Mining In-Class Social Networks for Large-Scale Pedagogical Analysis

Xiao-Yong Wei
Machine Intelligence Laboratory
College of Computer Science
Sichuan University
610065, Chengdu, China
cswei@scu.edu.cn

Zhen-Qun Yang
Machine Intelligence Laboratory
College of Computer Science
Sichuan University
610065, Chengdu, China
jessicayang@scu.edu.cn

ABSTRACT

Modeling the in-class student social networks is a highly desired goal in educational literature. However, due to the difficulty to collect social data, most of the conventional studies can only be conducted in a qualitative way on a small-scale of dataset obtained through questionnaires or interviews. We propose to solve the problems of data collection, social network construction and analysis with multimedia technology, in the way that we can automatically recognize the positions and identities of the students in classroom and construct the in-class social networks accordingly. With the social networks and the statistics on a large-scale dataset, we have demonstrated that the pedagogical analysis for investigating the co-learning patterns among the students can be conducted in a quantitative way, which provides the statistical clues about why prior studies reach conflicting conclusions on the relation between the students' positions in social networks and their academic performances. The experimental results have validated the effectiveness of the proposed approaches in both technical and pedagogical senses.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences

General Terms

Algorithms, Experimentation, Human Factors, Verification

Keywords

In-Class Social Networks, Pedagogical Analysis

1. INTRODUCTION

In educational literature, social interactions among learners is of great importance, for its effectiveness to promote the learning process by facilitating learner-to-learner co-construction of knowledge and the sharing of information and resources [1–5]. Therefore, a large amount of effort has been carried out to identify and investigate the social networks in which the learners are involved, in the hope of designing proper pedagogical strategies accordingly. However, even the proposal of the idea may date back to 1930s [1, 6], most of the investigations are still limited in a small scale which usually include less than 100 students (e.g., 25 students in [5], 11 students in [7]), due to the fact that the traditional way for collecting social data through questionnaires and interviews is time-consuming and expensive. The rapid development of online technologies seems to be a new opportunity to address this problem for its promise to collect the immense social data in an automatic manner. In this regard, online social network analysis has led to tremendous emerging researches in various disciplines such as sociology, pedagogy, psychology, and computer science.

In contrast with the overwhelming interest among practitioners, as commented by Boyd & Ellison [8], the majority of existing studies on online social networks still remain both conceptual and empirical, with a few of them having explored the link between social networks and education. Furthermore, researches along this line mainly focus on the "out-of-school" or "after-school" social practices of students within the "virtual" networking environment (e.g., [7, 9, 10]), where they may not behave naturally [8, 9]. By contrast, the study on the students' social learning patterns in their "real-life" at school has not been sufficiently conducted, although it is highly desired even before the birth of the term "social network" (by Jacob Moreno's work in 1930s). This is again due to the difficulty for obtaining social data from the real-life educational environment. Moreover, most existing educational researches have been conducted using "qualitative" analysis which samples a small-scale of students from the whole population for case studies [11]. It is thus easy to cause conflicting conclusions when the sampling are not representative in a statistical sense, e.g., Divjak *et al.* have reported in [12] that there is no relation between a student's position in student social network and her/his academic performance, while most of other researchers believe the existence of the relation [3, 4].

In this paper, we conduct a pilot study aiming to equip conventional pedagogical analysis with multimedia technology, for *automatic* and *large-scale* in-class social data collection and analysis. We first propose a novel image-based method to identify the in-class social networks which encapsulates both the student-to-student and student-to-teacher interaction-ships, on the basis of which we investigate their relations to the students' academic performances. The method can be easily applied to real-life classroom environment.

The framework of our method for in-class social network mining is shown in Figure 1, which consists of three parts:

MM'12, October 29–November 2, 2012, Nara, Japan.

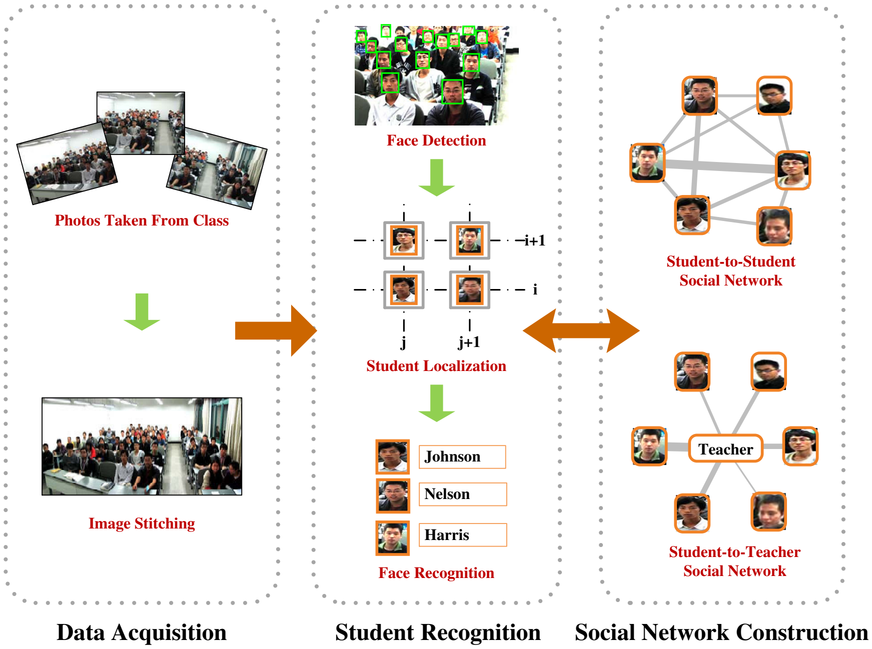

Figure 1: Framework of In-Class Social Network Mining[1].

data acquisition, student recognition, and social network construction. In data acquisition, a teacher takes a set of photos of the students in the class and the photos are combined into a single image using image stitching algorithm [13]. The faces in the image are then detected with their positions (i.e., row and column numbers in the classroom) identified by the student localization algorithm, on the basis of which face recognition is employed to identify the names of the students. In each course under investigation, we leave 9-13 weeks for the students to form teams for the course project, so as to encourage them to choose to sit and discuss with potential teammates. Therefore, at the end of the semester, the statistics on how each student sits together with her/his peers for co-learning and her/his preference of the sitting position can be obtained, for constructing the student-to-student and student-to-teacher social networks respectively. In the framework, the student recognition and the social network construction are two interdependent parts, in the way that the social network construction needs the information about the students' identities and their sitting positions, and the performance of face recognition can be improved if the social interaction-ship among students are known a priori. Therefore, we dynamically update the social networks after each class and the result will be used to help face recognition in the classes followed. There are several technical challenges which have been addressed in this framework, including

