Estimating Heights from Photo Collections:
A Data-Driven Approach

Ratan Dey, Madhurya Nangia, Keith W. Ross, Yong Liu

Dept. of Computer Science and Engineering, New York University

ABSTRACT

A photo can potentially reveal a tremendous amount of information about an individual, including the individual's height, weight, gender, ethnicity, hair color, skin condition, interests, and wealth. A *photo collection* – a set of inter-related photos including photos of many people appearing in two or more photos – could potentially reveal a more vivid picture of the individuals in the collection.

In this paper we consider the problem of estimating the heights of all the users in a photo collection, such as a collection of photos from a social network. The main ideas in our methodology are (i) for each individual photo, estimate the height differences among the people standing in the photo, (ii) from the photo collection, create a people graph, and combine this graph with the height difference estimates from the individual photos to generate height difference estimates among all the people in the collection, (iii) then use these height difference estimates, as well as an *a priori* distribution, to estimate the heights of all the people in the photo collection. Because many people will appear in multiple photos across the collection, height-difference estimates can be chained together, potentially reducing the errors in the estimates. To this end, we formulate a Maximum Likelihood Estimation (MLE) problem, which we show can be easily solved as a quadratic programming problem. Intuitively, this data-driven approach will improve as the number of photos and people in the collection increases. We apply the technique to estimating the heights of over 400 movie stars in the IMDb database and of about 30 graduate students.

1. INTRODUCTION

Many online social networks (OSNs) – such as Facebook, Instagram, WeChat, and Flickr – have access to vast collections of photos. For example, as of September 2013, more than 250 billion photos have been uploaded to Facebook, with 350 million photos being uploaded every day [8]. As of March 2004, more than 20 billion photos have been uploaded to Instagram with 60 million photos being uploaded every day [5]. A large fraction of these photos contain pictures of the users of these OSNs. A single photo can be mined to potentially reveal a tremendous amount of information about an individual, including the individual's height,

weight, gender, ethnicity, hair color, skin condition, interests, and wealth. A *photo collection* – a set of inter-related photos including photos of many people appearing in two or more photos – could potentially reveal a more vivid picture of the individuals in the collection.

In this paper we will consider the problem of estimating one such attribute from a photo collection, namely, the heights of *all* the people appearing in the photo collection. There are many applications of inferring heights from photos in social networks:

- **Targeted Advertising:** Users of social networks often do not explicitly provide their heights in their profiles. Yet many advertisers would like to have access to height information. For example, many clothing retailers have specific product lines and marketing strategies for tall/large people and for small people, and some clothing retailers (such as Destinaion XL [3]) are entirely devoted to these markets. Similarly, furniture retailers often create product lines specifically for tall/large/small people (such as King Size Direct [4] and Brigger Furniture [1]). Along with many other possible sectors, airlines would like to identify tall people so that they can attempt to sell them economy plus and business class seats.

- **Match Making:** Many individuals seek dating partners in specific height ranges. Height information extracted from photos can be used to help with match-making.

- **Forensics for Law Enforcement:** Law enforcement officials have traditionally used the height attribute in profile databases to help narrow down suspects in crimes. But for many individuals in these databases, height information is unavailable. By working directly with the OSN, or by scraping the photos themselves from the OSNs, law enforcement organizations can create world-wide profiles including valuable height information.

- **Height Validation:** Models, actors and athletes often make their heights known when marketing their services. But automatically estimating their height from photos, we can validate their advertised heights.

On the other hand, many people would consider the automatic extraction of users' height (and potentially other sensitive information) from photo collections as as a privacy violation. This paper also serves the role of putting the issue of extracting sensitive information from photo collections on the radar screen of privacy advocates.

Several research groups have previously attempted to estimate the height of a single person from a *single photo* with limited success. Some approaches require that the image include reference length in the background scene to indicate scale [20, 15, 14]. This approach has two main difficulties in practice: (i) a reference length may not always be available, and (ii) the results are strongly affected by camera perspective. Other approaches make use of human anthropometry, that is, estimating from the photo anthropometric ratios (such has neck height and head-to-chin distance, which are measured from the image) and anthropometric statistics [12]. This approach requires that the subject to be standing upright, be facing directly the camera, and that his/her upper body be present in the photo.

In this paper we minimally rely on image processing and computer vision techniques, and instead take a data-driven approach to the problem of height estimation. The approach not only has the potential of improving the estimates and relaxing the requirements of the traditional approaches, but also can provide estimates of thousands (potentially millions) of individuals. Specifically, in this paper we develop a novel approach for estimating the heights of all the users in a photo collection, such as a collection of photos from a social network. The main ideas in our methodology are (i) for each individual photo, estimate the height differences among the people standing in the photo, (ii) from the photo collection, create a people graph, and combine this graph with the height difference estimates from the individual photos to generate height difference estimates among all the people in the collection, (iii) then use these height difference estimates, as well as an *a priori* distribution, to estimate the heights of all the people in the photo collection. Due to camera angle and the possibility of some people slouching, for a single photo, some estimates may be overestimates and others may be underestimates. By chaining people across a large collection of photos, and exploiting the fact many people will appear in multiple photos, we will show that height estimation errors can be significantly reduced.

We formulate the estimation problem as a maximum likelihood estimation problem. Specifically, we find the height values of *all the subjects* in the photo collection that maximize the probability of the height-differences observations. The resulting problem becomes a quadratic programming problem, with constraints generated from an *a priori* mean and variance. We apply the technique to two data sets, for both of which we have ground-truth height information. The first data set consists of over 400 movie stars from 1,300 photos. The second data set consists of 30 graduate students from 25 photos. For both data sets, our technique was able

to significantly reduce the baseline errors, in both cases to average error of less than 1.5 inches.

In addition to providing a novel methodology to estimating heights from photos, this paper also shows the potential of mining data from social networks in which the same people appear in multiple, but different, photos. In addition to features like height and weight, it may be possible to use photo collections to mine a story around the community people in the photos, such as examining how off-line friends revolve over time.

This paper is organized as follows. In Section 2, we describe the methodology used in this paper, including the methodology to obtain height-difference estimates from individual photos and the maximum likelihood estimation procedure to obtain the height estimates from the photo collection. In Section 3 we describe the datasets, including the celebrity dataset and the student dataset. In Section 4 we present and discuss the results for the two datasets. In Section 5, we analyze the errors in greater detail. In Section 6 we describe related work and in Section 7 we conclude.

2. METHODOLOGY

2.1 Overview of the "Data-Driven Approach"

In this paper we examine estimating heights of people from collections of group photos. A group photo is a photo in which more than one person appears. Group photos can be collected from many different sources on the Internet, including social networks such as Facebook and Instagram, photo sharing sites such as Flickr, and stills from video sharing sites such as YouTube. The basic idea is to automatically estimate people's heights using computer-estimated height differences of the people in the photo collection, combined with a priori population height distributions.

As an example, consider the four photos in Figure 1, which were collected from the Internet Movie Database (IMDb). This collection of photos contains six distinct people. In the case of IMDb, the people in each of the photos are labeled. (We will discuss how people in photo collections can be labeled in Section 2.3.) We now consider the problem of estimating the heights of all six people in this collection of photos.

To this end, we first consider estimating the heights in each of the photos *separately*, for example, the heights of Person 0 (P0) and Person 1 (P1) in photo 1(b). We consider the following two-step procedure:

1. We first assign each person in the photo the same height, namely, the global population mean across the whole population. For example, if the global mean of height of movie stars is 171 cm, then we assign the height of 171 cm to both P0 and P1.

