

TOWARDS COMPUTATIONAL MODELS OF KINSHIP VERIFICATION

Ruogu Fang¹, Kevin Tang¹, Noah Snavely², Tsuhan Chen¹

¹Department of Electrical and Computer Engineering, Cornell University

²Department of Computer Science, Cornell University

ABSTRACT

We tackle the challenge of kinship verification using novel feature extraction and selection methods, automatically classifying pairs of face images as “related” or “unrelated” (in terms of kinship). First, we conducted a controlled online search to collect frontal face images of 150 pairs of public figures and celebrities, along with images of their parents or children. Next, we propose and evaluate a set of low-level image features that for use in this classification problem. After selecting the most discriminative inherited facial features, we demonstrate a classification accuracy of 70.67% on a test set of image pairs using K-Nearest-Neighbors. Finally, we present an evaluation of human performance on this problem.

Index Terms— Face recognition, inheritance, feature extraction

1. INTRODUCTION

In today’s digital work, a lot of recent work has used contextual information for recognition and for organization of image collection. Kinship could be another cue for these kinds of problems. So we look into this challenge by applying computer vision and machine learning techniques.

2. OVERVIEW

Human faces have abundant features that explicitly or implicitly indicate the family linkage, giving rich information about genealogical relations. We address the novel challenge by posing the problem as a binary classification task, and extracting discriminative facial features for this problem. Our method works according to the following steps:

Step 1: Parent-child database collection. The facial image database of parent-child pairs is collected through a controlled on-line search for images of public figures and celebrities and their parents or children. We collected 150 pairs using this method, with variations in age, gender, race, career, etc., to cover the wide distribution of facial overview of the facial image database collected.



Fig. 1. Parent-child Database

Step 2: Inherited facial feature extraction. We start with a list of features that may be discriminative for kinship classification (such as distance from nose to mouth or hair color). We first identify the main facial features in an image using a simplified pictorial structures model, then compute these important features and combine them into a feature vector.

Step 3: Classifier training and testing. Using the extracted feature vectors, we calculate the differences between feature vectors of the corresponding parents and children, and apply K-Nearest-Neighbors and Support Vector Machine methods to train the classifier on these difference vectors, as well as a set of negative examples (i.e., image pairs of two unrelated people)

The final outputs provide two different types of information: the most discriminative facial features for kinship recognition and trained classifier to differentiate between true and false image pairs of parents and children.

3. GROUND TRUTH DATA COLLECTION

Previous work in the field of face verification has typically used images of same people but in different lighting conditions, facial expressions, and environmental conditions. These different factors form challenges in the face verification problem, and have become the major focus of computer vision researchers in recent years. [1] Yet to the

best of our knowledge, the work presented in this paper represents the first effort to automatically verify parent-child relationships in image pairs through the analysis of facial features. For this purpose, we first carry out a controlled ground truth data collection through on-line image searching based on knowledge of public figures and family photos.

In order to collect our training and testing data set, we collected 150 pairs of parents and children images from the Internet (Fig 1). Most of the images were found on Google Images, using text queries such as ‘George W. Bush’ and ‘George H.W. Bush’. To ensure that the facial features extracted are of high quality, the face images are selected to be frontal and a neutral facial expression. The database includes around 50% Caucasians, 40% Asians, 7% African Americans, and 3% others; 40% of the 150 images are father-son pairs, 22% are father-daughter, 13% are mother-son, and 26% are mother-daughter. Therefore, it has a wide spread distribution of facial characteristics which depend on race, gender, age, career, etc.

4. FEATURE COMPUTATION

We propose a list of facial features that potentially encompass geological information passed down from parents to descendants, and build the final classifier from them.

4.1 Pictorial Structure Model

In total, we extract 22 features from each of the images. Since some of the features represent the color at a particular point, and some features are sub-windows of a particular facial part, the length of the entire feature vector is much greater than 22. From our experimental results, we have found that these facial features are able to effectively discriminate between related and unrelated parents and children.

To locate the facial parts for distance and extract facial parts, we modify the Pictorial Structure Model [4]. Pictorial structures work by representing an object by a collection of parts arranged in a deformable configuration. In our simplified variant, we consider the spring-like connection between a part and the average position of that part, as if we are fixing the spring on one end to a fixed location.

The appearance of each facial part is modeled and computed using normalized cross-correlation. Normalized cross-correlation works by first defining a template image that we would like to find as a sub-image within our given image. For our purposes, the templates consist of a template left eye, right eye, nose, and mouth. Then, using the templates, we can compute a match score at every position in our given image, which represents how well the template matches at each position. The templates for each part are generated by computing the average image of each part from a separate set of labeled face images.

The spring-like connections that model the deformable configuration are generated using a simple Euclidean distance function. Using the distance function, the cost of placing a part at a given position is computed by taking the distance of that position from the average position of the part. Again, the average positions of each part are computed from a separate set of labeled face images.