- *Face Recognition in Extreme Conditions*: Face recognition has been intensively studied for over four decades since 1970s [14] and still remains an active research topic nowadays. However, its performance is often limited when applied to real applications, due to the variations of illuminations, poses, facial expressions, motions, occlusions, and so on [15–17]. In additional to all these issues, as shown in Figure 2, we meet an even worse situation when applying face recognition to the photos captured in classroom, in the way that the faces are with various sizes and resolutions, sometimes being distorted seriously, and possibly with the majority of facial areas missing. According to our study, this makes most of the state-of-the-art face recognition approaches work awkwardly, because nearly all of them are developed on datasets that are acquired in well-controlled environments [16, 18]. We argue that conventional approaches mainly focus on the facial area of a person, ignoring the fact that the target's social context (e.g., the potential teammates he/she often sits with) is of great help to the identification process. To address this problem, we propose a *social inference scheme* which recognizes a student with the respect to the identifications of her/his socially related peers (learnt from the in-class social network), with the hope that, when the student's appearance in current photo is not visually distinguishable, her/his identity can still be inferred when her/his potential teammates are recognized at the neighboring seats;

- *Large-Scale Labeling Efforts*: Like all supervised learning approaches, before applying face recognition, we need to annotate a large scale of samples for training, which, in general, is costly. We bypass the difficulty by encouraging the students to label their faces by themselves. In particular, an "attendance checking" function for fun is embedded into all course homepages, which allows teachers to upload the photos and automatically stich the photos into a single one, on which the students (after login for downloading course materials) can tick their faces to confirm their attendances (if there are no machine-predicted labels available for their faces) or just to correct (if the predictions are wrong). The attendance records not only leave a good reference to the pedagogical analysis, but also accomplish the labeling at the same time. Fur-

[1] All portraits in this paper are used with the consents from the students.

Figure 2: Challenges for face detection/recognition in an uncontrolled environment.

thermore, the function itself is used for case-study in several courses related to Image Processing, Pattern Recognition, Data Mining, or Multimedia Computing. It has significantly boosted the students' interest;

- *Sitting Position Identification*: Detecting where a student sits in the classroom from the 2D image is not a trivial task, as the sitting heights of the students are diverse and it is impossible to obtain in advance. Moreover, the stitched images are with highly inconsistent view points and are often seriously distorted. We propose to use linear regression to model the seat arrangements in the classroom so as to estimate the positions of students and construct the in-class social networks. The details will be given in Section 3.2.

The contributions of this paper can be summarized as: the integration of multimedia technology into pedagogical analysis, the proposal of a practical approach for large-scale in-class social network construction, and the introduction of the social inference scheme into face recognition for working in extreme conditions.

2. RELATED WORK

The topic of this paper is related to several disciplines which may include social science, educational research, data mining, computer vision and computer graphics. Due to the highly interdisciplinary nature and the space limitation, we can only provide a brief review with the effort to cover the related studies to the largest extent.

It has been widely believed that social learning or community-centered learning within social networks has positive influence on a student's academic success [1–5]. Although the belief may have a much longer history than that of its formal proposal in 1930s [1, 6], however, due to the lack of large-scale data for social network modeling and quantitative validation, it has remained conceptual until the emergence of online social network sites (SNSs, e.g., Facebook, MySpace). SNSs provide not only a way to attract the students' participation, but also a tool to collect social data during the online communications (e.g., interaction-ship to their peers, material/information exchanged) , which in turn initiates a large amount of emerging researches for sociology, pedagogy, psychology, and computer science [8]. With the help of SNSs, educational researchers have conducted intensive studies to explore the relation between the students' academic performances and the social communities they are involved, with many positive results reported (e.g., researches by Ellison *et al.* [19] conducted on Facebook and Kraut *et al.* on general Internet communities [20]). However, despite the disagreements which also appear all the time, as stated by Boyd & Ellison [8], most of these studies are still conceptual and empirical in nature. Furthermore, many researchers (e.g., Walter *et al.* [21] and Boyd & Ellison [8]) have realized that the social interaction-ship revealed from the virtual network environment is significantly different from that of face-to-face communications in the students' real life. Therefore, most of SNSs proponents can only claim a "complementary" role for their findings to the traditional pedagogical analysis.

In contrast to the studies for the "out-of-school" SNSs, modeling the students' social interactions in "in-class" environment is always highly desired [22, 23], as it is one of the essentials of our pedagogical system. Due to the difficulty of collecting social data again, there are a few studies that have investigated the relation between the students' academic performance and their "in-school" social interactions. Examples along this direction include [12] which has investigated the relation in two social networks (of "teamwork" and "material exchanging"), and [5] which conducts a similar investigation within a social network constructed from social connections (e.g., friendship, collaboration-ship) reported by students. In these studies, social network analysis [24] is employed to measure the different types of networking positions, such as *betweenness*, *closeness*, and *degree* (which we will introduce in Section 4), for discovering the social ties or social capital (the leadership or controllership) in the communities. However, the data collecting process in these studies are still following the conventional fashion of questionnaires and interviews, resulting in none of them having directly constructed an "in-class" social network. Moveover, these investigations are still in a small-scale (e.g., 27-52 students in [12], 25 students in [5]), making the results not convincing enough in a statistical sense.

While the studies mentioned above are mainly focusing on the student-to-student social connections, the student-to-teacher interaction-ship have also attracted a lot of research attentions (e.g., [22, 25–28]), given the fact that most of our pedagogical systems are teacher-centered. These studies are mainly conducted by investigating the influence of seat arrangement in a teacher-centered classroom on the students' performances, in the sense that students sitting at different positions can have different chances to receive instructions from or to interact with the teacher, which determines their achievement at the end of the semester. The data of these studies are usually collected by fixing the students' sitting positions so as to investigate the effect of the arrangement on the students' academic performances [27, 28]. Even with evidential correlation between seat arraignment and academic performance reported, these studies are all conducted in a qualitative way, which cannot quantitatively answer questions like: how far (the exact distance) the student sitting from the teacher would be an advantage/disadvantage? Furthermore, the method for collecting data is questionable, because forcing a student to sit in a fixed position can distort her/his natural behaviors during learning.