2. We then adjust these heights by estimating the height difference between P0 and P1 from the photo. We will discuss a methodology for doing this in Section 2.2.

For example, suppose we estimate that P1 is 4 centimeters taller than P0. Then we can adjust our original estimates for P0 and P1 to be 171 - 4/2 = 169 cm and 171 + 4/2 = 173 cm.

Intuitively, the adjustment Step 2 can significantly improve the baseline mean estimate in Step 1. But the resulting estimates can be still far from the correct heights for two classes of reasons:

1. Because the photo population is very small (two in this example), there is a high probability that the photo-population mean differs greatly from the global population mean. If the two means differ significantly, the estimates will have large errors.

2. The estimates for the height differences may also be inaccurate. These inaccuracies can be due to people slouching, the ground level where the various people stand, the camera angle, and the heel sizes.

In order to respond to these two classes of error sources, we explore in this paper taking a "data-driven approach", namely, estimating people's heights from a large collection of inter-related photos. To this end, Figure 2 shows two network representations of the photo collection from Figure 1. There are a total of six people in the collection. The first network representation shows for the six persons, which photos they appear in. The second network representation has a node for each of the six persons, and a link between two nodes if the two corresponding persons appear in the same photo. Note that, for this example, this second network is fully connected, meaning that *it is possible to directly or indirectly compare the heights of all six persons throughout the collection*. Although in this example we are considering a small collection consisting of four photos, in practice we may have thousands of photos in a collection.

For the collection of photos, we now have a revised two-step procedure:

1. We first assign to each person in the collection the same height, namely, the global population mean calculated across the whole population. Because the number of people in the collection is much larger than the number of people in an individual photo, it is likely the collection mean will be relatively close to the global mean. In fact, as the number of individuals in the collection increases, the collection mean should approach the population mean due to the law of large numbers.

2. We then again adjust the estimates in Step 1 by examining the height differences in the various photos. But now, for any pair of persons, there can be multiple height difference estimates. For example, both P0 and P1 appear in Photo 1(a) and in Photo 1(b). So, for these two photos, we can obtain two height difference estimates for P0 and P1. As explained later in this section, we can obtain additional estimates by looking at
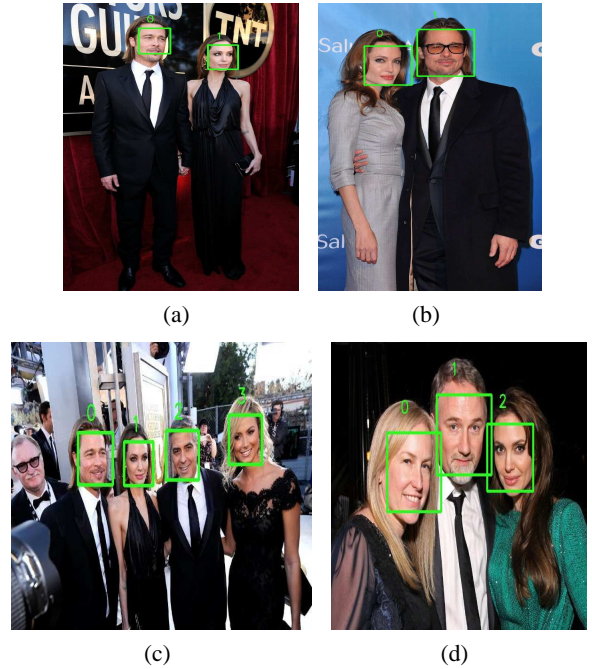

Figure 1: An example of a small collection of 4 photos with rectangles around faces

chains of people across photos. By aggregating the various height-difference estimates, errors due to slouching and camera angles will potentially be reduced due to averaging.

In the following subsections, we will formulate and solve the height estimation problem as a data-driven problem.

2.2 Height Difference Estimation from Individual Photos

A central part of the methodology is *intra-photo height-difference estimation*, that is, finding height difference estimates among pairs of people in the same photo. For each photo with multiple persons, we use the OpenCV tool to detect the faces of all persons in the photo. For each detected face, this tool provides the x and y coordinates of the top left corner of each rectangle as well as the height and width of every detected face. We assign indexes to each detected face based on their positions in the photo. Figures 1 and 3 show the boxes along with their assigned indexes on the detected faces. We then use the boxes to calculate the height differences among all persons in the photo. Specifically:

- For each pair of persons, we calculate the vertical distance between the mid-points of the two corresponding face boxes. This results in a height difference value in units of pixels. Pixel values are very much dependent on the resolution of the photo and are not very useful in comparing heights across different photos.

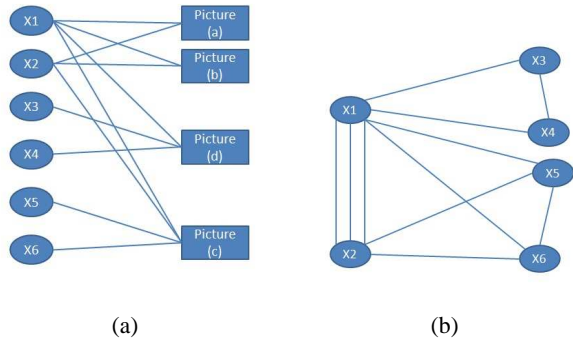- We then convert the height differences (in pixels) to average box size, which is the average height (in

Figure 2: Two network depictions of photo collection in Figure 1. The first network representation shows for each of the six persons, which photos he or she appears in. The second network representation has a node for each of the six persons, and a link between two nodes if the two corresponding persons appear in the same photo.

Figure 3: Example of a group picture of students with boxes around their faces

pixels) of all the detected head boxes in that photo. This results in height differences measured in units of "heads". For example, if person A is 35 pixels taller than person B and the average height of the boxes in the photo is 50 pixels, then person A is said to be $35/50 = 0.7$ heads taller than person B.

- In the end, we need to convert our height differences in "heads" to height differences in centimeters. One simple way of doing this is to assume a value in centimeters for the head height, and multiply this value by the number of heads taller. For example, suppose we assume that the average box head height corresponds to 20 cm. Then if person A is 0.7 heads taller than person B, we can estimate that person A is $0.7 \cdot 20 = 14$ cm taller than person B. (We will describe a more systematic way of determining head size in centimeters. But for now let us assume that we are working with a given reasonable value.)

We mention briefly that in many real scenarios, the images collected from a social network are taken with an indirect camera angle, for example, from the left and below of the faces. Such indirect angles can introduce errors into the height difference estimates in the the individual photos. It should be possible to reduce these errors by preprocessing the photos, for example, by using calibration [2]. In this paper, however, we do not preprocess, as we are interested in exploring how well a pure data-driven technique can perform: If each person appears in many photos, then errors from different camera angles could cancel out (as in the law of large numbers). We leave preprocessing of the photos, and its potential to reducing errors, to future research.

