Technical note

Recognition of children on age-different images: Facial morphology and age-stable features

Zuzana Caplova [a],[*], Valentina Compassi [a], Silvio Giancola [b], Daniele M. Gibelli [a], Zuzana Obertová [c], Pasquale Poppa [c], Remo Sala [b], Chiarella Sforza [a], Cristina Cattaneo [c]

[a] Università degli Studi di Milano, Dipartimento di Scienze Biomediche per la Salute, Via Mangiagalli 31, 20133 Milano, Italy
[b] Department of Mechanical Engineering, Politecnico di Milano, Via Giuseppe La Masa, 120156 Milano, Italy
[c] LABANOF, Laboratorio di Antropologia e Odontologia Forense, Dipartimento di Scienze Biomediche per la Salute, Sezione di Medicina Legale, Università degli Studi di Milano, Via Mangiagalli, 37, 20133 Milan, Italy

A R T I C L E   I N F O

A B S T R A C T

The situation of missing children is one of the most emotional social issues worldwide. The search for and identification of missing children is often hampered, among others, by the fact that the facial morphology of long-term missing children changes as they grow. Nowadays, the wide coverage by surveillance systems potentially provides image material for comparisons with images of missing children that may facilitate identification. The aim of study was to identify whether facial features are stable in time and can be utilized for facial recognition by comparing facial images of children at different ages as well as to test the possible use of moles in recognition. The study was divided into two phases (1) morphological classification of facial features using an Anthropological Atlas; (2) algorithm developed in MATLAB® R2014b for assessing the use of moles as age-stable features. The assessment of facial features by Anthropological Atlases showed high mismatch percentages among observers. On average, the mismatch percentages were lower for features describing shape than for those describing size. The nose tip cleft and the chin dimple showed the best agreement between observers regarding both categorization and stability over time. Using the position of moles as a reference point for recognition of the same person on age-different images seems to be a useful method in terms of objectivity and it can be concluded that moles represent age-stable facial features that may be considered for preliminary recognition.

© 2017 The Chartered Society of Forensic Sciences. Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

There are millions of missing persons at any given moment and a large number of these cases involve children younger than 18 years old - 250,000 children go missing every year in the EU [1,2]. Children go missing throughout the world as a result of human trafficking, organised crime, labour exploitation, sexual exploitation, illicit adoption and other reasons [1,3]. For instance, it is estimated that 1.2 million children are trafficked every year worldwide [1,2,4,5].

Different methods and approaches are used for the search of missing children – digital and paper versions of posters with photographs and details about the missing, telephone hotlines for reporting, comparison of biometric data, DNA samples, dental records, or images modified by forensic age progression [6]. Often the posters with photographs of missing children can be found on the streets, in media or newspapers and the family and police hope that strangers will recognise the missing child. The recognition of an unfamiliar person is of interest to many

research groups, as this process clearly differs from recognition of familiar persons. It is commonly thought that humans are good in recognising faces, especially familiar faces. However, human ability to recognise unfamiliar faces is poor [7]. Several studies showed the low recognition rates when dealing with recognition scenarios involving unfamiliar persons [8,9]. In case of the recognition of missing children the effect of prospective person memory plays a role – the memory in which a person has to remember to engage in an action when he or she encounters a particular cue in the environment. In order for members of the general public to identify a missing child, they must first commit to memory the faces of the children they are to be on the lookout for [10].

In cases in which a child has been missing for many years, the search and identification become very difficult because of changes due to growth and development that may dramatically alter the child's appearance [11]. Forensic age progression is a method that may provide important information concerning the issue of long term missing children, in which outdated photographs of the missing child are used to predict the child's current appearance. Age progression is performed by assessing craniofacial growth patterns and the appearance of the missing child's

* Corresponding author at: Via Mangiagalli 37, 20133 Milan, Italy.
E-mail address: zuzana.caplova@unimi.it (Z. Caplova).

Table 1
Sample description of the age and sex of each subject, who provided facial photographs.

		Sex	Age of the photographs			Sex	Age of the photographs
Image series	A	F	10 years 13 years 15 years 17 years 19 years	Image series	K	F	10 years 13 years 14 years 15 years 16 years 17 years 18 years
	B	M	11 years 12 years 14 years 16 years 19 years		L	M	13 years 14 years 15 years 17 years 19 years
	C	F	14 years 15 years 17 years 18 years 19years		M	F	10 years 11 years 13 years 14 years 20 years
	D	F	10 years 12 years 13 years 14 years 15 years 16 years 18 years		N	F	10 years 11 years 14 years 16 years 18 years 19 years
	E	M	12 years 13 years 14 years 15 years 16 years 17 years 20 years 21 years		O	F	14 years 15 years 16 years 17 years 18 years
	F	F	3years 4years 5years 6 years 7 years 8 years 9 years 10 years 11 years 12 years 13 years 14 years 15 years		P	F	10 years 12 years 14 years 16 years 18 years
	G	F	2years 3years 4years 5years 6years 7years 8years 9years 11years 13years		Q	F	10 years 12 years 16 years 17 years 18 years 19 years
	H	F	10 years 11 years 13 years 15 years 19 years		R	F	3 years 4 years 5 years 6 years 7 years 8 years 9 years 10 years 11 years 12 years 13 years 14 years 15 years

