uncontrolled, more realistic 'in the wild', circumstances. Examples are Labeled Faces in the Wild [16], HELEN [17], and Quis-Campi [18].

However, to date, the number of facial image datasets that are suitable for forensic research is limited. Even within the group of forensic type datasets, not every dataset is suitable for forensic evaluation of trace and reference material as described before. We identify three criteria that in our opinion determine the suitability of such a specific dataset. They are:

1. Representativeness of trace material.
2. Representativeness of reference material.
3. Availability of forensic features.

Representativeness refers to being typical of images encountered in forensic evaluation. With respect to the first criterion, real trace material typically consists of CCTV video

**Table 1** Contents of datasets

| Dataset | Subjects | Material | Forensic features |
|---|---|---|---|
| SCFace [19] | 130 | traces: seven surveillance cameras, three distances, one close-up surveillance | four landmarks |
| | | references: 5 images | |
| Chokepoint [20] | 29 | footage of three surveillance cameras | eye coordinates (per frame) |
| NIST Mugshot [21] | 1573 | grey scale mugshot | none |
| Morph (Academic) [22] | 13.618 | scanned and digital mugshots | eye coordinates |
| ATVS Forensic DB [23] | 50 | high-resolution, three distances | 21 landmarks |
| FRGC [12] | 568 | frontal images and 2.5D scans, taken under (un)controlled conditions with neutral or smiling subjects | 4 landmarks |
| Labeled Faces in the Wild [16] | 5749 | unconstrained face images | Identity label |
| HELEN [17] | 2330 | unconstrained face images | 199 landmarks |
| Quis-Campi [18] | 320 | traces: videos and images | eye coordinates (per frame) |
| | | references: registration images, gait, 3D model face | |
| ForenFace | 97 | traces: CCTV video and stills from six surveillance cameras of visible and partially occluded subjects | annotated facial parts |
| | | references: five images, 3D scan | |

footage and extracted stills of subjects that may have occluded parts of their face. The quality of trace material can vary between cases and depends for example on resolution and physical placement of the camera. Representativeness of reference material, the second criterion, are often high-resolution frontal, quarter profile, and profile images, and sometimes 3D scans are employed as well. Finally, the third criterion, which we believe is the most important one in the context of forensic evaluation, is the availability of forensic features that are typically used by a forensic facial practitioner. Exactly these features can be used to train and test specialised biometric classifiers.

In this paper, we introduce ForenFace, a forensic dataset designed with these three criteria in mind. It includes very rich manual annotation from which forensically relevant features can be extracted.

We note that some of the included annotations in some use cases might have been obtained by a computer vision algorithm. However, in general the poor image quality of trace material restricts the usability of such approaches. Moreover, facial part definitions are not always easily captured in an algorithm. For example, the proper detection of facial lines can be difficult.

Manual annotation is a very resource intensive process. Therefore, we restrict the annotation to three different forensic use cases that as a whole are representative of forensic case work. We define a forensic use case as a criminal act whose traces consist of distinct facial image types. The first very common forensic use case is a money robbery at a bank, shop or gas station. At those premises often CCTV surveillance cameras are mounted on a wall or ceiling. Since this is such an important use case, we have annotated two images of different resolution and illumination. Another use case is money withdrawal from an ATM using a stolen debit card. Here, the trace material is recorded by a small camera, often mounted near the keypad. The final use case that we address is when a customs or immigration officer suspects that the used identity document has not been tampered with; however, does not correspond to the person who is presenting it. These forensic use cases correspond to specific images. In each use case, the reference material consists of a high-resolution frontal image and its annotation is compared to annotation on trace images. We refer to this as the annotation scenario.

Although the annotation scenario forms the main reason of this dataset, there are several other research scenarios possible in which the annotation is not employed; however, for which this dataset is still of interest. We mention two of them here, other uses are discussed in Section 5.1. The first scenario is one in which stills extracted from video sequences are compared to a 3D image for forensic investigation. In the second scenario, two video sequences are compared to investigate whether the videos contain the same person. We present evaluation protocols for all three scenarios in Section 5.

In Table 1 we compare nine publicly available image datasets that can used in forensic research with the ForenFace dataset. Although the SCFace dataset has its merits and been used in numerous publications on low resolution face recognition, traces only consist of frontal surveillance camera images. The ChokePoint dataset is designed for 'person identification/verification under real-world surveillance conditions'. Since it does not contain reference images, it is not suitable enough for research within a forensic context. The NIST and Morph datasets only contains mugshots, and are mainly suited for longitudinal research. The ATVS Forensic DB only contains high-resolution mugshots. FRGC has been used in numerous face biometric studies, but it somewhat limited in its forensic relevance. Labeled Faces in the Wild is widely used to evaluate modern face recognition algorithms that can cope with uncontrolled settings. HELEN is mainly used for the training and evaluation of facial feature localisation algorithms on images taken under non-ideal conditions. Finally, Quis-Campi contains videos and stills taken from modern surveillance systems that typically have a higher resolution than those acquired by traditional systems. However, Quis-Campi lacks a good set of reference images.

The ForenFace dataset also comes with a set of three software tools. One tool can be used for the viewing and creation of manual annotation. Another tool can be used to setup a new dataset for annotation, and the third tool can be used to extract annotation in a flexible manner. The software tools are usable on any platform for which a Java virtual machine is available.

The dataset, annotation, and the toolset are available under a usage license. This usage license encompasses a privacy policy/statement, the right to use this dataset for research purposes, and the requirement to cite this paper whenever it is used in published research. A user guide is provided that contains any necessary details. More information can be found at [24].

The structure of this paper follows the three constituent parts of the ForenFace dataset. In Section 2 we present the data, in Section 3 we discuss the included annotation, and in Section 4 we show the accompanying toolset. In Section 5 we discuss potential uses, propose evaluation protocols, and give a baseline result. Finally, in Section 6 we present our conclusion.