2.3 Height Inference from a Photo Collection

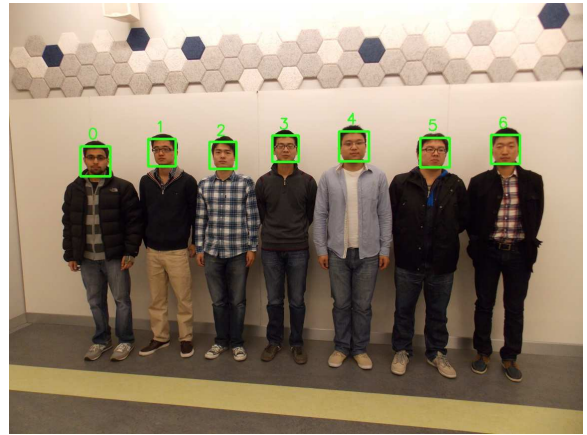Given a collection of photos, our goal is to infer the heights of the people in the collection based on all the height difference estimates within the individual photos. We refer to this as *inter-photo height estimation*. The first step is to extract group photos from the collection, that is, extract photos in which there are multiple people roughly standing next to each other. In this paper, we manually extracted such group photos from our collection; but it should be possible to find such photos automatically [17].

The second step is to label the people in the photos, so that we can link the same person in different photos to build graphs as in Figure 2. In many photo collections, the people appearing in the photos are explicitly labeled; for example, in the IMDb collection, each person in each photo is labeled. In other photo collections, not all the people in the photos are labeled, by a significant fraction may be. For example, in Facebook many photos are explicitly labeled via tagging. (Facebook encourages this by using face recognition to suggest tags.) Other photos, such as profile photos, are implicitly labeled. Even if the collection has no labels or tags, it should be possible to link many of the people across photos using face recognition technology. Stone et al. showed how social network contexts can enhance the performance of face recognition and hence can be used for automatically tagging millions of photos [22]. Linking people across photos using face recognition is an orthogonal problem and is not considered in this paper.

After making the associations between the same people in different photos, we can convert the height differences between positions into the height differences between actual people, thereby creating a graph as in Figure 2(b) with each link labeled with a height difference estimate in centimeters. Specifically, we create one node for each person who appears at least once in the photo collection. For each pair of persons (i, j) in picture p (without loss of generality suppose person i has lower ID than person j), we create a directed link from node i to node j, with the link weight d_{ij}^p being the estimated height difference of person i over person j in photo p. Note that if the pair (i, j) appear together in multiple group photos, they will be connected by multiple links

4

with different weights.

We can now describe our procedure for estimating the heights of people in the collection of photos. Let $\mathcal{P}$ be the set of photos, $\mathcal{I}_p$ be the subset of individual persons who appear in photo $p \in \mathcal{P}$, and $\mathcal{I} = \cup_{p \in \mathcal{P}} \mathcal{I}_p$ be the set of persons who appear in at least one of the photos. In each photo p, we obtain d_{ij}^p, $i, j \in \mathcal{I}_p$, the estimated height difference of person i over person j, as described in Section 2.2. (For now, we will assume that these values were calculated with a given and fixed box head size in centimeters.) Let x_i, $i \in \mathcal{I}$, be our estimated height of person i. Our task is to obtain good estimates of $\{x_i, i \in \mathcal{I}\}$ based on $\{d_{ij}^p, i, j \in \mathcal{I}_p, p \in \mathcal{P}\}$.

2.4 Maximum Likelihood Estimate

The most natural error model is to assume the estimated height difference between a pair of persons in the same photo is a noisy measurement of their real height difference, with the error following a zero-mean Normal distribution. Let D_{ij}^p be the random variable of the measured height difference of i over j in photo p. Then

$$D_{ij}^p = x_i - x_j + E_{ij}^p, \tag{1}$$

where E_{ij}^p is the random error, and $E_{ij}^p \sim \mathcal{N}(0, \sigma_{ijp}^2)$. Immediately, the density function of the height difference of i over j measured from photo p can be calculated as:

$$f(D_{ij}^p = d_{ij}^p | x_i, x_j) = C_{ij} \exp\left\{ -\frac{(d_{ij}^p - x_i + x_j)^2}{2\sigma_{ijp}^2} \right\},$$

where $C_{ij} = 1/(2\pi\sigma_{ijp}^2)^{1/2}$. For any pair of persons (i, j) in photo p, we can obtain their height difference d_{ij}^p using the procedure described in Section 2.2. Obviously the height differences are *dependent*, e.g., $D_{ij}^p = -D_{ji}^p$ and $D_{ij}^p + D_{jk}^p = D_{ik}^p$. We can select independent height differences by choosing a common reference point. Specifically, let p_0 be the person at the leftmost position in photo p. Any other pairwise height difference d_{ij}^p can be calculated as $d_{ip_0}^p - d_{jp_0}^p$. If we further assume that the error terms between different pairs are i.i.d., i.e., $E_{ip_0}^p$ is independent of $E_{jp_0}^p$, $\forall i \neq j$, and $\sigma_{ip_0 p}^2 = \sigma_p^2$, $\forall i \in \mathcal{I}_p$, then the joint density function of the measured height differences from photo p is:

$$f\left(\{D_{ij}^p = d_{ij}^p, \forall i, j \in \mathcal{I}_p, i \neq j\} \mid \{x_i, i \in \mathcal{I}_p\}\right) \tag{2}$$

$$= f\left(\{D_{ip_0}^p = d_{ip_0}^p, \forall i \in \mathcal{I}_p, i \neq p_0\} \mid \{x_i, i \in \mathcal{I}_p\}\right) \tag{3}$$

$$= \prod_{\forall i \in I_p, i \neq p_0} \frac{1}{(2\pi\sigma_p^2)^{1/2}} \exp\left\{ -\frac{(d_{ip_0}^p - x_i + x_{p_0})^2}{2\sigma_p^2} \right\}. \tag{4}$$

Finally, for the height differences observed from all photos $\mathcal{D} \triangleq \{\{D_{ij}^p = d_{ij}^p, i, j \in \mathcal{I}_p, i \neq j\}, p \in \mathcal{P}\}$, the joint distribution can be calculated as:

$$f(\mathcal{D} | \{x_i, i \in \mathcal{I}\})$$

$$= \prod_{p \in \mathcal{P}} f\left(\{D_{ip_0}^p = d_{ip_0}^p, i \in I_p, i \neq p_0\} | \{x_i, i \in \mathcal{I}_p\}\right)$$

$$= \prod_{p \in \mathcal{P}} \prod_{i \in I_p, i \neq p_0} \frac{1}{(2\pi\sigma_p^2)^{1/2}} \exp\left\{ -\frac{(d_{ip_0}^p - x_i + x_{p_0})^2}{2\sigma_p^2} \right\}$$

Then the Maximum Likelihood Estimate $\mathbf{x}_{ML}$ of $\mathbf{x} = \{x_i, i \in \mathcal{I}\}$ can be obtained as

$$\underset{\mathbf{x}}{\mathbf{argmax}} \quad f(\mathcal{D} | \mathbf{x})$$

$$= \underset{\mathbf{x}}{\mathbf{argmax}} \quad \prod_{p \in \mathcal{P}} \prod_{i \in I_p, i \neq p_0} \exp\left\{ -\frac{(d_{ip_0}^p - x_i + x_{p_0})^2}{2\sigma_p^2} \right\}$$

$$= \underset{\mathbf{x}}{\mathbf{argmin}} \quad \sum_{p \in \mathcal{P}} \sum_{i \in I_p, i \neq p_0} \frac{(d_{ip_0}^p - x_i + x_{p_0})^2}{2\sigma_p^2}$$

If we further assume that $\sigma_p^2 = \sigma^2$, $\forall p$. Then we have

$$\mathbf{x}_{ML} = \underset{\mathbf{x}}{\mathbf{argmin}} \quad \sum_{p \in \mathcal{P}} \sum_{i \in I_p, i \neq p_0} (d_{ip_0}^p - x_i + x_{p_0})^2 \tag{5}$$

To obtain the MLE, we need to numerically solve the optimization problem (5). Since variables $\{x_i\}$ always appear in the objective in the difference form of $x_i - x_j$, $\vec{x}$ and $\vec{x} + c$ have the same objective value. We need an additional constraint to obtain the appropriate estimates. We use the sample mean to introduce additional constraints for the MLE optimization. For the sample mean, we simply require the average of the height estimates be equal to the average height of the global population.

$$\text{Global Mean Constraint:} \quad \sum_{i \in \mathcal{I}} x_i = |\mathcal{I}|\bar{m}, \tag{6}$$

where $\bar{m}$ is the average height of the global population. In this case, the MLE is obtained by solving a quadratic programming problem, namely, minimize (5) subject to the global mean constraint (6). It can be easily and rapidly solved for tens of thousands of variables.

