

ARTICLE

Syndrome identification based on 2D analysis software

Stefan Boehringer^{*1}, Tobias Vollmar¹, Christiane Tasse¹, Rolf P Wurtz²,
Gabriele Gillessen-Kaesbach¹, Bernhard Horsthemke¹ and Dagmar Wieczorek¹

¹Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany; ²Institut für Neuroinformatik, Ruhr-Universität Bochum, Bochum, Germany

Clinical evaluation of children with developmental delay continues to present a challenge to the clinicians. In many cases, the face provides important information to diagnose a condition. However, database support with respect to facial traits is limited at present. Computer-based analyses of 2D and 3D representations of faces have been developed, but it is unclear how well a larger number of conditions can be handled by such systems. We have therefore analysed 2D pictures of patients each being affected with one of 10 syndromes (fragile X syndrome; Cornelia de Lange syndrome; Williams–Beuren syndrome; Prader–Willi syndrome; Mucopolysaccharidosis type III; Cri-du-chat syndrome; Smith–Lemli–Opitz syndrome; Sotos syndrome; Microdeletion 22q11.2; Noonan syndrome). We can show that a classification accuracy of >75% can be achieved for a computer-based diagnosis among the 10 syndromes, which is about the same accuracy achieved for five syndromes in a previous study. Pairwise discrimination of syndromes ranges from 80 to 99%. Furthermore, we can demonstrate that the criteria used by the computer decisions match clinical observations in many cases. These findings indicate that computer-based picture analysis might be a helpful addition to existing database systems, which are meant to assist in syndrome diagnosis, especially as data acquisition is straightforward and involves off-the-shelf digital camera equipment.

European Journal of Human Genetics (2006) 14, 1082–1089. doi:10.1038/sj.ejhg.5201673; published online 14 June 2006

Keywords: syndrome diagnosis; face; facial appearance; statistical discrimination; learning

Introduction

Making a diagnosis in a dysmorphic child requires a high degree of experience and expertise. Although some steps in finding the diagnosis are highly formalized, for example, database searches, others are recalcitrant to standardization. Among these are the physical examination and the evaluation of the overall impression of a patient by the examiner. Imprecise and nonstandardized nomenclature, especially of facial features, places a major difficulty for the

communication between clinical geneticists. Other more subtle problems involve the very process of perception, which is subject to influence from many psychological factors, like crossinfluences of perceptions.¹ In the wake of these problems, various attempts have been undertaken to characterize facial features on the basis of objective measurements; among them are photogrammetry and anthropometry, which are summarized elsewhere.^{2,3} One recent study applies photogrammetry to 3D scans.⁴ The study shows that landmarks can be positioned reliably, an assumption underlying 2D- and 3D-based systems for syndrome classification. Here we would like to focus on computer-based techniques, which promise to help the clinician in the diagnostic process of syndrome identification by providing an automated analysis of the face.

*Correspondence: Dr S Boehringer, Institut für Humangenetik, Universität Duisburg-Essen, Hufelandstr. 55, Essen 45122, Germany.
Tel: +49 201 723 4567; Fax: +49 201 723 5900;
E-mail: correspondence@s-boehringer.org
Received 7 February 2006; revised 13 April 2006; accepted 13 April 2006;
published online 14 June 2006

