

2012

## Family Relationship Analysis In Photos

Xiaolong Wang  
*West Virginia University*

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Wang, Xiaolong, "Family Relationship Analysis In Photos" (2012). *Graduate Theses, Dissertations, and Problem Reports*. 4936.

<https://researchrepository.wvu.edu/etd/4936>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

# Family Relationship Analysis In Photos

Xiaolong Wang

Thesis submitted to the  
Benjamin M. Statler College of Engineering and Mineral Resources  
at West Virginia University  
in partial fulfillment of the requirements  
for the degree of

Master of Science  
in  
Electrical Engineering

Guodong Guo, Ph.D., Chair  
Arun A. Ross, Ph.D.  
Xin Li, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia  
2012

Keywords: Family relationship analysis, familial traits, kinship, generation recognition, family photos, non-family photos, discriminative approach, facial features, unconstrained images, support vector machine (SVM).

Copyright 2012 Xiaolong Wang

# ABSTRACT

## Family Relationship Analysis In Photos

Xiaolong Wang

Family relationship analysis has many potential applications, ranging from homeland security through to image search and social activity analysis. In our work, we present five computational problems for family relationship analysis in face photos. Studying these challenging problems is important and useful for semantic image understanding and social context extraction. In our study, the familial traits are learned from pairs of salient local facial parts using discriminative approaches. It is motivated by human perception studies on kinship recognition and the existence of familial traits through genetic inheritance. Second, kinship verification is performed on a pair of faces by integrating the familial traits based on confidence measures. Then, the generation recognition and specific family relationship recognition are explored. Finally, the separation of family and non-family group photos is studied based on a decision that combines multiple pair-wise kinship detections. An image database consisting of both family and non-family group photos is collected, and labeled at different levels of details. Experiments are performed on the database for all five tasks, based on different representations of the facial parts. Preliminary results show that the proposed problems can be addressed with a reasonably good performance. Our encouraging results may inspire more effort from the computer vision and image processing research community.

## Acknowledgments

Firstly, I would like to thank my advisor, Dr. Guodong Guo for his guidance, care, encouragement and help. It is my pleasure to work with him, especially under his good supervision. The knowledge learned from him will benefit my future work.

Secondly, I would also like to thank my committee members: Dr. Xin Li and Dr. Arun A. Ross, for their guidance during my master's study in WVU and their help for the completion of my thesis.

Finally, I would like to thank the Lane Department of Computer Science and Electrical Engineering at West Virginia University for offering me with such a good research environment during my years as a graduate student.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview of Family Relationship Analysis . . . . .	2
1.2	Motivation . . . . .	4
1.3	Major Challenges . . . . .	4
1.4	Facial Parts Based Approach . . . . .	6
<b>2</b>	<b>Family Relationship Analysis</b>	<b>7</b>
2.1	Familial Trait Identification . . . . .	7
2.2	Kinship verification . . . . .	10
2.3	Generation Recognition . . . . .	12
2.4	Specific Family Relationship Recognition . . . . .	14
2.5	Family and Non-Family Photo Discrimination . . . . .	17
<b>3</b>	<b>Feature Descriptors</b>	<b>20</b>
3.1	Histogram of Oriented Gradients (HOG) . . . . .	21
3.2	Local Binary Patterns (LBP) . . . . .	23
3.3	Scale Invariant Feature Transform (SIFT) . . . . .	24
3.4	DAISY feature descriptor . . . . .	27
<b>4</b>	<b>Supervised Machine Learning</b>	<b>29</b>
4.1	Support Vector Machine . . . . .	29
<b>5</b>	<b>Experiments and Discussion</b>	<b>34</b>
5.1	Datasets Description . . . . .	34
5.2	Experimental Results . . . . .	35

**6 Concluding Remarks**

# Chapter 1

## Introduction

Image understanding is an active research topic in computer vision and image processing. The long-term goal of image understanding is to enable a computer to automatically extract semantic information and obtain knowledge from single image or image sequences. Although various techniques have been developed for digital image processing, it is still a great challenge to use machines to extract semantic meanings and gain knowledge from images, comparing with the human-level image understanding.

Given an arbitrary realistic photo, it's easy for the people to tell the story presented in the photo. For example, if we look at the photo shown in Fig 1.0.1, it's easy for the people to get the information as follows:

- (1) There are four persons in the photo.
- (2) They are in a family relationship.
- (3) The relationships in the photo include an infant, his/her mother, his/her grandmother, and his/her great-grandmother.



Figure 1.0.1: A family photo consists of an infant and his mother, grandmother, and great-grandmother, from source <http://en.wikipedia.org/wiki/Family>.

Is it possible for the computer to find all the information automatically? As far as we know, this is still a hard topic. In the first step, even face detection is not an easy job. It is even a harder job to estimate all the existing family relationships in a given photo. In this work, we mainly focus on the family relationship analysis for a given image in a higher level, and try to combine the estimation of kinship verification and family relationships together. We propose five computational problems related with family relationship analysis in face photos through five levels.

In Chapter 1, we reviewed the traditional family analysis problem and talked about several approaches in its subtopic – kinship verification approaches and proposed a new approach for kinship verification and investigated this problem in a comprehensive way.

In Chapter 2, we proposed five computational problems associated with family relationship analysis, and presented our method in details to deal with these problems.

Different feature descriptors for facial part encoding are discussed in Chapter 3.

After we represented each human image by some feature descriptor, we presented several feature descriptors in Chapter 3.

In Chapter 5, experiments and some discussion of family relationship analysis based on facial parts are presented.

## 1.1 Overview of Family Relationship Analysis

Family relationship is an important component of society. We can divide human beings into two general groups: kinship relationship and non-kinship relationship. Two people with kinship relationship present different appearance from non-kinship group.

We know humans are capable of recognizing family members since they are still very young. The study in human perception of kinship has been an active area of research in human perception, psychology and evolutionary studies [19, 8, 5, 4, 20]. From these studies, it shows that there exist kinship detection mechanisms in humans [20]. Psychologists believe that the resemblance between human faces is a good cue in recognizing the genetic relationship between parents and children and between siblings [8, 19]. Another finding is that the kinship perception in human is different from facial identity recognition. Martello and Maloney [4] reported experiments that aimed to determine what regions of the face contain the cues that signal kinship. They found that



(1) The upper half of the face contains more information about kinship than the lower half.

(2) The eye region contains only slightly more information about kinship than the upper half of the face outside of the eye region.

In parallel, there are some computational approaches to kinship verification. Fang et al. [10] proposed to extract some appearance features such as eyes and skin color and geometric features such as distances between facial parts for kinship verification on a pair of face images. Ghahramani et al. [13] used facial patches to extract family member resemblance. Wang et al. [30] used social context to infer familial relationships. Xia et al. [31] proposed to use young parents as an intermediate distribution to relate children and their old parents for kinship verification. Xia et al. [32] combine with the attribute information with the feature to do the kinship verification. Guo and Wang [17] developed a classification scheme based on Bayesian decision for kinship verification.

In summary, kinship recognition has become an active research topic in both human perception and computational recognition. However, only detecting kinship is not sufficient for family relationship analysis. To broaden the study towards family relationship analysis in face photos, we will present several related problems and propose corresponding approaches to solve each problem.

In this work, family relationship analysis is studied in face photos through five computational problems at different levels of details.

These include:

- (1) Familial trait identification.
- (2) Kinship verification.
- (3) Generation recognition.
- (4) Specific family relationship recognition.
- (5) Family and non-family photo discrimination.

The first four problems focus on a pair of face images, while the last problem is about a photo in which multiple faces appear. These problems are related to each other, but most of them can be studied independently. A systematic study of these problems is likely to advance the field of semantic image understanding and social context extraction.

I think through our work of studying a series of problems for family relationship analysis in face photos could inspire new research interests from the research community; We have developed appropriate methods to deal with each of the proposed problems;

We collect a database of family relationships and use it to perform an evaluation of our approaches.

## 1.2 Motivation

It's known that human inherit traits from their ancestors determined by the genetic inheritance. Parents pass down their inherited traits to their children. So in most cases they share the similar traits in their face. One of the good capabilities of humans is to recognize family members in a random photo. We could always hear the following phrases like "there is a son and a father or a daughter and a mother in the photo, and the boy in the photo has his father's eyes or the girl has her 'mother's mouth.'" Motivated by this situation, we consider the following question: Is it possible to develop a comprehensive system to verify kinship relationship and estimate the family relationship towards the people appearing in a given photo? If this idea works, our computational kinship measurement might have a big influence in real practice, such as finding lost children, child adoptions, trafficking/smuggling of children, and family photo and non-family photo classification. There are a lot of potential significances.

In our work, kinship is defined as a relationship between the family members with the blood relation, such as the relation between father and his son, daughter, sisters and brothers, grandmother and her grandchildren, grandfather and his grandchildren. Also the human presents many degrees of freedom (as illustrated by the joints in Fig. 2.1.1) to change its shape and perform different actions (e.g., in various two family members who have biological relation and blood connection. The relations of biological include parent-children, brothers and sisters, and grandparent-children. So adopted children is not included in biological kinship with their parents. Other relations such as grandmother-grandfather, cousin and husband-wife are not a biological kinship.

## 1.3 Major Challenges

Through the work of family relationship analysis, one of the most important parts is kinship verification. It is different from face identification or face verification. In face recognition, the problem is to recognize the same individual by matching the probe face

image with gallery images. The measure is about the “facial identity.” In kinship verification, the similar features are extracted from similar traits of different persons. Analyzing these challenges will help understand the problem deeply and provide a general guide to develop a computational system. The problem of kinship verification has its own specialty and should not be treated as the traditional face recognition problem. Three major challenges in kinship verification are present as follows:

**A. Identify the similarities across sex and age.** People belonging to the same verification relationship at different ages (e.g., grandfather-grandson), different sex (e.g., brother-sister), or combined them together (e.g., grandfather-granddaughter) could share the same family traits. We know in the work of face recognition, the effect brought by the aging or growth makes it a tough work. Here, the challenge still exists, which makes the facial similar traits among different family members difficult to match. Compared with the work of face recognition, sex difference also gives some difficulty in measuring the similarity of facial traits.

**B. Facial traits are not totally the same. They are individual.** Though the family members share the facial traits, the facial similarities are individual and specialized. Let’s take son - father relationship as example. It’s possible that the sons inherit traits from their parents individually. As shown in figure 1.3.1., one daughter has one nose like his father, while another daughter has a nose more similar to her mother. This characteristic could prove that the familial traits are personalized.

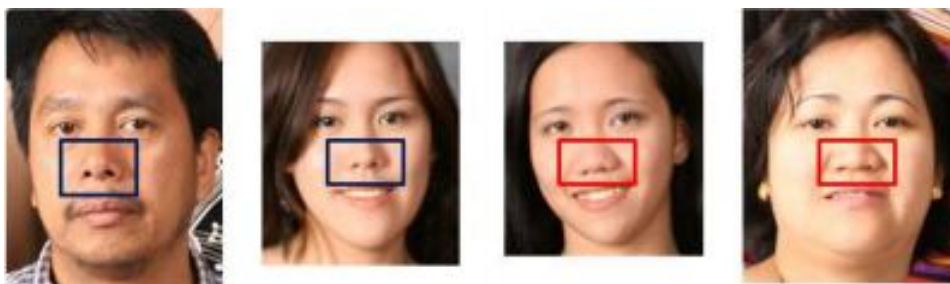


Figure 1.3.1: Familial features shared by parents and two daughters: One daughter has a similar nose to father (marked by black rectangles while another daughter’s nose is similar to mom(red rectangles))

**C. Similarities in Kinship Measures are Dynamic.** As described above, familial traits are individual. As a result, the kinship should be measured with a dynamic combination of familial traits. Under some sense, the shared traits between mother and sons can be arbitrary. In other words, salient facial features, e.g., mouth, eyes, nose, etc can be shared in an arbitrary manner for a kinship measure.

In sum, kinship verification in family relationship analysis has many advantages. Familial traits are the building basis for kinship measure. Our work is different from recent approaches, e.g [10]. They processed kinship more like a face verification problem.