If the genders of the people in the photos can be automatically identified, then an alternative is to introduce gender-based mean constraints. Let $\mathcal{I}^m$ be the set of males, and $\mathcal{I}^f$ be the set of females. In this case, we would have two gender-based constraints in the quadratic programming problem:

$$\sum_{i \in \mathcal{I}^m} x_i = |\mathcal{I}^m|\bar{m}_m, \quad \sum_{i \in \mathcal{I}^f} x_i = |\mathcal{I}^f|\bar{m}_f, \tag{7}$$

where $\bar{m}_m$ and $\bar{m}_f$ are the average heights for male and female population, respectively.

2.5 Head Size

Recall from Section 2.2 we measure height differences from the photos in units of heads. We then convert this height difference into units of centimeters by multiplying the

height difference in heads by an estimate of the head size in centimeters. Up until this point, to make this conversion, we have simply been using an assumed and fixed value for the box head size. In this section we present an adaptive approach for determining the head size. Let S be a variable denoting the head size in cm. Let b_{ij}^p in units of heads be the height difference between person i and j appearing in photo p. We now calculate the height difference (in centimeters) by multiplying the height difference b_{ij}^p by the head size variable S. Because we have added a new variable S, we also add a new constraint to automatically determine what the appropriate head size S should be. To this end, we now additionally require that the sample variance of the height estimates be equal to the a priori global population height variance. This technique, of course, requires that we have a good estimate of the global population variance. Using (5), the new optimization problem becomes minimizing:

$$\sum_{p \in \mathcal{P}} \sum_{\forall i \in I_p, i \neq p_0} (S \cdot b_{ip_0}^p - x_i + x_{p_0})^2 \qquad (8)$$

subject to

$$\text{Global Mean Constraint:} \quad \sum_{i \in \mathcal{I}} x_i = |\mathcal{I}|\bar{m},$$

and

$$\text{Global Variance Constraint:} \quad \sum_{i \in \mathcal{I}} (x_i - \bar{x})^2 = (|\mathcal{I}| - 1)\sigma^2,$$

where σ^2 is the height variance of the whole population. Similar to the gender-based mean constraints, we can also introduce gender-based variance constraints.

2.6 Evaluation and Baseline Methods

We used the *fmincon* optimization tool from Matlab to solve the constrained optimization problems and get MLE height estimates $\{x_i, i \in \mathcal{I}\}$ for all persons. To evaluate the performance of our technique, we compare the estimates with the ground-truth actual heights $\{h_i, i \in \mathcal{I}\}$. We calculate the Root Mean Squared Error (RMSE) as:

$$RMSE = \sqrt{\frac{\sum_{i \in \mathcal{I}} (x_i - h_i)^2}{|\mathcal{I}|}}$$

We will compare the RMSE error with the error of two baseline approaches. In **Baseline 0**, we assign everyone the same mean height, equal to the height of the global population. In **Baseline 0 with gender**, we also consider gender where we assign all males the average male population height and females the average female population height.

For Baseline 1, we estimate heights considering each of the photos separately using the height-difference techniques described in Section 2.2. *In Baseline 1, we take intra-photo height differences into account, but we do not the exploit inter-photo information from the collection.* Specifically, for the **Baseline 1 approach without gender**, for any given photo we first assign every person in the photo the same

average height for global population. We then adjust these height estimates using the height-difference estimates *only within that single photo*. This estimate will typically have an error, as it will not be exactly equal to the ground-truth height. The overall average error is given by the RMSE values of these individual errors. For the **Baseline 1 approach with gender**, for a given photo we first identify the males and females and assign them each their respective global average heights. We then adjust the heights of the males (respectively, females) using the intra-photo height differences of the males (respectively, of the females). We then again average the errors as just described.

3. DATA SET

Table 1: Basic Properties of Celebrity Dataset

Name	Number	Mean Heights (cm)	Standard Deviation (cm)
Sample Population	426	174.5	7.0
Sample Population: Male	230	180.9	7.5
Sample Population: Female	196	167.0	6.5
Global Population	1029	174.1	7.3
Global Population: Male	525	180.6	8.1
Global Population: Female	504	167.3	6.5

We tested the performance of our height estimation procedures using two datasets – the **celebrity dataset** and the **student dataset**, for both of which we have ground-truth height information. We also applied the baseline approaches to these two datasets.

3.1 The Celebrity Dataset

The Internet Movie Database (IMDb) is an online database which stores movie, TV, and celebrity content. It has information on more than 2.5 million titles and 5.2 million people, which can be accessed from the IMDB website. For this project, we used IMDb to obtain publicly available information on celebrities. Each celebrity has a profile page with his or her biography, awards and nominations, credits, occupation, pictures, personal details, height, and so on. IMDb also has photo galleries for movies, TV shows, events, and people, all of which is publicly accessible on the website. For each photo, the names of the people in the photo are displayed along with links to their profiles.

We wrote a web crawler in Python to collect information on celebrities and collect their group pictures from IMDb.com. IMDb ranks the celebrities. We downloaded profile pages, including heights, as well as the "event photos" they appear in, for 450 popular celebrities, giving rise to 1,300 photos which contain multiple (two or more) people standing next to each other. We then constructed a graph, with one node for each of the 450 celebrities, as in Figure 2(b). The resulting graph contained one large component with 426 nodes and several smaller components with a small number of nodes. Figure 4 shows the graph for the
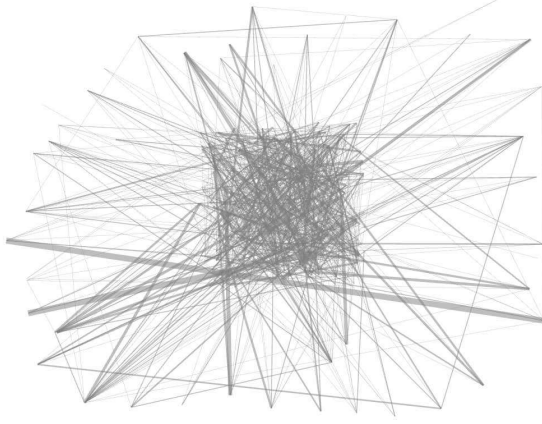
Figure 4: The largest connected component in the celebrity graph

largest component. Because we are interested in estimating the heights in large interconnected social groups, we estimate the height of these 426 celebrities in the large component only. We refer these 426 (230 are male and 196 are female) celebrities as the **sample population**. Although not easily seen in Figure 4, many of the node pairs have multiple links between them. For example, Angelina Jolie and Brad Pitt together appear in 15 photos.

To obtain global population statistics, we also collected heights and gender of other celebrities. All combined we collected gender and height information for 1,029 celebrities, which includes the 426 celebrities in our sample population. We refer to these 1,029 celebrities as our **global population**. The statistics of the celebrity dataset are summarized in Table 1.