Physicians, not particularly trained in dysmorphology, could potentially benefit by being able to follow a formalized procedure to establish diagnostic hypotheses, which certainly remain to be evaluated and verified by the specialist. Two approaches have emerged that try to automate the procedures of face analysis. These explore either 2D images² or 3D representation of faces.^{3,5} We have contributed to the 2D approach and shall show extended results later. We now briefly describe the technical similarities and differences between the systems and refer to the original publications for details. Both systems start by capturing a raw data set, which is acquired by either a standard 2D digital camera or by a 3D surface scanner. Next, both systems identify landmarks of the face (nose tip, lip edges, etc) with coordinates of points in the data sets (termed the correspondence). In the 2D approach, the coordinates of landmarks are enriched by a representation of the texture in a neighbourhood of the landmark, which is given by a Gabor wavelet transformation. In its current version, the texture of the face is ignored in the 3D approach. Rather, the 3D coordinates describing the surface of a face are used for discrimination purposes. Both approaches use landmarks to define corresponding regions in the face, which allows for a numeric representation of the face with numbers corresponding to well-defined features. Both methods seem to benefit from manual intervention in landmark placement.^{2,6} Therefore, both systems result in an objective description of the face, although quite different information is being used. An important question regarding automated procedures and their practical relevance is how well a sizeable number of syndromes can be dealt with simultaneously. Potentially, several 100 syndromic conditions with facial dysmorphisms could be relevant in clinical practice. Here we show an analysis of the classification of 10 syndromes to compare it with earlier results to approach this question. Another important question deals with the decision process during classification. It is reassuring to see a computer using criteria, which are similar to those of humans and even additional criteria that might be implicit to the human decision. We deal with this question by visualization of the decision process.

Materials and methods

Probands

An extensive set of pictures of individuals representing 10 syndromes was collected. Written informed consent was given by parents of the probands or by the probands themselves. Details about the cohorts can be found in Table 1. On average about 12 individuals are present in each group. The age of individuals varied from 1 to 40 years. Results of cytogenetic, molecular or biochemical analyses were obtained for all syndromes, except for Cornelia-de-Lange syndrome, for which no molecular test was available at the time of recruitment.

Table 1 Characterization of the data set

Condition	Age range	n
Microdeletion 22q11.2	1–12	23
Cri-du-chat syndrome	1–17	16
Cornelia de Lange syndrome	5–33	12
Fragile X Syndrome	4–14	12
Mukopolysaccharidosis III	5–25	10
Noonan syndrome	1–40	18
Prader–Willi syndrome	4–21	12
Smith–Lemli–Opitz syndrome	1–18	13
Sotos syndrome	1–20	18
Williams–Beuren syndrome	1–29	13

Picture acquisition conditions

Conditions for picture acquisition were standardized as much as possible. Three lighting sources providing soft light were used to illuminate probands. Ambient light was reduced to the extent possible, but could never be eliminated entirely as pictures were primarily taken at meetings of parent support groups in varying settings. A uniform background was used throughout. Standard digital cameras were used to take 2D still pictures (Nikon Coolpix 950, Nikon Coolpix 4500). Additionally, videos using a Panasonic NV-MX350EG were recorded from probands and still pictures were later extracted from the videos. Extracted pictures usually allow for selection of a pose that is exactly rotated into a frontal position, if a rotation of the face was recorded on the video.

Picture standardization and selection

From the collection of pictures for each proband, one picture was selected for the final analyses. The criteria used were sharpness of the image, pose of the face and facial expression. Sharp pictures with a face rotated into a frontal position were selected. By subjective judgement, pictures were further chosen to show as little emotional expression as possible. Pictures were then converted to grayscale and cropped to a size of 256 by 256 pixels. For the probands included in the previous study, we have used the same pictures, which are depicted elsewhere.² We give five examples of each group in Figure 1.

Picture analysis

Pictures have been analysed using custom software. Picture analysis is a two-step process. In the first step, a face is searched in the picture and nodes (points) are positioned on the face corresponding to a predefined pattern. Additionally, we have created a data set, for which correspondence of nodes to picture positions was established manually (manual node correspondence). A Gabor wavelet transformation was applied at each node.⁷ This transformation results in 40 coefficients per node, which can be used to locally approximate the picture texture in



Figure 1 Examples of the facial photographs of the 10 syndromes investigated. Each row comprises one syndrome: microdeletion 22q11.2, Cri-du-chat, Cornelia de Lange, fragile X, Mucopolysaccharidosis III, Noonan, Prader-Willi, Smith-Lemli-Opitz, Sotos, Williams-Beuren.

the neighbourhood of each node. Details of this process are given elsewhere.^{2,7}

Data set pruning

We have conducted the analysis based on data sets derived from both automatically and manually placed nodes.

