Int J Legal Med (2011) 125:301–306
DOI 10.1007/s00414-010-0446-4

TECHNICAL NOTE

A new atlas for the evaluation of facial features: advantages, limits, and applicability

Stefanie Ritz-Timme · Peter Gabriel · Zuzana Obertovà · Melanie Boguslawski · F. Mayer · A. Drabik · Pasquale Poppa · Danilo De Angelis · Romina Ciaffi · Benedetta Zanotti · Daniele Gibelli · Cristina Cattaneo

Received: 23 December 2009 / Accepted: 9 March 2010 / Published online: 6 April 2010

Abstract Methods for the verification of the identity of offenders in cases involving video-surveillance images in criminal investigation events are currently under scrutiny by several forensic experts around the globe. The anthroposcopic, or morphological, approach based on facial features is the most frequently used by international forensic experts. However, a specific set of applicable features has not yet been agreed on by the experts. Furthermore, population frequencies of such features have not been recorded, and only few validation tests have been published. To combat and prevent crime in Europe, the European Commission funded an extensive research project dedicated to the optimization of methods for facial identification of persons on photographs. Within this research project, standardized photographs of 900 males between 20 and 31 years of age from Germany, Italy, and Lithuania were acquired. Based on these photographs, 43 facial features were described and evaluated in detail. These efforts led to the development of a new model of a morphologic atlas, called DMV atlas ("Düsseldorf Milan Vilnius," from the participating cities). This study is the first attempt at verifying the feasibility of this atlas as a preliminary step to personal identification by exploring the intra- and interobserver error. The analysis yielded mismatch percentages from 19% to 39%, which reflect the subjectivity of the approach and suggest caution in verifying personal identity only from the classification of facial features. Nonetheless, the use of the atlas leads to a significant improvement of consistency in the evaluation.

Keywords Forensic science · Forensic anthropology · Personal identification · Photo comparison · Facial traits · Morphological features

Electronic supplementary material The online version of this article (doi:10.1007/s00414-010-0446-4) contains supplementary material, which is available to authorized users.

S. Ritz-Timme · P. Gabriel · Z. Obertovà · M. Boguslawski · F. Mayer
Institut für Rechtsmedizin des Universitätskinikums,
University of Düsseldorf,
Moorenstraße 5,
Dusseldorf, Germany

A. Drabik
Zentrum für Klinische Studien, University of Münster,
Von-Esmarch-Strasse 62,
Münster, Germany

P. Poppa · D. De Angelis · R. Ciaffi · B. Zanotti · D. Gibelli · C. Cattaneo (✉)
LABANOF, Laboratorio di Antropologia e Odontologia Forense,
Dipartimento di Morfologia e Scienze Biomediche,
Sezione di Medicina Legale, Università degli Studi di Milano,
Via Mangiagalli, 37,
20133 Milan, Italy
e-mail: cristina.cattaneo@unimi.it

Introduction

Criminal offenses which have been captured on video-surveillance systems (e.g., robberies, murders, and assaults) often require verification of the identity of the persons involved and therefore call for personal identification based on facial or other somatic traits. Identification by facial morphology has a long history. The first approaches were performed by Alphonse Bertillon in France [1] and subsequently by Salvatore Ottolenghi in Italy in 1907 [2]. The initial and recurrent idea is to propose a set of features and classify parts of the face according to such features by the use of an atlas. Specific sets of facial and head features for the classification of faces has been proposed—after the initial *bertillonage*—by the International Criminal Police

Identification (Interpol) in the Disaster Victim Identification (DVI) form which includes sketches of head and ear shape, and other facial features (e.g., nose shape) [3]. Subsequently, other authors proposed further similar classification systems and studied their applicability [4, 5].

Classification of the face according to preset features is a preliminary step towards identification (when the face is visible). Final positive identification can subsequently be performed by other more specific methods in the field of anthroposcopy, anthropometry, 2D/2D or 2D/3D superimposition of images, 2D/3D image comparisons [6–14], along with other factors such as height estimation [15] and gait analysis [16, 17]. These methods are widely applied, in spite of the fact that extensive and conclusive validation tests for these methods are lacking nor do accredited guidelines exist apart from an attempt by the German AGIB group [18].