biological relatives at the target age to make an estimation of the current appearance of the child. Effective child age progressions can be hand drawn, software-based or a combination of both. However, in real-life cases the effectiveness of age progressed images seems to be insufficient and fails to provide an advantage for identification [11–13]. Therefore, other methods using facial images are required such as the morphological comparison of two different images of the same person taken at different ages in order to examine whether and how facial features change with age and whether the children represented in different images are the same person.

In past years, the identification of the living through the study of photographic images has become a very important topic due to the increase of the use of photography and video surveillance cameras [e.g. 14–16]. Although this activity seems to be intuitive and simple, judicial developments are far from simple. So far, the methodology of comparing the morphological and metric traits of the face on different images is not entirely conclusive for positive identification purposes [17].

There are three general approaches to facial identification: (1) morphological classification of facial features according to standardized classification systems, (2) photoanthropometry using direct distances between landmarks, angles and indices, and (3) superimposition of the whole face or facial features [17].

Identification by facial images has some innate limitations. Images may be of poor quality, features may not be visible or covered by hair, glasses or clothes. The quality of the camera and lenses, different distances between the individual and the camera, direction of lighting and pixel resolution also needs to be considered. A good-quality image is required for identifying as many details as possible [18]. Recognition accuracy also depends on face positioning and facial expressions. In real-life situations, it is almost impossible to find facial images that are taken under exactly the same conditions, with the same parameters and in the same poses.

When comparing age-different facial images of a child it is important to consider the morphological changes associated with growth at different developmental phases [19–22]. The face continues to change also during adulthood but such changes are slow and less dramatic than those during childhood [23]. The growth of the face is a complex process, not merely a process of size increase. The child's face is not the miniature of the adult face, each of its parts mature differently [13].

A number of longitudinal and cross-sectional studies assessed changes of the face during growth. Sforza et al. [19,20,23] provided information concerning changes in linear distances, areas and volumes related to age and sex of the lips, orbital region, nose and ear from childhood to the adult age, finding that age significantly influenced all analysed measurements. Kau and Richmond [24] also reported that differences in the timing of changes between males and females are significant. Nute and Moss [25] studied the age gradient of change in facial width with the result that little or no increase occurs in the upper face, while small changes were observed in the mid-face and greater increases in the lower face over time.

In addition to facial morphology, soft biometric traits such as moles, scars or freckles may provide additional information about the identity and may facilitate facial recognition. Park and Jain [26] proposed to use demographic information in combination with soft biometric traits for face matching. In case of long-term missing children it is important to consider that also soft biometric traits such as moles, follow a general pattern of evolution through a person's life. Siskind et al. [27] observed the tendency to increase in size among raised but not among flat moles, although there are many exceptions to this pattern. Usually large moles (flat or raised) do not disappear [27].

The aims of this study were to test whether a morphological atlas of facial features can be used as a tool for improving observer subjectivity in the categorization of features, and thus aid personal recognition; to examine whether and which morphological features of the face are assessed by the observers as stable over time, and the potential usefulness of such features for the recognition of missing children on age-different images; and to assess the utilization of the position of moles as a method facilitating recognition of the same person on age-different photographs by using an automatic computational approach (an algorithm specifically developed for this purpose).

2. Material and methods

The study was divided into two different phases. In the first phase, the Anthropological Atlas of Female Facial Features [28] and the Anthropological Atlas of Male Facial Features [29] were used for categorization of facial features. The features were evaluated in age-different images and the performance of different observers was tested. In the second phase, the MATLAB software was used to develop an algorithm to be used for the comparison of age-different images and to explore the possibility to use moles as a method facilitating recognition.

A series of good-quality photographs of the face in frontal or slightly lateral view taken at different ages were collected from 16 Italian subjects (13 females and 3 males), who signed an informed consent for the use of images. At least one mole was visible on each of the provided facial photographs. Table 1 shows the age distribution of the photographs provided by each person. In total, 110 images were acquired.

Table 2
Ten landmarks used in the computation of the homography.