3.2 Student Dataset

Table 2: Basic properties of Student Dataset

Name	Number	Mean Heights (cm)	Standard Deviation (cm)
Sample Population	30	172.3	5.6
Sample Population: Male	25	174.6	5.1
Sample Population: Female	5	160.8	8.1
Global Population	94	171.1	6.3
Global Population: Male	73	173.9	6.1
Global Population: Female	21	162.5	6.9

In addition to studying a relatively large dataset, as in our celebrity dataset, is of interest to study the performance of our methodology for smaller dataset. From our university graduate students, we asked for volunteers to appear in group photos. We asked each student who volunteered to tell us his/her height. Thirty students volunteered to appear in 25 photos. We refer to these students as the **sample population.** The graph for these 30 students is shown in Figure 5. Note that the graph has two components. To obtain population statistics, we additionally asked for the heights and genders

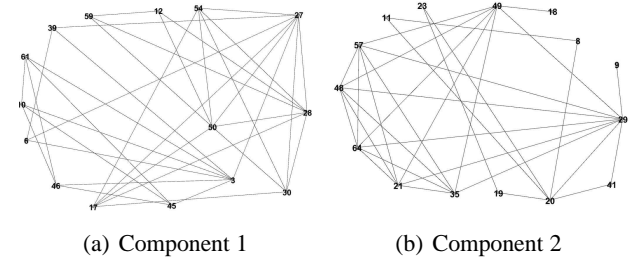(a) Component 1   (b) Component 2

Figure 5: Components of Student Networks

of other students. Totally, we obtained height and gender data for 94 students, including the 30 students in the sample population. We refer to these 94 students as our **global population**. The basic properties of our student dataset are shown in Table 2.

4. RESULTS

We now describe the results for our data-driven approach, which uses the constrained optimization problem (8) resulting from the MLE procedure. In the baseline and data-driven approaches, we initially consider the case where we cannot identify gender in the photos. For this case, we use aggregate statistics in the global mean and variance constraints. We then also consider the case when gender is available, which we call the baseline approach with gender and the data-driven approach with gender. In this latter approach, we use two gender-based mean constraints but one aggregate variance constraint.

With the celebrity dataset, many of the actors (female and male) wear high heels, with the heel sizes varying from actor to actor, and also possibly varying for the same actor across different photos. To deal with this issue, we make the rough assumption that women's heels are six centimeters higher on average than men's heels. This is accounted for by adding six centimeters to each women's ground truth height, then calculating global population means and variance accordingly. With the student dataset, the female students generally wear smaller heel lifts. We make the rough estimation that female students' heel sizes are two centimeters higher than men's heel sizes.

4.1 Results for the Celebrity Dataset

Table 3: Results found for celebrity dataset: Without Gender

Approach Name	RMSE Overall	RMSE Male	RMSE Female
Baseline 0	9.9	10.1	9.6
Baseline 1	7.0	7.0	7.1
Data-Driven	3.8	4.0	3.6

Table 3 and 4 show the root mean square error (RMSE) for the celebrity dataset. We calculate RMSE for all of the celebrities, only for male celebrities, and only for female celebrities. Note that our Baseline 1 approach significantly

Table 4: Results found for celebrity dataset: With Gender

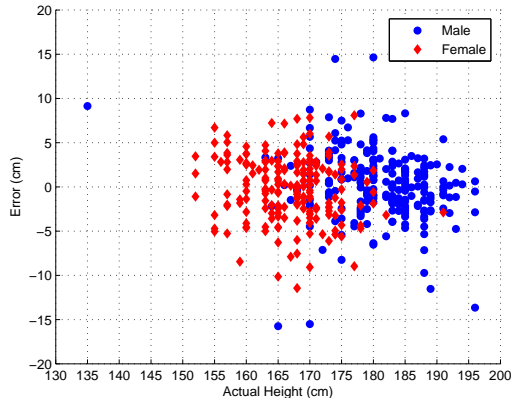| Approach Name	RMSE Overall	RMSE Male	RMSE Female
Baseline 0	7.0	7.6	6.4
Baseline 1	5.7	5.9	5.4
Data-Driven	3.9	4.1	3.6

Figure 6: Actual height versus error using data-driven approach in celebrity dataset

reduces the error of the naive Baseline 0 approach, for when gender is identified and gender is not identified in the photos. Recall that our Baseline 1 approach takes advantage of intra-photo height differences but does not exploit inter-photo information.

Our data-driven approach exploits inter-photo information as well is intra-photo height differences. We see from Table 3 that when gender is not available, our data-driven approach gives a dramatic reduction in the error. The Baseline 1 error is further reduced from 7.0 cm to 3.8. In the case when gender can be identified (Table 4, our data-driven approach reduces the RMSE from 5.7 cm to 3.8 cm or 5.7 cm (baseline 1) to 3.9 cm. In both cases, our data-driven approach gives a substantial improvement over the baseline approaches. (Because there may be ground-truth errors for some of the celebrities as reported on IMDb, the actual estimation errors are likely lower than reported in this section; see Section 5.)

Note that when gender is available, the separate gender constraints do not improve the performance of our data-driven approach. We also note that all the estimation procedures (baseline and data-driven) are more accurate for females than for males. This is likely due to two reasons. First, the standard deviation for female heights is less than it is for male heights in the global population (see Table 1), making male heights more difficult to estimate. Second we believe there is more variation in male celebrity's heel sizes than in female celebrity's heel sizes. Note that for the data-driven approach with and without gender, we get the same optimal head size, namely, $S = 21.5$cm.

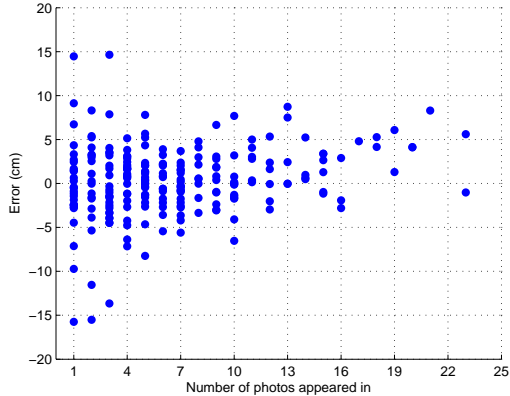Figure 6 shows a scatter diagram for the data-driven ap-

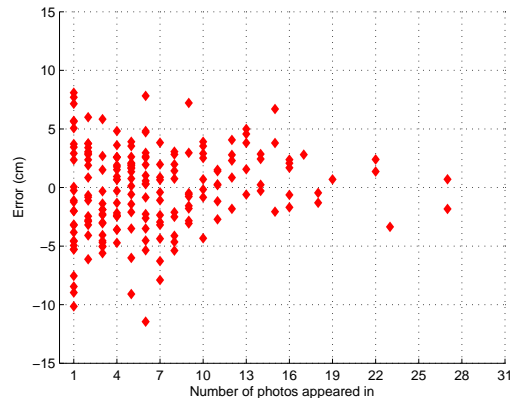Figure 7: Celebrity Male: Number of photos appeared in vs error for data-driven approach

Figure 8: Celebrity Female: Number of photos appeared in vs error for data-driven approach

proach, with actual heights on the x-axis and errors on the y-axis. If the difference between estimated height and actual height is positive, then we are overestimating the actual height; and if it is negative, we are underestimating the actual heights.