Regardless of which may be the best final method of positive identification, most authors will agree that the classification of facial features is one of the first operations towards identification, in order to exclude identity or reach a verdict of compatibility of two faces with a similar shape. In order to do this properly, two fundamental factors are necessary. Firstly, a classification system needs to be adopted, such as an atlas—with detailed versions of facial types. Up to now, Vanezis' atlas edition [5] is based on 23 facial features, while the DVI form by Interpol [3] includes only six traits: frontal, profile, forehead height and width, forehead inclination, nose, and ear lobes. Secondly, every specific atlas needs to know its inter- and intraobserver error in order to be used with the appropriate caution.

In the course of a project funded by the European Union called "Optimisation of methods for identification of persons on photographs (photoidentification): a contribution to combat and prevention of crime in Europe," shared by researchers from Germany, Italy, and Lithuania, a new model of atlas was developed, called DMV (from the Universities of Düsseldorf, Milan, and Vilnius which participated in the project). Reference photographs of 900 males between 20 and 31 years of age in Germany, Italy, and Lithuania were acquired, and based on these photographs, facial features were described and evaluated. These efforts led to the first edition of an anthropological atlas published in 2007 [19], which includes detailed descriptions, drawings, and photographic examples (Figs. 1 and 2) of 43 morphological features of male faces (Fig. 3) and is the most detailed atlas produced up to now. In 2008, the atlas of male facial features was revisited, and population frequencies for the facial features were added [20]. In 2009, an atlas of female facial features was developed [21]. These two atlas editions include the descriptions, drawings, and photographic examples of 45 facial traits and allow a morphological assessment of both male and female faces.

In order to explore the intra- and interobserver error, tests were performed in two countries, Italy and Germany, by experienced and non-experienced researchers in the field of visual personal identification.

Material and methods

Photographs of 900 males (300 from Italy, 300 from Germany, and 300 from Lithuania) aged between 20 and 31 years, were taken with a digital camera. For each subject, five standardized photographs were acquired with the focus on seilion and the distance between seilion (the deepest point of the nasal root depression) and the camera 1.5 m. The head of the test person was oriented in the Frankfurt plain and photographed in the following positions: left lateral (90°), 45° left, frontal, 45° right, right lateral (90°). This database allowed the authors to devise a more elaborate atlas, which considered several types of forms for every facial character.

Based on these 900 photographs, 43 facial features were described and evaluated, which led to the compilation of the DMV atlas [19]. The applicability and feasibility of this atlas was then tested in two countries—Italy and Germany. In Italy, a subsample of 270 males was randomly selected from the total Italian sample and was evaluated by two observers, a post-graduate student in science (observer A) and a PhD anthropology student (observer B). Prior to the evaluation, both observers were thoroughly trained by an expert in the field of forensic identification of the living. In the evaluation, the DMV atlas [19] was used for the description of the facial features. The results were used for the calculation of the intra- and interobserver error for each facial feature.

The intraobserver error was calculated for both operators using 60 subjects randomly chosen from the 270 subjects; thus, these 60 subjects were evaluated twice by the same observer. The interval between the two evaluations by one observer was circa 2 months. The interobserver error was based on the total number of 270 subjects. The intra- and interobserver errors are depicted as mismatch percentages for individual features, i.e., the number of cases from the total sample, in which the classification differed between the first and second observation of one observer, or between the evaluations of the two observers.

In Germany, a subsample of 25 males was randomly selected from the total German sample and was evaluated by an expert and a non-expert in the field of visual identification of the living. The non-expert (observer A) was an assistant physician, while the expert (observer B) had many years of experience in the identification of persons on photographs. The observers performed four evaluation rounds. The first and second evaluation took place without the DMV atlas [19]. During this evaluation stage, only a list of features and