Additionally, we have constructed data sets for which the nodes on the rim of the face have been left out. The rationale behind this step is to reduce noise by removing potentially uninformative nodes, as hair is expected not to be a stable trait for many individuals and rim nodes also include information from the background. Using all permutations, we conducted all analyses with four data sets. Unless noted otherwise, results are reported for manually labelled data with all nodes included.

Statistical analysis

The data set was analysed using several classification techniques (software package R, version 2.1.0; <http://r-project.org>). Specifically, these were linear discriminant analysis (LDA), support vector machines (SVM) and k th nearest neighbours (kNN). These techniques were compared to jet voting (JV), a technique used in a previous paper,² which performs a nearest-neighbour classification at each node and classifies the syndrome taking a majority vote over all nodes. To evaluate classification accuracy, we have performed cross-validation procedures. We used 10-fold cross-validation, the accuracy estimates of which were averaged over 20 runs. We have performed both simultaneous classification and pairwise classification of syndromes. The simultaneous classification serves to evaluate the problem of assigning a syndrome to an unknown face, that is, the problem of diagnosis. Pairwise comparisons of syndromes can be used to evaluate similarity of syndromes and to compare the performance achieved with the current data set with respect to other data sets published thus far.

Before performing LDA, SVM and kNN, we performed a principal component analysis (PCA). PCA transforms a data set into a new coordinate system such that the variances of the new coordinates decrease with increasing rank. The PCA was used to reduce the amount of covariates resulting from picture analysis. In the following analyses, we have used a contiguous block of principal components (PCs) starting with the first PC and including a variable number of subsequent PCs. The number of PCs included was chosen as to optimize classification results. For SVM, we used polynomial kernel functions and optimized classification results varying the degree of polynomials. For kNN, the number of neighbours k was varied. We will refer to these procedures as model selection.

To assess the validity of the classifiers resulting from the statistical procedures, we investigated the linear discriminant (LD) functions resulting from LDA. We have used the LD functions to produce diagrams that display the importance of individual wavelet coefficients in the classification decision. Details are given in Appendix A. These visualizations can be used to exclude artificial classification results, which might, for example, stem from slightly differing acquisition conditions. Additionally, criteria of classifiers can be compared with clinical characteristics of syndromes.

We provide a supplementary document to describe statistical methods in an informal way.

Results

Simultaneous classification

Results for simultaneous classification decisions are reported in Table 2, which demonstrates accuracies after performing model selection. Overall accuracies are shown as well as a breakup of accuracies for individual syndromes. The best overall accuracy of 76% was achieved by LDA. Both SVM and kNN performed worse at 70 and 68%, respectively. The model selection revealed that the simplest choice of classifiers for SVM and kNN with a polynomial degree of 1 for SVM and the number of neighbours also 1 for kNN yielded this result. This reflects the high complexity of the data set, which seems to dictate robust, that is, simple classifiers. JV performed considerably worse at 55%. This result is not surprising as JV does not make use of joint information at different nodes, whereas kNN does. It therefore seems that JV does not have the potential to handle a large number of syndromes. Figures 2 and 3 shows an example of the model selection procedure for LDA showing results for individual syndromes. In general, for 20 included PCs classification rates are better than 70% and change slowly from that point on. However, the joint accuracy is optimal for 36 PCs, indicating that subtle information is contained in PCs in the range from 20 to 40. This diagram also suggests that classifiers could be optimized for syndromes individually and later be combined.⁸ One apparent exception from the good results for most syndromes is Cri-du-chat syndrome, for which performance dropped to 28%.

Using a completely automatically analysed data set, the overall accuracy for LDA drops to 52% (Table 2, column 2).

This is largely owing to failures in localizing the face in some of the pictures from the automatic process resulting in strong noise signals in the data set.

Data sets with excluded rim nodes performed very similar to the corresponding, full data sets. We therefore do not give any detailed results, but report the overall accuracies. For the manual data set, performance was as follows: LDA 75.8% (40 PCs), SVM 68.7% (degree = 1, 18 PCs), kNN 62.2% ($k=1$, 19 PCs) and JV 55%. For the automatically labelled data set, we obtained: LDA 52.8% (12 PCs), SVM 52.4% (degree = 1, 11 PCs), kNN 48.6% ($k=1$, 10 PCs) and JV 50%.