In this paper, we propose to address the social data collecting problem with multimedia technology, with the hope that the social data can be collected in an automatic manner, and the in-class networks can be captured in a natural way (the students are allowed to choose where to sit freely). The method will not bias the students' social preferences and is applicable to large-scale investigation at the same time. Technically speaking, our methods for student localization and recognition are related to techniques of image stitching/alignment [13] and face recognition [15–17] respectively.

Image stitching is developed to compose images with varying degrees of overlap into a single one for providing a wider view of the target. To this end, nearly all approaches utilize the assumption that one image is the result of a linear transformation from another image within the Homogeneous Coordinate System, which can be modeled by a transformation matrix. Therefore, once the matrix is learnt, we can transform one image back to the same plane where another image lies in (this has been sophisticatedly developed in Computer Graphics) so as to compose the two image into a single one. The transformation matrix is learnt by finding correspondence between points in the overlapping areas of two images. Various features have been proposed for detecting the correspondence (matching points) [13], among which the Scale Invariant Feature Transform (SIFT) by Lowe's [29] is considered one of the state-of-the-art. We adopt the same feature in this study. Due to space limitation, the reader is referred to [29] for more details.

Face recognition is one of the most intensively studied technologies in computer vision, with new approaches emerging every year and encouraging results reported. According to recent survives [15–17], however, most existing studies are conducted on data captured in controlled environment. It makes the successes of those approaches limited to real applications, where the variations of illuminations, poses, facial expressions, motions, occlusions and so on always raise big challenges for both face detection and recognition. As most of those challenges are still considered as open questions, generally, there is no conclusion on which approach(es) can perform the best. Nevertheless, there are some approaches are more popularly employed than others. For example, the Viola-Jones detector [30] is considered the start-of-the-art for face detection in a survey [17] mainly written for Viola-Jones detector and its variations. For face recognition, the methods using Eignfaces [31], Fisherfaces [32] and Sparse Coding [33] are popularly adopted as baselines. In this paper, our aim is not to solve the open questions of face recognition in a general sense, but instead to bypass those difficulties in a classroom sitting with the help of the contextual information obtained from the in-class social networks. Since in our framework the student recognition and the construction of the in-class social networks are two interdependent processes, we will introduce the recognition first and then turn to social network construction.

3. STUDENT RECOGNITION

To stitch multiple photos, we adopt the method proposed in [10], which automatically find the matching points between two photos using SIFT features, and compose them into a single image accordingly. The resulting image will be used for face detection, student localization and recognition.

3.1 Face Detection

We detect the faces in the stitched image mainly using the Viola-Jones detector [30] which is considered the state-of-the-art in a recent survey [17] and has been implemented in OpenCV. However, due to the difficulties introduced in Section 1 and the fact that the OpenCV implementation is mainly trained on frontal faces, it can only detect 78% of the faces in our experiment. To deal with this problem, we implement the simple skin-color based detector [34] and pool the results from both detectors. The teachers will help to filter out the non-facial areas (less than 3% of those in the pool) with tiny effort before the student localization and recognition being carried out.

Figure 3: Student localization by linear regression and image alignment ($\frac{a|b}{c|d}$): (a) faces in the stitched image for the first class (the source image) are manually associated to row and column numbers, on the basis of which a line indicator (yellow line) is trained for each row or column using linear regression; (b) the matching points between a newly taken image (the target image) and the source image are detected (as connected with green lines); (c) the source image is aligned to the target with all line indicators being transformed; (d) RANSAC algorithm is applied on each transformed line indicator to obtain its new model in the target.

3.2 Student Localization

Prior to face recognition, we detect where (i.e., row and column numbers) each student sits in the classroom. There are mainly two difficulties, namely the diversity of the students' sitting heights and the inconsistency of view-points across the stitched images. To tackle the first issue, we manually annotate the row and column numbers of each student in the image taken in the first class (call source image hereafter), and use linear regression for the centers of faces appearing at the same row (column). As shown in the upper-left of Figure 3, the result is a line (the yellow ones) for each row (column) indicating the places where the face centers will likely to appear, so that, under the assumption that all the lines can be correctly detected in a newly taken image (call target image hereafter), we can assign a face to the nearest line(s) to identify its row (column) number(s). To detect the lines in the target image, we use the method [29] again (which we employ for image stitching) to find matching points between the source and target images (see the green lines from the upper-left to upper-right in Figure 3), and to calculate the coordinate transformation matrix, on the basis of which we transform the source image into the coordinate plane of the target (as shown in the bottom-left of Figure 3). This process is called image alignment in literature [13], which also transforms each row (column) indicator (i.e., line) into the target coordinate plane, so that we can

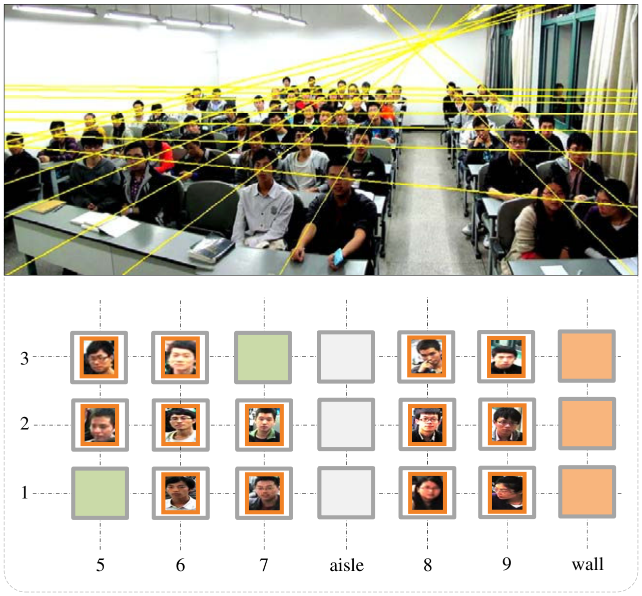

Figure 4: Result of student localization and a partial view of the face graph $\mathcal{G}$.