Note that for the data-driven approach, the errors are roughly uniformly distributed over the entire height range, whereas for the baseline 0 approaches, intuitively the taller and shorter people have the largest errors. For the data-driven approach, more than 85% of the celebrities are within an error of margin of 5 cm (2 inches), whereas for the baseline 0 approach, only 55% of the celebrities are within an

Table 5: Results found for male celebrities who appeared in small number of photos

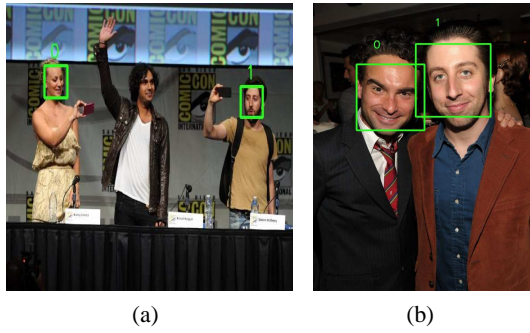| Number of Photos (n)	1	2	3	4	5
Baseline 1: Average error over people with n photos	7.6	6.3	5.1	4.9	4.6
Data-Driven: Average error over people with n photos	3.9	4.2	3.7	2.6	2.4

(a)   (b)

Figure 9: Error analysis for Johnny Galecki and Simon Helberg

Figure 10: Actual height versus error for the student dataset using the data-driven approach

error margin of 5 cm.

For the data-driven approach, most of the large errors (more than 10 cm) are for males. We believe that males may have larger errors because males may have a larger variation in heel size.

The scatter diagrams 7 and 8 investigate the error, for the data-driven approach, as a function of the number of photos a person appears in. In these diagrams, we show overestimation and underestimation of height for each celebrity vs. number of photos he/she appears in. We can clearly see a general trend of accuracy improving as the number of photos increases. Thus this confirms that as the data gets "bigger," the data-driven approach provides better results.

Many of the larger errors are simply due to individuals only appearing in 1-2 photos which have sloping camera angles or people not standing upright. For example Johnny Galecki and Simon Helberg have large errors, 15.8 cm and 15.6 cm, respectively. Galecki appears in one photo only, appearing with Helberg; Helberg appears in only one other photo, appearing with Kaley Cuoco. These two photos are shown in Figure 9. These two celebrities (Galecki and Helberg) form a hanging branch in the graph. The photo 9(a) with Cuoco and Helberg is problematic: due to the camera angle Helberg appears to be much shorter than Cuoco, which isn't the case. In photo 9(b), due to Galecki leaning into Helberg, Galecki also appears to much shorter than Helberg, which also isn't the case. The error for Galecki is therefore particularly large due to the cascading errors. If these two actors appeared in more photos, and if they also form a cycle rather than just a hanging branch, the errors would likely be significantly reduced.

But even for people who appear in a small number of photos, our data-driven approach does better than our Baseline 1 approach, as height comparisons are implicitly made over more people people. Table 5 compares the error averaged over male celebrities who appear in the same number of photos (1 to 5) for Baseline 1 and for the data-driven approaches. We can clearly see that performance of data-driven approach is significantly better. We find similar results in the case of female celebrities.
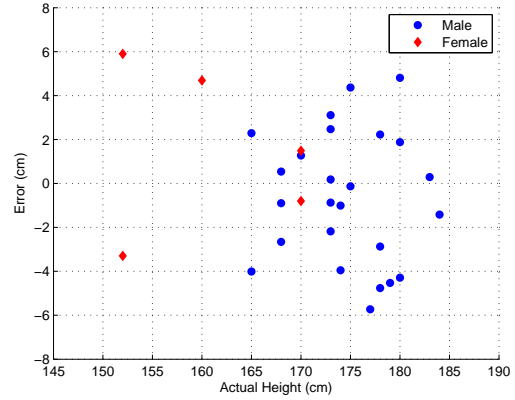
4.2 Results for Student Dataset

In Table 6 and 7, we have shown the results in terms of root mean square error (RMSE) for student dataset. As with the celebrities, our baseline 1 approach performs significantly better than the naive baseline 0 approach, and our data-driven approach performs better than our baseline 1 approach.From Table 7, we can see that when gender is considered, the error is reduced from 5.8 to 3.1, which is significant. Note that a significant portion of this error is due to the sample mean differing from the global mean by 1.2 cm. By using a larger photo collection, this error source would be eliminated, reducing the RMSE error to less than an inch. Also note that the error with the student dataset is less than it is with the celebrity dataest, even though the photo collection for the celebrities is much larger. This is likely because most students did not wear high heels, and the photos were taken in a more controlled environment. We found the same head size, $S = 14.25cm$ for the data-driven approaches (both with gender and without gender). Figure 10 shows the errors as a function of height.

Table 6: Results found for the student dataset: Without Gender

Approach Name	RMSE Overall	RMSE Male	RMSE Female
Baseline 0	7.7	5.7	13.8
Baseline 1	4.7	4.8	4.3
Data-Driven	3.1	3.0	3.8

Table 7: Results found for the student dataset: With Gender

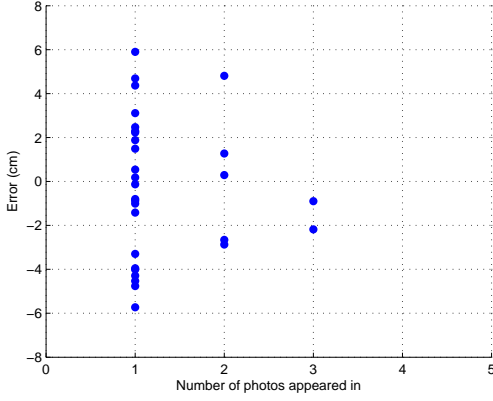| Approach Name	RMSE Overall	RMSE Male	RMSE Female
Baseline 0	5.8	5.2	8.2
Baseline 1	4.0	3.6	5.9
Data-Driven	3.3	3.1	3.9

9

Figure 11: Student: Number of photos appeared in vs error for data-driven approach

Figure 11 investigates the error, for the data-driven approach, as a function of the number of photos a student appears in. In this diagram, we show overestimation and underestimation of height for each student vs. number of photos he/she appears in. We can clearly see a general trend of accuracy improving as the number of photos increases.

4.3 Head size analysis

Recall that in our data-driven approach (8) we are choosing a head size S that matches the *a priori* global variance. We also examined the *optimal head size*, namely, the head size that gives the least RMSE. We did this for both the celebrity and student datasets. To this end, we solved the optimization problem using only the global mean constraint. In particular, we did not use the global variance constraint, but instead solved the problem repeatedly for different values of S ranging from 5 to 30. We then chose the value of the head size S that gives the lowest RMSE. For the celebrity dataset, the lowest RMSE was 3.6 (reduced from 3.8) with an optimal head size value 18 cm (instead of 21.5). For the student dataset, the lowest RMSE was 3.0 at head size value 13 cm. These RMSE values are only slightly smaller than what we obtain in Tables 3 and 6, for which we do not use a specific head size (which is difficult to obtain *a priori*).