## 1.4 Facial Parts Based Approach

Based on the analysis above, we know people could recognize the kinship mainly based on the facial part , e.g., mouth, nose and eyes. So we want to represent the face using robust and meaningful features. But there is still not a conclusion about which kind of feature is good to characterize the facial parts to represent family relation. We try to use both the shape and appearance to represent the facial part. Commonly used feature descriptors are utilized, including SIFT, Daisy, LBP, HOG, and edges. So we can perform a comprehensive evaluation of various schemes in our problems. We use edge to characterize gradient and shape information. Canny edge descriptor is used. In recent years, SIFT is widely used in computer vision for its robust to the appearance with scale and rotation invariance. In our work, we use dense sampling to extract feature within each facial part. Daisy is another feature descriptor. Compared with SIFT, Daisy descriptor implementation is faster and more efficient. We also use LBP and HOG operators. These two features characterize the texture and shape information respectively. We divide the whole image into fixed dense grids, and use dense sampling scheme as using SIFT. Towards the facial part representation using different feature descriptors, family relationship analysis is conducted.

# Chapter 2

## Family Relationship Analysis

In this Chapter, we will talk about our five computational problems within the range of family relationship analysis in five sections.

### 2.1 Familial Trait Identification

Ancestors pass down their familial traits to their offspring. Family members look alike because they share familial traits in common. We define "familial trait" as a characteristic shared by different members of a family, such as eyes, mouth, nose, etc. The figure below shows one example including a mother and her daughter sharing one similar familial trait (as illustrated in Fig. 2.1.1).

Familial trait is one good cue to analyze the kinship relationship in a given photo. And it could be used to describe the similarity among family members. Our computational scheme derived from the familial traits obtained from a pair of family members' face images.

As we analyze above, age progression takes a main effect in young face. So identifying the familial trait between a parent and a child should consider the significant facial changes caused by aging. For example, a mother's nose is bigger than her daughter's, though their nose shapes are very similar. During the phase of face matching, the alignment of face is usually based on the two eyes. Under common sense, it's not sufficient to deal with the aging variation. Another progression is occlusion, such as in the realistic photos, some faces are occluded with each other. To deal with these problems, we proposed to use

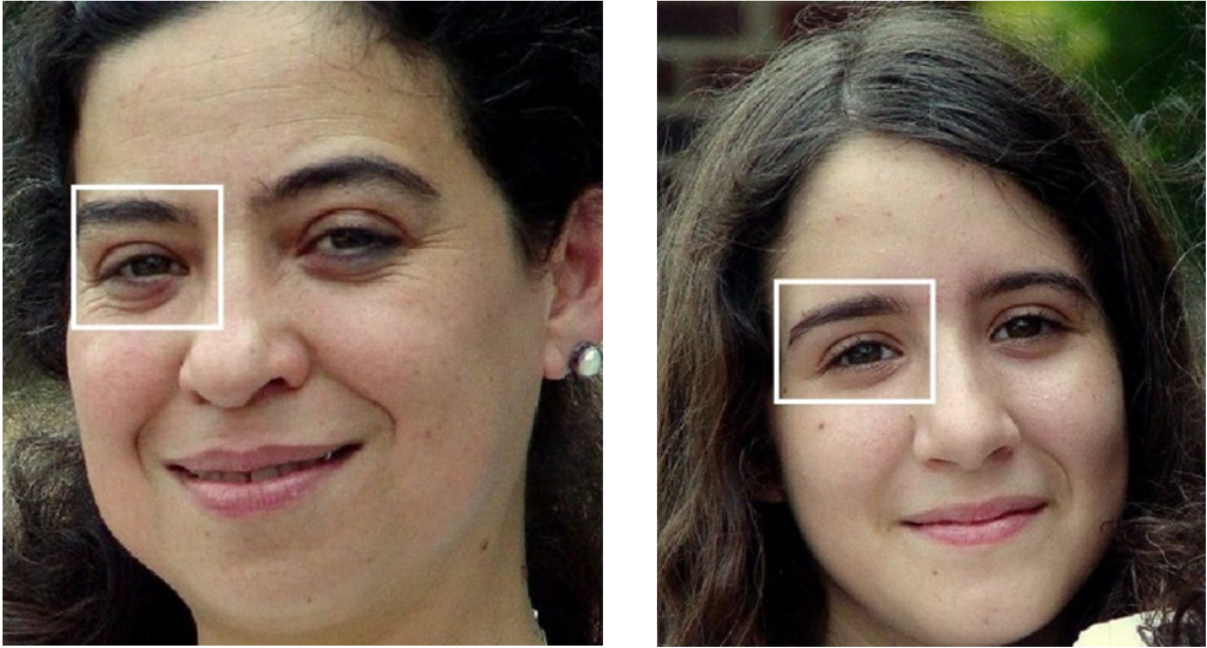


Figure 2.1.1: An illustration of the many degrees of freedom for articulation in a human body.

Modified Hausdorff distance which is similar to Hausdorff distance. This algorithm is helpful for finding the corresponding matching points between two faces. It could reduce the influence of non ideal alignment.

Given the patch, we densely sample the feature, eg., SIFT [21] or other features at each densed point. The features sampled at one point (denoted as  $A$ ) will be matched to another point sampled at other locations from the corresponding part from other face (denoted as  $B$ ). From all the calculated distance from each sampled point in face  $A$  to all the sampled points in face  $B$ , we choose the pair with minimal distance calculated by 2.1.1. For example  $A = \{\tilde{a}_1, \dots, \tilde{a}_n\}$  and  $B = \{\tilde{b}_1, \dots, \tilde{b}_m\}$  denote two sets of features. Then the Hausdorff distance [9] is defined as

$$H(A, B) = \max(h(A, B), h(B, A)), \quad (2.1.1)$$

where

$$h(A, B) = \max_{\vec{a} \in A} \min_{\vec{b} \in B} \|\vec{a} - \vec{b}\|, \quad (2.1.2)$$

In our work, the modified Hausdorff distance ( $MHD$ )[9] is defined as

$$h(A, B) = \frac{1}{A} \max_{\vec{a} \in A} \min_{\vec{b} \in B} \|\vec{a} - \vec{b}\|, \quad (2.1.3)$$

Using the averaging algorithm could diminish the influence brought by the outlier. We use MHD as the measure criteria of similarity between two facial parts. Each set contains features sampled at the finite sample of points. This measure algorithm could characterize the shape similarity of a facial part, e.g., the nose, for a pair of faces containing different variations, including pose variation, facial expressions and aging.

It's known that the original (modified) Hausdorff distance measure[9] deals with edge pixels. Here we adapt it to various features for facial part similarity. In order to deal with the aging effect and potential occlusions and other variations. The MHD-based measure is the basic method for our similarity measure in identifying familial traits.

Learning-based discriminative method is used to identify familial traits. Based on the manually labeled familial traits and non-familial traits, the discriminative method such as the support vector machine (SVM) can be used to learn the classification function.

The distance  $H(A, B)$  in Equ.2.1.1 is a scalar value. To get a better result using feature vectors, several different norms are used, such as 1-norm, 2-norm and p-norm respectively, with  $p = 1.5$ . Thus, three dimension vectors are used to measure the similarity and dissimilarity between a pair of facial parts. SVM is trained for classification. This approach is called "Discriminative 1".

We also use another way to construct a "vectorized" distance measure. In the first step, we use the Euclidean distance to find the most similar feature point from set  $B$  to all the feature point in set  $A$ , i.e.,  $\vec{b}_0 = \min_{\vec{b} \in B} \|\vec{a} - \vec{b}\|$ , where  $\vec{b}_0$  is the vector in set  $B$  that is close to  $\vec{a}$ . Then the component-wise vector difference is computed as  $\vec{a}^* = |\vec{a} - \vec{b}_0|_c$ , where  $|\cdot|_c$  is the component-wise absolute value. Then a vector  $\vec{a}^*$  is derived from the feature vector  $\vec{a}$  and  $\vec{b}$ . Their dimensions are the same. Equ. 2.1.3 is changed to a vectorized Hausdorff distance (VHD), defined by

$$\vec{h}_{vec}(A, B) = \frac{1}{A} \sum_{\vec{a} \in A} \vec{a}^*, \quad (2.1.4)$$

The above equation is a vector average (i.e., a component-wise average of a vector) over all points in set  $A$  with their minimum distances to set  $B$ . To make it symmetric,

both  $A$  to  $B$  and  $B$  to  $A$  are computed, and then Equ. 2.1.1 is changed to

$$\vec{H}(A, B) = \max\_c \sum \vec{a}^*, \quad (2.1.5)$$

where  $\max\_c$  means component - wise max operation, the output is not a scalar value. It is a vector. The distance vector  $\vec{H}$  is calculated from a number of training example pairs to train a SVM for identifying familial trait . We call this approach "Discriminative 2", in comparison with the approach "Discriminative 1".

Compared with these discriminative approaches, a generative approach was proposed in [17], based on Bayes decision. It learns a probabilistic distribution for each familial trait and does classification based on the Bayes rule [17]. The discriminative approaches learn the decision boundary and can utilize vectorized distance measures, e.g., Equ.2.1.4. From the experiments below, we find that the discriminative approach, especially our discriminative 2, is better than the generative approach in terms of the recognition accuracy for familial trait identification (see experiments).

The significance of recognizing familial traits has two purposes:

- (1) Familial traits are used to recognize a kinship for a pair of face images.
- (2) From the Familial traits, we can interpret more details about the reason why a kinship is detected in a pair of face images.

Meanwhile, there are some researchers trying to investigate the familial traits between twins [33, 34]. Some of them also analyzed the performance of face recognition between twins [35]. Actually, twins belong to the same family, but their familial traits are more similar than other family members. I think the research conducted in familial traits is very useful.

## 2.2 Kinship verification

From the analysis above, we can identify the familial traits from a set of face pairs. Based on the detected familial traits, a comprehensive scheme to verify kinship is proposed.



The familial traits are learned separately. To identify kinship relationship between two faces, a comprehensive decision criteria is necessary to learn. We propose a "soft" decision on familial traits (rather than a hard classification), and use confidence value measures to combine familial traits to determine a kinship. Assuming we use  $l$  facial parts as potential familial traits. For each part  $j$ , with  $j = 1, 2, 3 \dots, l$ , we approximate the confidence value by

$$s_j = \sum_{i=1}^p a_i y_i K(x_i, x^j) + b \quad (2.2.1)$$

This equation is based on kernel SVM classifier for  $x^j$ , which is the feature vector for  $j$ -th facial part computed from a pair of faces, i.e., by Equ.2.1.5,  $K(\cdot, \cdot)$  is the radial basis function (RBF) kernel, while  $x'_i$ s are the support vectors,  $a_i$  is the corresponding weight, and  $y_i$  is the label. These two factors are learned by SVM classifier [28]. In a regular SVM classifier - a sign function,  $\text{sign}(\cdot)$ , is used to get a "simple" classification result (0 or 1). In our work, SVM classifier is used to derive confidence values, then combine several potential familial traits to form a kinship decision. The confidence values of all familial parts can be sorted in a descending order. Without any loss of generality, let us assume  $s_1 \geq s_2 \dots \geq s_l$ .

In a general family, the members do not have all their corresponding facial parts alike. So to identify kinship relation, we need to find  $p$  similar facial parts from all the  $l$  facial parts based on the ranking of confidence value from all the members in a given photo. In our study, we choose three traits with the top confidence values to make the decision about a kinship relation. We propose the following rule to determine the kinship as follows:

$$\left\{ \begin{array}{l} \text{Kinship,} \quad \text{if } s_1 > \delta_1, \\ \quad \text{and } (s_1 + s_2) > \delta_2 \\ \quad \text{and } (s_1 + s_2 + s_3) > \delta_3 \\ \text{No Kinship} \quad \text{otherwise} \end{array} \right. \quad (2.2.2)$$

The parameters—  $\delta_i$ ,  $i = 1, \dots, 3$  are small positive constants to adjust the confidence measures for different trait combinations. The rationale for this decision rule includes:

(1) Compared with the faces without any kinship, a pair of faces with a kinship should have more "familial traits".

(2) There are some variations in the degrees of similarity for different facial parts even though there is a kinship.

For the most similar facial part, the corresponding confidence value  $s_1$  needs to exceed  $\delta_1$ . This means that a "high confidence" value should be found for a pair of facial parts if there is a kinship. The similarity of the top two traits,  $s_1 + s_2$  also needs to be greater than  $\delta_2$ , where  $\delta_2 < \delta_1$ . Note that the first condition  $s_1$  is necessary in the addition of  $s_1 + s_2$ . The reason is that there are many possible variations in the procedure of image matching.  $s_2$  could be slightly below zero. From the value range of  $s_1$ , we could see the extent of confidence value, such as if  $s_1$  is much larger, i.e., a very higher degree of confidence. Which means the facial part corresponding to  $s_1$  is much similar, using the sum, the first value  $s_1$  could "rise" the confidence for the second most similar part, based on the summation. Similarly, the third most similar part is also considered in our problem. It is in conjunction with the first two in making a decision kinship. The three thresholds satisfy the following constraint:  $\delta_1 > \delta_2 > \delta_3$ .