use Random Sample Consensus (RANSAC) [35] to find the regression model for each of the transformed row (column) indicator and then apply to the target image. To avoid being biased by noises or outliers when using all points for regression, RANSAC automatically searches a set of points which can represent the distribution of the majority of the data, so that, in our case, using RANSAC for regression can filter out the points lying in the distorted parts of each transformed row (column) indicator (see the bottom-left of Figure 3). The result of RANSAC is shown in the bottom-right of Figure 3. Since the image alignment is a comparably matured technique, we skip the details here for space limitation. The reader is referred to [29] for more details. Note that the alignment has solved the inconsistency issue of view-points at the same time. As shown in Figure 4, the result of student localization can be represented as a face graph $\mathcal{G}$ for further analysis, where three types of extra nodes, i.e., the "empty seats", "aisle" and "wall", are automatically added for providing more precise neighboring information. To traverse the face graph conveniently, we define a function $\mathcal{N}(\boldsymbol{x})$ on $\mathcal{G}$, which returns a set of faces that are the neighbors for any given $\boldsymbol{x}$ within an 8-neighbor system.

3.3 Face Recognition within A Social Context

Our scheme for recognizing the identities of students include two stages: a *local classification* stage which recognizes each student individually by only considering her/his visual features, and a *social inference* stage which infers the identity of a student regarding to those of her/his neighbors in current classroom, with the hope that further considering the social context of a student can help improve the performance of weak classifiers in the local stage. This can be formulated as an iterative process as

$$\boldsymbol{p}^{(i+1)}(\boldsymbol{x}) = \lambda\, Social^{(i)}(\boldsymbol{x}) + (1-\lambda)\, Local(\boldsymbol{x}) \qquad (1)$$

where the superscript $(i+1)$ indicates the iteration number, $\boldsymbol{x}$ represents an observation of a face, $\boldsymbol{p}^{(i+1)}(\boldsymbol{x})$ is a vector encapsulating a posterior probability distribution $p(l_k|\boldsymbol{x})$ at the $(i+1)$-th iteration over a set of identity labels (e.g., names, student IDs) $\mathcal{L} = \{l_k\}_{k=0}^{|\mathcal{L}|}$, $Local(.)$ and $Social(.)$ are functions for local classification and social inference respectively, and $\lambda \in [0,1]$ is a factor to control speed of converging and to balance the influence of the two functions at the same time. The process will converge when $\|\boldsymbol{p}^{(i+1)}(\boldsymbol{x}) - \boldsymbol{p}^{(i)}(\boldsymbol{x})\| \rightarrow 0$. There are two strategies which have been considered to utilize the social context of $\boldsymbol{x}$ for improving student recognition. First, the process iteratively fuses the result of social inference into that of the local classification, so that the classification results of the weak local classifiers can be improved through collaborative learning with the neighboring classifiers. Second, the process will be performed on the face graph $\mathcal{G}$ (see Figure 4) which defines the neighboring interaction-ship. This enables the effort of collaboration to be propagated to the whole classroom, with the hope that, once the process converges, the identity labels assigned to faces are the most consistent one to both the neighboring interaction-ship in the face graph and the social context encapsulated in the social network. In the rest of this section, we describe our methods for implementing the local classification and social inference functions.

3.3.1 *Local Classification Function* $Local(.)$

The local classification can be implemented by employing any state-of-the-art face recognition approach developed in Computer Vision. In our case, we choose the one proposed by Wright *et al.* [33], which picks several sample images from each student to compose a dictionary, and represents each instance $\boldsymbol{x}$ as a linear combination of the samples in the dictionary by associating each sample with a coefficient through Sparse Coding. This results in a set of coefficients with a minimized number of non-zero entities assigned to the samples that are most similar to $\boldsymbol{x}$. Therefore, we can calculate the posterior probability that $\boldsymbol{x}$ is from a given student by accumulating the coefficients of her/his sample images. By repeating this for all the students, the output of $Local(.)$ is a vector $\boldsymbol{q}(\boldsymbol{x})$ which encapsulates another posterior probability distribution $q(l_k|\boldsymbol{x})$ over $\mathcal{L}$.

3.3.2 *Social Inference Function* $Social(.)$

To utilize the in-class social context for student recognition, a basic belief is that a student would likely to appear if her/his potential teammates appear at the neighboring seats. Let us temporarily assume this co-occurrence can be learnt form the in-class student-to-student social network (the details of the network construction will be given later in Section 4). The interaction-ship of any label pair (l_k, l_n) can thus be represented as a conditional probability $\pi(l_k|l_n)$ indicating how likely l_k will be observed when l_n appear at its neighboring seats. The pair-wise interaction-ship can be encapsulated into a matrix $\mathbf{R}_{|\mathcal{L}|\times|\mathcal{L}|}$ where each entity equals to $\mathbf{R}_{k,n} = \pi(l_k|l_n)$. The social inference function is then implemented as

$$Social^{(i)}(\boldsymbol{x}) = \frac{1}{|\mathcal{N}(\boldsymbol{x})|} \sum_{\boldsymbol{y}\in\mathcal{N}(\boldsymbol{x})} \mathbf{R}\cdot\boldsymbol{p}^{(i)}(\boldsymbol{y}) \qquad (2)$$

where $\mathcal{N}(\boldsymbol{x})$ is the neighboring function we define in Section 3.2. Eq. (2) is intuitively a function which summarizes the recognition results of $\boldsymbol{x}$'s neighbor nodes on $\mathcal{G}$. Later in Eq. (1), the summarized result will be iteratively prorogated to $\boldsymbol{x}$ after fusing with the output of $Local(.)$.

It is worth mentioning that this iterative recognition scheme might be closely related to the Random Walk Model, and also possible to be extended to Markov Random Field or Conditional Random Field. To keep this pilot study as simple/straightforward as possible, we leave these for future investigation.

Once the iteration is converged, we can obtain a vector $\hat{\boldsymbol{p}}(\boldsymbol{x})$ for each face, indicating its identity potentials over the label set $\mathcal{L}$. Now we need to make a decision by assigning $\boldsymbol{x}$ to a determined identity label, which can be represented as an unit vector $\boldsymbol{e}$ over the label set $\mathcal{L}$. The components of $\boldsymbol{e}$ are all zeros excepting the k-th one, indicating the assignment of $\boldsymbol{x}$ to the k-th label in $\mathcal{L}$. The rationality of the assignment can then be measured by the inner product $< \hat{\boldsymbol{p}}(\boldsymbol{x}), \boldsymbol{e} >$, on the basis of which we can use Dynamic Programming to maximize the sum of rationalities of all faces, and therefore complete the recognition process.