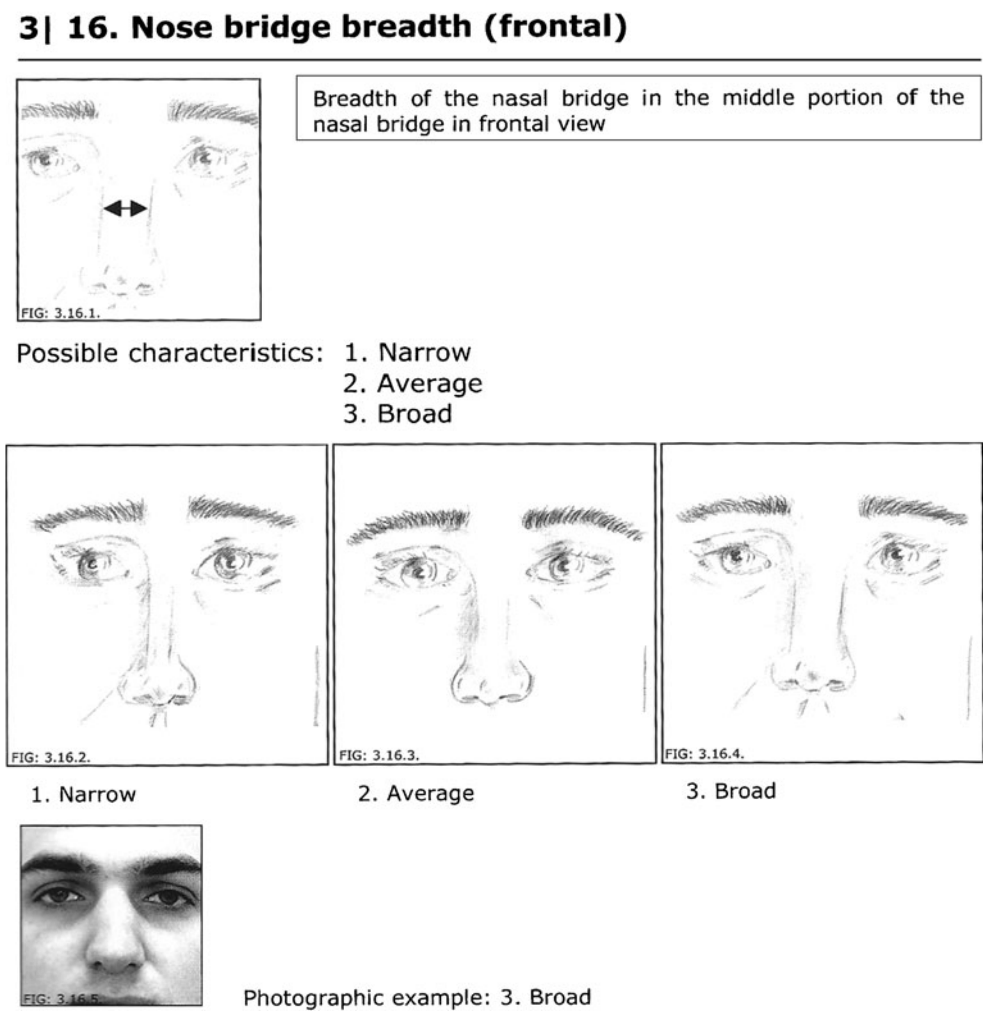

Fig. 1 Example of the written and graphical description of facial feature Nr. 16: nose bridge breadth (frontal) from the DMV atlas [19]

classifications (without detailed descriptions or illustrations) was available to the observers. The third and fourth evaluation proceeded with the DMV atlas. The interval between the two evaluations by one observer was 1 month. As for the Italian sample, inter- and intraobserver errors were calculated in terms of mismatch percentages.

The statistical analysis was performed by SPSS and Microsoft Excel.

Results

The DMV atlas [19] includes the descriptions, drawings, and photographic examples of 43 facial features. These features were selected based on the experience of the investigators considering visibility, consistency, and usability in the forensic practice.

The first step in exploring the feasibility of this atlas as a tool in the process of visual personal identification was to obtain the intra- and interobserver error rates. The intra- and interobserver mismatch percentages were established based on the classifications of 43 facial features for 270 (interobserver) and 60 (intraobserver) Italian males (Table 1) and 25 German males (Table 2).