Thus, we can model the energy function for our pictorial structures framework as follows:

$$E(x, y) = Match(x, y) - Dist(x, y)$$

Where the Match function is the match score computed from normalized cross-correlation, and the Dist function is the distance from the position to the average part position. Searching the image for the position that maximizes the energy function gives us the most likely position for a particular part. Shown in Fig. 2 are surface plots of the energy function for the Match and Dist functions for the left eye part of a sample image. Combining the two functions, we obtain the total energy function, which is shown in Fig. 3.

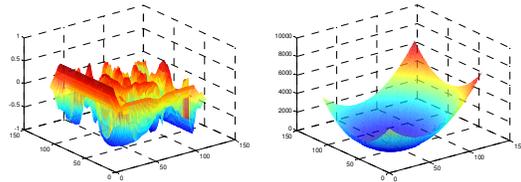


Figure 2 Energy functions for Match and Dist

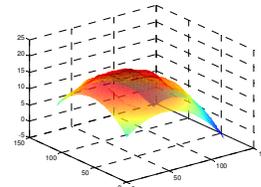


Figure 3 Total energy function

4.2 Feature Extraction

Below, we will briefly discuss general methods used to extract these features.

Color: For features such as eye color, the central position of the facial part is found, and the color at this point is used. For skin color, the center of the nose was used. For hair color, a sub-window of the top of the image was taken, and a mode filter is applied to this sub-window to obtain the most commonly occurring color in this region.

Facial Parts: Facial parts are found by first detecting the central position as well as the boundaries of each part. With these image coordinates, the sub-window for each facial part can be extracted. The length and width are each normalized by the length and width of the entire face, so that faces of different sizes due to changes in pose and position have equivalent representations.

Facial Distances: We use Euclidean distance between the centers of facial parts normalized by the length and width of the face.

Gradient Histograms: For Histogram of Gradients feature, the horizontal and vertical gradient for each image is computed by convolving with gradient filters. Then, these gradients were combined to find the orientation using the arctan function.

5. EXPERIMENTAL RESULTS

To classify the parent-child pairs into two true and false categories, we first create training and test sets by pairing up images in our ground truth set. The positive examples are the true pairs of parents and children and negative examples are each parent with a randomly selected child from the children images who is not his/her true child. In this data preparation method, we are able to create 150 positive examples and an equal size of negative examples. Each example for the training stage is a pair of parent and child from the database, as shown in Figure 5.

For each face image I , we extract the output of k inherited facial feature candidates $f_{i=1\dots k}$ and concatenate these vectors to form a large feature vector $F(I) = \langle f_1(I), \dots, f_k(I) \rangle$. To decide if two face images I_1 and I_2 are of parent-child relationship, we find the difference between their feature vector and feed into the final classifier D which defines our verification function v :

$$v(I_1, I_2) = D(F(I_1) - F(I_2))$$

5.1 Parent-child Data Set

In order to classify the image pairs into true and false parent-child pairs, we use two machine learning methods: K-Nearest-Neighbors with $K = 11$ and Euclidean distance, and Support Vector Machine with a radial basis function (RBF) kernel and the LibSVM package [2]. In our experiments, we classify the data with five-fold cross-validation where 150 positive example pairs and 150 negative example pairs are used as the cross-validation dataset. We performed this classification on the difference between feature vectors of potential parent and child with each of the 44 compiled facial features. The best 14 features in these cross-validation tests are shown in Table 2. The classification performance of these features is fairly stable: the average standard deviation of the classification accuracy across features and over the 150 runs is 1.5892 (min = 1.24, max = 2.03).

In order to combine the individual features, we use a hybrid of a filter-based and wrapper-based approach,



Fig. 5 Positive (left) and negative (right) examples

Similar to [5]. We only consider the features that individually perform above 50% (the chance accuracy rate). We first pick the inherited facial feature which classifies the data the best. The next selected feature is the one that classifies the data the best in conjunction with the first selected feature, and so on. A 6-basic-feature, 10-dimensional feature vector is thus formed. The feature vector is restricted to 6 component due to relatively small the number of image pairs in ground truth (150) and in order to prevent overfitting.

Table 1 Classification accuracy of top 14 features

Feature	Classification Accuracy	F1 Score
righteye-color	61.43%	0.6387
lefteye-color	60.50%	0.6216
skin-gray	59.83%	0.6137
skin-color	59.70%	0.631
lefteyewindow-gray	59.70%	0.6649
lefteyewindow-color	58.13%	0.6154
righteye-gray	57.50%	0.5909
eye2nose-distance	56.53%	0.5762
mouthwindow-color	56.37%	0.5777
righteyewindow-gray	55.97%	0.573
eye2mouth-distance	55.80%	0.5705
HoG-magnitude	55.77%	0.5672
righteyewindow-color	55.63%	0.5671
eye-distance	55.47%	0.5681

The selected features in order of their classification performance after being combined with the previously

selected features are: right eye RGB color (64%, $\sigma=1.28$); skin gray value (63.33%, $\sigma=1.27$); left eye RGB color (65%, $\sigma=2.03$); nose-to-mouth vertical distance (65.67%, $\sigma=1.15$); eye-to-nose horizontal distance (66.67%, $\sigma=1.36$); left eye gray value (70.67%, $\sigma=1.25$). An overall classification accuracy of 70.67% is thus obtained. A support vector machine with a radial basis kernel is trained in a similar procedure with final classification accuracy of 68.60%.