## 2.3 Generation Recognition

When we talk about the structure of each human family, we usually talk about the generations. Generation recognition is useful for social structure analysis and semantic image understanding. When two people get married and give birth to a baby, then a new generation is formed. So generation recognition is a significant research in addition to the kinship verification results. We could get more useful detail information in the family relationship analysis. That means we can determine the recognized kinship is within which generation. Are they within the same generation or between different generations? We can't get the specific information just from the kinship verification result. Under the fact that facial similarity reduction along with the increase of generation gap, also the difficulty in collecting data, in our work, we consider three generations. We can classify kinship into one of the three relationships as below:

- (1) Sibling, i.e., within the same generation.
- (2) Parent-child, i.e., two consecutive generations.

(3) Grandparent-grandchild, i.e., across two generations.

As far as we know, less work has been well conducted in the problem of generation recognition. For different generations, the age gaps are different. So our study is focusing on developing a model based on the "age gap" of a pair of faces for computational generation recognition. We refer the statistical work on humans and families in order to develop a knowledge-based mathematical model for generation analysis. Based on the work [23] and [22], we got the mean age of mothers when they have their first child birth in several countries. The statistical data for the French is shown in Fig.2.3. From this statistical data, we find that the age gap between the mother and her children can be modeled by a Gaussian distribution. Here we want to build a model to model the kinship with one generation difference, i.e., parent-child. Based on a detailed analysis of the statistical data in work [22] and [23], we derive a Gaussian distribution for modeling the parent-child relationship,  $N_1(\mu_1, \sigma_1)$ , where  $\mu_1 = 28$  is the mean age gap between parent and children, and  $\sigma_1 = 4.67$  is the standard deviation, manually computed from the statistical data.

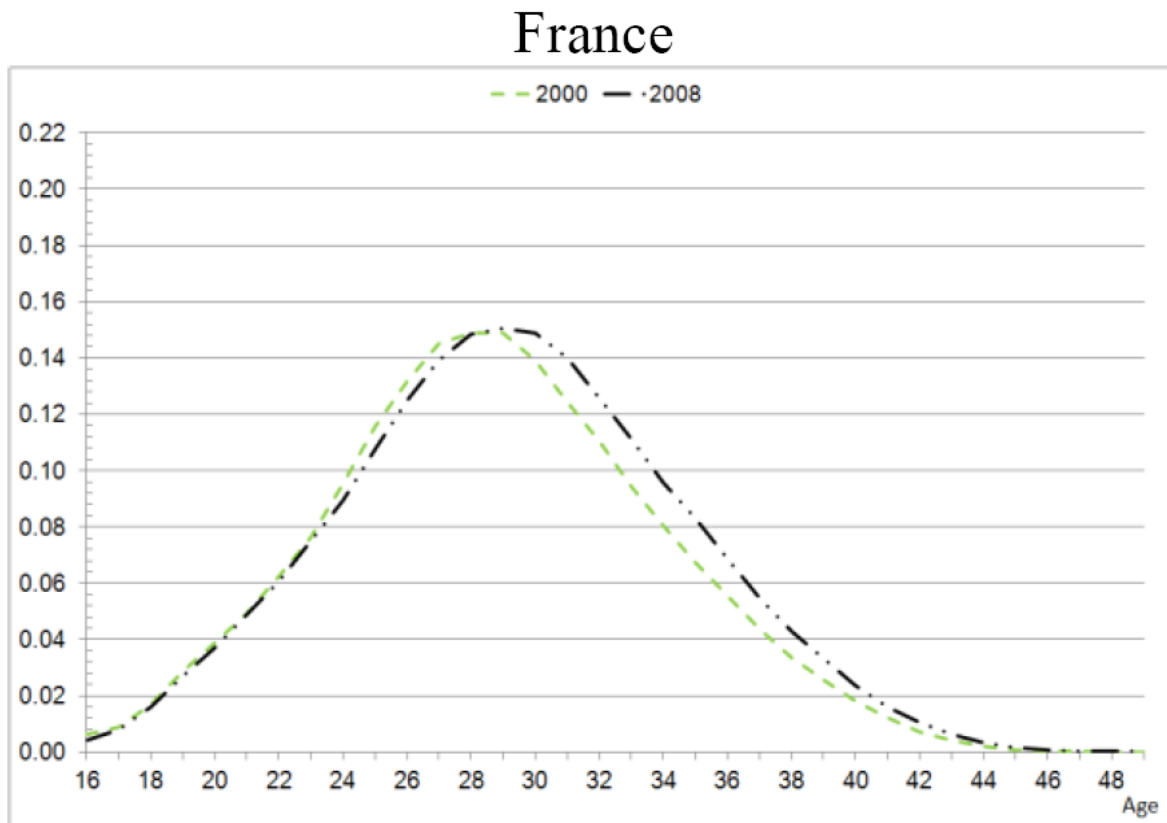


Figure 2.3.1: Mean age of mothers at the first childbirth (France). Source: [21]

From the statistical work conducted in this field, we found that approximately 72% of Americans have their first grandchild when their ages are over 50 [29] [7]. The average age is between 49 and 53 years. Based on this knowledge, we use another Gaussian distribution to model the age gap for the grandparent-grandchild relationship,  $N_2(\mu_2, \sigma_2)$ ,  $u_2 = 52$  represents the mean age gap between the grandparent and grandchild, and  $\sigma_2 = 3.0$  is the standard deviation. Within the same generation, i.e., sibling, the age gap can also be modeled by a Gaussian distribution,  $N_0(\mu_0, \delta_0)$ . We set  $u_2 = 5.6$  as the mean age gap between a pair of siblings, and  $\sigma_2 = 3.0$  as the standard deviation.

So the problem of generation recognition can be performed by measuring the age gap between a pair of faces that have a kinship, and put the age gap value into three Gaussian distributions. A pair of faces with a detected kinship will be classified into one of the three generation-based relationships. The result is based on the closeness to each of the three Gaussian distributions.

Now how to obtain the age gap between two faces is a key problem to solve the generation classification. In our work, in the first step, we estimate the age of each face and then calculate the age difference.  $D = |age_1 - age_2|$ ,  $age_1$  and  $age_2$  are the estimated ages of two faces, respectively.

The problem of age estimation itself is very challenging. We have developed a technique for age estimation using the biologically-inspired features (BIF) [16]. The idea of BIF is motivated by hierarchical visual cortex models, and is invariant to small changes in faces. Here we adapted the method proposed in [16] to a cross-database age estimation (see experiments for details).

## 2.4 Specific Family Relationship Recognition

From our work above, we can tell which generation that the given face pair belongs to the same generation or across generation. As far as we know, there are many specific family relationships, e.g., mother-son. This problem could not be solved from the generation recognition only. In order to achieve the specific family relationship recognition, we propose a method further. In real society, there are eleven specific family relationships in total, including father-son, father-daughter, mother-son, mother-daughter, brother-brother, sister-sister, brother-sister, grandfather-grandson, grandfather-granddaughter, grandmother-grandson, and grandmother-granddaughter. The relations among kinship,

generation recognition, and specific family relationship recognition analysis are shown in 2.4. From the common knowledge (obtained from source <http://en.wikipedia.org/wiki/Family>), we can find the genetic kinship degrees of the family relationships are different according to the generation. For instance, the father-son kinship has 50% genetic similarity.

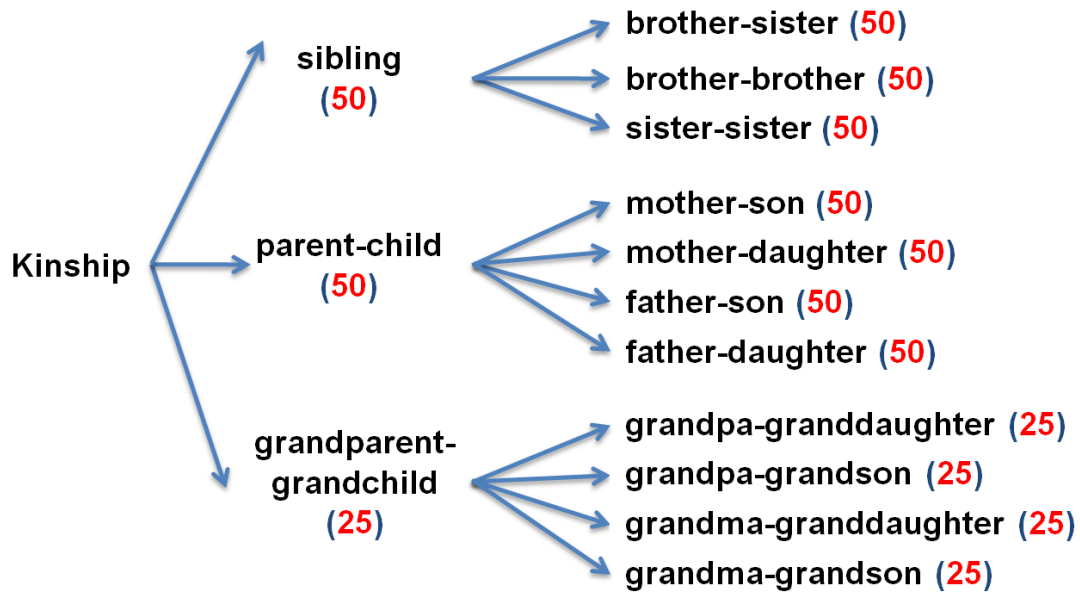


Figure 2.4.1: A kinship can have different levels of relationships in a family structure. The relationships in the middle is determined by the generations, which can be further divided into specific family relationships, as shown in the right column. The genetic kinship degree of relationship is marked with red numbers by percentage (%). These relationships can be analyzed in faces.

Except for the relations listed in our figure, other relations might be recognized as well, for example, uncle-nephew, aunt-nephew, uncle-niece, and aunt-niece. In our current work, we do not consider them here for two reasons. Firstly, such kinds of photos are not very easy to collect, such photos are much less than those from direct family members. Secondly, the work of labeling these kinds of relations is difficult. In our computational approach, without biological relationship, we could not recognize a family relationship, such as husband-wife and grandfather-grandmother, and parents with their adopted children.

Our framework of recognizing the specific family relationships is illustrated in Fig.2.4.2. Given a pair of faces with a kinship, we perform gender and age group classification. Based on our experiment result, we found that this group classification will improve both age estimation and gender classification performance [15]. Then within each gender and

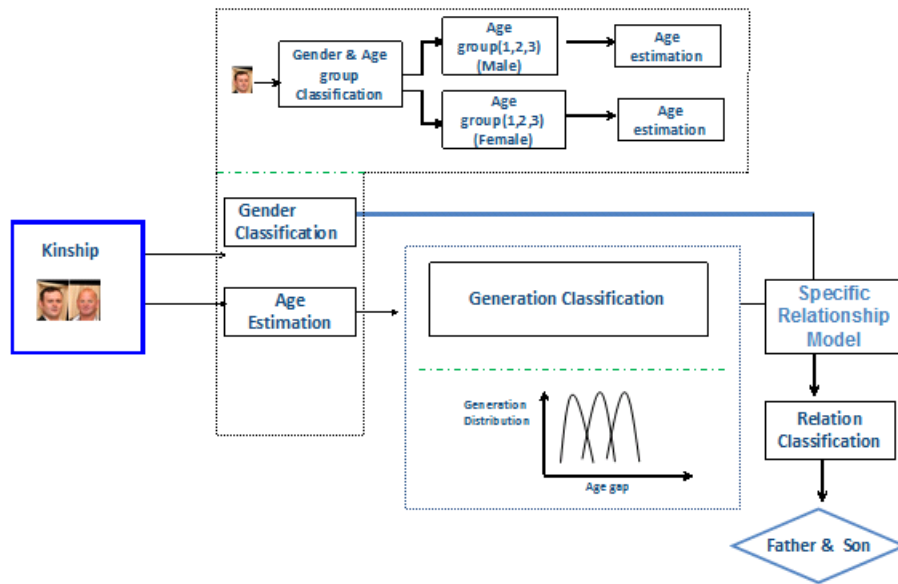


Figure 2.4.2: Proposed framework for generation recognition and specific family relationship recognition.

age group, we conduct age estimation and gender recognition separately. The estimated ages of the pair of faces will be used to estimate the generation level. In the last step, we combine the generation recognition and gender classification results to recognize the specific family relationship from a pair of faces.