In the Italian sample, the mean intraobserver mismatch percentage for all features obtained by operator A (undergraduate student) was 30%, i.e., in 70% of the cases, the classification from the first evaluation was repeated in the second evaluation round. The lowest mismatch percentage (2%) obtained by operator A was found in the description of the fold above upper eyelid (feature Nr. 11) and the nose tip incisure (feature Nr. 21). The highest mismatch percentage (53%) for this operator was observed in the description of the alar wing length (feature Nr. 24) and the ear breadth (feature Nr. 38). The mean intraobserver mismatch percentage for all features obtained by operator B (PhD student) was 28% between the first and the second evaluation. The lowest percentage of error (8%) for this observer occurred in the description of the nose tip incisure (feature Nr. 21), whereas the highest percentage of error (68%) was found in the description of the eyebrow shape (feature Nr. 8).

The mean interobserver mismatch percentage obtained when comparing the classifications for 270 Italian males was 39%. The highest percentage of mismatch was 65% in the

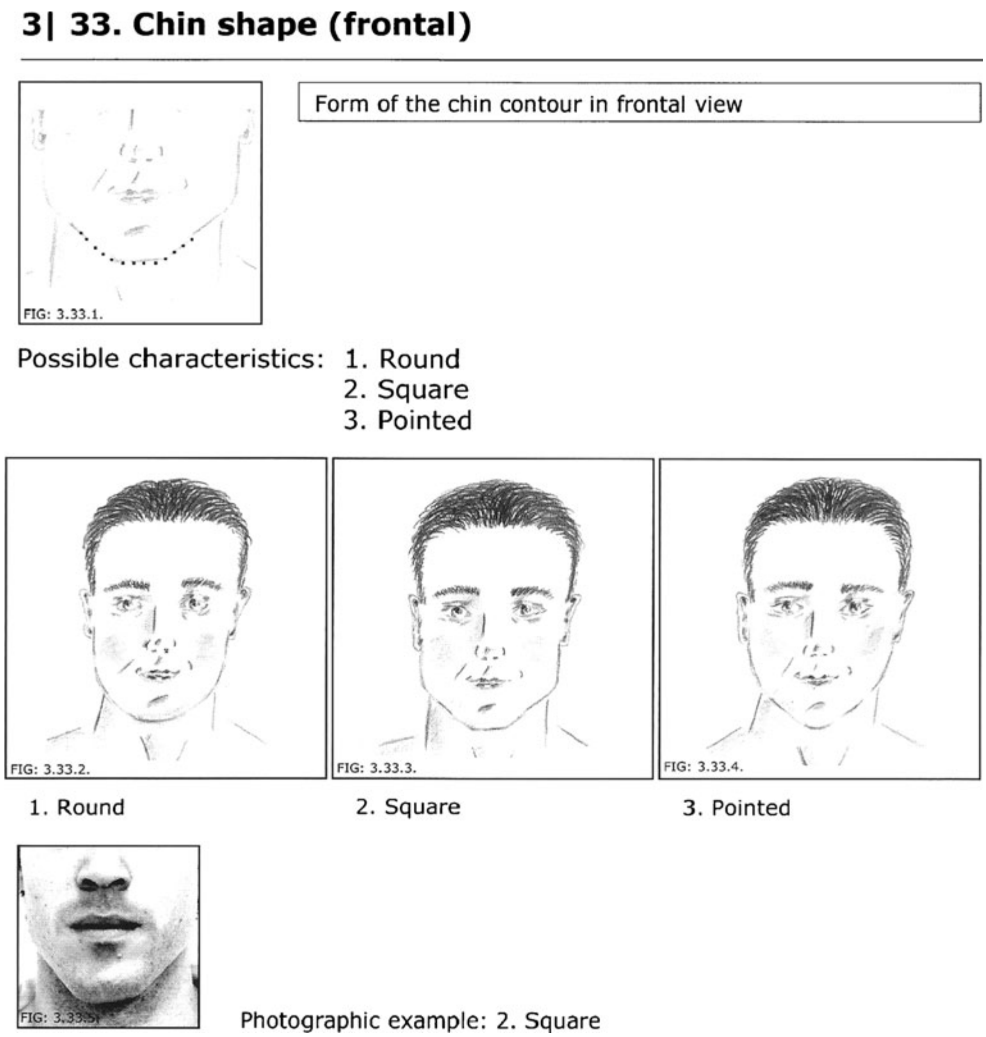

Fig. 2 Example of the written and graphical description of facial feature Nr. 33: chin shape (frontal) from the DMV atlas [19]

description of the extent of protrusion of the nose (feature Nr. 22). The lowest value (8%) was found in the description of the visibility of the fold above upper eyelid (feature Nr. 11).