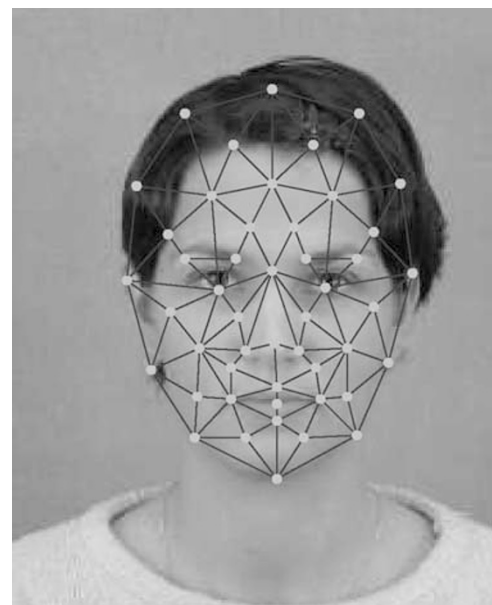


Figure 2 Example result of the automatic procedure matching a graph in a given picture.

Table 2 Classification accuracies of four classification methods with different data sets

Condition	LDA	LDA	SVM	kNN	JV
Data set	Manual	Auto	Manual	Manual	Manual
No. of PCs	36	15	20	21	—
Overall	75.7%	52.1%	70.4%	68.1%	55.8%
22q-	77.4%	96.0%	84.8%	88.6%	90.5%
5p-	27.5%	< 0.1%	26.3%	24.2%	< 0.1%
Cornelia de Lange	83.7%	60.4%	85.8%	60.0%	64.6%
Fragile X	99.9%	99.6%	91.7%	75.0%	79.2%
Mucopolysaccharidosis III	94.2%	95.8%	41.7%	81.7%	50.0%
Noonan	78.4%	1.2%	70.9%	79.4%	78.1%
Prader-Willi	77.9%	85.4%	89.6%	91.2%	60.4%
Smith-Lemli-Opitz	59.0%	< 0.1%	54.5%	64.5%	7.5%
Sotos	72.3%	0.3%	52.7%	34.7%	30.0%
Williams-Beuren	88.1%	89.6%	81.9%	76.9%	59.6%

The first row indicates whether nodes were placed automatically or manually. The second row lists which number of PCs was chosen from the model selection procedure (see Statistical analysis). The degree of polynomials for SVM was chosen as 1. The number of neighbours for kNN was chosen as 1. Bold figures represent overall results.

Pairwise classification

Table 3 shows results for pairwise comparisons of syndromic conditions. These results were achieved using the first seven PCs after selecting the pair of syndromes. Most pairwise comparisons allow for accuracy well above 90%. Problems to discriminate Cri-du-chat syndrome in the joint analysis are reflected by the fact that comparisons with Noonan (79%), Smith–Lemli–Opitz (82%) and Sotos syndromes (81%) are at the lower end of the accuracies achieved for pairwise comparisons. One notable exception is the comparison of Sotos and Noonan syndromes, for which accuracy is 31%, indicating that the correct facial features discriminating these syndromes could not be learned by the classifier. This example is further scrutinized by graphical means below.