Reference points	Description
Exocanthion (*ex* R., *ex* L.)	Outer corner of the eye
Endocanthion (*en* R., *en* L.)	Inner corner of the eye
Alare (*al* R., *al* L.)	The most lateral point on each alar contour
Gnathion *gn*	The most inferior midline point on the mandible
Sides of philtrum (*ph* R., *ph* L.)	The most lateral part of subnasal depression where the philtrum margins meet the upper lip vermilion border
Centre of the cupid's bow *lp* (lip tubercle)	Centre of the shape created by the upper lip vermilion border within the philtrum peaks

Fig. 1. Projection of landmarks from one image to another and selection of the mole.

2.1. Evaluation of facial features by using an Anthropological Atlas

A questionnaire was developed to test inter-observer differences in classifying facial features of children by using an Anthropological Atlas. In addition, age-different photographs of the same person were used to test whether and which facial features are assessed as stable over time, i.e., are assigned the same description at different ages by the observers. The evaluation of the features on age-different images was used to assess the inter-observer agreement and to test, which facial features were perceived as being stable over time by all observers.

The study was undertaken at the Institute of Legal Medicine, University of Milan, Italy. Six observers with a degree in biology or medicine completed the test. Each observer received a copy of The Anthropological Atlas of Male Facial Features, a copy of The Anthropological Atlas of Female Facial Features, the series of facial images and answer sheets. The observers were asked to classify all observable facial features based on the illustrations and descriptions found in the Atlases. Each Atlas includes descriptions, drawings and photographic examples of 43 facial features for males and 45 facial features for females. For the

Table 3

Example of the categorization pattern of facial features in one photograph provided by three observers.

Facial feature	Evaluation		
	Observer 1	Observer 2	Observer 3
Head shape	Wedge	Pentagon	Wedge
Frontal height	Average	High	High
Frontal breadth	Broad	Average	Average
Frontal hairline	Convex	Straight	Straight
Eyebrow height	Average	Slight. curve	Slight. curve
Eyebrow density	Bushy	Bushy	Bushy
Eyebrow shape	Slight. curve	Slight. curve	Slight. curve
Mono-brow	None	Present	None
Upper eyelid-eyebrow distn.	Average	Average	Average
Upper eyelid	Visible	Visible	Visible
Inner eyebase	Wide	Average	Average
Lid axis	Straight	Straight	Straight
Nasal root	Narrow	Average	Narrow
Nose bridge length	Long	Average	Average
Nose bridge breadth	Broad	Average	Average
Nose bridge process	Symmetrical	Downwards broader	Symmetrical
Nose tip shape	Bulbous	Bulbous	Round
Nose tip cleft	Not clefted	Not clefted	Not clefted
Nasal breadth	Broad	Average	Broad
Philtrum height	Average	Average	Low
Philtrum depth	/	Shallow	Shallow
Philtrum shape	Divergent	Divergent	Parallel
Upper lip notch	/	Straight	Wavy
Labial breadth	/	Average	Broad
Mouth corner	/	Straight	Slightly up
Upper vermillion	/	Thin	Thin
Lower vermillion	/	Average	Thin
Chin shape	Pointed	Pointed	Pointed
Chin height	Height	Average	Height
Chin transition	Transition	No transition	No transition
Chin dimple	Absent	Absent	Absent
Ear protrusion	Close fitting	Slightly pronounced	Slightly pronounced
Transition head neck	Neck clearly narrowed	Neck clearly narrower	Neck slightly narrower
Pronunciations of cheek bones	Strongly pronounced	Strongly pronounced	Strongly pronounced

/ - not evaluated.

Table 4

Percentage (minimum, maximum, mean) of miss-match between three observers evaluating line-up of photographs of two female and two male subjects.

Feature	Minimum	Maximum	Mean	Category
Head shape	52%	64%	58.00%	shape
Frontal height	36%	64%	50.00%	size
Frontal breadth	52%	80%	66.00%	size
Frontal hairline	44%	92%	68.00%	shape
Eyebrow height	36%	44%	40.00%	size
Eyebrow density	8%	44%	26.00%	shape
Eyebrow shape	12%	36%	24.00%	shape
Mono-brow	4%	36%	20.00%	shape[a]
Upper eyelid-eyebrow dist.	12%	48%	30.00%	size
Upper eyelid	24%	44%	34.00%	shape
Lid axis	28%	48%	38.00%	shape
Nasal root	56%	64%	60.00%	size
Nose bridge length	56%	100%	78.00%	size
Nose bridge breadth	20%	64%	42.00%	size
Nose bridge process	8%	36%	22.00%	shape
Nose tip shape	20%	72%	46.00%	shape
Nose tip cleft	0%	0%	0.00%	shape[a]
Nasal breadth	28%	52%	40.00%	size
Philtrum height	20%	48%	34.00%	size
Philtrum depth	32%	84%	58.00%	size
Philtrum shape	32%	72%	52.00%	shape
Upper lip notch	40%	72%	56.00%	shape
Labial breadth	52%	88%	70.00%	size
Mouth corner	64%	80%	72.00%	shape
Chin shape	28%	68%	48.00%	shape
Chin transition	44%	68%	56.00%	shape
Chin dimple	8%	28%	18.00%	shape[a]
Ear protrusion	28%	48%	38.00%	shape
Transition head neck	32%	40%	36.00%	shape
Pronunciations of cheek bones	52%	84%	68.00%	shape[a]

[a] Presence of "absent" category.

purpose of this study, 34 facial features were assessed from the Atlases, since the facial features visible only in lateral view could not considered due to the frontal position of the photographs utilized in this test.

