

Automatic Verification of Parent-Child Pairs from Face Images

Tiago F. Vieira*, Andrea Bottino, and Ihtesham Ul Islam

Politecnico di Torino,
Corso Duca degli Abruzzi, 24 - 10129 Torino, Italy
{[tiago.figueiredo](mailto:tiago.figueiredo@polito.it), [andrea.bottino](mailto:andrea.bottino@polito.it), [ihtesham.ulislam](mailto:ihtesham.ulislam@polito.it)}@polito.it
<http://www.polito.it/cgvg>

Abstract. The automatic identification of kinship relations from pairs of facial images is an emerging research area in pattern analysis with possible applications in image retrieval and annotation, forensics and historical studies. This work explores the computer identification of pairs of kins using different facial features, based on geometric and textural data, and state-of-the-art classifiers. We first analyzed different facial attributes individually, selecting the most effective feature variables with a two stage feature selection algorithm. Then, these features were combined together, selecting again the most relevant ones. Experiments shows that the proposed approach provides a valuable solution to the kinship verification problem, as suggested by the comparison with a different method on the same data and on the same experimental protocol.

Keywords: Kinship verification, SVM, Random Forests, mRMR, SFS.

1 Introduction

The analysis of 2D or 3D facial images is a main research topic in pattern analysis and computer vision. Automatic Kinship Verification (KV) has recently received attention from the research community. KV aims at recognizing the degree of kinship of two individuals from their facial images and has possible applications in historic and genealogic research, automatic management and labeling of image databases, forensics and finding missing family members. This is a challenging problem, which should deal with different degrees of kinship and variations in age and gender.

Automatic KV was first introduced by Fang et al. [1], who analysed a database of 150 pairs of parent-child images. Features were extracted for each face with a simplified Pictorial Structure Model and the best classification achieved 70.69%, outperforming the 67.19% obtained by a panel of human raters on the same data. In order to identify parent-child pairs considering the influence of age factor, Xia et al. [2] proposed an extended Transfer Subspace Learning (TSL), which

* The author is also with the Dept. of Electronics and Systems from the Federal University of Pernambuco, Brazil.

is meant to simplify the recognition task by transferring the knowledge learnt from the similar, but easier task, of recognizing the same parent-child pairs but using images of parents in youth. Classification based on geometric and textural features provided a 60% accuracy. Somanath et al. [3] analyzed the multi-class problem of identifying both parent-child and siblings using Metric Learning, providing 75% and 80% accuracies for, respectively, siblings and parent-child pairs. In [4], Lu et al. proposed a new neighborhood repulsed metric learning (NRML) method for kinship verification. Working on different degrees of kinship, they obtained an average class accuracy of 76.5%.

Recently, in [5] we presented a work on sibling identification. Different facial attributes, related to geometric, holistic and textural features, were first extracted and then combined together. Support Vector Machines (SVM) classification, with the contribution of a Feature Selection process, outperformed the recognition capabilities of a panel of human raters.

This paper extends the main ideas of our previous work in order to analyze the capabilities of different facial attributes to recognize other degrees of kinship and, specifically, of parent-child relationships. The contribution of this work is twofold. First, identifying the facial attributes more fit to tell parent-child pairs from unrelated individuals. Second, comparing on these attributes the accuracy of two state-of-the-art classifiers, namely, SVM and Random Decision Forests (RDF), with previous results in the literature. Our experiments show that some of the facial attributes, when considered individually, are indeed able to classify pairs of parent-child images with performances better than previous works, and, most of all, that the combination of attributes of different natures improves the performance of the final classifier.

The remaining of the paper is organized as follows, in Section 2, we describe the database of parent-child images we used in our experiments. Section 3 details the algorithm we used for tackling the kin verification problem. Results are presented and discussed in Section 4 and in Section 5 we draw the conclusions.

2 Image Database

The recent interest into the KV problem led to availability of some databases of facial images of individuals related by different kinship degrees, such as those used in the papers referenced in the Introduction. They are all composed by an heterogeneous set of images, mostly collected through the Internet, and are characterized by non-uniform illumination, background, pose, expression and different age range and ethnicity of the depicted individuals.