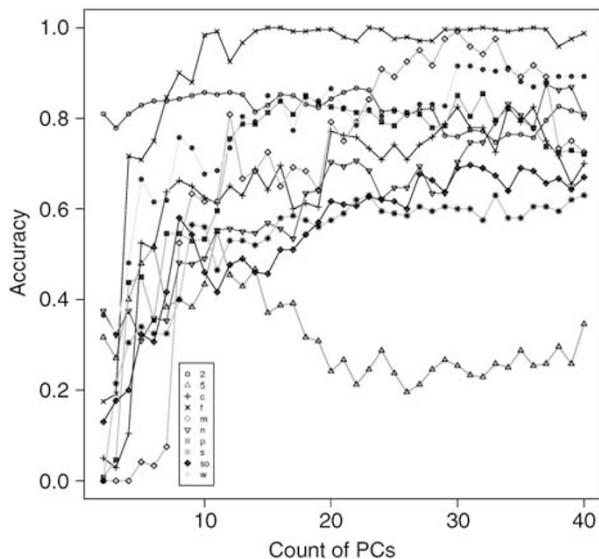


Figure 3 Discrimination accuracy graphs for individual syndromes using LDA illustrating the model selection procedure. On the *x*-axis, the number of PCs used is given and the *y*-axis shows classification accuracies. Syndromes are abbreviated as follows: Microdeletion 22q11.2 (2), Cri-du-chat (5), Cornelia de Lange (C), fragile X (F), Mucopolysaccharidosis III (M), Noonan (N), Prader–Willi (P), Smith–Lemli–Opitz (S), Sotos (So) and Williams–Beuren (W).

Visualization

In general, it is important to verify classification results by scrutinizing decision rules as these rules might pick up certain characteristics in the data set that are not associated with the classification goal, for example, background information in the case of face classification. We have therefore opted to visualize classification rules resulting from a pairwise LDA classification as described above. Figure 4 shows examples from these comparisons, which summarize the important features that can be learned from these visualizations. Figure 4a demonstrates that the decision rule to distinguish fragile X from Cornelia de Lange syndrome mainly relies on the eye, eyebrow and lower nose region. Additionally, features detected at the lower edges of the ears seem to be important. In general, the pairwise comparisons often highlight the eye region for fragile X and the eyebrow region for Cornelia de Lange syndrome, which consequently overlap when this pair is visualized. These two signals allow for a perfect discrimination of these two syndromes. Figure 4b demonstrates the comparison of Cornelia de Lange with Mucopolysaccharidosis type III. Still signals are located in the eyebrow region. The eyes, however, do not play a role for the decision between Cornelia de Lange and Mucopolysaccharidosis III. In general, the decision rule seems to integrate information from the entire face, sparing the hair and chin region. Discrimination between Smith–Lemli–Opitz and Noonan, depicted in Figure 4c, shows that information from nodes on each side of the nose as well as the mouth region is most important. The examples shown so far correspond well with clinical expectation. The worst pairwise classification rate of Smith–Lemli–Opitz and Noonan, however, which is shown in Figure 4d, visualizes a classifier that does not seem to have learned relevant clinical traits. Especially, the fact that the weights seem to be far from symmetric about the vertical axis seems to be a hint that the data set did not comprise enough information to achieve a reasonable discrimination between Smith–Lemli–Opitz and Noonan syndrome.

Table 3 Pairwise classification accuracy of syndrome discrimination using LDA (seven PCs)

	22q–	5p–	Cornelia de Lange	Fragile X	Mucopolysaccharidosis III	Noonan	Prader–Willi	Smith–Lemli–Opitz	Sotos
5p–	0.92	—	—	—	—	—	—	—	—
Cornelia de Lange	0.99	0.83	—	—	—	—	—	—	—
Fragile X	1.00	0.87	0.95	—	—	—	—	—	—
Mucopolysaccharidosis III	1.00	0.89	0.65	0.95	—	—	—	—	—
Noonan	0.89	0.79	0.94	0.99	0.88	—	—	—	—
Prader–Willi	0.99	0.91	0.87	0.88	0.84	0.99	—	—	—
Smith–Lemli–Opitz	0.96	0.82	0.80	0.95	0.77	0.76	0.96	—	—
Sotos	0.89	0.81	0.88	0.94	0.88	0.31	0.90	0.77	—
Williams–Beuren	0.98	0.92	0.89	0.97	0.88	0.91	0.78	0.80	0.89