## 2 Data

### *2.1 Video sequence acquisition*

Data acquisition took place at the Netherlands Forensic Institute in The Hague, Netherlands, over a period of four days. The location is an open space between the staircase, offices, and a corridor. The arrangement of the six surveillance cameras is depicted in Fig. 2. Fig. 3 is a still image extracted from Camera 3 footage showing a subject on position C. The location and pose parameters $(p_1, p_2)$ of the cameras are shown in Table 2. Here, the parameter $p_1$ denotes

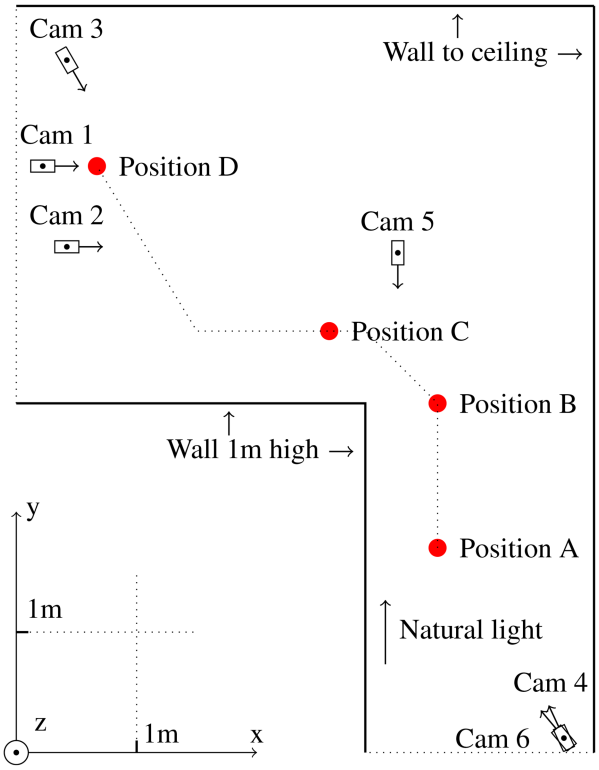


**Fig. 2** *Abstract top view of the layout of the experiment showing camera locations, subject positions, and subject paths. Objects present at the physical setup like tables and a closet are omitted for clarity*

**Fig. 3** *CCTV footage from Camera 3 when subject is standing at position B wearing a baseball cap. Camera 2 can be seen at the lower right-hand side corner and the pole on which Cameras 4 and 6 are attached is visible at the top, slightly left-hand side from the middle*

**Table 2** Surveillance cameras setup

| Camera | 3D coordinates | Pose (°) |
|---|---|---|
| Camera 1 | (0.22, 4.87, 1.00) | (90, 45) |
| Camera 2 | (0.42, 4.20, 1.60) | (90, 0) |
| Camera 3 | (0.42, 5.75, 2.40) | (150, −25) |
| Camera 4 | (4.55, 0.12, 2.00) | (335, −25) |
| Camera 5 | (3.17, 4.15, 1.60) | (180, 0) |
| Camera 6 | (4.55, 0.12, 2.60) | (320, −50) |

the compass bearing (clockwise from North = 0°) and assumes that the positive $y$-axis corresponds to the Northern bearing (Table 3). The parameter $p_2$ is the angle with the constant $z$-plane. A positive (resp. negative) angle means the camera is pointing upwards (resp. downwards). The camera types are shown in Table 4. In addition, the location of positions A-D are shown in Table 3.

The location was artificially illuminated. Using the compass bearing described in Table 2, natural light was coming out of the 180° direction. Subjects were asked to (i) stand at position A facing 0°, look around, (ii) stand at position B facing 0°, (iii) stand at position C facing 270°, look around, (iv) stand at position D facing

**Table 3** Positions A-D

| Position | 2D coordinates |
|---|---|
| Position A | (3.50, 1.70) |
| Position B | (3.50, 2.90) |
| Position C | (2.60, 3.50) |
| Position D | (0.67, 4.87) |

**Table 4** Surveillance camera types

| Camera | Model | Type |
|---|---|---|
| 1 | Watec WAT-230A | BW pinhole |
| 2 | Sanyo VCC-6580P | narrow angle |
| 3 | Panasonic WVP480 | wide angle |
| 4 | Vista VEC30H-DN | low light |
| 5 | Sony SSC-D372 | narrow angle |
| 6 | Dallmeier DDF3000A | dome |

270°, look down into camera 1, look around, look up into camera 3, (v) stand at position C facing 90°, look around, (vi) stand at position B facing 180°, look around, and finally (vii) stand at position A facing 180°, look around. This procedure was executed twice (with/without baseball cap) and leads to 12 video sequences. Frontal facial images were extracted and a selection is shown in Fig. 4. A selection of reference data is shown in Fig. 5.

### *2.2 3D scan and other image acquisition*

In an adjacent room three (half profile/half frontal left and right, and frontal) 3D scans were obtained using a Minolta VIVID910 scanner, after which they were merged into one collection of polygons (ply format). Five (profile left and right, half profile/half frontal left and right, and frontal) reference images were acquired by a Canon EOS 10D. Finally, the identity document type images were taken from employee cards. These passport style photographs were taken several months or years before.

### *2.3 Image and scan contents*

For each subject a number of video sequences, images, and 3D scans are available. Details are given in Tables 5 and 6. In Table 5, ≪ sid ≫ refers to the subject id, and ≪ camera ≫ to the camera number ∈ {1, …, 6}. Also, a (resp. b) refers to footage and images in which the subject does not wear (resp. does wear) a baseball cap. Finally, IPD is the interpapillary distance measured in pixels. In addition, in Table 6 in the Canon EOS 10D entry, $f$ refers to frontal, $p$ to profile (right/left), and $q$ to quarter profile (right/left).