However, some of these databases were not found suitable for our approach. For instance, the parent-child dataset collected by Jiwen Lu et al. [6] is composed by faces cropped in a way that precludes the very first step of our method, *i.e.* the automatic detection of facial landmarks (see Section 3). The one used in [7] contains many grayscale images, thus hampering the use of color-based textural features which have proven to be effective to identify siblings [5]. The database we found more suited to our work was the one collected by Fang et al. [1]. This

dataset consists of 288 individuals' images¹ (144 parent-child pairs). The colour photographs depict several public celebrities with different age, gender and race in slightly different poses (mostly frontal), illumination conditions and expressions (some neutral but often smiling). The individuals are 50% Caucasians, 40% Asians and 10% of other ethnicities; 40% of the samples are father-son pairs, 22% are father-daughter, 13% are mother-son, and 26% are mother-daughter.

3 Algorithm Outline

In the literature, different representations of the information conveyed by faces have been experimented for different tasks. In our previous work on the automatic identification of siblings [5], we found that the discriminative power of separate facial attributes, related to geometric, textural and holistic features, is substantially improved by that of the integration of information of different nature. Based on this consideration, in our work we first analysed individually the contribution of different attributes to the parent-child recognition problem, and then we evaluated different combinations of them.

The outline of the proposed classification algorithm is the following. For each individual, we normalized his/her image and we extracted different feature vectors, one for each feature extraction technique considered. When an image is characterized combining different attributes, their corresponding feature vectors are concatenated. Then we constructed a pair dataset containing all the positive (kin) and an equal number of randomly chosen negative (non-kin) pairs. For each attribute (or attribute group) and each pair, a representative vector is built. Finally, the most relevant pair feature variables were selected and used to train and test a classifier.

3.1 Image Normalization

Image normalization is aimed at reducing the influence of different illumination, background and orientation of the faces. First, 76 facial landmarks were automatically identified using the Active Shape Models (ASM) technique [8] (Fig. 1). Second, the ellipse best fitting the 15 landmarks around the chin was used to segment the face and discard the image background. Third, images were geometrically aligned by making the external corners of the eyes coincident with two reference position. Geometric normalization involved translation, rotation and isotropic scaling of the original images. The size of the final normalized images is 100 by 100 pixels.

3.2 Features Extraction and Characteristic Vectors

The choice of the facial feature used in this work takes into account the lessons learned in our previous experience on sibling verification. In particular, we found

¹ Although the authors reported the use of 150 pairs, 300 images, the online version contains only 144 pairs.



Fig. 1. Examples of parent (1st row) - child (2nd row) pairs. For each individual, we show the original image, the detected landmarks and the normalized image

that the contribution of holistic attributes and of some of the geometric and textural attributes experimented in [5] was negligible. Conversely, more discriminative attributes were found in our preliminary experiments on the parent-child dataset, whose detailed results are not reported here for the sake of brevity. In the following, we briefly summarize the characteristics of the chosen attributes.

In order to extract geometric attributes, we first created a dense reference net composed by 184 segments for each face, which was obtained from the Delaunay Triangulation (DT) of the average position of each ASM landmark over the database images. Then, we computed the following attributes: **SEGS**, the 184 lengths of the DT segments, **ANGLES**, the 342 angles of the triangles obtained from DT, and **RATIOS**, the set of 862 ratios of pairs of DT segments sharing the same vertex. Each pair is considered only once to compute a ratio, *i.e.* when a ratio is computed, its inverse is not considered.

Two image descriptors were used to characterize our samples. The first (**CLID**) is based on color local image descriptors. Their general idea is to encode, for a reference point, the Scale-Invariant Feature Transform (SIFT) [9] descriptors computed separately on each image channel. This allows to obtain a representation of the point neighbourhood which is invariant to several image variations (e.g. light intensity change and shift, light color change and shift; see [10] for details). Among the color descriptors surveyed in [10], we choose C-SIFT [11] since in our preliminary experiments it performed slightly and consistently better. To characterize a sample with this attribute, we computed a C-SIFT descriptor (a vector of 384 components) on each of the 76 facial landmarks.

The second textural attribute is the Weber local Descriptor (**WLD**) [12] which is based on the Weber's law. It states that a just-noticeable difference in a stimulus is proportional to the magnitude of the original stimulus. Translating this concept into image intensities, WLD first characterizes a pixel with (i) the *differential excitation*, computed from the sum of differences of intensity with its neighbors later divided by its intensity, and (ii) the orientation of the pixel gradient. Then, the WLD features, computed using a multi-scale analysis onto each image pixel of the intensity image, are encoded into a histogram containing 2.880 elements.

Each of the described attributes summarizes a facial image into a *characteristic vector*. When more attributes are considered, the *characteristic vectors* of

the images are obtained by simply concatenating the different attributes. For each attribute, or attribute group, the *characteristic vector* $v^{(ab)}$ for a pair of individuals a and b is given by the vector of Euclidean distances, in their respective n -dimensional space, of the corresponding elements of the *characteristic vectors* of a and b . Thus, the *characteristic vectors* of a pair are commutative, *i.e.* $v^{(ab)} = v^{(ba)}$.