4.4 Estimating males and females separately

Table 8: Statistics and results for male-only estimation

Name	Value
# of Images	289
# of Male Celebrities in Big component	160
RMSE for Data-Driven method	5.2
RMSE found for Data-Driven method in Table 4	4.1

To gain further insight into the problem, we considered estimating the heights of the males and females separately in the celebrity dataset. Specifically, we extract from the photo dataset only those photos that have only males in them, and
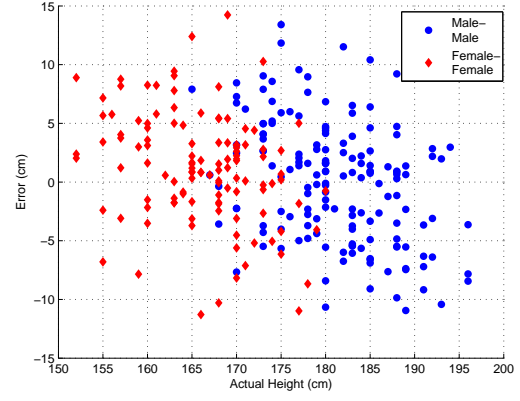

Figure 12: Actual height versus error: Estimating males and females separately

then apply the data-driven approach to that subset of photos. We also do the same for the females. In this way, we can minimize the effects of heel size differences across genders.

For males, we have 289 photos in which only male celebrities appear, with 160 male celebrities in those photos. The statistics and results are summarized in Table 8. For females, we have 213 images where only female celebrities have appeared and in these images 119 female celebrities appear. The statistics and results are summarized in Table 9. We see that these RMSE values are significantly higher than for when the entire photo collection is used (that is, including photos in which males and females appear together). To gain further insight, in Figure 12 we show the scatter plots for actual heights versus errors for the case of estimating males and females separately. We see that, as compared with the case when using the entire photo collection, significantly more people have estimated height errors that are more than 5 cm. These results further argue for the data-driven approach: by using a larger dataset, which includes photos mixing males and females together, we can significantly reduce the error.

Table 9: Statistics and results for female-only estimation

Name	Value
# of Images	213
# of Female Celebrities in Big component	119
RMSE for Data-Driven method	4.9
RMSE found for Data-Driven method in Table 4	3.6

5. GROUND-TRUTH ANALYSIS

Simon Helberg is one of the celebrities for which our estimate has a large error. IMDb says he is 170 cm tall whereas we estimate him to be 154.5 cm. With more than 15 cm estimation error and examining photos of him in our collection and outside of our collection, we became suspicious about his ground truth height as reported by IMDb. We then found that several other sites say he is 163 cm tall. Assuming his
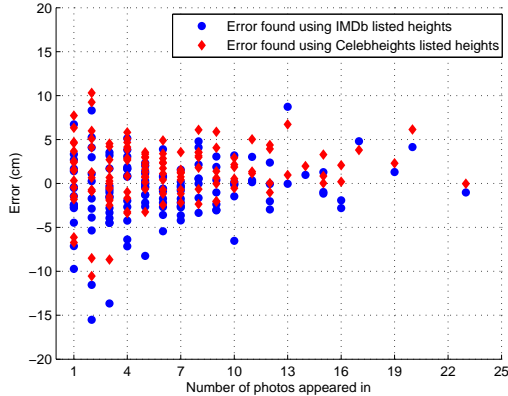
Figure 13: Celebrity Male for whom mismatch found in ground-truth heights: Number of photos appeared in vs error

Figure 14: Celebrity Female for whom mismatch found in ground-truth heights: Number of photos appeared in vs error

real height is indeed 163 cm, the estimation error for Helberg is reduced from 15.5 cm to 8.5 cm.

We then compared the IMDb values with height values from another site http://www.celebheights.com, from which we downloaded the heights of all 426 celebrities. We find mismatches for 244 celebrities, most of which are small but some rather big (such as for Simon Helberg). If we change the ground truth heights for these 244 celebrities, as given in celebheights.com, the RMSE error for males goes down from 4.0 cm to 3.8 cm, but stayed the same for females at 3.6 cm. More importantly, with the new ground-truth values, many of the larger errors significantly decrease. Figures 13 and 14 show the errors, calculated using the two sources of ground truth heights, for these 244 celebrities (i.e., with mismatches) as a function of the number of photos a person appears in. From these figures, we see that with the new ground-truth data, all men (in the mismatch set) have estimation errors less than 11 cm and all women have estimation errors less than 8 cm. By simply eyeballing many of the photos, we believe that the site http://www.celebheights.com is somewhat more accurate than IMDb. Thus the estimation methodology given in this paper can be used to crosscheck the ground-truth information given in various height datasets. The methodology can also be used to dispel myths about peoples' heights, whether they are famous or not.

6. RELATED WORK

6.1 Estimating Height

Several research groups have previously attempted to estimate the height of a single person from a *single photo* with limited success. One approach is to make use of human anthropometry, that is, estimating from the photo anthropometric ratios (such as neck height and head-to-chin distance) and combining these estimates with anthropometric statistics to estimate height [12] [19]. This approach requires that the subject be standing upright, be facing directly
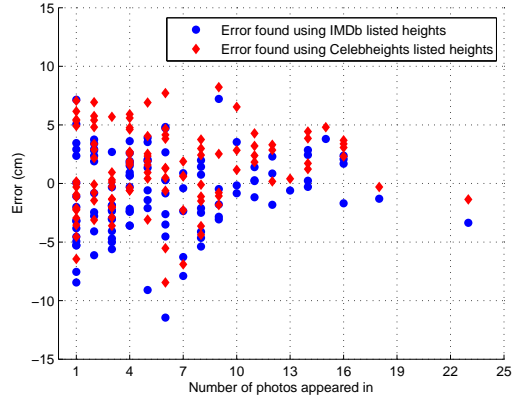
the camera, and his/her upper body be present in the photo. BenAbdelkader and Yacoob [12] estimated height from full body images and from images which contain only the upper part. They evaluated their methodology on synthetic data (generated randomly) and real images. Their real image dataset consisted of 108 high-resolution full body photos of 27 adults. They compared there estimates with the same baseline estimates we use in this paper, namely, estimating each male's height simply as the male population mean and each female's height simply as the female population mean. In contrast, our methodology does not require the photos to be calibrated. For example, our celebrity dataset contains photos directly downloaded from IMDb, with photos taken from a variety of camera angles and subjects who are occasionally slumping or not facing the camera directly. They also manually located different body landmarks (top of head, chin, left corner etc) to measure anthropometric parameters, whereas we automated the measurements using off-the-shelf face detection tools.

Our data-driven approach gives significantly better results. As shown in Table 10 their methodology estimated heights for less than 60% of the cases within a 5 cm error, whereas our methodology estimated heights for 85% of the celebrities and 93% of the students within 5 cm. Similarly, their methodology estimated heights for less than 83% of the cases within a 10 cm error, whereas our methodology estimated heights for 98% of the celebrities and 100% of the students within 10 cm.

Table 10: Comparison of the results for the anthropometry approach with our data-driven approach for the celebrity and student datasets.

Error	Anthropometry (at most)	Celebrity (at most)	Student (at most)
Within 3 cm	40%	63%	60%
Within 5 cm	60%	85%	93%
Within 8 cm	80%	96%	100%
Within 10 cm	83%	98%	100%

Another approach is to exploit a known reference length in the background image to indicate scale [20, 15, 14]. This approach has two main difficulties in practice: (i) a reference length may not always be available, and (ii) the results are strongly affected by camera perspective. Criminsi et al. [15, 14] presented an algorithm which estimated height from a single known physical measurement in an image. They measured the vanishing line and vanishing plane directly from the image. Their methodology requires the reference height of a length in the vertical direction in the scene. They evaluated their methodology on a small number of subjects and achieved very small estimation errors. But this methodology is not applicable in our case as it requires reference heights of the images. Their evaluation dataset is also very small compared to our student and celebrity datasets.