The video sequences were converted from a Dallmeier proprietary format to MPEG4 by using the PStream Convert conversion tool [25]. The Dallmeier SMAVIA viewer software [25] was used to manually select and extract still images from the CCTV footage. The 3D scans can be viewed with several open source software packages, such as MeshLab [26].

## 3 Annotation

### *3.1 Forensic features*

As indicated in the introduction, forensic facial practitioners use anthropomorphical features during forensic case work. We have selected a large set of these features to be included in this dataset, which are presented in Figs. 6 and 7. Most FISWG characteristic descriptors are contained in or can be determined from this set. For example, with respect to Fig. 1, the eyebrow shape is contained in the set, the A position can be determined from the eyebrow shape, whereas the B and C positions can be determined from the eyebrow and fissure outline.

**Fig. 4** *From top left clockwise, stills from camera (subject position): 1 (D), 2 (B), 4 (C), 3 (D), 6 (A), 5 (B), 4 (B), and 3 (B)*

**Fig. 5** *From top left clockwise: identity document, frontal reference, 3D scan, and half profile reference*

**Table 5** Available video sequences and extracted images

| Source | Description | Format | Avg. IPD (px) | # Wearing no cap/cap |
|---|---|---|---|---|
| Camera 1–6 | Video sequence | ≪ sid ≫c ≪ camera ≫ {a,b}.mpeg | N/A | 97/97 |
| Camera 1 | Position D | ≪ sid ≫ c1{a,b}7.bmp | 65 | 89/86 |
| Camera 2 | Position B | ≪ sid ≫ c2{a,b}3.bmp | 27 | 97/96 |
| Camera 3 | Position B | ≪ sid ≫ c3{a,b}3.bmp | 11 | 97/97 |
| | Position C | ≪ sid ≫ c3{a,b}8.bmp | 15 | 97/97 |
| | Position D | ≪ sid ≫ c3{a,b}16.bmp | 38 | 90/90 |
| Camera 4 | Position C | ≪ sid ≫ c4{a,b}2.bmp | 13 | 97/97 |
| | Position B | ≪ sid ≫ c4{a,b}7.bmp | 15 | 97/96 |
| | Position A | ≪ sid ≫ c4{a,b}12.bmp | 23 | 95/95 |
| Camera 5 | Position B | ≪ sid ≫ c5{a,b}3.bmp | 68 | 97/97 |
| Camera 6 | Position A | ≪ sid ≫ c6{a,b}3.bmp | 38 | 93/94 |

**Table 6** Other images and 3D scans

| Source | Description | Format | Avg. IPD (px) | # |
|---|---|---|---|---|
| Canon EOS 10D | Reference | ≪ sid ≫{f, lp, lq, rp, rq}.jpg | 370 | 97 |
| Unknown camera | Identity document | ≪ sid ≫a.jpg | 35 | 97 |
| Minolta | 3D scan | ≪ sid ≫.ply | N/A | 93 |

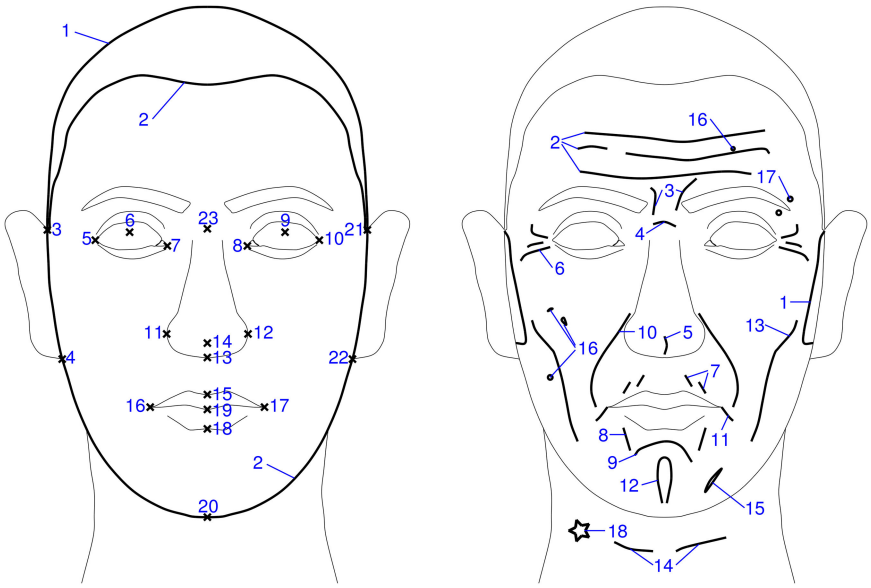


**Fig. 6** *Left-hand side: Face from a holistic perspective. Right-hand side: face from detailed perspective. Prefix H (resp. D) refers to holistic (resp. detailed) perspective. (H1) Cranial Vault, (H2) Shape of Face, (H3-H23) 21 landmarks (upper/lower connection ears to face (H3, H4, H21, H22), inner/outer corners eyes (H5, H7, H8, H10), pupils (H6, H9), alae (H11, H12), below nose (H13), nose tip (H14), upper/lower lip (H15, H18), mouth corner (H16, H17), mouth (H19), chin (H20), and nasal root (H23). (D1) Facial Hair Outline, (D2) Forehead Creases, (D3) Vertical Glabellar, (D4) Nasion Crease, (D5) Bifid Nose Crease, (D6) Periorbital Creases, (D7) Upper Circumoral Striae, (D8) Lower Circumoral Striae, (D9) Mentolabial Sulcus, (D10) Nasolabial Creases, (D11) Marionette Lines, (D12) Cleft Chin, (D13) Buccal Creases, (D14) Neck wrinkles, (D15) Scars, (D16) Facial Marks, (D17) Piercing, and (D18) Tattoo*