3.3 Building the Classifiers

In order to assess the capabilities of the different attributes, or attribute groups, to tell kin from non-kin pairs, we compared two state-of-the-art classifiers, namely, SVM [13] and Random Decision Forests RDF [14], which are widely recognized for their classification performances by computer scientists and machine learning researchers.

Concerning SVM, we used a radial basis kernel, optimizing its parameters by means of a grid search as suggested in [15]. Before applying SVM, each feature variable was linearly scaled to the range [0,1]. This avoids the variables in larger scales to dominate those in smaller ranges and reduces numerical problems in the computation of the SVM kernels.

Since SVM classifiers are likely to be affected by overfitting, being in most experiments the number of features much greater than the number of samples, we applied the two-step feature selection (FS) process described in [5]. First, the features are ranked for relevance according to the min-Redundancy Max-Relevance (mRMR) method [16]. Then, the set including the top 50 mRMR features is further reduced to its optimal size (*i.e.*, that optimizing the SVM classification accuracy) with a Sequential Forward Selection (SFS) scheme. For feature vectors obtained as combination of different attributes, the FS selection was first performed separately on each attribute and then repeated on the aggregation of the selected feature variables.

As for RDF, we first optimized their parameters (*e.g.*, number of trees, tree depth and so on) with a grid search over the parameter space, choosing the set with the lowest out of bag error on the training set. Since at each split node a single feature variable is selected for decision, scaling the feature vectors is not relevant in this case.

Due to their working mechanism, RDF are relatively unaffected by overfitting. Nevertheless, selecting the most relevant features can improve their overall accuracy. To this end, once the optimal parameters were found, we sorted the features according to their Variable Importance [14] and we trained again the RDF with a subset of these variables. The size of this subset was chosen, heuristically, by iteratively increasing the number of candidates in the feature set until the global optimum was found.

4 Results and Discussion

In our experiments, we first computed the accuracies based on the classification of each individual attribute. Then, we evaluated the accuracies obtained by

characterizing each facial image with three different groups of attributes: **GEO-METRIC**, grouping the geometric attributes, **TEXTURE**, combining textural information, and **ALL**, concatenating all the described attributes. The classification results obtained are summarized in Table 1 and organized by attribute, or attribute group, and by classification algorithm (SVM vs. RDF). Results were assessed using stratified five-fold cross-validation (CV), and, hence, we report the average classification rates of each classifier over the different CV rounds.

The following remarks can be drawn:

- concerning the individual attributes, textural features have a higher discriminative power than the geometric ones, with WLD obtaining the best performances (78.0% with SVM);
- the more heterogeneous the information, the better the accuracies. As a matter of facts, grouped attributes performed consistently better than their single components, and the best accuracies were obtained for both algorithms considering all attributes together achieving 81.8% and 77.5% for, respectively, SVM and RDF;
- as for the classification techniques, SVM, in combination with a proper selection of the most relevant features, provides, in this specific problem, consistently better performances than RDF.

One expected result, not shown in Table 1, is that Feature Selection (FS) always provides a significant classification improvement (between 6% and 14%, for SVM, and 1% and 12% for RDF). Concerning the selection process, it is also interesting to analyse the distribution of features surviving the FS pruning for composite attributes and different classifiers (see Figure 2), which could provide some insights into the more relevant facial characteristics to recognize kins.

We can observe the following: (i) RATIOS is the most relevant geometric attribute, suggesting its good descriptive capabilities; (ii) as for textural features, WLD is more relevant than CLID in the TEXTURE groups, but CLID features contribute reasonably when geometric features are added; (iii) when attributes are grouped, features from all attributes are chosen to compose the final vector;

Table 1. Accuracy results. For each attribute and each classification algorithm, we show the percentage of correct classifications and, in brackets, the optimal number of variables selected by the FS process.