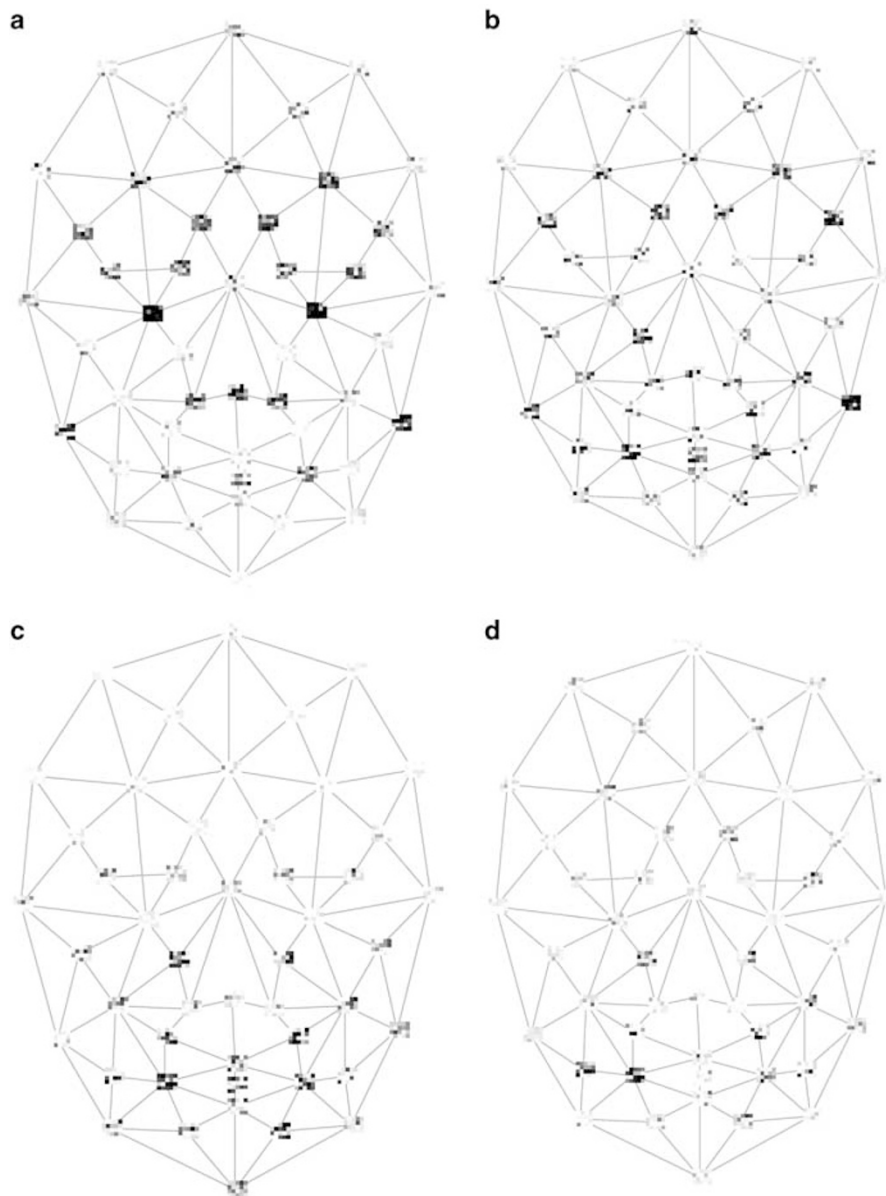


Figure 4 Visualization of pairwise discrimination rules derived from LDA. The boxes at the graph nodes show the collection of coefficients extracted there, and dark colors represent a strong influence of a node/coefficient on the classification decision. For details see text. (a) Fragile X vs Cornelia de Lange, (b) Mucopolysaccharidosis III vs Cornelia de Lange, (c) Noonan vs Microdeletion 22q11.2 and (d) Sotos vs Noonan.