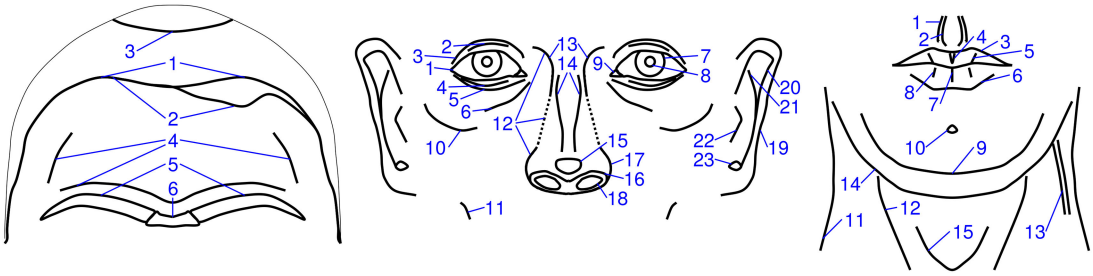

**Fig. 7** *Left-hand side: Upper facial parts. Middle: Middle facial parts. Right-hand side: lower facial parts.Prefix U (resp. M and L) refers to the upper (resp. middle and lower) parts. (U1) Forehead hairline, (U2) hairline/forehead boundary, (U3) Cranial baldness, (U4) Ridge structures, (U5) Eyebrows Outline, and (U6) Unibrow. (M1) Fissure Outline, (M2) Upper Folds, (M3) Superior Palpebral Furrow, (M4) Lower Folds, (M5) Inferior Palpebral Furrow, (M6) Infraorbital Furrow, (M7) Iris, (M8) Pupil, (M9) Caruncle Outline, (M10) Cheekbone, (M11) Dimple Cheek, (M12) Nose Outline, (M13) Nasal Root, (M14) Nasal Body, (M15) Nasal Tip, (M16) Nasal Base, (M17) Alae, (M18) Nostrils, (M19) Outer Helix, (M20) Inner Helix, (M21) Anti-Helix, (M22) Tragus, and (M23) Anti-Tragus. (L1) Philtrum Ridges, (L2) Philtrum Furrow, (L3) Upper Lip Outline, (L4) Upper Lip Tubercle, (L5) Upper Lip Creases, (L6) Lower Lip Outline, (L7) Lower Lip Median Sulcus, (L8) Lower Lip Creases, (L9) Chin Outline, (L10) Chin Dimple, (L11) Neck Boundaries, (L12) Musculature, (L13) Veins, (L14) Double chin, and (L15) Laryngeal*

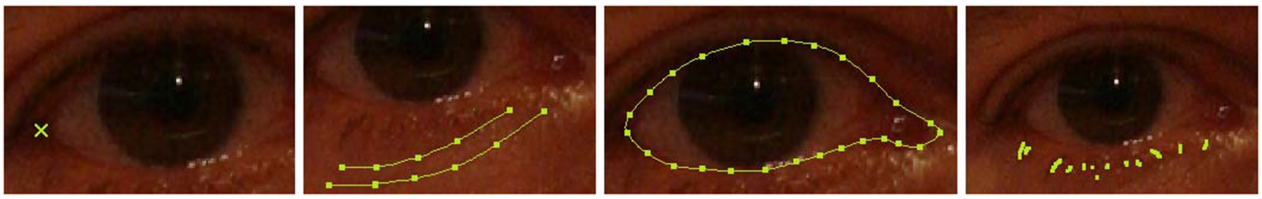

**Fig. 8** *Examples of the four annotation types. From left- to right-hand side: outer eye (landmark), lower folds (open shape), fissure (closed shape), and lower eye lashes (point cloud type)*

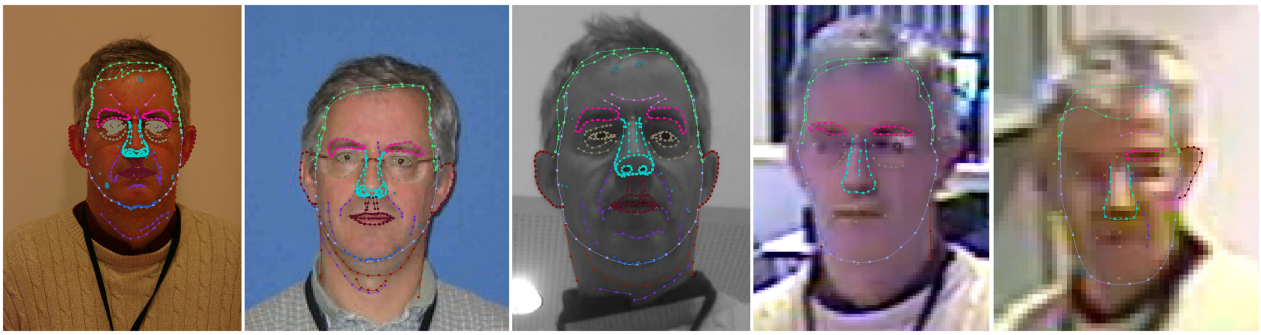

**Fig. 9** *Example images that have been annotated. From left- to right-hand side : Reference image high-res and four trace images, respectively mid-res, bw-down, near-frontal low-res 1, and near-frontal low-res 2. Short image names are defined in Table 7*

### *3.2 Manual annotation*

In our dataset, the forensic features can be extracted from the manual annotation. We use four annotation types. Examples are shown in Fig. 8.