The *rules* of classification for specific family relationship determination are described here:

(1) Within the same generation:

If the given two faces are both males, their relationship will be termed as brother-brother; if both faces are estimated as females, the relationship will be sister-sister, if the two faces have different genders, the relationship will be brother-sister;

(2) Across one generation:

The two faces are within the same gender:

- The genders are males. The relationship will be classified as father-son;

- Both faces are females. It will be termed as mother-daughter;

The two faces' genders are different:

- The older face is male, while the younger is female, the relation will be father-daughter;
- The older face is female, while the younger is male, it is mother-son;

(3) Across two generation:

Four relationships are needed to be considered, including grandfather-granddaughter, grandmother-granddaughter, grandfather-grandson, and grandmother-grandson. The rules used are similar to (2).

In the work of gender classification, we propose to use the biologically inspired features (BIF) to encode the face images, then use SVM [28] for classification. In the work [14], we have found that the BIF encoding performs better than the local binary pattern (LBP) and histogram of oriented gradient (HOG) in gender classification. The method adapted in work [14] is used in gender classification to recognize our specific family relationship .

## 2.5 Family and Non-Family Photo Discrimination

In our previous work, the recognition problem is only performed on two compared faces. Those problems are related but different from another problem: family and non-family photo discrimination. This problem is to classify a given photo including many persons into one of two classes, i.e., family or non-family. It's known that a family photo is a photo that contains several persons belonging to the same family, whereas a non-family photo contains several individuals that have no familial relationships. In real world, a family photo usually has more than two family members. To best illustrate this problem, we use two example photos, they are shown in Fig.(2.5.1). In this example, we can see there are several faces in each photo, but the two photos have quite different social contexts. The first one (upper one) is a family photo, while the below one is a non-family photo. It's known that humans can judge this result without much difficulty. Our question is: how to make a computer to solve this problem more effectively.

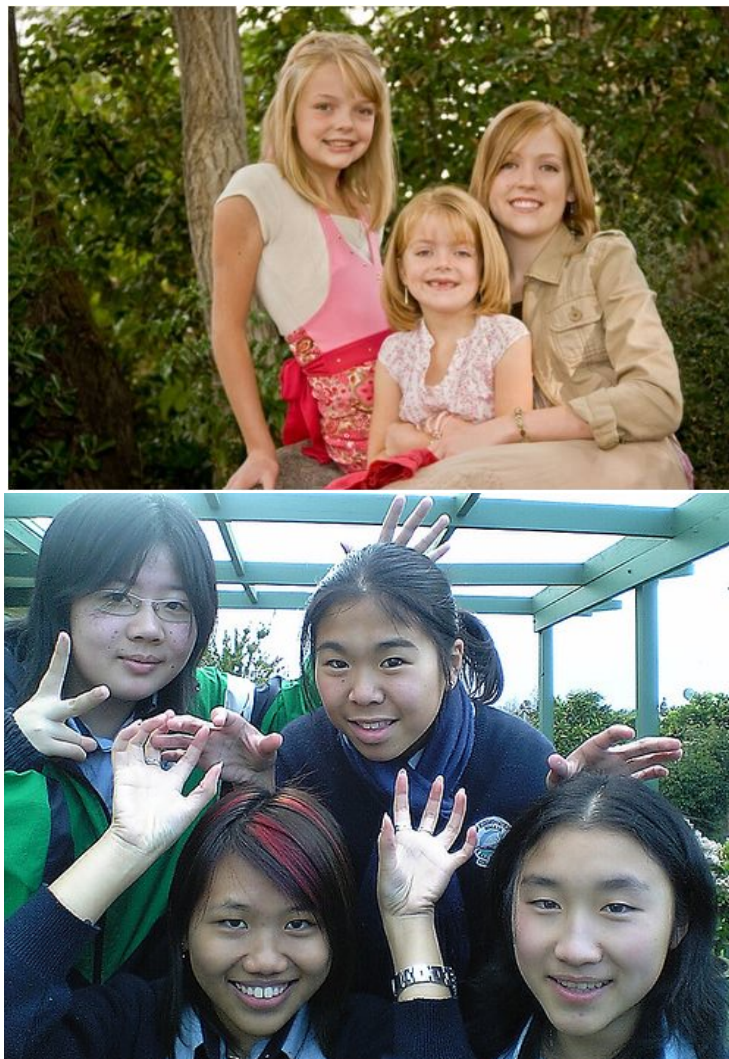


Figure 2.5.1: A family photo (upper) and a non-family photo (below). Both images contain multiple people but with different relationships (or social contexts).

As far as we know, the problem of classifying family and non-family photo has not been fully addressed before. Here we want to show that based on the kinship analysis of pair-wised faces and a combinational analysis of all faces in a photo, a computational method could be developed to perform this classification task. It will help photo categorization and organization, social context extraction and social media understanding.

In the next step, let us discuss our computational method. That is, we know given  $t$  faces in a group photo  $f$ , there are  $\frac{t(t-1)}{2}$  pairs of faces in  $f$ . The kinship determination of each pair can use the method described in Section 2.2. So now our problem is how to



use  $\frac{t(t-1)}{2}$  kinships pairs to determine whether  $f$  is a family photo or not ?

Based on our kinship definition, some face pairs in a family photo have no kinship, such as husband and wife. However, if there are three people in a photo (e.g., a couple and their child). This photo can be classified to a family photo based on our kinship definition. The child may share familial traits with his or her parents. So there will be two pairs out of three with the kinship. As a result, we can find that the majority rule is good at determining a family photo.

Based on the above analysis, we propose a decision rule to separate family photos from arbitrary group photos.

Given a photo  $f$  with  $t$  faces,  $t \geq 2$ , a decision to classify it into a family photo, denoted as  $Fa$ , or a non-family photo, denoted as  $NFa$ , we use the following criteria to deal with this problem.

$$\begin{cases} f \in Fa, & \text{if } \#kinPairs \geq k \cdot C_t^2, t > 2; \text{ or} \\ & \text{if } \#kinPairs \geq 1, t = 2; \\ f \in NFa, & \text{Otherwise} \end{cases} \quad (2.5.1)$$

In the equation above,  $\#kinPairs$  denotes the number of pairs of faces with kinships,  $C_t^2 = \binom{t}{2} = \frac{t!}{2!(t-2)!}$ , we set  $k = 1/3$  in our experiments.

## Chapter 3

# Feature Descriptors

After obtaining the normalized body parts, we need to extract some information from the normalized images to represent those images. In computer vision, the concept of feature is used to denote the piece of information which is useful for solving certain problems. Feature descriptors provide a method for comparing images or image regions. They are used in many different applications, such as object recognition, image alignment, 3D reconstruction, Motion tracking, etc. The most frequently used image feature descriptors are histograms of oriented gradients (HOG), scale invariant feature transform (SIFT), shape descriptors and color descriptors.

Since there is no previous work that has studied the articulated body or body parts based gender recognition, it is unknown about what kinds of features are useful and discriminative. To discover this in our study, we present some features that might be useful, compare the performance of these representations, and then find what representations are good for gender recognition in articulated body images.

In this section, we will have a brief review of the histogram of oriented gradient (HOG) [6], local binary patterns (LBP) [1], scale-invariant feature transform (SIFT) [21], and RGB colors features (where the histogram with 32 bins is computed for each color channel and concatenated). These features will be extracted from the normalized images to represent the body parts. These features were originally proposed for other computer vision problems. We will evaluate whether they are useful for our problem in section 5.

### 3.1 Histogram of Oriented Gradients (HOG)

Navneet Dalal and Bill Triggs proposed Histogram of Oriented Gradient descriptors in their June 2005 paper [6], in their work, this feature vector had been proved to be an effective algorithm in detecting the pedestrian in static images. This feature makes use of the distribution of local intensity gradients or edge direction to characterize the local object appearance and shape, without precise knowledge of the corresponding gradients or edge positions. So this character is very important in the problem of pedestrian descriptors, because the images associated with the pedestrian detection problem are always not clear and less detail included. Dalal and Bill Triggs also expanded their tests in human detection in film and video, this feature also could be applied to the common animals and vehicles in static imagery.

In practice, the steps extracting the HOG descriptors from the image can be described as follows:

(1) Divide the whole image into small patches, there are overlap between these patches. We call the divided these small patches as cells. There are two types of cells: rectangular R-HOG blocks and circular C-HOG blocks.

R-HOG blocks: The grids are square, represented by three parameters: cells' number per block, pixels' number per cell, and the number of channels per cell histogram. In the author's work, the optimal parameters were found to be  $3 \times 3$  cells blocks of  $6 \times 6$  pixels with 9 histogram channels. Applying the Gaussian spatial window within each patch before calculating histogram votes is used to weight pixels around the edge of the patch. The R-HOG patches seem to be similar to the scale invariant feature transform. They compute R-HOG blocks in dense grids at some single scale.

C-HOG blocks: Four parameters are involved in these C-HOG blocks – the number of angular and radial bins, the central bin's radius, and the expanding factor for the radius of additional radial bins. In their work, Dalal and Triggs found that expansion factor of 2 provided the best performance in their experiments.

(2) Within each cell, obtain a histogram of gradient directions or edge orientations for the pixels, and combine the values from all patches into one vector to represent the image.

(3) In order to deal with the changes in illumination or noise, a normalization scheme has been applied in the patches of images. And use the energies to normalize all the cells in the image. There are several normalized schemes to deal with this problem.

L2-norm:

$$f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (3.1.1)$$

L2-hys: L2-norm followed by limiting the maximum value of  $v$  to 0.2 and renormalized.

L1-norm:

$$f = \frac{v}{\sqrt{\|v\|_1 + e}} \quad (3.1.2)$$

L1-sqrt:

$$f = \sqrt{\frac{v}{\sqrt{\|v\|_1 + e}}} \quad (3.1.3)$$

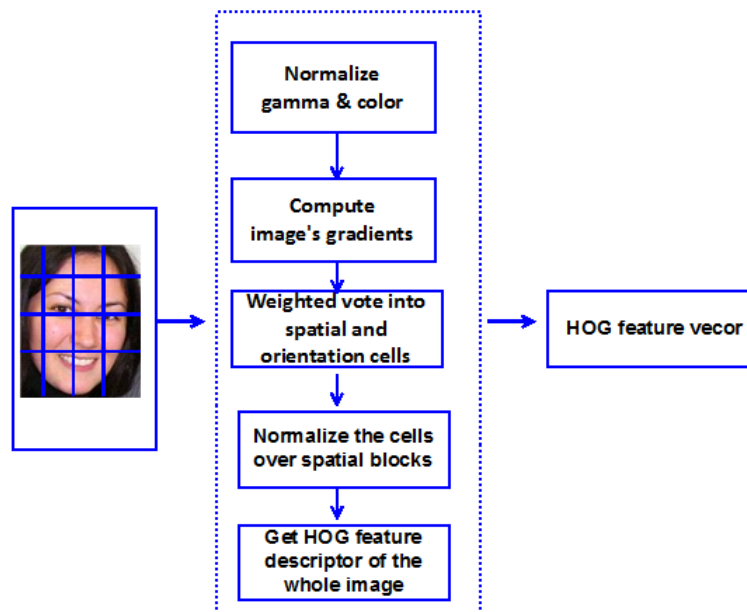


Figure 3.1.1: General process of getting HOG descriptor

In the process of calculating the gradient of the image intensity function, 1-D centered, discrete filter is applied in the horizontal and vertical direction. The author used

$[-1, 0, 1]$  filter kernel in the corresponding horizontal direction, and  $[-1, 0, 1]^T$  in the vertical direction. Other filters also could be applied, such as  $3 \times 3$  Sobel operator. Dalal and Triggs also tried to smooth the image using Gaussian function before applying the discrete filter. This method achieved a better result.

## 3.2 Local Binary Patterns (LBP)

T. Ojala et al.[1] proposed Local Binary Patterns (LBP) to deal with texture classification in 1994, compared with traditional algorithms which quantify the texture measures by single values (means, variances etc.). This work focused on the distributions of feature values. Generally speaking, LBP is calculated using the difference information between the central pixel value with the its surrounding pixel values, then encoded the result as a binary number.