The German subsample was evaluated four times by both observers: the first and second time without the DMV atlas [19], and the third and fourth time by consulting the DMV atlas for feature descriptions and drawings. The mean intraobserver mismatch percentage for all features obtained by operator A (assistant physician) in the evaluation round without the atlas was 37%, i.e., in 63% of the cases, the classification from the first evaluation was repeated in the second evaluation. The mean intraobserver error for this operator when using the atlas was 29%. The mean intraobserver mismatch percentage for all features obtained by operator B (expert in visual personal identification) was 25% between the first and the second evaluation and 19% between the third and fourth evaluation, during which the atlas was used.

The chi-square test was applied to the German data to explore the differences between the evaluation of the facial features by an experienced and inexperienced observer and to assess the contribution of the DMV Atlas [19] to the evaluation. The experienced observer showed a significantly higher congruence in his evaluations, without (p=0.0000) and with the atlas (p=0.0000) compared to the inexperienced observer. The use of the atlas led to a significantly higher consistency between the two evaluations of the expert (p=0.0001) and the non-expert (p=0.0001). The lowest and highest mismatch percentages are reported only for the German evaluation round with the atlas to enable the comparison with the Italian data (Fig. 4).

The lowest mismatch percentage (8%) obtained by German operator A was observed in the description of the fold above upper eyelid (feature Nr. 11). The highest mismatch percentage (64%) for this operator was found in the description of the eyebrow shape (feature Nr. 8). The lowest percentage of error (0%) for the German expert (operator B) occurred in the description of the nose profile (feature Nr. 18), orientation of the nose tip (feature Nr. 19), the nose tip incisure (feature Nr. 21), and the transition head/neck (feature Nr. 42). The highest percentage of error (52%) for this observer was found in the description of the eyebrow shape (feature Nr. 8).

N°	Feature
01.	Head shape (frontal)
02.	Frontal height (frontal)
03.	Frontal breadth (frontal)
04.	Frontal hairline (frontal)
05.	Forehead bias (lateral)
06.	Eyebrow height (frontal)
07.	Eyebrow density (frontal)
08.	Eyebrow shape (frontal)
09.	Mono-brow (frontal)
10.	Distance upper eyelid - eyebrow (frontal)
11.	Upper eyelid (frontal)
12.	Lid axis (frontal)
13.	Lower eyelid fold (frontal)
14.	Nasal root (frontal)
15.	Nose bridge length (frontal)
16.	Nose bridge breadth (frontal)
17.	Nose bridge process (frontal)
18.	Nose profile (lateral)
19.	Inclination of the columella (lateral)
20.	Nose tip shape (frontal)
21.	Nose tip incisure (frontal)
22.	Nose protrusion (lateral)
23.	Nasal breadth (frontal)
24.	Alar wing length (lateral)
25.	Alar wing height (lateral)
26.	Nostrils (lateral)
27.	Philtrum height (frontal)
28.	Philtrum depth (frontal)
29.	Philtrum shape (frontal)
30.	Upper lip notch (frontal)
31.	Labial breadth (frontal)
32.	Orientation of mouth corner (frontal)
33.	Chin shape (frontal)
34.	Chin transition (frontal)
35.	Chin protrusion (lateral)
36.	Chin dimple (frontal)
37.	Ear height (lateral)
38.	Ear breadth (lateral)
39.	Ear lobe size (lateral)
40.	Ear lobe attachment (lateral)
41.	Ear protrusion (frontal)
42.	Transition head - neck (frontal)
43.	Pronounciation of cheek bones (frontal)

Fig. 3 List of facial features described in DMV atlas [19]

The mean mismatch percentage between the German operators A and B was 42% without the DMV atlas and 39% with the atlas (Fig. 4). The highest percentage of mismatch was 70% in the description of the eyebrow shape (feature Nr. 8). The lowest value (14%) was found in the description of the protrusion of the ear (feature Nr. 41).