The first annotation type is the landmark type, which is a well-defined, fiducial point on the facial image. We identified in total 21 landmarks, which can be used to determine the overall face/head composition (Fig. 6*a*, (H3)-(H23)).

The second and third annotation types are used to annotate shapes. Often shapes are represented by a polygon defined by a collection of landmarks. A disadvantage of this approach is that parts of the shape with a high curvature need significantly more landmarks than almost linear parts of the shape. Therefore, we propose a more flexible and compact solution using Hermite splines. A Hermite spline is a piecewise third order polynomial parametric curve [27]. It is defined by the interpolation of the landmarks, and, in our work, by assuming that the tangent at a landmark is given by the directional vector that interpolates the neighbouring landmarks. This approach has several advantages over the polygon approach. First, it needs a reduced number of landmarks to capture a rich variation in shapes. Second, if needed, it can be subsampled into a set of landmarks of arbitrary resolution. More details on the subsampling process are given in Section 4. The second (resp. third) type is the open (resp. closed) facial shape, in both cases represented by a Hermite spline. The nose outline is an example of an open shape, whereas the eyebrow shape is an example of a closed shape.

The fourth annotation type is the point cloud type, which describes multiple points belonging to the same feature, without performing an interpolation. Although these points could also have been represented by open or closed curves, it is more efficient to utilise this type. Typical examples are eye lashes or lip creases.

The trace and reference image names and annotation properties are summarised in Table 7. As expected, the average number of annotated facial parts depends inversely on the interpapillary distance. This is caused by the fact that a significant portion of the considered forensic features are detailed to very detailed, and therefore are only discernible in good quality images with a relative large interpapillary distance. Example annotations are shown in Fig. 9. Out of 97 subjects, 87 subjects have all five images available, yielding in total 435 annotated images.

### *3.3 Annotation acquisition*

Three paid participants were recruited for the annotation. The participants had prior knowledge and experience, as they had participated in another forensic annotation experiments. Prior to the instruction, the instructor discussed instruction details with the NFI. The instruction was given in a single session of a day. Image sets were prepared such that participants annotated a subject exactly once in a session of 87 images. After a week the annotations were evaluated, and feedback was given to the

**Table 7** Annotated trace and reference images

| Forensic use case illustration | Short image name | Trace material | Avg. IPD (px) | Avg. # annotated facial parts |
|---|---|---|---|---|
| |ID card | mid-res | ≪ sid ≫a.jpg | 35 | 51 |
| Debit Card | bw-down | ≪ sid ≫c1a.bmp | 65 | 50 |
| Robbery 1 | near-frontal low-res 1 | ≪ sid ≫c4a12.bmp | 23 | 27 |
| Robbery 2 | near-frontal low-res 2 | ≪ sid ≫c3a3.bmp | 11 | 19 |
| Reference | high-res | ≪ sid ≫f.jpg | 370 | 74 |

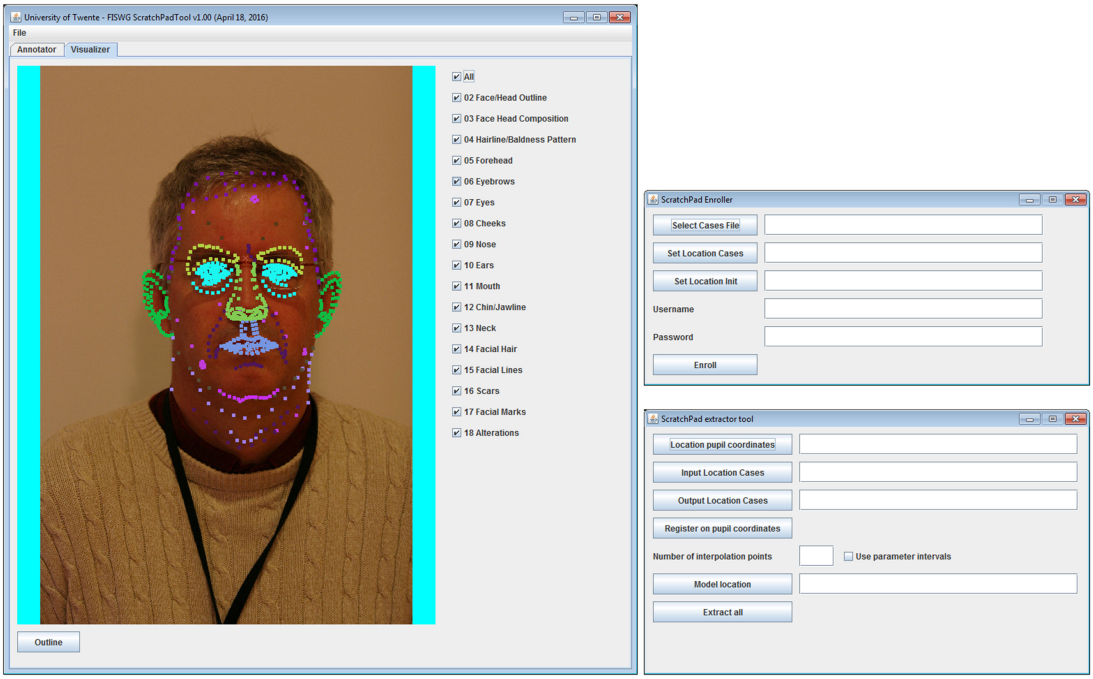


**Fig. 10** *Provided software tools. Left-hand side : ScratchPadTool for viewing and creation of annotation. Top right-hand side: ScratchPadEnroller for the preparation of a new dataset. Bottom right-hand side: ScratchPadExtractor for the extraction of features from the annotation*

annotators. The duration of an annotation session varied between two to four weeks. The annotation was facilitated by an application that provides basic drawing tools and a visualiser that gives feedback to the participant. Moreover, the participant could specify either he/she could not determine an annotation or state his/her confidence of the annotation on a five-point scale (very unconfident, unconfident, neutral, confident, and very confident). The annotation was assessed and approved by the instructor.