There are several steps to extract the LBP feature vector. These could be described as follows:

1. Divide the given image into several patches. There are always overlapping between adjacent patches.
2. For one pixel within each patch, after setting the radius of circle, compare its intensity value with its surrounding. Compared with SIFT which uses 3D histograms, in this histogram, two dimensions are used to represent the image spatial dimensions and one additional ding neighbors. That is, if the intensity value of the central location is larger than its neighbor's pixel value, then set the value to 1, else assign it to 0.
3. Then for each local patch, its texture is represented by eight elements, each of which has 0 or 1, obtained from the neighborhood surrounding the central pixel. The author calculate the histogram within each patch. Then the author count the occurring times of each "number" in the range of whole image,
4. Normalize the histogram. Then concatenate the histograms of all patches to the feature vector of the given image.

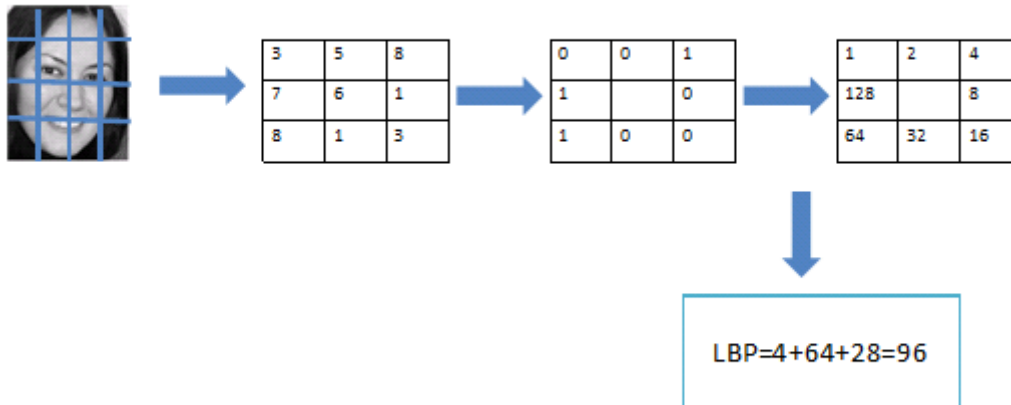


Figure 3.2.1: Illustration chart of extracting LBP descriptor

Because LBP is very efficient in face photo analysis. This feature descriptor has become a very popular approach in the field of computer vision.

### 3.3 Scale Invariant Feature Transform (SIFT)

In 2004, D. G. Lowe advocated Scale Invariant Feature Transform (SIFT) [21] to deal with the problem of matching between different views of an object or scene. SIFT has been widely used in many topics for its good characteristics. It is invariant to different image scales and rotation, robust matching across the substantial range of affine distortion, noise illumination variations and changing in 3D viewpoint. D.G.Lowe also used SIFT feature in the object recognition work. This feature had been proved to be very effective for object identification among clutter and occlusion.

The algorithm for extracting the SIFT descriptors could be summarized in four steps generally:

1. Scale-space extrema detection.

The author's work (Lowe, 1999) used scale-space extreme in difference-of-Gaussian

function convolved with the image.

$$\begin{aligned} D(\mathbf{x}, \boldsymbol{\sigma}) &= [G(\mathbf{x}, k\boldsymbol{\sigma}) - G(\mathbf{x}, \boldsymbol{\sigma})] * I(\mathbf{x}) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3.3.1)$$

Where  $D(\mathbf{x}, \boldsymbol{\sigma})$  is the Gaussian function,  $I(\mathbf{x})$  is the intensity function of the image.  $\mathbf{x}$  is the location of the pixel,  $\sigma$  is the scale,  $k$  is the varying scale factor. This equation deals with the smoothed images keeping the images' key characteristic. Local maximum or minimum value in the  $D(\mathbf{x}, \boldsymbol{\sigma})$  is, the value is maximum or minimum is determined by comparing its value with its surrounding eight pixel values.

## 2. Keypoint localization.

After locating the key point candidates, the author further use a detailed fitting to remove those points that have low contrast (easily affected by the noise) or the points that are located along the edge. This step is to find the stable key points. Then the author referred to the work proposed by Brown and Lowe [35], they used a scheme for fitting a 3D quadratic function. This function interpolated the location maximum in local sample points. In their work, the Taylor expansion (up to the quadratic terms) of  $D(\mathbf{x}, \boldsymbol{\sigma})$  is expressed as

$$T(\mathbf{x}) = D + \frac{\partial D^T}{\partial x} \mathbf{t} + \frac{1}{2} \mathbf{t}^T \frac{\partial^2 D^T}{\partial \mathbf{x}^2} x \quad (3.3.2)$$

Where  $D$  and its derivatives are calculated at the sample point, and  $x = (x, y, \sigma)^T$  is the offset from  $\mathbf{x}$ . Then compared with SIFT which uses 3D histograms, in this histogram, two dimensions are used to represent the image spatial dimensions and one additional taking the derivative of this function and set to zero, giving

$$\tilde{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D^T}{\partial \mathbf{x}}. \quad (3.3.3)$$

If  $\tilde{\mathbf{x}}$  is larger than some threshold (e.g., 0.5), it implies that the extreme's location is closer to some other neighboring sample point, this point could be used to calculate  $\frac{\partial D}{\partial \mathbf{t}}$  and  $\frac{\partial^2 D}{\partial \mathbf{t}^2}$ . Under this situation, the sample point is changed and its new neighbor will be used to perform the interpolation and get the new interpolation for the keypoint.

### 3. Eliminating edge responses

After rejecting the keypoints with low contrast, this step is used to eliminate the false keypoints that locate along the edge, because the difference of Gaussian function have a strong response along the edges, even some edges are not stable and sensitive to the noise. The principal curvature along the edge of those poorly defined peak is bigger than the value in the perpendicular direction.

To eliminate these false keypoints, we compute the principal curvature from the  $2 \times 2$  Hessian matrix,  $H$ , computed at each location and scale of the keypoint:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (3.3.4)$$

Where  $D$  represents the principal curvature.

Then the largest magnitude  $\alpha$  and second magnitude value  $\beta$  are calculated. Get the trace and product from the determinant:

$$\begin{aligned} Tr(H) &= D_{xx} + D_{yy} = \alpha + \beta \\ Det(H) &= D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \end{aligned} \quad (3.3.5)$$

The ratio  $r$  between these two values could identify if the keypoint is true or false.

$$r = \frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} \quad (3.3.6)$$

Let  $k$  represents the ratio between  $\alpha$  and  $\beta$ , where  $k = \frac{\alpha}{\beta}$ , then

$$r = \frac{Tr(H)^2}{Det(H)} = \frac{(k + 1)^2}{k} \quad (3.3.7)$$

In the author's paper,  $k = 10$ , so if  $r$  is bigger than the threshold, then we will eliminate it, or else keep it.

### 4. Orientation assignment.

When determining the orientation of the keypoint, the histogram of orientation is computed from the gradient orientations of neighbors around the detected keypoint. Then the author tried to get the highest peak in the histogram and any other local biggest value that is within 80% of the highest peak is used to create the keypoint in that orientation.



5. Get the feature vector of each local image.

Firstly, the author calculated the gradient magnitude and orientation at each sampling image point around the keypoint location. Then use a Gaussian window to weight them, and accumulate the samples into orientation histograms summarizing the contents over  $4 \times 4$  subregions.

6. Formalizing the descriptor representation

Through the processing above, each key point descriptor has been computed as a set of orientation histograms over  $4 \times 4$  subregions.  $0$  to  $2\pi$  are divided into 8 parts and used to form the orientation histograms. Each chosen window is divided into  $4 \times 4$  subpatches. Within each patch, gradient direction in 8 orientations is computed. So for each keypoint, SIFT feature is represented by concatenating the gradient histogram of  $4 \times 4$  sub-regions. The length of SIFT feature vector is  $4 \times 4 \times 8 = 128$ . Then the feature vector will be normalized to enhance its stability to changes in illumination.

In our work, we extract SIFT feature on a densely sampled grid of locations under the fixed scale and orientation, so that means that step 1 to step 3 to extract the SIFT descriptors have been skipped.

### 3.4 DAISY feature descriptor

Engin Tola .etc proposed Daisy feature descriptor to compute dense depth and occlusion maps from wide-baseline image pairs, and this feature descriptor achieves very good result. In the author's work, Daisy feature descriptor proved to be very effective against the photometric and geometric transformation.

Compared with SIFT which uses 3D histograms, in this histogram, two dimensions are used to represent the image spatial dimensions and one additional dimension is used to represent the image gradient direction. These dimensions are calculated over local regions, so all the pixels in the local all contribute to the histogram, the confidence that they contribute to final result is determined by the pixel's location in the local region, by the orientation and the norm of the image gradient at its location. That means each bin is comprised by the weight (the weight is inversely to the distance to the key point). As

a result, each bin is comprised of the weighted sum of the norms of the image gradients around the key point, the weight is roughly determined by the distance to the bin center.

To extract DAISY feature descriptors, the author use the convolutions of the gradients with the Gaussian filters in specific direction to replace the weighted sums of gradient norm.

The major steps of extracting DAISY Descriptor can be summarized as follows:

(1) Calculate the orientation maps.

Given an input image, compute  $H$  orientation maps,  $G_i$ , where  $i = [1, H]$ ., each orientation map corresponds to one quantized direction.

(2) Obtain convolved orientation maps.

Each orientation map is convolved with Gaussian kernel.

$$G_o^\Sigma = G_\Sigma \otimes \left(\frac{\partial I}{\partial o}\right)^+ \quad (3.4.1)$$

where  $G_\Sigma$  is a Gaussian kernel, and there are several different Gaussian kernels differed by  $\Sigma$  values.  $\Sigma$  values are used to control the size of the map region.

To fasten the computational speed, the author used consecutive convolution with small Gaussian kernels to replace the convolution with a large Gaussian kernel.

$$G_o^{\Sigma_0} = G_{\Sigma_0} \otimes \left(\frac{\partial I}{\partial o}\right)^+ = G_\Sigma * G_{\Sigma_1} * \left(\frac{\partial I}{\partial o}\right)^+ = G_\Sigma * G_o^{\Sigma_1} \quad (3.4.2)$$

where  $\Sigma = \sqrt{\Sigma_2^2 - \Sigma_1^2}$ . ( $\Sigma_2 > \Sigma_1$ )

So each vector in DAISY feature descriptor is consist of the values from the convolution orientation maps located on the concentric circles centered in the selected pixel. The series of the Gaussians just like a flower, that's where the feature descriptor name coming from.

# Chapter 4

## Supervised Machine Learning

Supervised machine learning is one learning algorithm for recognition. The task of Supervised machine learning is used to predict the unknown sample's label based on the training samples. It's the key characteristics that mainly focus on the process of learning the classification model. Each training sample is labeled with its class (the label definition of the training sample is based on the specific problem, such as when do the gender classification. The label information is consisted of "male" and "female"). Each training sample is also represented by a feature vector extracted from the original image. The task of this algorithm is to predict a classification function that could give the right label of the testing sample.

The generalized procedure of supervised machine learning is shown in Fig 4.0.1.

There are many different kernels in support vector machine (SVM) recent years. In our work, RBF kernel is used in our work to do kinship verification and family relationship analysis. In the next section, we will introduce the general scheme of SVM.

### 4.1 Support Vector Machine

A support vector machine (SVM) is first introduced by Cortes and Vapnik, this classification method has been widely used in data analysis and pattern recognition. Recently, SVM has been widely used as a classification tool in the field of computer vision, and gets so much popularity.