4. SOCIAL NETWORK CONSTRUCTION AND PEDAGOGICAL ANALYSIS

In this section, we construct two in-class social networks, namely the student-to-student and student-to-teacher networks respectively, on the basis of which we perform pedagogical analysis following some of the educational literature (e.g., [5,12,22,25–28], where the analysis are conventionally conducted by questionnaires/interviews).

4.1 Student-to-Student Social Network

With the positions and identities recognized in last section, the construction of student-to-student network can be performed by using students as vertexes and connecting two vertexes according to the frequency (or co-occurrence) of the corresponding students sitting in neighboring seats (i.e., the joint probability $\pi(l_k, l_n)$). We also record the frequency of the attendance for each vertex as it can be used as the prior probability for each label (i.e., $\pi(l_k)$). This will make the calculations of $\pi(l_k|l_n)$ and $\pi(l_n|l_k)$ in Section 3.3 be easily carried out using Bayes' formula. Note that, for the first class, when the position indicators and the face detectors have not been trained yet, these probabilities are calculated from the results of manual annotation (i.e., positions identified by teachers and labels provided by the students during attendance checking).

Following the main stream of pedagogical studies using social network analysis (SNA) [24], to measure the "sense of community" of each student, we calculate the *betweenness*, *closeness* and *degree* (which are commonly adopted measures in educational research, e.g., [5, 12]) for each vertex in the student-to-student social network. Linear regression analysis can then be applied on these measures to investigate their relation to students' performance (the final grades).

Let us denote $\mathcal{V}$ the set of vertexes in the network. The *betweenness* (usually called *betweenness centrality* in literature) is defined as

$$Betweenness(v) = \sum_{s \neq v \neq t,\ s,v,t \in \mathcal{V}} \frac{g_{st}(v)}{g_{st}} \tag{3}$$

where g_{st} is the number of shortest paths from $s \in \mathcal{V}$ to $t \in \mathcal{V}$, and $g_{st}(v)$ is the number of shortest paths from s to t that $v \in \mathcal{V}$ lies in. In SNA theory, the *betweenness* measures a student's connectivity by further considering the "neighbors" she/he is indirectly connected with, which gives a higher value to those bridging social groups. A "between" student is the cutpoint in the shortest path connecting two other students, and thus might control the flow of information/resource exchange. The larger *betweenness* a student has, the higher centrality she/he is in the community.

The *closeness* is measured by

$$Closeness(v) = \left[\sum_{v \neq t,\ v,t \in \mathcal{V}} d(v,t) \right]^{-1} \tag{4}$$

where $d(v,t) = d(t,v)$ is the shortest distance (called *geodesic distance* in SNA) between two vertexes $v \in \mathcal{V}$ and $t \in \mathcal{V}$. The *closeness* is intuitively the inverse of the sum of geodesic distances from a students to her/his peers, which reflects the ability of a student to access information through other community members.

The $Degree(v)$ of a student is defined as the number of students with whom v is directly connected. It is a local measure of centrality of a student within the small social group she/he is directly involved.

4.2 Student-to-Teacher Social Network

As introduced in Section 2, in educational literature, the student-to-teacher interaction-ship is usually inferred from the students' preference of sitting positions. The corresponding network is thus a star shaped graph with the teacher at the center, connected by the students through edges weighted by their in-class distances to the teacher (as shown in Figure 1). With the student localization method proposed in Section 3, we can calculate the distances more accurately, rather than a qualitative measure by "far" or "close" (indicated by which row a student sits) in conventional studies [22, 25–28]. Furthermore, once the distances are determined, we can employ linear regression analysis again to investigate the "quantitative" relation between the students' sitting preference and their academic performances, which is not easy to be obtained in conventional studies.

Suppose a student is sitting at the i-th row and j-column in the classroom. With two additional "virtual rows" added in front of the first row and every aisle being considered as a "virtual column", the student's distance to teacher can be estimated by her/his geometric distance to the center of the podium (where most teachers usually stand at) as

$$Distance(i,j) = \sqrt{i^2 + (j - \frac{c}{2})^2} \tag{5}$$

where c is the total numbers of columns, and the podium is assumed to be placed at the 0-th row and the $(\frac{c}{2})$-th column.

5. EXPERIMENT-I: DOSE THE APPROACH TECHNICALLY WORK?

In this section, we validate the practicability of the proposed framework with a technical perspective, while the pedagogical effectiveness will be evaluated in next section. All the experiments are conducted on a dataset collected from 6 courses (related to Image Processing, Pattern Recognition, Data Mining, or Multimedia Computing) at two universities across two semesters, which include 379 students aging from 20 to 24 with 5,040 faces captured at 132 classes. Discussions between students are highly encouraged in class so as to observe the co-learning patterns. The ground-truth is composed of labels provided (when no machine labels

Figure 5: Results of student localization in a sparse, normal and dense classroom respectively.

are predicted) or corrected (when machine predictions are wrong) by the students themselves through the "attendance checking" website. The dataset can be downloaded at our demonstration page[2].

5.1 Student Localization

With method introduced in Section 3.2, we are able to detect the positions of students accurately. Figure 5 shows three examples form the sparse, normal, and dense class respectively. It is easy to see that the localization method works well for the students located at the central-front areas, even when the density is considerably high. In general, columns are easier to be correctly detected than rows. To further evaluate the accuracy, we manually check the localization results for 11 classes of the most dense course, where 631 positions in total need to be recognized. The precision is as high as 83.77%, and 94.17% of the errors are caused by mis-assignments of the student with only one row shift to front or back. Due to the 8-neighbor system adopted in face recognition (cf. Section 3.3), the influence of the one-row mis-assignment errors is not significant for the processes after localization.