Discussion

Classification results presented in this study are promising with respect to the problem that was posed initially: Can a computer be helpful in analysing faces if more than a few syndromes are involved? However, there are several caveats in drawing further conclusions, which we shall discuss. It is surprising how few examples per class are sufficient to maintain a stable distinction between syndromes even when the number of syndromes is sizeable. Also, the classifiers seem to learn features similar to those described by clinicians in most cases. As two experienced clinicians

reviewed the pictures (GG-K, DW), arguably the data set might be biased towards 'typical' appearances. Unless the computer is instructed as to which examples are typical and which are not, the computer is going to treat all examples equally in the learning process, which might hamper accuracy for unseen examples. Conclusively, larger numbers of examples per class seem desirable but not essential to achieve accurate classification decisions even for a large number of syndromes. For real-world application additional challenges like ethnicity, age and mimics have to be accounted for. Whereas ethnicity and age can be

handled by enriching the study with appropriate probands, the influence of mimics seems to be best accounted for by an explicit model, on account of the tremendous flexibility of mimics. It has to be noted that the description of mimics is an open research problem, which is tackled independently.⁹

One specific result to be discussed is that for Cri-du-chat syndrome. An explanation could be the evolution of the facial phenotype in Chri-du-chat syndrome, which shows marked differences of facial features comparing individuals of different ages. Also this is in accord with an often difficult diagnosis of Cri-du-chat syndrome on clinical evaluations alone.¹⁰ For the other syndromes, it is interesting to see that classification results are generally robust against age variation. Keeping in mind the results for Cri-du-chat syndrome, it is certainly desirable to explicitly account for age, say by regression analysis, but such an analysis has to be postponed until larger data sets are available. We have therefore opted to include probands irrespective of age, as age ranges were roughly similar. Another point concerning the learning algorithms is that the methods had to be changed as compared with the earlier study, which was extended here.² This however is expected in light of the fact that a larger data set contains more information that might be best explored by different methods. As syndrome numbers grow, we expect the statistical methodology to be changed again to sustain reasonable classification accuracy. At the moment LDA, a very simple and robust method, performs best, which is reassuring with respect to the validity of the results. Validity was verified by graphical means and again reassured our findings. If classifiers become more complex (eg nonlinear), visualization might become more difficult, however, one possibility is to use subsets of data sets in the future and establish graphical validation as described here. The large discrepancy of classification results between manually and automatically obtained node correspondences seems to indicate that manual steps cannot be excluded entirely from any facial analysis software that intends to extract as much information as possible. However, finding landmarks in the face was not the primary goal of this study. There are several ways to improve automatic processing of data including optimizing our current methods¹¹ or using additional heuristic methods to find the localization of the face in a given picture (Kalina, in preparation).

Compared with the previous study, accuracy dropped slightly from 80 to 75%. Taking into account the number of syndromes chosen, the relative accuracy (dividing accuracy by *a priori* accuracy) increased from 4 to 7.5, a fact that is promising with respect to the further extension of this study. Further improvements seem to be possible by integrating spatial information from coordinates of landmarks and side views. Preliminary results indicate

that classification can indeed be improved using this information.

Running the statistical analysis employed here on the previously used data set² results in a classification accuracy of >90%.

The most recent paper describing 3D methodology demonstrates an accuracy of 89% for a joint discrimination of five syndromes.⁵ Pairwise comparisons seem to be similar for both approaches. It should be noted that the analyses are not directly comparable as statistical analyses differ somewhat as well as sample sizes. On account of sample size within syndromes, results for the 3D approach should be more accurate. As for 3D analysis, texture information is captured simultaneously with spatial information, it would be very interesting to see how the combination of 3D and texture information would perform. Whereas capturing of 3D information results in a richer data set and allows for excellent visualization as demonstrated recently,⁵ 2D analysis has several advantages in practical use: equipment is cheap and it is easy to handle. It has to be noted that neither 2D nor 3D methods have direct applicability in clinical practice yet, as the number of syndromes is still very small. However, recent results demonstrate that automated technologies have the potential in amending and enriching the process of finding the diagnosis. Finally, integration of clinical information seems to be critical to establish usable databases. In clinical practice, the relative importance of face and other clinical information varies from syndrome to syndrome and from patient to patient. It is our goal to contribute to this integration.