The photographs were divided into two groups: 1) Individuals A-H (58 facial photographs in total), and 2) individuals K-R (52 facial photographs in total; Table 1). Each group was evaluated by three observers. The division was done in order to make the study feasible in terms of the time the observers needed to invest to evaluate the high number of features on each of the photographs (maximum $34 \times 58 = 1972$ features). The test consisted of series of age-different facial images of a given person printed on one sheet and corresponding answer sheets. There were no time restrictions for completing the test. The observers were not asked to state if photographs belong to the same child.

The categorisation of facial features was compared between observers, and between the age-different images of one subject. Mismatch percentages were calculated by facial feature for three pairs of observers for the series of 25 age-different photographs of two female and two male subjects.

2.2. Moles as age-stable features

The photographs used in the first phase were also the basis of this test. From the full sets of age-different photographs, photographs with similar head position (mainly in frontal view) and well-visible mole were purposefully selected, in order to limit the effect of head orientation and image quality on the result. In addition, the pairs of photographs for comparisons between different individuals were purposefully selected to represent persons of the same sex who have moles in approximately the same position or at least in the same facial region, and for comparisons between age-different photographs of the same person and also of different persons pairs of photographs with the largest possible age gap between the youngest and the oldest face were selected.

For analysis, the final sample of 68 pairs of images of 16 subjects was divided into three groups:

A. 22 photographs were used for comparisons with their copies, among others to test the precision in setting the landmarks (intra-observer error);
B. 27 pairs of age-different photographs of the same subject were used to examine the change of the position of the mole due to the growth; and
C. 19 pairs of age-different photographs of different subjects were used to test the sensitivity and specificity of the approach.

An algorithm was developed to plot and measure the change of the position of the moles by using the Computer Vision toolbox of MATLAB. The comparison between the images was possible by using a face detection algorithm based on Viola-Jones algorithm [30]. The Viola-Jones algorithm provides real-time image-based object detection, such as face detection. The method learns a classifier from a large number of labelled
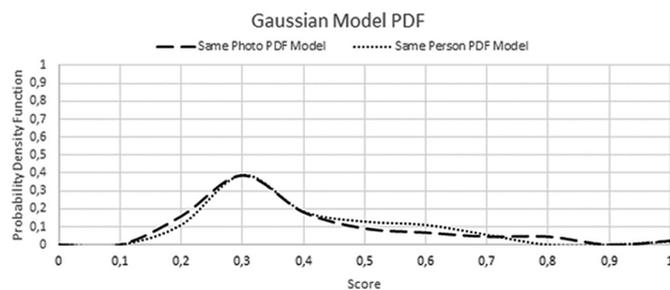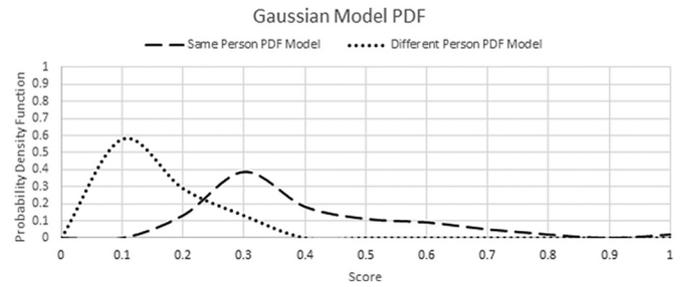

Fig. 3. Probability Density Function (PDF) of Group A + Group B vs Group C: the different shapes of the similarity score curves illustrate that it is possible to discriminate between the same person and different persons by the position of their moles.

data. The training is based on Haar features found on integral image cleverly selected with an Adaboost algorithm that finds the best classifier as a linear combination of weaker classifier. The images of the faces were cropped and scaled in the training process, in order for the output of such a classifier to be normalized.

In order to identify the orientation of the person's face, fixed landmarks were manually selected on pairs of facial images. The landmarks were set by two operators. *Exocanthion, Endocanthion, Alare* and *Gnathion* were selected as they were shown to be points with low inter-observer dispersion [31]. In total, ten reference points were used to identify the orientation of the face in the photographs (Table 2). These reference points were used to find the homographic transformation between pairs of two-dimensional face representations in order to create a univocal mapping of the faces' skin surfaces. The moles were then manually identified and their positions were compared in pairs of images. The distance in pixel was computed between a mole in a reference photograph and a mole in a comparative photograph projected onto the reference one. The shape of the moles as such was not considered. Fig. 1 illustrates the ten reference points used for the homography of a face with another, with the mole indicated as a cross.