	SVM	RDF
SEGS	68.2 (18)	60.1 (40)
RATIOS	73.1 (13)	59.3 (175)
ANGLES	68.9 (30)	57.2 (100)
CLID	74.1 (14)	66.3 (62)
WLD	78.0 (27)	70.6 (250)
GEO-METRIC	74.3 (8)	65.4 (175)
TEXTURE	80.1 (18)	76.1 (150)
ALL	81.8 (29)	77.5 (150)

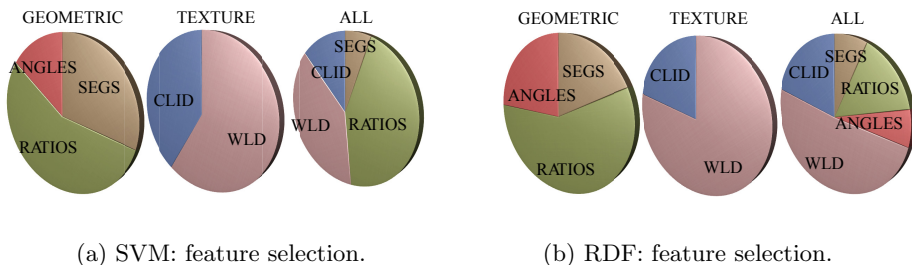


Fig. 2. Feature selection applied to (a) SVM and (b) RDF: distribution of feature variables per type for different attribute groups

the only exception was the ALL group with SVM where ANGLES were discarded (which could be expected since they convey an information similar to SEGS); (iv) when geometric and textural features are combined (ALL group), the latter are preferably selected to compose the final dataset, in particular with RDF.

Finally, we can assess the quality of our results by comparing the classification accuracies of our experiments with that obtained on the same dataset by Fang et al. [1]. The performance of their approach (70.69%) and that of a panel of human raters on the same data (67.19%) are already improved by that obtained in our work with several individual attributes, and outperformed by our best result (81.8%), achieved with the integration of all attributes.

Concluding, the experimental results show that our approach, based on the integration of geometric and textural features, together with a proper selection of the feature variables, is indeed a valuable solution to the automatic KV process.

5 Conclusion and Future Work

We presented an approach for automatically identifying pairs of parent-child images through the extraction and selection of several features from face images. Different attributes, related to geometry and texture data, have been first analysed individually and then combined together to provide higher classification performances.

Simulation results using state-of-the-art classification algorithms show that our combination of features, together with a proper selection of the feature variables, is indeed a valuable solution to the automatic KV process, obtaining high classification accuracies (81.8%) and outperforming previous approaches on the same data.

As future work, we are planning to address the multi-class problem of identifying the specific degree of kinship (e.g., parent-child, parent-grandchild and so on), a problem which has not been yet thoroughly investigated. Another interesting point to be considered is how factors such as gender and age influence a kinship predictor, and possible approaches to alleviate such influences.

References

1. Fang, R., Tang, K.D., Snavely, N., Chen, T.: Towards computational models of kinship verification. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 1577–1580. IEEE (September 2010)
2. Xia, S., Shao, M., Fu, Y.: Kinship verification through transfer learning. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI 2011, vol. 3, pp. 2539–2544. AAAI Press (2011)
3. Somanath, G., Kambhamettu, C.: Can faces verify blood-relations? In: IEEE International Conference on Biometrics: Theory, Applications and Systems, BTAS (2012)
4. Lu, J., Hu, J., Zhou, X., Shang, Y., Tan, Y.P., Wang, G.: Neighborhood repulsed metric learning for kinship verification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2594–2601 (2012)
5. Vieira, T.F., Bottino, A., Laurentini, A., DeSimone, M.: Detecting Siblings in Image Pairs. *The Visual Computer* (to appear, 2013)
6. Lu, J., Hu, J., Zhou, X., Shang, Y., Tan, Y.P., Wang, G.: Neighborhood repulsed metric learning for kinship verification (2012), Database available at <https://sites.google.com/site/elujiwen/download>
7. Xia, S., Shao, M., Luo, J., Fu, Y.: Understanding kin relationships in a photo. *IEEE Transactions on Multimedia* 14(4), 1046–1056 (2012), Database available at <http://www.ece.neu.edu/~yunfu/research/Kinface/Kinface.htm>
8. Milborrow, S., Nicolls, F.: Locating facial features with an extended active shape model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part IV*. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008)
9. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
10. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(9), 1582–1596 (2010)
11. Abdel-Hakim, A.E., Farag, A.A.: CSIFT: A SIFT Descriptor with Color Invariant Characteristics. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2006, vol. 2, pp. 1978–1983 (2006)
12. Chen, J., Shan, S., He, C., Zhao, G., Pietikäinen, M., Chen, X., Gao, W.: Wld: A robust local image descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(9), 1705–1720 (2010)
13. Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* 20(3), 273–297 (1995)
14. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
15. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 27:1–27:27 (2011)
16. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(8), 1226–1238 (2005)