Discussion

The anthroposcopic or morphological method is the most frequently used method in visual identification, even

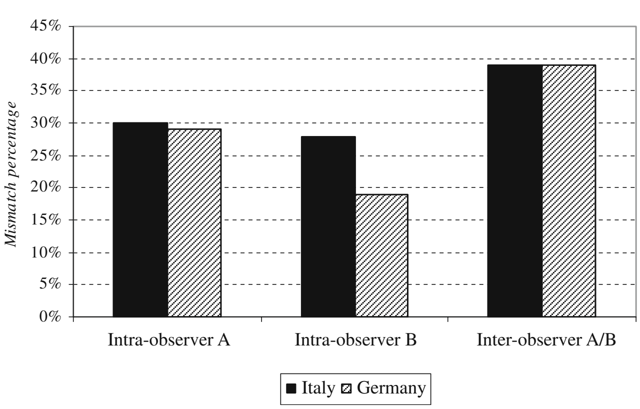

Fig. 4 Mean values of the intra- and interobserver mismatch percentages obtained by Italian and German operators A and B for all facial features by using the DMV atlas [19]

though a specific set of applicable features has not yet been agreed on. Until now, the description of morphological features in this procedure is largely based on subjective opinions of the forensic experts since there is not enough scientific data to validate the method [5, 21]. Naturally, the results of the identification attempt are strongly influenced by several factors, such as the quality of the images, the positioning of the video-surveillance camera, and the use of disguises. In the present work, these were not taken into consideration since the main goal was to verify the reliability of the system in the best possible conditions. Obviously further research needs to be performed now for less optimal conditions.

The DMV atlas [19] includes detailed descriptions, drawings, and photographic examples of 43 morphological features of male faces. To verify the user-friendliness of this atlas, error rates between different observers in Italy and Germany were acquired. In Germany, the tests were carried out in two steps: first, facial features were recorded and described without the atlas, and in the second step, the DMV atlas [19] was used in the classification process. In Italy, the evaluation proceeded only with the atlas.

The intraobserver mismatch percentages for two Italian and two German observers based on all features when using the DMV atlas [19] ranged from 19% to 30%. The interobserver mismatch percentages for the German observers was 42% without the atlas and 39% with the atlas, and the Italian examiners differed by 39% when using the DMV atlas. Thus, in two independent tests performed in two countries by four different observers and on different numbers of photographs (25 in Germany and 270 or 60 in Italy), the interobserver error was quite constant—approximately 39%. The intraobserver error was similar for the less experienced observers (28–30%), while the German expert in visual personal identification achieved the lowest error rate (19%).

Considering the error percentages for the individual facial features, a wide variability was observed; every observer showed a specific recognition pattern for the individual facial features. Moreover, the mismatch percentages for individual features differed between the German and Italian groups. Thus, it seems that the morphologic assessment of faces is affected also by cultural variables. Nevertheless, training in visual personal identification plays a relevant role in the facial assessment, as shown by the lower percentages of mismatch of the skilled observers, both in Italy and in Germany. The main goal of this work was to verify intra- and interobserver error. Results steer towards a great caution in the use of such traits, which should be performed only by experts. Since the aim is not to identify via these traits, but to exclude or find compatibilities, the authors did not dwell further on the possible effects on overall identification of rates of the error, which however will be the object of future studies.

The intra- and interobserver test results suggest that caution must be applied when conducting morphological comparison of facial features in the assessment of identity from an image comparison process. Nonetheless, the use of an atlas, such as the DMV atlas, which is at the present the most detailed edition of a morphologic atlas available, decreases the influence of the subjective description of individual facial features.