Acknowledgements

We thank all families who took part in this study. This work is supported by grants of the Deutsche Forschungsgemeinschaft (DFG): BO 1955/2-1 and WU 314/2-1. We thank Beate Albrecht, Christian Grünenberg, Yorck Hellenbroich and Peter Meinecke for helping with taking pictures. We thank Bärbel Maus for help with manually labelling pictures and helpful discussions. Additionally, we thank Roxana Moselehi for critically reading the manuscript.

References

- 1 Webster MA, Kaping D, Mizokami Y, Duhamel P: Adaptation to natural facial categories. *Nature* 2004; **428**: 557–561.
- 2 Loos HS, Wieczorek D, Wurtz RP, von der Malsburg C, Horsthemke B: Computer-based recognition of dysmorphic faces. *Eur J Hum Genet* 2003; **11**: 555–560.
- 3 Hammond P, Hutton TJ, Allanson JE et al: 3D analysis of facial morphology. *Am J Med Genet A* 2004; **126**: 339–348.
- 4 Aldridge K, Boyadjiev SA, Capone GT, DeLeon VB, Richtsmeier JT: Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *Am J Med Genet A* 2005; **138**: 247–253.
- 5 Hammond P, Hutton TJ, Allanson JE et al: Discriminating power of localized three-dimensional facial morphology. *Am J Hum Genet* 2005; **77**: 999–1010.

- 6 Hutton TJ, Buxton BF, Hammond P, Potts HWW: Estimating average growth trajectories in shape-space using kernel smoothing. *IEEE Trans Med Imag* 2003; **22**: 747–753.
- 7 Wiskott L, Fellous JM, Krüger N, von der Malsburg C: Face recognition by elastic bunch graph matching. *IEEE Trans Pattern Anal Mach Intell* 1997; **19**: 775–779.
- 8 Freund Y, Schapire R: A decision-theoretic generalization of online learning and an application to boosting. *J Comp Syst Sci* 1997; **55**: 119–139.
- 9 Tewes A, Wurtz RP, von der Malsburg C: A flexible object model for recognising and synthesising facial expressions. in Kanade T, Ratha N, Jain A (eds): *Proceedings of the International Conference on Audio- and Video-based Biometric Person Authentication*, Lecture Notes in Computer Science Berlin: Springer, 2005, pp 81–90.
- 10 Wilkins LE, Brown JA, Nance WE, Wolf B: Clinical heterogeneity in 80 home-reared children with cri du chat syndrome. *J Pediatr* 1983; **102**: 528–533.
- 11 Heinrichs A, Muller MK, Tewes A, Wurtz RP: Graphs with principal components of Gabor wavelet features for improved face recognition. In *Proceedings of the Fifth International Workshop on Information Optics, Toledo 2006*; in press.
- 12 Hastie T, Tibshirani R, Friedman J: *The elements of statistical learning*. New York: Springer, 2001.
- 13 Hair JF, Anderson RE, Tatham RL, Black WC: *Multivariate data analysis*. London: Prentice Hall, 1998.

Appendix A

Visualization

LDA results in a set of decision rules of the form: $\delta_k(x_1, \dots, x_n) = a + b_1x_1 + \dots + b_nx_n$, where δ_k is the discriminant function for class k . The values of discriminant functions for observation $x = (x_1, \dots, x_n)$ can be used to make a classification decision. For pairwise comparisons a single discriminant function is sufficient to determine the decision by deciding for the first class if, say $\delta(x) > 0$ and against it otherwise.¹² The coefficients b_i weigh individual components of x to form the final decision. PCA results in a matrix M that contains the loadings of PCs in the space of wavelet coefficients.^{8,13} If we denote the loading matrix for the first i PCs with M_i (with loadings given column wise), the weights on the wavelet coefficients is given by $w = M_i b$, with $b = (b_1, \dots, b_n)$. We take the absolute values of components of w and standardize w to have the largest component to be 1. In the visualization of pairwise decision rules, each square represents one component of w , with white signifying a standardized weight of 0 and black representing a standardized weight of 1.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)