5.2 Face Recognition with Social Inference

To evaluate our social network (SN) based face recognition scheme, we use three popularly employed baselines: sparse coding method (*Sparse*) [33], *Eigenfaces* [31], and *Fisherfaces* [32]. All baselines are integrated into the proposed scheme to create three network-based versions as *Sparse+SN*, *Eigenfaces+SN*, and *Fisherfaces+SN* (with the the λ in Eq. (1) empirically set to 0.85). Since in our study, the students will provide the labels for theirs faces after each class, the number of training examples for each student is then in-

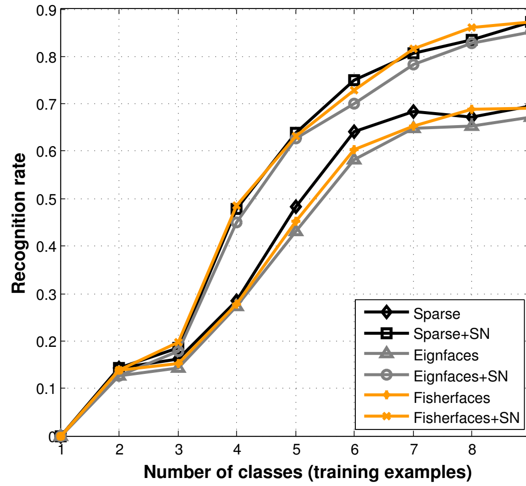

Figure 6: Comparison of recognition rates.

creasing periodically, and the corresponding classifier will be retrained with the same manner.

The result of performances comparison is shown in Figure 6, where the recognition rate is defined as the rate of students being correctly recognized across all classes at the point of time. The maximal number of classes in Figure 6 is 9 because the minimal number of classes for a course is 10 in our dataset. It is easy to see that, with the help of the in-class social network, the recognition rates of the baselines are significantly improved by 28.28%$\pm$18.56, 32.79%$\pm$16.33, and 34.74% $\pm$ 18.75, for *Sparse*, *Eignfaces* and *Fisherfaces*, respectively. The accuracies of the network-based versions are all approaching 65% after the 4-th class (i.e., there are generally 4 training examples for each student), which technically means that more than 3/5 of the students no longer need to label their faces. It has largely reduced the annotation effort.

For the three baselines, *Sparse*, *Eignfaces* and *Fisherfaces* have all demonstrated moderate performances with the final recognition rates less than 70%, which are much lower than those reported in literature (e.g., generally over 80% for all of them in [33]). It might confirm the hypothesis in several recent surveys [15–17] that most of the conventional approaches may work awkwardly in an uncontrolled environment. Nevertheless, *Sparse* still appears to work slightly better than the other two, which may attribute to its "example-based" nature. In a scenario where the training faces are scarce and are with high diversity, "summarizing" the examples with holistic features may even cause confusion, which sequentially affects the performances of *Eignfaces* and *Fisherfaces*. By contrast, without the modeling process, example-based methods like *Sparse* may work better in this case, with the hope that a new instance can be matched to some of the training examples with high probability when the corresponding student occasionally sit at the similar position in the classroom where she/he sits previously, so that her/his pose to the camera, even might be a difficult case for conventional face recognition, is similar to those of the training examples. However, the slight favorite of *Sparse* disappears when all the baselines are upgraded to their network-based versions. This again confirms that the employment of social inference is the biggest contributor to the recognition rate improvement.

[2] http://www.cs.cityu.edu.hk/~xiaoyong/in.class.sn/

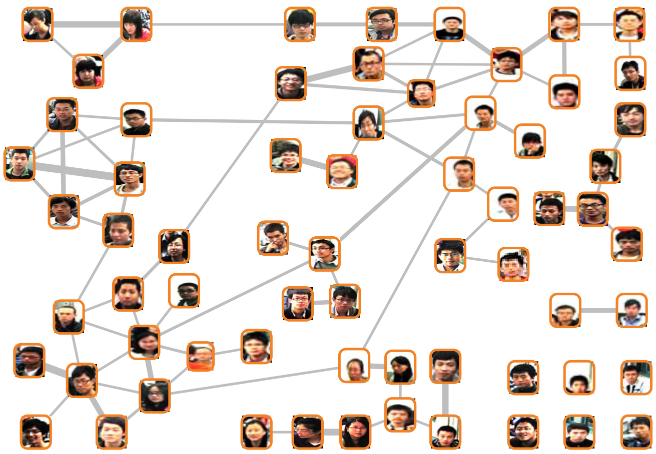

Figure 7: Student-to-student social network constructed from the course with the highest density. The edges with weights smaller than 0.3 are removed for better visualization.

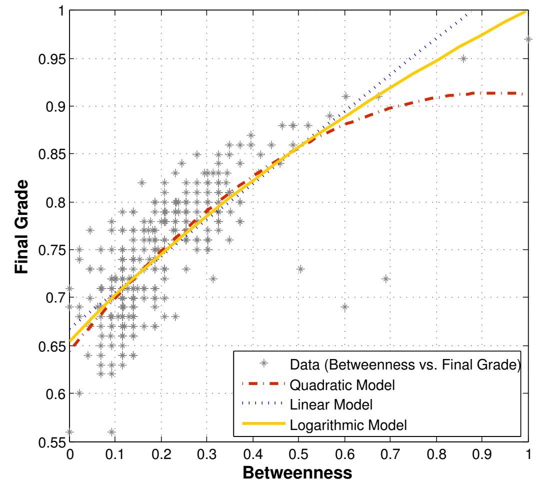

Figure 8: Regression analysis on the relation between the students' *betweenness* and final grades.

5.3 Social Network Construction

With the students' positions localized and their faces recognized, the construction of the in-class social networks are straightforward by following the methods introduced in Section 4. Figure 7 shows the student-to-student social network for the most dense course including 66 students, where the students who are closely related in the network are those frequently sitting near to each other, while the students apart from the center of the community are those who choose their seat randomly. The student-to-teacher social network has also been successfully constructed, which is simply a star shaped graph as shown in Figure 1. We skip its presentation here for space limitation. The whole graph can be found on our demonstration page[2].

To validate the ability of the social network to model the students' relationship in real-life, a quick email survey has been conducted involving all the students by asking them to a) rate the top-5 classmates she/he most communicates with; b) give classmates (up to 5) she/he builds friendship with. Till the time of writing, 324 students have responded. Using the data as ground-truth, we observe that the student-to-student social network can predict the in-class interaction-ship with mean average precision (MAP) as high as 96.67% and the real-life friendship with 87.23%.