The general scheme is analyzed as follows:

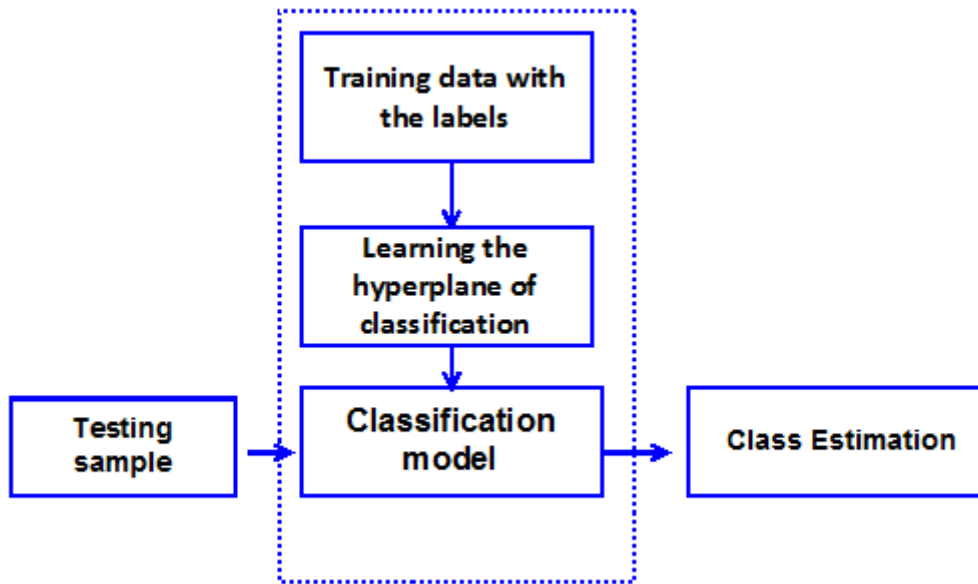


Figure 4.0.1: General procedure illustration of Supervised Machine Learning

Given  $n$  input data, say  $x_i \in R^N$ ,  $x$  is  $(n \times M)$  matrix,  $n$  is the number,  $M$  is the feature dimension, and corresponding binary prediction  $y_i \in \{-1, 1\}$ , to maximize the margin between hyper-plane in a higher dimensional space, the requirement of the support vector machine is as follows:

$$\begin{cases} x_i \cdot w + b \geq 1, & \text{if } y_i = 1 \\ x_i \cdot w + b \leq -1, & \text{if } y_i = -1 \end{cases} \quad (4.1.1)$$

1

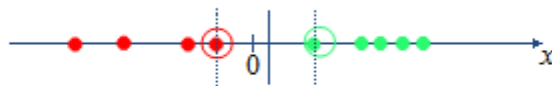


Figure 4.1.1: Linearly separable

---

<sup>1</sup>Reference from Andrew Moore

To find the maximum-margin hyper-plane, we should find the appropriate  $w$  and  $b$  that minimize (  $P(w, b) = w^T w / 2$  ) .

We can simplify the above equation to the following format:

$$y_i[w^T x_i + b] \geq 1 \quad (4.1.2)$$

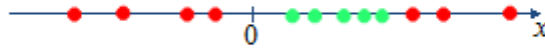


Figure 4.1.2: Complicated dataset distribution

Based on this idea, we can solve many problems such as in 4.1.1. But meanwhile under some situation, the training samples with opposite labels mixed with each other, there is not a hyperplane to separate them 4.1.2, so we need to find a optimized hyperplane dealing with this problem:

We need to minimize

$$P(w, b, \xi) = w^T w / 2 + C \sum_{i=1}^n \xi_i \quad (4.1.3)$$

s. t.

$$y_i[w^T x_i + b] \geq 1 - \xi_i$$

where  $\xi_i \geq 0$  ,  $\sum_{i=1}^n \xi_i$  is the total number of training errors, it determines the limit of errors.  $C$  is a parameter balancing the class distance and errors.

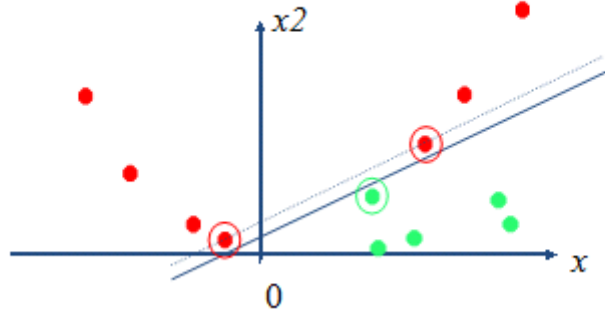


Figure 4.1.3: Seperable result got by Nonlinear

There are two major classification structures. One is linear model and another is non-linear model. Linear model is simple. It could deal with many problems. But the distribution of some data with different classes is not easy to classify using linear model. Under this situation, kernels models are proposed to deal with these problems. Kernel functions have been widely used to prove no-linear structure work very well. Kernels are introduced to project original data to a higher dimensional space, so that the projected data are linear separable in the projected space, as illustrated in 4.1.3 . There are several commonly used as follows:

(1) Linear Kernel:

$$K(x, z) = x^T z + c \quad (4.1.4)$$

;

(1) Gaussian Radial basis function (RBF):

$$K(x, z) = e^{-\|x-z\|^2/2\sigma^2} \quad (4.1.5)$$

(2) Polynomial Kernel:

$$K(x, z) = (c_1 x^T z + c_2)^p \quad (4.1.6)$$

;

(3) Chi-Square Kernel:

$$K(x, z) = 1 - \sum_{i=1}^n \frac{2(x_i - z_i)^2}{(x_i + z_i)} \quad (4.1.7)$$

The final classification rule is very similar with the linear model described as follows:

$$\hat{y}_t = b + \sum_{s \in S} w_s K(x_t, x_s) \quad (4.1.8)$$

where  $b$  is the constant value,  $w_s$  is the weight vector,  $S$  is the set of support vectors,  $x_t$  is the testing feature descriptor,  $x_s$  is one of the support vector in the training set.  $\hat{y}_t$  is the predicted classification result of the testing sample  $x_t$ .

In this work, we have tried different kernels in our experiments and realized that the RBF kernel gave the best result in our problem.

# Chapter 5

## Experiments and Discussion

### 5.1 Datasets Description

We applied our family relationship analysis on our collected data sets. The database is collected from Flickr and Google Images using keywords like "family" and "groups." The photos downloaded by us are just those that are allowed by the owners. The dataset collected by us includes about 1,000 photos from the initial response, and then we filtered the photos manually to select appropriate images for evaluating our methods.

In the work of kinship verification, the individuals with non-biological relatedness are not considered. Note that adopted children do not have a kinship with their parents. So in our study, we don't consider the relationship of husband-wife and grandmother-grandfather.

To label the kinship relation of our photo correctly, five people are involved in our work to filter the initial collection of images. The filtering criterion in our work includes:

(1) Each labeled family photo contains face pairs that have evident kinship relationship;

(2) Each labeled non-family group photo doesn't contain any kinship among all faces within the same photo;

(3) If one is not 100% sure to label an image as family or non-family, just discard the image.

After our manual filtering, five people are asked to vote for each image. The image labeled can be selected only if all five people give the same vote: family or non-family.



The number of images selected for each category is slightly above 100. To make two classes balanced, we chose exactly 100 family photos and 100 non-family group photos. In sum, there are 810 faces in the selected database, 348 faces are included in family photos and 462 faces are contained in non-family group photos. On average, there are 3.5 faces in each family group photo, 4.6 faces in each non-family photo.

Furthermore, five individuals are also asked to label facial familial traits in the selected family photos. They were told to label the similarity with respect to important facial parts, including left eye, right eye, nose and mouth. A familial trait is said to be true only if all the five subjects have an agreement. To reduce the ambiguity and maintain consistency, only a pair of faces in each family photo is labeled with the familial traits. Also the kinship was labeled for all pairs of faces when all five individuals voted "yes" to each family photo.

It's known that some people may have similar facial parts, such as mouth or nose, even though they don't belong to the same family. So here we propose to use 100 non-family group photos as the "negative" examples to prove our algorithm of kinship detection in family photos. So similar to the work above, dissimilarity in terms of salient facial parts in each non-family group photo is also labeled in our work. The criterion of labeling the non-family parts is also that all five individuals voted the same.

In sum, there was a significant amount of labeling work done on our database. The ground truth labeling work has different levels of details. The first level is the familial (and non-familial) traits in faces; the second level is the kinship (or non-kinship) relation for a pair of face images; The third level is the family photo and non-family group photos. Furthermore, specific family relationships and generations were also labeled for testing the developed methods.

## 5.2 Experimental Results

Our experiments are conducted on our family image dataset, which includes 100 family photo and 100 non-family group photos separately. Because our collected images are selected from an uncontrolled environment, we have to deal with the variations existing in the faces, such as pose changes, illumination variations, facial expressions, and some facial occlusions. In our work, we will evaluate our proposed algorithm on the five tasks:

- (1) Familial trait identification;

- (2) Kinship verification;
- (3) Generation recognition;
- (4) Specific family relationship recognition;
- (5) Family and non-family photo classification.

Ten-fold cross validation is used in our work to test the accuracy of familial trait identification, kinship verification, and family photo recognition. Five different feature schemes for facial part representation are evaluated, that is SIFT [21], DAISY [27], LBP [1], HOG [6], and Edges[3]. The DAISY descriptor has been adopted in our previous study [17] , and it is used here for comparisons.

Face detection algorithm from the OpenCV package is used in each photo. For those faces that cannot be detected because of the significant head pose variations in the unconstrained images, we manually labeled the locations of those faces in order to evaluate our methods. The two eyes were detected using a procedure similar to face detection algorithm. Other facial parts , nose, mouth are estimated with the prior knowledge based on the relative positions of the eyes. As we analyzed above, the modified Hausdorff distance measure could handle the misalignment issues in the facial parts matching.

### 5.2.1 Familial trait identification

We evaluate two discriminative classification schemes and compare with the Bayes decision scheme [17]. Meanwhile, we explore five different representations for each facial part. Through our comprehensive evaluation, then we could find which scheme is better and which feature representation is more effective in learning familial traits. The basic idea is to learn the facial part similarity and dissimilarity from two sets of labeled samples, and perform a two-class classification, as discussed in Section2.1. Considering the aging effect and other facial variations, the vectorized Hausdorff distance measure in 2.1.4 was used in all feature representations. The accuracies of familial trait recognition are shown in Table 5.1.

Matching	Representation	Familial Trait Identification Accuracy			
		Left Eye	Right Eye	Nose	Mouth
Bayes	SIFT	0.783	0.871	0.800	0.783
	DAISY	0.775	0.814	0.733	0.750
	HOG	0.692	0.721	0.617	0.567
	LBP	0.625	0.693	0.641	0.733
	Edge	0.492	0.493	0.542	0.483
Discri.1	SIFT	0.830	0.860	0.855	0.705
	DAISY	0.750	0.790	0.705	0.610
	HOG	0.620	0.650	0.560	0.560
	LBP	0.685	0.600	0.585	0.595
	Edge	0.610	0.585	0.550	0.550
Discri.2	SIFT	0.830	0.860	0.855	0.705
	DAISY	0.775	0.825	0.735	0.615
	HOG	0.620	0.685	0.595	0.570
	LBP	0.790	0.775	0.745	0.670
	Edge	0.525	0.630	0.575	0.595

Table 5.1: Familial Trait Identification Accuracy.

From the result in the table, one can see that the SIFT based feature representation performs the best among all these five representations. The Daisy operator is close to the SIFT and better than LBP and HOG. Furthermore, our discriminative scheme 2 is consistently better than the discriminative 1 scheme, and performs better compared with the Bayes decision scheme [17] in some cases.

To summarize the result of familial trait identification, we draw a bar graph and display it in Fig.(5.2.1). It visually shows the performance of learning each familial trait with different matching schemes. It visually shows the performance of learning each familial trait with different matching schemes.