Since the homographic transformation introduces an error in the projection of a pixel in another face, the accuracy of such mapping was assessed. To do so, the Homography Distance was introduced, which indicates the accuracy of the superimposition and is given as the mean distance between the ten reference points that had been taken into account for mapping between two images. Since the images were normalized, the mean distances given in pixels were comparable. In addition, the Mole Distance estimated the distance between two moles (located in approximately the same position or similar area in the faces, for example on the right cheek) in the pair of images projected onto each other as explained previously. Consequently, the similarity score represents a percentage of similarity in the position between two moles taking into account the accuracy of the homography between one face and another. Since the Homography Distance is constant and

Fig. 2. Probability Density Function (PDF) of Group A vs Group B: The similar shape of the similarity score curve illustrates that moles do not change position with facial growth, so the person can be recognized as such on age-different images by the position of the mole.
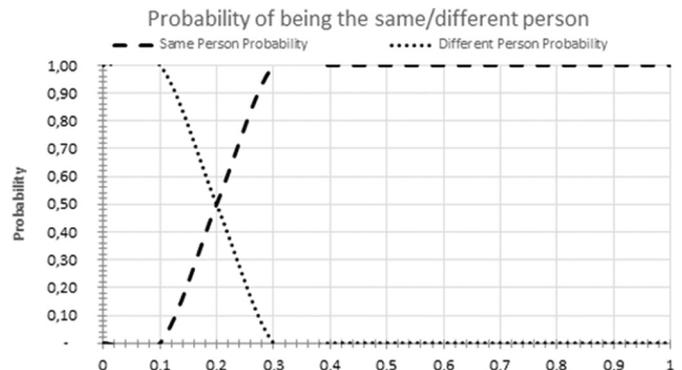

Fig. 4. Probability scores plotted against the probability of classification as the same person or not: the threshold similarity score is 0.22.

Table 5
Facial features grouped into three categories based on low, medium, and high mismatch percentages among observers.

LOW Min 0%–max 48% (mean 25% and lower or 5 best performing)	Eyebrow shape Mono-brow Nose bridge process Nose tip cleft Chin dimple	
MEDIUM Min 20%–max 72% (mean 26%-59%)	Frontal Height Upper eyelid Nose bridge breadth Nose tip shape Philtrum depth Upper lip notch Chin transition Transition head neck Upper eyelid-eyebrow dist.	Eyebrow density Lid axis Eyebrow density Nasal breadth Philtrum shape Chin shape Ear protrusion Philtrum height Head shape
HIGH Min 44%–max 100% (mean 60% and higher or 5[a] worst performing)	Frontal breadth Frontal hairline Labial breadth Pronunciations of cheek bones (Nasal Root)	Nose bridge length Mouth corner

[a] There are actually seven because in two instances the same mismatch percentage was found for two features

negligible, the smaller the Mole Distance the higher the similarity score and vice versa.

The similarity scores were obtained from two observers who assessed all subjects in every category using equation:

$$\text{similarity score} = \frac{\text{Homography Distance}}{\text{Mole Distance} + \text{Homography Distance}}$$

The performance of the algorithm was assessed by probability density functions, and probabilities of belonging to one group or the other as a function of a given similarity score. To increase the sample size, the similarity scores obtained by the two observers were combined in these analyses.

3. Results

3.1. Evaluation of facial features by using the Anthropological Atlas

Not all facial features in the frontal view could be evaluated due to the poor quality of the images. On average 63% of all facial features listed in Atlas were observable and possible to evaluate in one photograph of a given individual. The classification patterns for the complete set of features on a single image and for age-different images of one person varied considerably among observers, i.e. none of the patterns was the same for all three observers. Table 3 shows an example of the categorization pattern of individual facial features in one photograph provided by three observers. This example illustrates that the overall pattern of categorization differed between observers and also that only a small number of features were classified into the same category by more than two observers even when using the Atlases (in this case 7 out of 34). The mismatch percentage calculated by facial feature for three pairs of observers for the series of 25 age-different photographs of two female and two male subjects are presented in Table 4. The mean inter-observer mismatch percentage was 45.6% (calculated as a mean from all mismatch percentages derived from ($25 \times 3 \times 30$) 2250 comparisons). The mean mismatch percentage for the 11 features

describing size was 53%, while the mean mismatch percentage for the 19 features describing shape was 41%. The nose tip cleft was the only feature that showed complete agreement among observers in the classification and also was evaluated as stable over time in all age-different images.