## 4 Toolset

Three graphical applications bundled in a toolset are made available. The ScratchPadTool is primarily used to create the annotation, and it is also possible to view or create the annotation. Moreover, the ScratchPadTool can be used to annotate any another dataset. For this, the ScratchPadEnroller must be used to prepare a dataset for use by the ScratchPadTool.

The final tool, ScratchPadExtractor, has two functions. Since the annotation uses the coordinate system of the image it belongs to, the annotation must be registered to a common coordinate system prior use. Therefore, ScratchPadExtractor provides a function to register the annotation on pupil coordinates. The tool is also used to extract points from the annotation. As discussed in Section 3.2, the annotation is stored as a collection of points that define a Hermite spline. The tool can sample these Hermite splines to create a dense collection of points that represent a shape. The user can provide some parameters, such as the number of sampling points and the manner in which the sampling is performed. As an example, two sets of sampled Hermite splines are provided. These points can be used directly as a feature (e.g. eyebrow shape) or indirectly in a feature (e.g. the angle of the eye fissure).

The tools are shown in Fig. 10 and are described in more detail in the user guide.

## 5 Potential uses, evaluation protocols, and an example

### *5.1 Potential uses*

We envision that this dataset is particularly useful for forensic research. The chosen forensic features are line with the characteristic descriptors found in [5]. We can extract a large number of these characteristic descriptors from the annotation. For example, from the eyebrow and eye fissure annotation we can derive multiple characteristic descriptors: the eyebrow shape, eye fissure shape and angle and for example three particular relative positions A, B, and C as shown in Fig. 1 in the introduction. These features can then in turn be used by biometric classifiers. Other uses include the matching of 3D images with video sequences and video sequences versus video sequences.

We realise that the size of this dataset is small from a biometric perspective. For example, the FRGC dataset [12] contains more than 39,000 images of 568 subjects. Still, we believe that the availability of a rich annotation could aid research on facial parts that have had little attention before. This is in line with a growing interest in the biometric community to fuse soft biometric facial features with highly discriminating biometric features to enhance performance in non-ideal situations. A typical example is the periocular region complementing iris images. An earlier study by Zeinstra *et al.* [7] has shown that using information captured in annotated images of the periocular region performs comparably to more texture based approaches described for example in the work by Park *et al.* [28].

Another potential use of this dataset is the evaluation of computer vision algorithms for the extraction of facial features. In this respect, the annotation contained in this dataset can serve as a ground truth for the evaluation of such an algorithm.

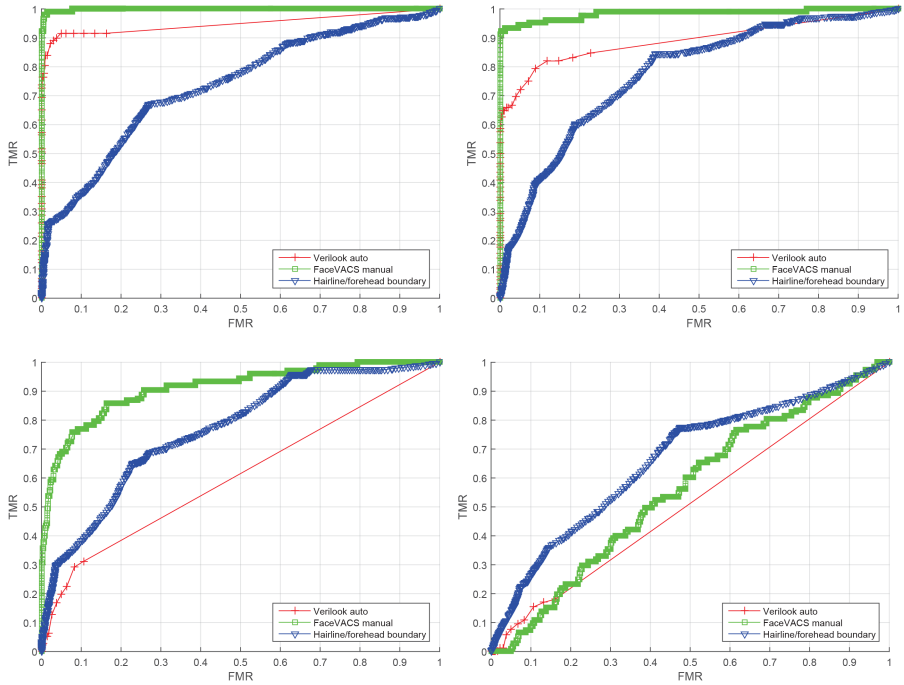


**Fig. 11** *Receiver operator characteristic curves for Verilook, FaceVACS, and the hairline/forehead boundary in the four use cases. Top row, left-hand side: mid-res versus high-res, right-hand side: bw-down versus high-res. Bottom row, left-hand side: near-frontal low-res 1 versus high-res, right-hand side: near-frontal low-res 2 versus high-res*

### *5.2 Evaluation protocol 1: annotation scenario*

We propose the following evaluation protocol for the annotation scenario. Since the number of subjects is limited, we provide 50 random different partitions of 67 train and 20 test subjects. This particular choice is a trade-off between having enough train and test subjects. The test results of each of the 50 partitions are collected in a single test result set. The performance of a system using this scenario should be reported on this aggregated set. Note that the test results are those of a family of very related classifiers, rather than a single classifier. More details can be found in the user guide.

### *5.3 Evaluation protocol 2: video versus 3D scenario*

We propose the following two evaluation protocols for the video versus 3D scenario.