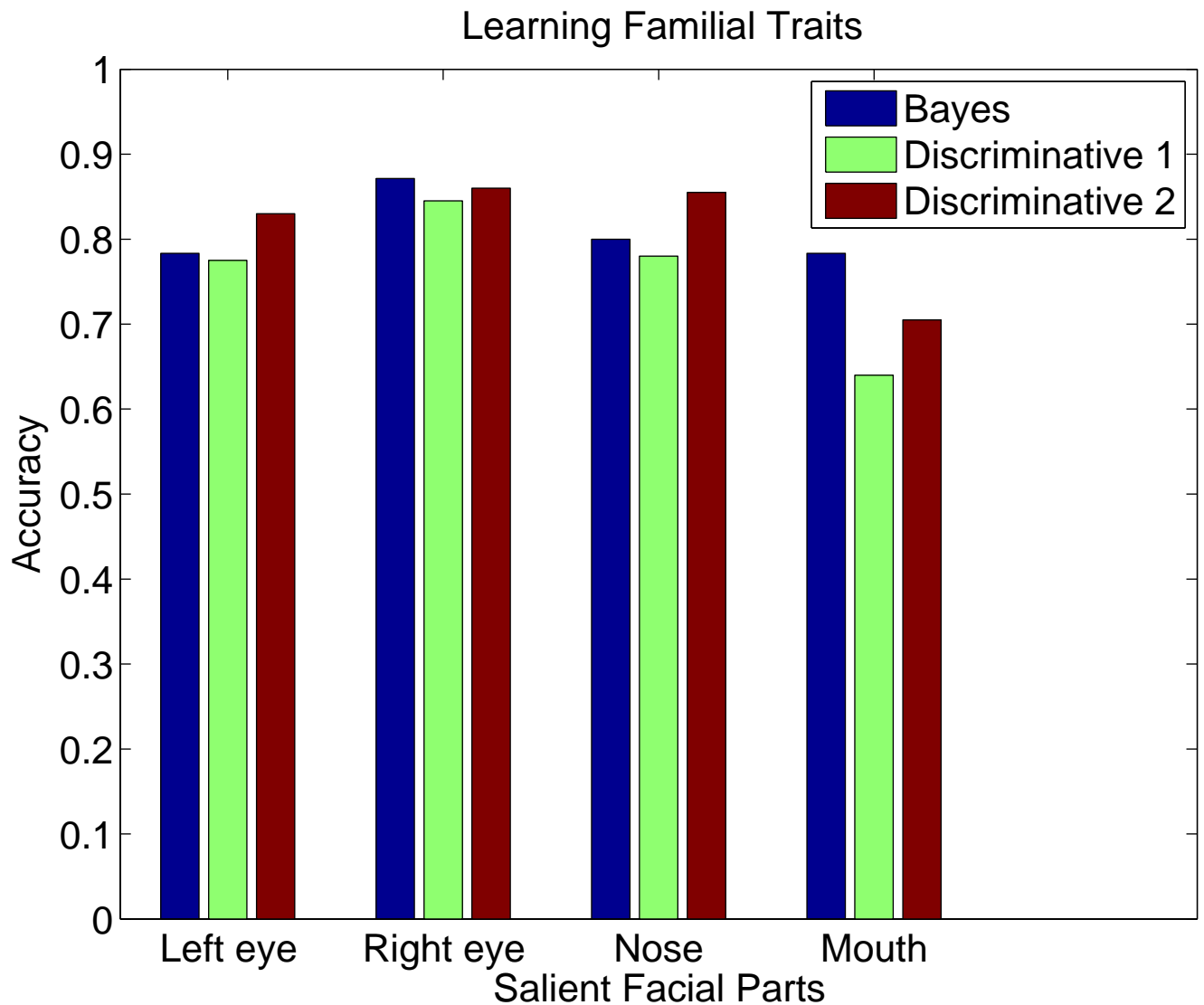


Figure 5.2.1: Familial trait identification using different matching schemes on each of the four traits, with the SIFT based representation

### 5.2.2 Kinship verification

Based on the previous work, our next step is to recognize a kinship for a pair of faces, this step will use the familial traits learned previously. As described in Section (2.1), Equ.(2.1.3) is used to combine familial traits dynamically (by sorting) and probabilistically (measured by confidence values) to determine the kinship. Small positive constants are

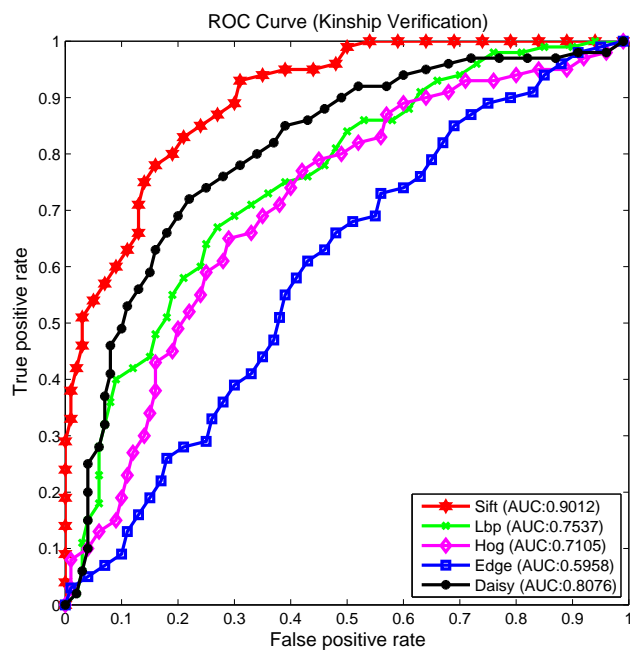
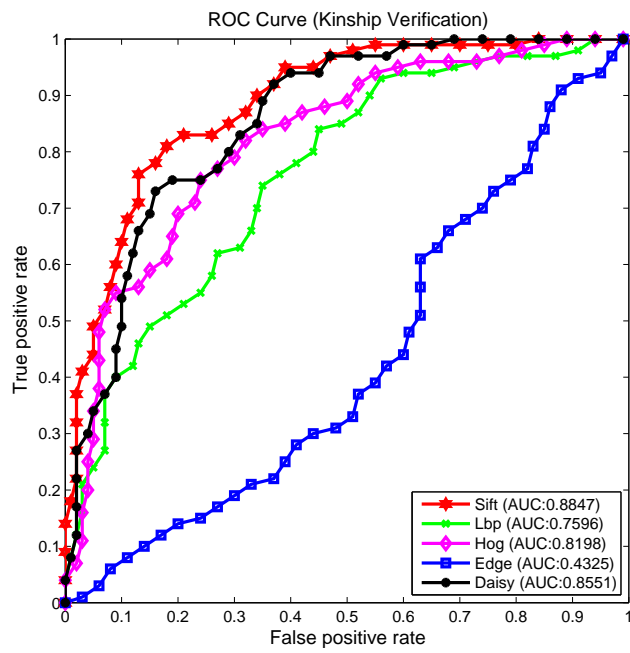
selected,  $\delta_1 = 0.02$  ,  $\delta_2 = 0.01$  , and  $\delta_3 = 0.005$  , to set the threshold for confidence measures. The setting of these values is from the analysis of the experiments. The underlying meanings are that the facial parts of faces with a kinship look similar but with different degrees of similarity, as discussed in Section (2.2). We use ten-fold cross validation to compute the accuracies using different facial part representations. The results are shown in Table 5.2. From the table, we find that SIFT based feature gives the highest accuracy of 84.5% in kinship verification.

Representation	Kinship Verification Accuracy		
	Bayes	Discriminative 1	Discriminative 2
SIFT	0.800	0.820	0.845
DAISY	0.770	0.750	0.825
HOG	0.740	0.675	0.735
LBP	0.675	0.705	0.825
Edge	0.535	0.615	0.650

Table 5.2: Kinship Verification Accuracy

Based on the discriminative 2 scheme, compared with this scheme, discriminative 1 scheme achieves 82.0%, this is a little lower than discriminative 1 scheme. Both schemes are better than 80.0% got by the Bayes decision scheme [17], work [17] uses likelihood ratios to make the decision. The DAISY based representation achieves lower than SIFT, but higher than the HOG and LBP, much higher than the edge based method. In sum, our matching scheme based on discriminative 2 is the best among all the three proposed approaches.

It's known that the representations and discriminative matching schemes have a direct influence in computing the confidence values in Equ. 2.2.1 , which also further affect the kinship decision in Equ.2.2.2 . Our kinship verification scheme could achieve 80% or higher, this is an amazing result. To best represent the performance of different matching schemes and various representations, we draw the ROC curves based on our experiment result, this is shown in Fig.(5.2.2). To compare with the previous work [10], we also implemented the algorithm according to the illustration in this work. Then we performed kinship verification using our database. The result is not good, just could achieve 59.5% accuracy. I think it's because the database in work [10] has low-resolution images and is not sufficient for learning the familial traits using our methods.



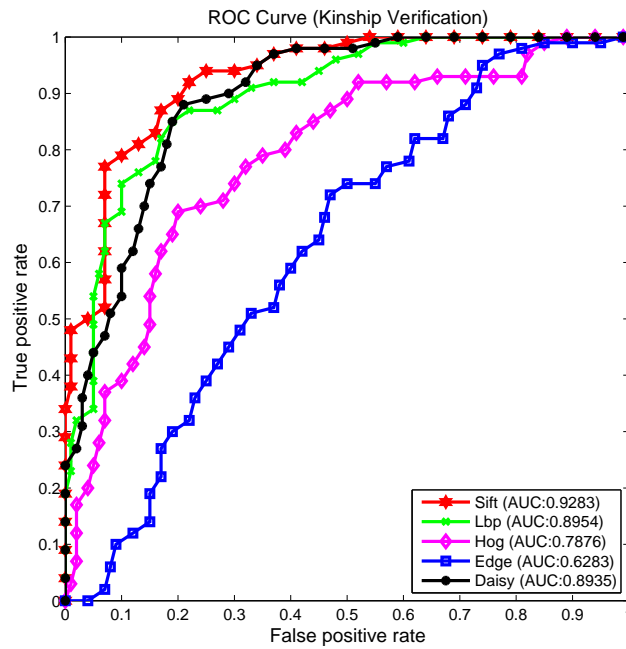


Figure 5.2.2: ROC curves of three different matching schemes (from left to right: Bayes, Discriminative 1, and Discriminative 2, respectively) in kinship verification, using five different facial part representations.

The comparison illustration is shown in Fig. (5.2.3). We apply our discriminative 2 scheme to various representations. From the graph, it's easy to find that all representations in our approach outperform the method proposed in [10].

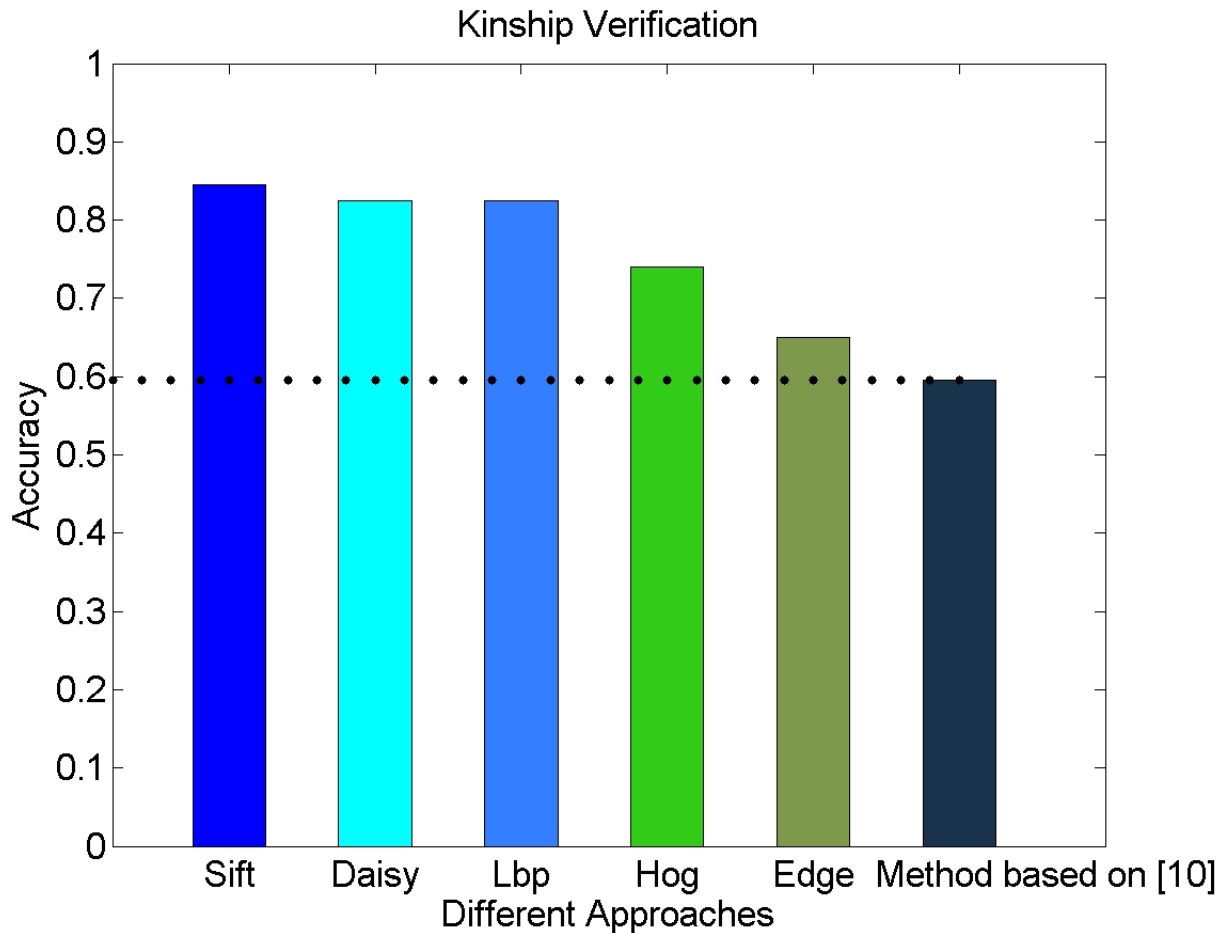


Figure 5.2.3: Bar graph display of kinship verification, comparing various representations under the proposed matching scheme discriminative 2 with the method based on [10]

Compared with our scheme, no familial trait learning has been performed in [10]. From Table 1 in [10], we can see the top feature used in [10] include eye colors, skin colors, and distance between facial parts, these features are not used in our database. The diversity of races is not very discriminative, because our database mainly contains the Caucasian people with only a small diversity of race. So we can't judge if the compared faces have the kinship relationship from their eyes and skin colors.