There was a difference in perceiving the facial features as stable over time among the three observers: for two observers the percentage was 87% and 74%, respectively, while for the third one the percentage was 32%. When only the image series in which features were reported as stable over time were considered, in 47.5% of cases at least two observers agreed on the classification of the feature. Overall, all three observers agreed on the classification of all age-different images of one person in 14 cases (12%) concerning nine different variables. Interestingly, in all these cases the features were also perceived as being stable over time. Apart from the nose tip cleft, three observers agreed on the classification and the stability over time for the chin dimple for three image series and two observers were in agreement for the fourth one.

3.2. Moles as age-stable features

The intra-observer error in setting the reference points was 5% for each of the observers, which is classified as good or acceptable error.

For the comparison of facial images of the same subject with a mole at the same age and of the same subject at different ages (Group A vs Group B), the mean similarity score is 35%, with a standard deviation of 18% and a range of 10% to 94%, meaning that the data from age-different images are too similar so it does not seem possible to discriminate the age of a subject as a function of the position of her/his mole, indicating that the moles are age-stable features (Fig. 2).

Fig. 3 shows the similarity scores of the comparisons among the same and age-different images of the same person and those of different persons (Group A + Group B vs Group C). In this comparison, two different statistical trends were observed: the first corresponds to the same-person similarity score, with a mean value of 0.35 (SD = ±0.18), while the second corresponds to the different-person similarity score, with a mean value of 0.1 (SD = ±0.07). The difference

between the means of the similarity scores indicate that it seems to be possible to differentiate between the same and a different person, and thus to classify whether a new entry belongs to the given (same) person or not. Consequently, Fig. 4 shows the probability of a facial image with a mole being of the same or different person with similarity scores below 0.1 indicating that two images represent different subjects; while scores above 0.3 indicating that images represent the same person based on the position of their mole. The similarity score of 0.22 represents the threshold of discriminating between the same and a different person. By using this threshold, 78% of facial images returned a true positive result, 8% a false negative and 92% a true negative when using the position of the mole as reference. In other words, the sensitivity of the method was 78%, the specificity 92%, and the miss rate was 8%.

4. Discussion and conclusion

This study set out to test how different observers categorize facial features in age-different photographs of the same person by using an Anthropological Atlas, and which facial features are perceived as stable over time. Depending on the results, the method could be used for the assessment of age-different photographs of missing children and aid their recognition over time by observers unfamiliar with the child. In addition, the position of moles on facial images was compared by using a specifically developed algorithm to verify whether the position of moles changes with age and whether this method can aid recognition of the same person over time and also differentiate a given person from other persons with moles.

When using the Anthropological Atlases for the categorization of facial features, this study showed that the inter-observer agreement in evaluating the features was poor; the mean mismatch percentage among observers being 46%. In comparison, Ritz-Timme et al. [32] reported that when using one of the Atlases used in the present study, "The Anthropological Atlas of Male Facial Features" the mean inter-observer mismatch percentage was approximately 39%, while the intra-observer mismatch percentages ranged from 19% to 30%. In general, none of the classification patterns for the complete set of features on a single image and for age-different images of one person was the same for all three observers.

The mismatch percentages for individual facial features ranged from 0% to 100%. To illustrate the usefulness of individual facial features for comparisons, the assessed features were grouped into three categories based on low, medium and high mismatch percentages among observers (Table 5). The lowest mismatch percentages were observed for facial features that included the category of the element being absent and described shape: the nose tip cleft, the chin dimple and the mono-brow.

The five "best-performing" (with low mismatch percentages) features were all features describing shape. Among the "worst-performing" (with high mismatch percentages) features, four were describing size and three were describing shape (frontal hairline, mouth corner and pronunciation of cheekbones). Considering the three features describing shape that showed high mismatch percentages among observers, frontal hairline may change with age due to changes in hairstyle (when observed on images), mouth corner is highly dependent on facial expression and pronunciation of cheekbones may be difficult to assess on frontal images. On average, the mismatch percentages were higher for features describing size than for those describing shape, which may be due to insufficiencies in categorization and descriptions of features relating to size. Although it would be expected that using an Anthropological Atlas with descriptions and illustrations of the respective categories of facial features would minimise the inter-observer error rates, the findings of this study indicate that this is not the case when using the two particular Atlases. Consequently, the high mismatch percentages may be the result of the inherent limitations of the Atlases used, and different results may be possibly achieved when using a different Atlas.

Since the human face (face in general and its elements, for example the nose) is growing faster and changes more significantly in length than in width [25], the features describing lengths may be expected to have changed (or would be perceived as having changed) due to the changes in facial proportions occurring with natural growth and development (depending on the age range of the individual image series). On the other hand, other features, such as the absence of a chin dimple, would be expected to be stable over time. In this study, the nose tip cleft and the chin dimple, which showed the highest agreement among observers regarding both the categorization and the stability over time in age-different images features, were indeed features that had the option to be categorized as absent. A more detailed analysis of the changes over time in given features was precluded by the finding that the reported patterns differed among observers, so the essential condition for the method to be used, which is sufficient agreement in the classification among observers, has not been met in the first place.