The first evaluation protocol (2A) is that this dataset is only used for the evaluation of video versus 3D algorithms that are trained on other datasets. Both identification and verification modes of operation are possible. Any camera sequence (12 = 6 cameras × wearing no cap/cap) can be matched against all 3D reference shapes.

The second evaluation protocol (2B) is similar to evaluation protocol 1. We provide 50 random different partitions of 73 train and 20 test subjects. For each of the 12 camera sequences, the test results of each of the 50 partitions are collected in a single test result set. The performance of a system using this scenario should be reported on this aggregated set.

### *5.4 Evaluation protocol 3: video versus video scenario*

We propose the following two evaluation protocols for the video versus video scenario.

The first evaluation protocol (3A) is that this dataset is only used for the evaluation of video versus video algorithms that are trained on other datasets. Both identification and verification modes of operation are possible. Any camera sequence (12 = 6 cameras × wearing no cap/cap) can be matched with any other camera sequence, giving a total of $12 \times 11/2 = 66$ possible combinations.

The second evaluation protocol (3B) is similar to evaluation protocols 1 and 2B. We provide 50 random different partitions of 77 train and 20 test subjects. For each of the 66 camera combinations, the test results of each of the 50 partitions are collected in a single test result set. The performance of a system using this scenario should be reported on this aggregated set.

### *5.5 Example protocol 1: baseline versus hairline/forehead boundary*

In this section, we present baseline results and compare those with what can be achieved by using only the hairline/forehead boundary. All experiments use the proposed evaluation protocol of Section 5.2.

For the baseline experiment we use Neurotec Verilook 6.0 [29] and Cognitec FaceVACS 9.1 [30]. These systems use the full face. Prior to the experiment, we let Neurotec Verilook 6.0 automatically determine the pupil coordinates and we provide FaceVACS with manually determined pupil coordinates. We consider four cases: mid-res versus high-res, bw-down versus high-res, near-frontal low-res 1, and near-frontal low-res 2. The results are shown in Fig. 11.

We compare these results with what can be achieved by only using the hairline/forehead boundary. Prior to comparison, we register the annotation on pupil coordinates in order to introduce a common coordinate system. We then subsample the hairline/forehead boundary with 100 equidistant points. We use a shape similarity score function to compare two shapes. If $X = \{\boldsymbol{x}_i \in \mathbb{R}^2 | i = 1, \ldots, N_x\}$ and $Y = \{\boldsymbol{y}_i \in \mathbb{R}^2 | i = 1, \ldots, N_y\}$ are two shapes, we define the shape similarity score function $s_s : \mathbb{R}^{2 \times N_x} \times \mathbb{R}^{2 \times N_y} \to \mathbb{R}$ as:

$$s_s(X, Y) = -\frac{1}{N_x} \sum_{i=1}^{N_x} d_{pc}^2(\boldsymbol{x}_i, Y) - \frac{1}{N_y} \sum_{i=1}^{N_y} d_{pc}^2(\boldsymbol{y}_i, X), \quad (1)$$

where $d_{pc}$ measures the minimal distance between a point $\boldsymbol{w} \in \mathbb{R}^2$ and a point cloud $Z = \{\boldsymbol{z}_i \in \mathbb{R}^2 | i = 1, \ldots, N\}$: $d_{pc}(\boldsymbol{w}, Z) = \min_{i=1,\ldots,N} \| \boldsymbol{w} - \boldsymbol{z}_i \|$.

We compare the results with the hairline/forehead boundary shape as an illustration of its relative robustness against severe image degradation.

The results are shown in Fig. 11. We can make several observations. First of all, both commercial systems clearly outperform the hairline/forehead boundary shape-based system in

the comparison of mid-res versus high-res and bw-down versus high-res. The commercial systems are typically designed to cope with these image types and conditions. The situation changes when we consider the near-frontal low-res 1 and 2 images. In the case of near-frontal low-res 1 versus high-res we notice that Verilook is performing worse than the hairline/forehead boundary shape, but FaceVACS is still the best. However, we observe that hairline/forehead boundary shape performs better than both commercial systems in near-frontal low-res 2 versus high-res. Another observation is that the hairline/forehead boundary shape is not very discriminating, but has some robustness under different comparisons. Note, however, that a forensic facial practitioner takes all available comparison results into account during an assessment of evidential value.

## 6 Conclusion

In this paper, we have presented ForenFace, a novel forensic facial video and image dataset. It contains CCTV footage, extracted still images, reference images, and 3D scans. Its novelty with respect to other forensic facial datasets in the forensic domain is twofold. Inspired by the FISWG characteristic descriptors, it is the first dataset that includes a rich forensically relevant annotation of almost 19.000 facial parts on 435 images of five different image types of varying quality. Moreover, it comes with a toolset of three complementary software tools that can be used on other datasets as well. We believe that these two factors lead to a dataset that has an added value in the field of forensic face datasets.

We proposed evaluation protocols and showed that in the annotation scenario the baseline performance of commercial systems on the severest case is less than a system that is only using the hairline/forehead boundary shape.

By making this dataset available to the research community, we hope to encourage research especially in the forensic domain. As can be seen from the baseline experimental results, face recognition in a realistic forensic setting is still not a solved issue.

## 7 Acknowledgments

The authors thank the volunteers at the Netherlands Forensic Institute for their participation in the creation of the dataset, and the annotators for their time investment. Finally, the authors thank Neurotechology and Cognitec Systems GmbH. for supporting our research by providing the VeriLook and FaceVACS software. Results obtained for VeriLook and FaceVACS were produced in experiments conducted by the University of Twente, and should therefore not be construed as a vendor's maximum effort full capability result.