### 5.2.3 Generation recognition

As far as we know, just one study try to study the problem of generation recognition. Here, we have developed a knowledge-based decision rule to determine if an identified kinship is



within the same generation, across one generation, or across two generations. Generation recognition work is essential to make the understanding of family relationship more deeply. We use three Gaussian distributions modeling the realistic distribution of generation, the input parameter is the age gaps, age gap is calculated based on the estimated age of each given face. To build the model of aging estimation, we combine the MORPH [25] and FG-NET [11] databases. Because of the character of age range in MORPH, the age range is narrow (16 to 67 years), many samples are contained in this database. On the other hand, compared with MORPH, the FG-NET contains less number of face samples, but its age range is wider, an age span from 0 to 69 years old. FG-NET contains more individuals whose ages are below 16. Biologically inspired features combined with SVM classifiers are learned to estimate age [16], but we conducted on a different database. Then we use the learned aging estimation function to apply in our newly collected family photos. The result of generation recognition is shown in the first row of Table 5.3.

	All Kinship Pairs	Kinship Pairs Recognized by			
		HOG	LBP	SIFT	DAISY
Generation Recognition	0.52	0.50	0.55	0.52	0.53
Specific Relationship Recognition	0.41	0.38	0.40	0.42	0.44

Table 5.3: Accuracies of generation recognition and specific family relationship recognition on kinship pairs.

In Table 5.3, we perform the generation recognition work in two cases. In the first case, we use all pairs of faces with the known kinship (ground truth), then we test the accuracy of generation recognition. We obtain an accuracy of 52% shown in the first column. In the second case, we use the recognized kinship pairs based on the matching scheme of discriminative 2 with four different representations (edge representation is not performed here), as shown in the last four columns. The accuracies range from 50% to 55%. The number of pairs of faces used here is usually smaller than using all the pairs, since some kinship pairs are not identified. However, the differences between these two cases are not significant. We can get accuracies over 50% in generation recognition in each case. Though these accuracies are higher than a random guess (about 33%), they are not very high. From this result, we can see the problem of age estimation is still a challenging problem. A more developed age estimation procedure may improve the performance of the generation recognition.

### 5.2.4 Specific family relationship recognition

We can get the accuracies of specific family relationship recognition from the second row of Table 5.3. More similar to the generation recognition, the recognition are also performed in two cases. Firstly, all pairs of faces labeled with the kinship are used to test the accuracy of specific family relationship recognition on these pairs. An accuracy of 41% is achieved in the first column. Secondly, we use the recognized kinship pairs (the same as that used for generation recognition) to perform the experiment, the final result is shown in the last four columns. The accuracies range from 38% to 44%. The differences between these two cases are pretty well. We can get accuracies above 40% in specific family relationship recognition in either case. Compared with the random guess (about 9%), the accuracies got by our scheme perform much better, but still lots of space could be improved. Seen from the result demonstrated, we can see that the problems of age estimation and gender classification are still very challenging in the unconstrained face images.

From the four recognition problems discussed above, more information could be obtained from a pair of faces, instead of just a kinship. As shown in Fig.(5.2.4)., except the kinship, our methods could also report the familial traits with confidence value measures, a sibling relation, and more specifically, the brother-sister relationship.



Figure 5.2.4: Given a pair of face photos (automatically aligned by the eyes), our system can report recognition results at several levels of details.

### 5.2.5 Recognizing family photos.

The experiment in this section separates the family photos from non-family group photos. We can extend the kinship recognition based on a pair of faces to multiple faces in a family photo. Based on the decision rule described in Equ. 2.5.1, we could discriminate family photos from non-family group photos. This problem has not been studied in previous research. To get a quantitative measure of the performance, we conduct ten-fold cross-validation experiments on our database. The results are shown in 5.4, this result is based on different matching schemes and facial part representations for kinship verification. In this problem, the DAISY operator performs slightly better than the SIFT when the Bayes decision is used for kinship recognition. But when our discriminative schemes are used, the SIFT based representation has higher accuracies than DAISY. The highest accuracy of 86.0% is obtained by the SIFT representation under the matching scheme of discriminative 2 for kinship verification. This result is quite encouraging. We show some examples of correct and incorrect recognition results in Fig.5.2.5., the result in the figure is based on the SIFT representation, discriminate 2 classification scheme, and pair-wise kinship combination. We checked the incorrect recognition results Fig. 5.2.5., we found the majority of this problem is caused by the facial variations, for example, pose, illumination, expression, aging, and partial occlusion. Though we have taken these problems into consideration in our work, but we have to admit that it is still very challenging to separate family photos from other groups in the unconstrained images.

	Family/Non-Family Classification Accuracy		
	Bayes	Discriminative 1	Discriminative 2
SIFT	0.770	<b>0.810</b>	<b>0.860</b>
DAISY	<b>0.800</b>	0.800	0.820
HOG	0.660	0.645	0.755
LBP	0.730	0.650	0.770
Edge	0.505	0.650	0.680

Table 5.4: Results of family & non-family photo recognition.



Figure 5.2.5: Some recognition results. Top left: a family photo recognized correctly; Top right: a non-family photo recognized correctly; Bottom left: a family photo recognized incorrectly as non-family; Bottom right: a non-family photo recognized incorrectly as family.

# Chapter 6

## Concluding Remarks

In our work, five computational problems have been presented in family relationship analysis in face photos, and have explored appropriate methods to deal with these problems. We also have shown that the familial traits can be learned from pairs of local, salient facial parts. Then we integrate the learned familial traits to form a decision for kinship verification. We also furthered our work into generation recognition based on the modeling the age gaps, also specific family relationship recognition based on the generation recognition and gender classification. By extending kinship recognition from the pair-wise to multiple pairs, we have studied how to identify the family photos and non-family groups from the given images.

To solve these problems, two discriminative schemes dealing with the familial trait identification, kinship verification and family photo recognition are proposed to compare with the result using a Bayes decision. Various features are proposed to represent facial parts combined with various matching schemes. Through our comprehensive evaluation, we have found that the SIFT descriptor performs better than DAISY, LBP, HOG and Edges. From the analysis of our experiment result, we can see that SIFT is more powerful in terms of scale and affine invariance, considering the possible variations in the unconstrained family and non-family database we collected. Our work is good explore in the generation recognition and specific family relationship recognition based on face images. The encouraging results based on our validation result could inspire further research on family relationship analysis for semantic image understanding and social context analysis from photos.

In our future research, it will be interesting to explore more familial traits to further

improve the performance of our algorithm. Such as one of the interesting works might be used to investigate the impact of familial traits on face recognition. Considering the character of the different family members share common familial traits, how to recognize a person in the family photo will be a very interesting work in the future.

# Bibliography

- [1] T. Ojala, M. Pietikäinen, and D. Harwood (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582 - 585.
- [2] A. Alvergne, R. Oda, C. Faurie, A. Matsumoto-Oda, V. Durand, and M. Raymond. Cross-cultural perceptions of facial resemblance between kin. *J. of Vision*, 9(6):1–10, 2009.
- [3] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.
- [4] M. F. Dal Martello and L. T. Maloney. Where are kin recognition signals in the human face? *J. of Vision*, 6:1356–1366, 2006.
- [5] M. F. Dal Martello and L. T. Maloney. Lateralization of kin recognition signals in the human face. *J. of Vision*, 10(8):1–10, 2010.
- [6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conf. on Comput. Vision and Pattern Recognit.*, pages 886–893, 2005.
- [7] C. Davies and D. Williams. The grandparent study 2002 report. Washington, DC: AARP, 2002.
- [8] L. DeBruine, F. Smith, B. Jones, S. Roberts, M. Petrie, and T. Spector. Kin recognition signals in adult faces. *Vision Research*, 49:38–43, 2009.

- [9] M.-P. Dubuisson and A. K. Jain. A modified hausdorff distance for object matching. In ICPR, pages 566–568, 1994.
- [10] R. Fang, K. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In ICIP, 2010.
- [11] FGNET. The fg-net aging database. In <http://www.fgnet.rsunit.com/>,
- [12] Y. Fu, G.-D. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(11):1955–1976, 2010.
- [13] M. Ghahramani, W. Yau, and E. Teoh. Family facial patch resemblance extraction. In ACCV, 2010.
- [14] G.-D. Guo, C. Dyer, Y. Fu, and T. S. Huang. Is gender recognition affected by age? In *IEEE International Workshop on Human-Computer Interaction*, pages 2032–2039, October 2009.
- [15] G.-D. Guo, G. Mu, Y. Fu, C. Dyer, and T. S. Huang. A study on automatic age estimation on a large database. In *IEEE International Conference on Computer Vision*, pages 1986–1991, 2009.
- [16] G.-D. Guo, G. Mu, Y. Fu, and T. S. Huang. Human age estimation using bio-inspired features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 112–119, 2009.
- [17] G.-D. Guo and X. Wang. Kinship measurement on salient facial features. *IEEE Trans. Instrumentation & Measurement*, 2012. In press.
- [18] L. Hogben. The genetic analysis of familial traits I. Single gene substitutions. *J. of Genetics*, 25(1):97–112, 1931.
- [19] G. Kaminski, S. Dridi, C. Graff, and E. Gentaz. Human ability to detect kinship in strangers’ faces: effects of the degree of relatedness. *Proc.R. Soc. B*, 276:3193–3200, 2009.
- [20] D. Lieberman, J. Tooby, and L. Cosmides. The architecture of human kin detection. *Nature*, 445:727–731, 2007.



- [21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [22] T. J. Mathews and B. E. Hamilton. Delayed childbearing: More women are having their first child later in life. US Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2009.
- [23] OECD. The Organization for Economic Co-operation and Development (OECD) Family Database, OECD, Paris, 2011. <http://www.oecd.org/dataoecd/62/49/41919586.pdf>.
- [24] Narayanan Ramanathan and R. Chellappa. Modeling age progression in young faces. In *IEEE CVPR*, pages 387–394, 2006.
- [25] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *IEEE Conf. on AFGR*, pages 341–345, 2006.
- [26] D. A. Roark, A. J. O’Toole, and H. Abdi. Human recognition of familiar and unfamiliar people in naturalistic video. In *IEEE Intl. Workshop on Analysis and Modeling of Faces and Gestures*, 2003.
- [27] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Trans. Pattern Anal. and Mach. Intell.*, 32:815–830, 2010.
- [28] V. N. Vapnik. *Statistical Learning Theory*. John Wiley, New York, 1998.
- [29] R. E. Vonderohe. What is so “grand” about grandparenting? *NebGuide*, 2008.
- [30] G. Wang, A. Gallagher, J. Luo, and D. Forsyth. Seeing people in social context: Recognizing people and social relationships. In *Eur. Conf. on Comput. Vision*, 2010.
- [31] S. Xia, M. Shao, and Y. Fu. Kinship verification through transfer learning. In *IJCAI*, pages 2539–2544, 2011.

- [32] Xia, S. and Shao, M. and Fu, Y. Toward Kinship Verification Using Visual Attributes. International Conference on Pattern Recognition (ICPR), 2012.
- [33] Z. Sun, A. Paulino, J. Feng, Z. Chai, T. Tan, and A. Jain, “A study of multibiometric traits of identical twins,” in Proc. SPIE Defense, Security, and Sensing: Biometric Technology for Human Identification, 2010, pp. 1–12.
- [34] Bibliography B. Klare, A. Paulino, and A. Jain. Analysis of facial features in identical twins. In International Joint Conf. Biometrics, 2011.
- [35] Pruitt, M.T. and Grant, J.M. and Paone, J.R. and Flynn, P.J. and Bruegge, R.W.V. Facial recognition of identical twins. In IJCB, pages 1–8, 2011.
- [36]