The algorithm developed to test whether the position of moles may be useful for matching age-different facial images of the same person (are stable over time) or differentiating different persons proved to perform well. By comparing images of the same subject at the same age with respect to the ones at different ages, no differences in the pattern of results were noted; therefore moles seem to be stable in their position over time. Using the position of the mole as a reference point, similarity scores below 0.1 indicating that two images represent different subjects while scores above 0.3 indicating that images represent the same person. The similarity score of 0.22 was identified as the threshold of discriminating between the same and a different person. When using this threshold, the sensitivity of the approach was 78%, the specificity 92%, and the miss rate was 8%. Despite of these promising results, the range of similarity scores for the same and different persons was wide, so the methodology needs to be improved. This may be done by working with better-quality images (which may be difficult to achieve in real-life scenarios, but with the increasing use of high-resolution photography and video not unrealistic), by comparing 3D and 2D facial images, and by minimising human intervention (concerning landmark setting) within an automated system. In addition, a better understanding of the anatomical evolution of moles with growth and development, and potentially also the choice of reference landmarks would be beneficial.

In conclusion, from the two types of morphological assessment of the face in age-different images applied in this study, the first, using the categorization of a number of facial features according to Anthropological Atlas showed great inter-observer differences and thus failed the primary requirement of repeatability that is essential for a method to be acceptable for use in the forensic context. Although there were a few features, such as the chin dimple and particularly the nose tip cleft that were recognized as stable over time and showed low mismatch percentages among observers, their use for recognition and identification purposes would be limited since overall classification patterns of the sets of facial features varied among observers for one image and for age-different images of one person. The second approach, which included an algorithm developed specifically for the purpose of comparing the position of moles across age-different images of one person and among different persons, may have potential for the recognition of missing children over time, since the position of moles seems to be an individualizing feature. However, this study focused on the position of single moles (not their shape) in a limited sample, so further research undertaken on larger samples and with focus on a set of moles on one face regarding their position individually and in relation to each other, possibly also including the shape of the moles would likely strengthen the individualizing potential of the feature. In addition, more research is needed on changes of moles through time and what types of moles are suitable to be considered individualizing. As with any morphological feature, the question remains how many moles would be considered

individualizing "enough", would it be feasible, for instance, to use the position of a mole or a number of moles as a biological indicator of identity solely in association with the biological profile of a person? Although the present study is far from being able to answer this question, research within other forensic fields, such as tool mark and material analysis may provide inspiration, including the debate about uniqueness in forensic sciences, the type of most appropriate statistical approaches for data presentation and the best practice in interpretation and presentation of the results related to identification.

References

[1] J. Koudelová, J. Dupej, J. Brůzek, P. Sedlak, J. Velemínská, Modelling of facial growth in Czech children based on longitudinal data: age progression from 12 to 15 years using 3D surface models, Forensic Sci. Int. 248 (2015) 33–40.

[2] Missing children Europe, facts and figures, http://missingchildreneurope.eu/figures 2016 accessed 07.03.2016.

[3] International comission on missing persons, who are the missing? http://www.icmp.int/the-missing/who-are-the-missing/ 2014 (accessed 07.03.2016).

[4] J. O'Connell Davidson, Telling tales: child migration and child trafficking: stories of trafficking obscure the realities for migrant children, Child Abuse Negl. 37 (2013) 1069–1079.

[5] A. Abu-Ali, M. Al-Bahar, Understanding child survivors of human trafficking: a micro and macro level analysis, Procedia Soc. Behav. Sci. 30 (2011) 791–796 (2nd World Conference on Psychology, Counselling and Guidance).

[6] National center for missing and exploited children, missing children, http://www.missingkids.org/Missing 2016 (accessed 07.03. 2016).

[7] P.J.B. Hancock, V. Bruce, A.M. Burton, Recognition of unfamiliar faces, Trends Cogn. Sci. 4 (2000) 330–337.

[8] A.M. Magreya, A.M. Burton, Unfamiliar faces are not faces: evidence from a matching task, Mem. Cogn. 34 (2006) 865–876.

[9] R. Kemp, N. Towell, G. Pike, When Seeing should not be Believing: photographs credit cards and fraud, Appl. Cogn. Psychol. 11 (1997) 211–222.

[10] J.M. Lampinen, J. Arnal, J.L. Hicks, The Effectiveness of Supermarket Posters in Helping to Find Missing Children, J. Interpers. Violence 24 (2009) 406–423.