In conclusion, the comparative atlas approach based on facial morphology can be considered appropriate as an initial screening method in order to find traits which may exclude identity or verify concordances. Classification of the morphology of a face however should be considered a preliminary approach of identification, and the definitive result should be achieved with the aid of additional methods, such as comparison of facial silhouettes or 3D modeling.

Acknowledgment The study was supported by the European Commission (grant number JLS/2005/AGIS/054).

Ethical standards The authors declare that the experimental project comply with the current laws of the countries in which they were performed.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bertillon A (1896) Signaletic instructions including the theory and practice of anthropometrical identification. National Library of Medicine, Chicago
2. Ottolenghi S (1907) Polizia Scientifica Identificazione fisica e psichica. Investigazioni giudiziarie. Società Poligrafica Editrice, Roma
3. Interpol DVI (Disaster Victim Identification) form. Available at http://www.interpol.int/Public/DisasterVictim/forms/ID-FormEngUpd.pdf
4. Roelofse MM, Steyn M, Becker PJ (2008) Photo identification: facial metrical and morphological features in South African males. Forensic Sci Int 177:168–175
5. Vanezis P, Lu D, Cockburn J, Gonzales A, McCombe G, Trujillo O, Vanezis M (1996) Morphological classification of facial features in adult Caucasian males based on an assessment of photographs of 50 subjects. J Forensic Sci 41:786–791
6. Catterick T (1992) Facial measurements as an aid to recognition. Forensic Sci Int 56:23–27
7. Yoshino M, Matsuda H, Kubota S, Imaizumi K, Miyasaka S (2000) Computer-assisted facial image identification system using a 3-D physiognomic range finder. Forensic Sci Int 109:225–237
8. Porter G, Doran G (2000) An anatomical and photographic technique for forensic facial identification. Forensic Sci Int 114:97–105
9. Halberstein RA (2001) The application of anthropometric indices in forensic photography: three case studies. J Forensic Sci 46:1438–1441
10. Yoshino M, Noguchi K, Atsuchi M, Kubota S, Imaizumi K, Thomas CDL, Clement JG (2002) Individual identification of disguised faces by morphometrical matching. Forensic Sci Int 127:97–103
11. Fraser NL, Yoshino M, Imaizumi K, Blackwell SA, Thomas CD, Clement JG (2003) A Japanese computer-assisted facial identification system successfully identifies non-Japanese faces. Forensic Sci Int 135:122–128
12. De Angelis D, Sala R, Cantatore A, Grandi M, Cattaneo C (2009) A new computer-assisted technique to aid personal identification. Int J Legal Med 123:351–356
13. Goos MI, Alberink IB, Ruifrok AC (2006) 2D/3D image (facial) comparison using camera matching. Forensic Sci Int 163:10–17
14. Kleinberg KF, Vanezis P, Burton AM (2007) Failure of anthropometry as a facial identification technique using high-quality photographs. J Forensic Sci 52:779–783
15. De Angelis D, Sala R, Cantatore A, Poppa P, Dufour M, Grandi M, Cattaneo C (2007) New method for height estimation of subjects represented in photograms taken from video surveillance systems. Int J Legal Med 121:489–492
16. Larsen PK, Simonsen EB, Lynnerup N (2008) Gait analysis in forensic medicine. J Forensic Sci 53:1149–1153
17. Lynnerup N, Vedel J (2005) Person identification by gait analysis and photogrammetry. J Forensic Sci 50:112–118
18. Rösing FW (1999) Standards für die Identifikation lebender Personen aufgrund von Bilddokumenten. Rechtsmedizin 9:152–154
19. Aßmann A, Nohrden D, Schmitt R, Gabriel P, Ritz-Timme S (2007) Anthropological atlas of male facial features. Verlag für Polizeiwissenschaft, Frankfurt
20. Ohlrogge S, Nohrden D, Schmitt R, Drabik A, Gabriel P, Ritz-Timme S (2008) Anthropological atlas of male facial features. Verlag für Polizeiwissenschaft, Frankfurt
21. Ohlrogge S, Arent T, Huckenbeck W, Gabriel P, Ritz-Timme S (2009) Anthropological atlas of female facial features. Verlag für Polizeiwissenschaft, Frankfurt