6. EXPERIMENT-II: DOSE THE PEDAGOGICAL ANALYSIS MAKE SENSE?

In this section, we investigate the students' positions in the two in-class social networks in related to their academic performances. All measures (i.e., *betweenness*, *closeness*, *degree*, *distances*, and final grades) are normalized into the range of $[0, 1]$. Regression analysis is adopted to study the relations.

6.1 Student-to-Student Interaction-ship

Figure 8 shows the students' final grades over their *betweenness* to investigate whether the *betweenness* can be a predictor for the academic performance (as in [5,12]). While a positive correlation between the two variables can be seen apparently, we conduct a linear regression following the work in [5]. The model fits the data well, further confirming the positive correlation between the performance and the *betweenness*. Note that in [5], the authors have observed a negative correlation between these two variables. This is not conflicting to our result, in the way that the student social network in [5] is constructed from the communication logs (of an online forum) where the students having high *betweenness* might be those who have spent comparably more time online and who are enthusiastic about replying the posts of others. Our result further suggests the difference between the students' behaviors in their "virtual" and "real" lives. Compared with the result reported in [12] where the network is learnt from the questionnaires in which the students rate their interaction-ship to others, our result is consistent to the first conclusion the authors have made that the successful students are those with high *betweenness* because they are willing to communicate with others and benefit from the co-learning. However, the authors have also concluded that the students with high *betweenness* are not necessarily those successful in school, because one exception has been observed in their study. On the basis of the fact that the conclusions in [12] are made on only 4 students with high *betweenness*, we argue that the one exception is possibly one of those outliers (can also be found in bottom-right of Figure 8) which makes the conclusion statistically unfair. Nevertheless, we treat our result not conclusive. It needs investigations in larger scale and interdisciplinary collaborations. In addition, we have also conducted the regressions with logarithmic and quadratic models, as shown in Figure 8, in which the logarithmic model demonstrates the best ability to fit the data than the other two.

For the *closesness* and *degree*, the regression results shown in Figures 9 and 10 have demonstrated similar relations. This suggests that a student is tend to success in school if she/he has an averagely shorter path to access the knowledge of others, or if she/he has more "friends" to provide immediate information for learning. The results is consistent to those reported in [12] and [5]. Moreover, the linear, logarithmic and quadratic have demonstrated similar abilities for regression analysis in both Figures 9 and 10.

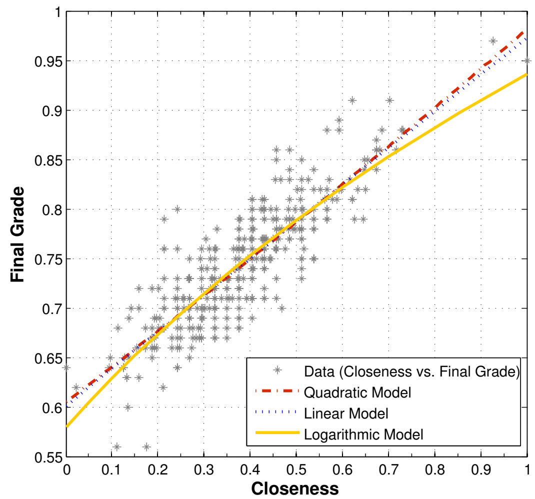

Figure 9: Regression analysis on the relation between the students' *closeness* and final grades..

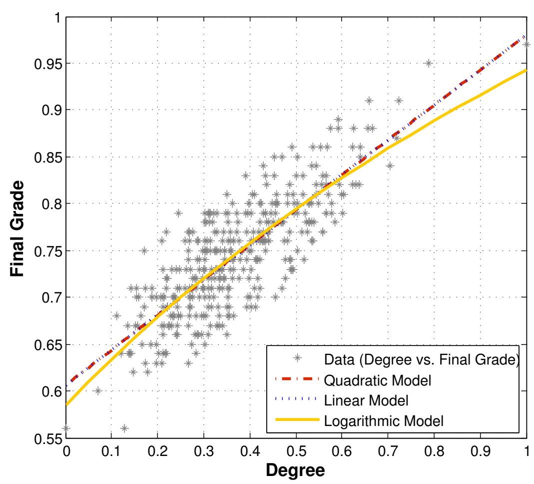

Figure 10: Regression analysis on the relation between the students' *degrees* and final grades.

6.2 Student-to-Teacher Interaction-ship

Figure 11 depicts the relation between the students' final grades and their average distances to the teachers. Compared to the relations observed in the student-to-student social network, the data are more sparsely distributed, which makes none of the linear, quadratic, and logarithmic regression model it well. At the first glance, this seems contradictory to the conventional belief among educational researchers that successful students sit at the central-front of the classroom. However, after investigating 10 students (randomly picked) who achieve high grades but do not always sit at the central-front, we find 80% of them prefer to communicate with the teachers by asking questions after the classes. According to the study by Waller [1], those student can represent the self-starters who do not depend too much on teachers for learning, so that they may not always choose the central-front to sit. In this case, we perform a RANSAC algorithm [35] to search the best model which explains the majority of the data. The resulting model named "Majority Model" in Figure 11 indicates that in a general sense, the closer the students sit to the teachers, the better

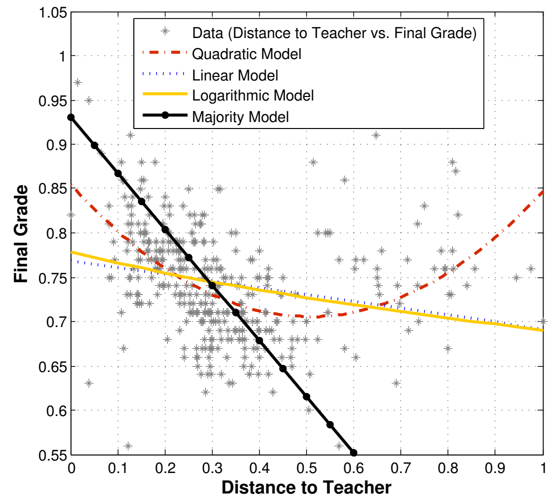

Figure 11: Regression analysis on the relation between the students' *distance-to-teacher* and their final grades.

achievement they obtain in school. In fact, Figure 11 provides a statistical explanation to the dispute about whether the seat preference determines the academic success, in the way that using qualitative approaches on only several samples, it is easy for conventional studies to reach the conclusions that there is no correlation between the seat preference and the academic performance (like in [28] and our first attempt) when one models the whole population with linear model, and that the two are closely related (like in conventional belief and those reported in [22,25–28]) when samples are mainly picked from the main body of the population. In other words, in conventional studies, it is thus easy to be confused when the statistical evidences on a large-scale population are not available.