5.2 Human Performance on Parent-child Data Set

While many works have been done human evaluation on face verification of same person [1] or kin recognition signals between siblings [3], there are few published results about how well people perform on the task of recognize kinship between parents and children. Furthermore, it is unknown from which parent that the kin recognition signal is more perceivable. To this end, we conducted several experiments on human verification where 16 participants decided the authenticity of kinship on an example of size 20 pairs which are randomly selected from 150 facial image pairs database.

The experiment result shows that the average accuracy of human performance on the task of kinship verification on Parent-child Data Set is 67.19%, which is 4.9% lower than automatic algorithm that we have designed for this challenge. The standard deviation of human performance is 10.16, much higher than the average standard deviation of 1.5892 of the automatic algorithm. The highest accuracy achieved by human is 90% and lowest 50%.

We also separated the dataset into Father-Son (FS), Father-Daughter (FD), Mother-Son (MS), Mother-Daughter (MD) subsets and evaluated the human performance of kinship verification with different relationship. The average accuracy for FS relationship is 72.94%, FD 54.55%, MS 73.81% and MD 61.29%. It is interesting to notice that the verification accuracy for parents and son is above average while that for parents and daughter is below average. These preliminary results suggest that kin recognition signal is less obvious for human to perceive from daughters than sons, which is in accordance with earlier psychological research results [6].

6. CONCLUSION AND FUTURE WORK

In this paper, we propose a lightweight facial feature extraction algorithm and forward selection methodology to tackle the challenge of kinship verification and find the most discriminative inherited facial features. We have first conducted a controlled on-line image search to collect 150 facial images of parents and children pairs of public figures. The database was collected such that it has a wide spread distribution on facial characteristics which depend on race, gender, age, career, etc. Next we propose 22 low-level features including color cues, distances between features,

facial parts, and global features such as HoG to characterize the kin relationship in human faces. In order to find the most discriminative inherited facial features, we have applied the forward selection methods to kinship verification. Based on the collected ground truth data, we have automatically selected 6 features and form a 10-dimensional feature vector. The classification accuracy into authentic or fictitious is 70.67% using KNN as a classifier. We will make the Parent-child database available publicly for further improvement and evaluation of the data. Finally we make a human performance evaluation on Parent-child dataset with classification accuracy of 67.19% on the whole database and varying accuracy across gender of parents and children.

In order to develop more accurate kinship verification models, we plan on increasing the number of images in our ground truth database and conduct a larger scale human performance evaluation, both on kin relation verification ability and key inherited facial feature identification. Additional future work includes exploring genealogical models to characterize kinship, investigating where in face the cues that signal kinship falls by blacking out facial parts and showing the remaining face, and shedding light on which features of our kinship verification model may be universal vs. gender-dependent, and assessing the influence of age, race etc. in kinship verification so as to form a general classification model. Finally, we also plan to develop novel kinship verification and key inherited facial feature locating user interface to assist solving social problems of lost children searching and historical consanguinity identification.

7. REFERENCES

- [1] Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, Shree K. Nayar, "Attribute and Simile Classifiers for Face Verification," *IEEE International Conference on Computer Vision (ICCV)*, Oct, 2009.
- [2] C.Chang and C. Lin. LIBSVM, "a library for support vector machines," <http://www.csie.ntu.edu.tw/~djlin/libsvm/>, 2001.
- [3] Maria F. Dal Martello, Laurence T. Maloney, "Where are kin recognition signals in the human face?" *Journal of Vision*, pp1356–1366, June 2006
- [4] Pedro F. Felzenszwalb, Daniel P. Huttenlocher, "Pictorial Structures for Object Recognition," *Intl. Journal of Computer Vision*, 61(1), pp. 55-79, January 2005.
- [5] R. Datta, D. Joshi, J.Li, and J. Wang, "Studying aesthetics in photographic images using a computational approach," *Lec. Notes. in Comp. Sci.*, 3953:288,2006.
- [6] Platek, S. M., Raines, D. M., Gallup, G. G. Jr., Mohamed, F. B., Thomson, J. W., Myers, T. E., et al. "Reaction to children's faces: Males are more affected by resemblance than females are, and so are their brains." *Evolution and Human Behavior*, 25, 394–405. 2004.