[11] J. Lampinen, J.T. Miller, H. Dehon, Depicting the missing: prospective and retrospective person memory for age progressed images, Appl. Cogn. Psychol. 26 (2012) 167–173.

[12] S.D. Charman, R.N. Carol, Age progressed images may harm recognition of missing children by increasing the number of plausible targets, J. Appl. Res. Mem. Cogn. 1 (2012) 171–178.

[13] K.T. Taylor, Forensic Art and Illustration, CRC Press LLC, Florida, 2001 (608 pp.).

[14] R.A. Halberstein, The application of anthropometric indices in forensic photography: three case studies, J. Forensic Sci. 46 (2001) 438–441.

[15] R. Moreton, J. Morley, Investigation into the use of photoanthropometry in facial image comparison, Forensic Sci. Int. 212 (2011) 231–237.

[16] N. Lynnerup, M. Andersen, H.P. Lauritsen, Facial image identification using Photomodeler, Legal Med. 5 (2003) 156–160.

[17] C. Cattaneo, A. Cantatore, R. Ciaffi, D. Gibelli, A. Cigada, D. De Angelis, R. Sala, Personal identification by the comparison of facial profiles: testing the reliability of a high-resolution 3D-2D comparison model, J. Forensic Sci. 57 (2012) 182–187.

[18] J.P. Davis, T. Valentine, R.E. Davis, Computer assisted photo-anthropometric analyses of full-face and profile facial images, Forensic Sci. Int. 200 (2010) 165–176.

[19] C. Sforza, G. Grandi, F. Catti, D.G. Tommasi, A. Ugolini, V.F. Ferrario, Age- and sex-related changes in the soft tissue of the orbital region, Forensic Sci. Int. 185 (2009) 115.e1–115.e8.

[20] C. Sforza, G. Grandi, M. De Menezes, G.M. Tartaglia, V.F. Ferrario, Age- and sex-related changes in the normal human external nose, Forensic Sci. Int. 204 (2010) 205.e1–205.e9.

[21] S.E. Bishara, J.R. Jakobsen, T.J. Hession, J.E. Treder, Soft tissue profile changes from 5 to 45 years of age, Am. J. Orthod. Dentofac. Orthop. 114 (1998) 698–706.

[22] L. Vahdettin, Z. Altug, Longitudinal soft-tissue profile changes in adolescent, J. Orofac. Orthop. 6 (2012) 440–453.

[23] C. Sforza, G. Grandi, M. Binelli, C. Dolci, M. De Menezes, V.F. Ferrario, Age- and sex-related changes in three-dimensional lip morphology, Forensic Sci. Int. 200 (2010) 182.e1–182.e7.

[24] H.C. Kau, S. Richmond, Three-dimensional analysis of facial morphology surface changes in untreated children from 12 to 14 years of age, Am. J. Orthod. Dentofac. Orthop. 134 (2008) 751–760.

[25] S.J. Nute, J.P. Moss, Three-dimensional facial growth studied by optical surface scanning, J. Orthod. 27 (2000) 31–38.

[26] U. Park, A.K. Jain, Face Matching and Retrieval Using Soft Biometrics, IEEE Trans. Inf. Forensics Secur. 5 (2010) 406–415.

[27] V. Siskind, S. Darlington, L. Green, A. Green, Evolution of melanocytic nevi on the faces and necks of adolescents: a 4 y longitudinal study, J. Invest. Dermatol. 118 (2002) 500–504.

[28] S. Ohlrogge, T. Arent, W. Huckenbeck, P. Gabriel, S.S. Ritz-Timme, Anthropological Atlas of Female Facial Features, Verlag für Polizeiwissenschaft, Frankfurt, 2009 (62 pp.).

[29] S. Aßmann, D. Nohrden, R. Schmitt, P. Gabriel, S. Ritz-Timme, Anthropological atlas of male facial features, Verlag für Polizeiwissenschaft, Frankfurt, 2007 (62 pp.).

[30] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, Conference on Computer Vision and Pattern Recognition, 2001 (9 pp.).

[31] M. Cummaudo, M. Guerzoni, L. Marasciuolo, D. Gibelli, A. Cigada, Z. Obertová, M. Ratnayake, P. Poppa, P. Gabriel, S. Ritz-Timme, C. Cattaneo, Pitfalls at the root of facial assessment on photographs: a quantitative study of accuracy in positioning facial landmarks, Int. J. Legal Med. 127 (2013) 699–706.

[32] S. Ritz-Timme, P. Gabriel, Z. Obertová, M. Boguslawski, F. Mayer, A. Drabik, P. Poppa, D. De Angelis, R. Ciaffi, B. Zanotti, C. Cattaneo, A new atlas for the evaluation of facial features: advantages, limits, and applicability, Int. J. Legal Med. 125 (2011) 301–306.