6.3 Test of Generalizability

We have conducted two leave-one-out cross validation experiments to validate the generalizability of the results in Sections 6.1 and 6.2. Given one class for testing and the rest of classes for training, we check a) if the distribution learnt from the training is statistically form the same distribution of the class for testing using two-sample Hotelling T-Square test; b) if the regression model learnt from the training can explain that of the class for testing. The answers for a) are all YES at significance level 0.01, with the answers for b) are all YES as well indicated by the goodness-of-fit measured by R^2 all above 86.34%.

7. CONCLUSION

We have conducted a pilot study for equipping conventional pedagogical analysis with multimedia technology. With the image-based approach proposed in this paper, we are able to collect the large-scale in-class social data in an automatic manner, and conduct the pedagogical analysis in a quantitative way. The experimental results have validated both the technical and pedagogical effectiveness of the approach. In addition, we have demonstrated that, with the support of the large-scale statistics, it is possible to find why some educational researchers reach conflicting conclusions on the relation between the students' positions in social networks and their academic performances. Furthermore, compared with the student-to-teacher interaction-ship, the student-to-student interaction-ship seems a more reliable predictor for the students' academic performances. However, since the focus of this paper is mainly on the practicability of utilizing multimedia technology for educational studies, the results of the pedagogical analysis reported in between are thus not conclusive. It needs more interdisciplinary collaborations to explain. Moreover, we believe that, with more sophisticated data mining techniques, there are more valuable information which can be explored from the in-class social networks constructed in this paper.

8. ACKNOWLEDGMENTS

The work described in this paper was supported by a grant from the National Natural Science Foundation of China (No. 61001148) and the Open Projects Program of National Laboratory of Pattern Recognition, China.

9. REFERENCES

[1] Willard W. Waller. *The sociology of teaching*. New York: Willey, 1932.

[2] Kaveri Subrahmanyam and Patricia Greenfield. Online communication and adolescent relationships. *The Future of Children*, 18(1):119–140, 2008.

[3] David W. Johnson. Student-student interaction: The neglected variable in education. *Educational Researcher*, 10(1):5–10, 1981.

[4] James C. McCroskey and Rod W. McVetta. Classroom seating arrangements: Instructional communication theory versus student preferences. In *Annual Meeting of the International Communication Association*.

[5] Shane Dawson. A study of the relationship between student social networks and sense of community. *Educational Technology & Society*, 11(3):224–238, 2008.

[6] John Dewey. *Experience and education*. Macmillan, 1938/1997.

[7] Christine Greenhow and Beth Robelia. Old communication, new literacies: Social network sites as social learning resources. *Journal of Computer-Mediated Communication*, 14(4):1130–1161, 2009.

[8] D. M. Boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):119–140, 2007.

[9] Grace Salas and Julie S Alexander. Technology for institutional enrollment, communication, and student success. *Directions for Student Services*, 2008(124):103–116, 2008.

[10] Brigid Barron. Interest and self-sustained learning as catalysts of development: A learning ecology perspective. *Human Development*, 49(4):193–224, 2006.

[11] A. Martínez, Y. Dimitriadis, B. Rubia, E. Gómez, and P. de la Fuente. Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, 41(4):353–368, 2003.

[12] Blaženka Divjak and Petra Peharda. Relation between student academic performance and their position in a social network. In *Central European conference on information and intelligent systems*, pages 65–71, 2010.

[13] Richard Szeliski. Image alignment and stitching: A tutorial. Technical report, Microsoft Research, 2004.

[14] M. D. Kelly. Visual identification of people by computer. Technical report, AI-130, Stanford AI Project, 1970.

[15] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

[16] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725–1745, Sep. 2006.

[17] Cha Zhang and Zhengyou Zhang. A survey of recent advances in face detection. Technical report, MSR-TR-2010-66, Microsoft Research, 2010.

[18] Xiaozheng Zhang and Yongsheng Gao. Face recognition across pose: A review. *Pattern Recognition*, 42(11):2876–2896, 2009.

[19] N. B. Ellison and *et al.* The benefits of facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.

[20] Robert Kraut and *et al.* Internet paradox revisited. *Journal of Social Issues*, 58(1):49–74, 2002.

[21] J. B. Walther and M. R. Parks. Cues filtered out, cues filtered in: Computer mediated communication and relationships. *The Handbook of Interpersonal Communication*, pages 529–563, 2002.

[22] İkram Çinar. Classroom geography: Who sit where in the traditional classrooms? *Journal of International Social Research*, 3(10):200–212, 2010.

[23] Christopher M Johnson. A survey of current research on online communities of practice. *Internet and Higher Education*, 4(1):45–60, 2001.

[24] S. Wasserman and K. Faust. *Social Network Analysis: Methods and applications*. New York: Cambridge University Press, 1994.

[25] Ed Baines, Peter Blatchford, and Peter Kutnick. Changes in grouping practices over primary and secondary school. *International Journal of Educational Research*, 39(1-2):9–34, 2003.

[26] Mary Ellen Benedict and John Hoag. Seating location in large lectures: Are seating preferances or location related to course performance. *Journal of Economic Education*, 35(3):215–231, 2004.

[27] Katherine K. Perkins and Carl E. Wieman. The surprising impact of seat location on student performance. *The Physics Teacher*, 43(1):30–33, 2005.

[28] Steven Kalinowski and Mark L. Taper. The effect of seat location on exam grades and student perceptions in an introductory biology class. *Journal of College Science Teaching*, 36(4):54–57, 2007.

[29] M. Brown and D. G. Lowe. Recognising panoramas. In *The Ninth IEEE International Conference on Computer Vision*, pages 1218–1225, 2003.

[30] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 511–518, 2001.

[31] M. Turk and Alex Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[32] Peter N. Belhumeur and *et al.* Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[33] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–22, 2009.

[34] Peter Peer, Jure Kovac, and Franc Solina. Human skin colour clustering for face detection. In *EUROCON 2003 - International Conference on Computer as a Tool*, pages 144–148, 2003.

[35] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.