A third approach for for estimating height requires a sequence of images from a single camera or multiple cameras (for example, from videos) [11]. Height estimates are obtained by segmenting the person from the background and fitting the height to a time-dependent model of stride. They evaluated their methodology on a dataset of frontal parallel sequences taken in a outdoor environment with 45 people. For each person, they used 4 samples of walking a 5 meter long path. Their height estimates had an average error of 3.5 cm, on the same order as the methodology presented in this paper. But this stride-matching approach requires not only a sequence of photos but also that the person is walking upright with a constant velocity for at least 3-4 seconds and that the camera is calibrated with respect to the ground plane. When obtaining a collection of photos from a social network, it is clearly not possible to meet all these conditions.

6.2 Estimating demographics and social information from group photos

Although to our knowledge there is no previous work on using photo collections to predict height, there is some previous work on using group photos and photo collections to infer demographic and social information from the photos. For example, Gallagher et al. [16] examined how positions of people in a group photo can be used to infer the genders and ages of the individuals in the photo. Shu et al. [18] proposed a methodology for classifying the type of group photo based on spatial arrangement and the predicted age and gender from the faces in the image. They demonstrated that face arrangement, when combined with age and gender prediction, is a useful clue in estimating the type of photo (family photo, sports team photo, friends hanging out, and so on).

6.3 Photos in Social Networks

Although we believe this is the first paper to use inter-photo techniques to mine information from online social networks, there has been significant work in using photos in social networks. Yardi el al. [24] have used Facebook graph and tagged photos to build web based authentication system. White [23] has discussed how travel photos posted by Facebook friends may influence the travel decisions of those who view these photos. Acquisti el al.[9] has mapped match.com users' profile to their corresponding Facebook profile using only photos posted in both websites and face recognition technology. Sedar et al. [21] have shown that intensity of smiling in Facebook photos may predict future life satisfaction. Other researchers also discuss privacy concerns associated with sharing and tagging photos in social networks. Facebook photos are already used by employers [7] to investigate job seekers and by law enforcement agencies [6] to resolve criminal cases. Besmer et al. [13] described various privacy concerns of Facebook users related to tagged photos and Ahern et al. [10] discussed Flickr users' privacy concerns and managing policies about uploaded photos.

7. CONCLUSION

In this paper we consider the problem of estimating the heights of all the users in a *photo collection*, such as a collection of photos from a social network. The main ideas in our methodology are (i) for each individual photo, estimate the height differences among the people standing in the photo (intra-photo estimation), (ii) from the photo collection, create a people graph, and combine this graph with the height difference estimates from the individual photos to generate height difference estimates among all the people in the collection (inter-photo estimation), (iii) then use these height difference estimates, as well as an *a priori* distribution, to estimate the heights of all the people in the photo collection. Because many people will appear in multiple photos across the collection, we showed that the height of each person can be compared with a much larger set of people (sometimes with most of the people in the collection), to significantly improve the estimates.

We formulated the problem as a MLE estimation problem, which when combined with a priori global mean and variance information, gives rise to a quadratic programming problem that can be easily solved with thousands of variables (people) at once. We applied the technique to estimating the heights of over 400 movie stars in the IMDb database and the heights of about 30 graduate students. Our results indicate that this data-driven approach can give good estimates, with significantly lower errors than the baseline approaches. We further found that as the photo collection increases, with people appearing in multiple photos, accuracy generally improves.

In addition to providing a novel methodology to estimating heights from photos, this paper also shows the great potential of mining data from photo collections in which the same people appear in multiple, but different, photos. In addition to features like height and weight, it may be possible to use photo collections to mine a story around the community people in the photos, such as determining kinship relationships and also current and past romantic relationships.

8. REFERENCES

[1] Brigger Furniture - Comfort Seats Made to Fit Your Body. http://www.briggerfurniture.com/.

[2] Camera Calibration With Opencv. http://docs.opencv.org/doc/tutorials/calib3d/camera_calibration/camera_calibration.html#cameracalibrationopencv.

[3] Destination XL - Men's Big and Tall Superstore. http://www.destinationxl.com/mens-big-and-tall-store/.

[4] King Size - The Big and Tall Experts. http://www.kingsizedirect.com/.

[5] Instagram reports 200m users, 20b photos shared, Gigacom, March 25, 2014. http://gigaom.com/2014/03/25/instagram-reports-200m-users-20b-photos-shared/.

[6] Walking a New Beat: Surfing Myspace.com helps cops crack the case, Newsweek, April 17, 2006. http://www.informationliberation.com/?id=9643.

[7] Face it: book no secret to employers, The Wastington Times, July 17, 2006. http://www.washingtontimes.com/news/2006/jul/17/20060717-124952-1800r/?page=all.

[8] A Focus on Efficiency, White Paper, September 23, 2013. http://www.informationliberation.com/?id=9643.

[9] A. Acquisti, R. Gross, and F. Stutzman. Faces of facebook: Privacy in the age of augmented reality. *BlackHat USA*, 2011.

[10] S. Ahern, D. Eckles, N. Good, S. King, M. Naaman, and R. Nair. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In *CHI*, volume 7, pages 357–366, 2007.

[11] C. BenAbdelkader, R. Cutler, and L. Davis. Person Identification using Automatic Height and Stride Estimation. In *16th International Conference on Pattern Recognition*, volume 4, pages 377–380, 2002.

[12] C. BenAbdelkader and Y. Yacoob. Statistical Body Height Estimation from a Single Image. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–7, 2008.

[13] A. Besmer and H. Richter Lipford. Moving beyond untagging: photo privacy in a tagged world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1563–1572. ACM, 2010.

[14] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *Internation Journal on Computer Vision*, 40(2):123–148, Nov. 2000.

[15] A. Criminisia, A. Zisserman, L. Van Gool, S. Bramble, and D. Compton. New Approach to Obtain Height Measurements from Video. In *Proceedings of SPIE- The International Society for Optical Engineering*, volume 3576, 1999.

[16] A. Gallagher and T. Chen. Understanding Images of Groups of People. *IEEE Conference on Computer Vision and Pattern Recognition*, 0:256–263, 2009.

[17] A. C. Gallagher and T. Chen. Finding Rows of People in Group Images. In *IEEE international conference on Multimedia and Expo*, pages 602–605, 2009.

[18] H. C. Henry Shu, Andrew Gallagher and T. Chen. Face-graph Matching for Classifying Groups of People. In *International Conference on Image Processing*, 2013.

[19] K. Kato and A. Higashiyama. Estimation of Height for Persons in Pictures. *Perception and Psychophysics*, 60(8):1318–1328, 1998.

[20] N. Saitoh, K. Kurosawa, and K. Kuroki. Study on Height Measurement from a Single View. In *IEEE International Conference on Image Processing*, pages 523–526, 1999.

[21] J. P. Seder and S. Oishi. Intensity of smiling in facebook photos predicts future life satisfaction. *Social Psychological and Personality Science*, 3(4):407–413, 2012.

[22] Z. Stone, T. Zickler, and T. Darrell. Autotagging facebook: Social network context improves photo annotation. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008.

[23] L. White. Facebook, friends and photos: A snapshot into social networking for generating travel ideas. *Tourism informatics: Visual travel recommender systems, social communities and user interface design. Hershey, PA: IGI Global*, pages 115–129, 2010.

[24] S. Yardi, N. Feamster, and A. Bruckman. Photo-based authentication using social networks. In *Proceedings of the first workshop on Online social networks*, pages 55–60. ACM, 2008.