## 8 References

[1] Spaun, N.A.: 'Forensic Biometrics from Images and Video at the Federal Bureau of Investigation. In Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007'. First IEEE Int. Conf. on, September 2007, pp. 1–3

[2] Prince, J.P.: 'To examine emerging police use of facial recognition systems and facial image comparison procedures', 2012. Available at www.churchilltrust.com.au/media/fellows/2012_Prince_Jason.pdf, accessed 22 April 2014

[3] FISWG website. Available at https://fiswg.org, accessed 22 April 2014

[4] FISWG Guidelines for Facial Comparison Methods. Available at https://fiswg.org/FISWG_GuidelinesforFacialComparisonMethods_v1.0_2012_02_02.pdf, accessed 09 January 2017

[5] FISWG Facial Image Comparison Feature List for Morphological Analysis. Available at https://fiswg.org/FISWG_1to1_Checklist_v1.0_2013_11_22.pdf, accessed 09 January 2017

[6] Zeinstra, C.G., Veldhuis, R.N.J., Spreeuwers, L.J.: 'Towards the automation of forensic facial individualisation: comparing forensic to non forensic eyebrow features'. Proc. of the 35th WIC Symp. on Information Theory in the Benelux, Eindhoven, Netherlands, Enschede, May 2014, pp. 73–80, Centre for Telematics and Information Technology, University of Twente

[7] Zeinstra, C.G., Veldhuis, R. N.J., Spreeuwers, L.J.: 'Beyond the eye of the beholder: on a forensic descriptor of the eye region'. 23rd European Signal Processing Conf., EUSIPCO 2015, Nice, IEEE Signal Processing Society, September 2015, pp. 779–783

[8] Tome, P., Fierrez, J., Vera-Rodriguez, R., *et al.*: 'Identification using face regions: Application and assessment in forensic scenarios', *Forensic Science Int.*, 2013, **233**, (13), pp. 75–83

[9] Tome, P., Fierrez, J., Vera-Rodriguez, R., *et al.*: 'Facial soft biometric features for forensic face recognition', *Forensic Sci. Int.*, 2015, **257**, pp. 271–284

[10] Kanade, T.: 'Picture processing system by computer complex and recognition of human faces'. Doctoral dissertation, Kyoto University, November 1973

[11] Turk, M., Pentland, A.: 'Eigenfaces for recognition', *J. Cogn. Neurosci.*, 1991, **3**, (1), pp. 71–86

[12] FRGC website. Available at http://www.nist.gov/itl/iad/ig/frgc.cfm, accessed 22 April 2014

[13] Jonathon Phillips, P., Moon, H., Rizvi, S.A., *et al.*: 'The FERET evaluation methodology for face-recognition algorithms', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (10), pp. 1090–1104

[14] Phillips, P.J., Scruggs, W.T., O'Toole, A.J., *et al.*: 'FRVT 2006 and ICE 2006 large-scale experimental results', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **32**, (5), pp. 831–846

[15] Face Recognition Vendor Test (FRVT) – Performance of Face Identification Algorithms, NIST Interagency Report 8009. Available at http://biometrics.nist.gov/cs_links/face/frvt/frvt2013/NIST_8009.pdf, accessed 07 April 2016

[16] Huang, G.B., Ramesh, M., Berg, T., *et al.*: 'Labeled faces in the wild: a database for studying face recognition in unconstrained environments'. Technical Report 07-49, University of Massachusetts, Amherst, October 2007

[17] Le, V., Brandt, J., Lin, Z., *et al.*: 'Interactive facial feature localization', in Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (Eds.): '*Computer vision ECCV 2012*' (Springer, Berlin Heidelberg, 2012), (LNCS, **7574**), pp. 679–692

[18] Neves, J.C., Santos, G., Filipe, S., *et al.*: '*Quis-Campi: extending in the wild biometric recognition to surveillance environments*' (Springer International Publishing, Cham, 2015), pp. 59–68

[19] Grgic, M., Delac, K., Grgic, S.: 'SCFace - surveillance cameras face database', *Multimedia Tools Appl.*, 2011, **51**, (3), pp. 863–879

[20] Wong, Y., Chen, S., Mau, S., *et al.*: 'Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition'. IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops, IEEE, June 2011, pp. 81–88

[21] NIST Mugshot Identification Database. Available at http://www.nist.gov/srd/nistsd18.cfm, accessed 25 April 2016

[22] Ricanek, K.Jr., Tesafaye, T.: 'MORPH: a longitudinal image database of normal adult age-progression'. Proc. of the 7th Int. Conf. on Automatic Face and Gesture Recognition, FGR '06, Washington, DC, USA, 2006, IEEE Computer Society, pp. 341–345

[23] Vera-Rodriguez, R., Tome, P., Fierrez, J., *et al.*: 'Analysis of the variability of facial landmarks in a forensic scenario'. Biometrics and Forensics (IWBF), 2013 Int. Workshop on, April 2013, pp. 1–4

[24] ForenFace website. Available at http://scs.ewi.utwente.nl/downloads/show,ForenFace/, accessed 08 June 2016

[25] Dallmeier website. Available at http://www.dallmeier.com, accessed 04 January 2016

[26] MeshLab website. Available at http://meshlab.sourceforge.net, accessed 08 January 2016

[27] Bartels, R.H., Beatty, J.C., Barsky, B.A.: '*An introduction to splines for use in computer graphics and geometric modeling*' (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987)

[28] Park, U., Jillela, R., Ross, A., *et al.*: 'Periocular biometrics in the visible spectrum', *IEEE Trans. Inf. Forensics Sec.*, 2011, **6**, (1), pp. 96–106

[29] Neurotechnology Verilook website. Available at http://www.neurotechnology.com/verilook.html, accessed 10 May 2016

[30] Cognitec FaceVACS website. Available at http://www.cognitec.com/products.html, accessed 29 December 2015