Estimating Age, Gender, and Identity using First Name Priors

Andrew C. Gallagher
Carnegie Mellon University
Pittsburgh, Pennsylvania
agallagh@cmu.edu

Tsuhan Chen
Carnegie Mellon University
Pittsburgh, Pennsylvania
tsuhan@cmu.edu

Abstract

Recognizing people in images is one of the foremost challenges in computer vision. It is important to remember that consumer photography has a highly social aspect. The photographer captures images not in a random fashion, but rather to remember or document meaningful events in her life. The culture of the society of which the photographer is a part provides a strong context for recognizing the content of the captured images.

We demonstrate one aspect of this cultural context by recognizing people from first names. The distribution of first names chosen for newborn babies evolves with time and is gender-specific. As a result, a first name provides a strong prior for describing the individual. Specifically, we use the U.S. Social Security Administration baby name database to learn priors for gender and age for 6693 first names.

Most face recognition methods do not even consider the name of the individual of interest, or the name is treated merely as an identifier that provides no information about appearance. In contrast, we combine image-based gender and age classifiers with the cultural context information provided by first names to recognize people with no labeled examples. Our model uses image-based age and gender estimates for assigning first names to people and in turn, the age and gender estimates are improved.

1. Introduction

Face recognition is one of the most important, yet difficult tasks in computer vision. Current methods focus on measuring the similarity between face images and consequently are of little value when attempting to recognize an entirely new never-before seen person. Yet, in some situations, humans are able to do just that.

We contend that recognizing people in consumer images is far more than solely a face recognition problem. To best understand the semantics of who is in an image, we need to understand people, their culture, and the social aspects of their interactions. To illustrate this point, consider Figure 1, which shows two images, each containing a pair of people. Given the first names of the people in each image, most people familiar with American first names will be able to correctly assign the first names to all four faces. If the names were merely labels that contain no information, we would expect to properly assign only two names to the correct people (by random chance). Yet humans gain an understanding of their culture that allows them to easily perform complex recognition tasks such as illustrated here. Specifically, humans learn to associate first names with appearance, age, and gender. The apparent age or gender affects the likelihood that a person has a particular name. Likewise, a person's first name allows us to better estimate their age and gender. We model and exploit this association in our paper.

Figure 1. Is it possible to recognize people for which no labeled examples exist? We describe the use of contextual information related to first names to recognize people in consumer images. (Left) An image of Sierra and Patrick. By recognizing the gender of the people and names, we can confidently conclude that Patrick must be the man on the right, while Sierra is the woman. (Right) This image contains Mildred and Lisa. Mildred, a first name popular in the early 20$^{\text{th}}$ century, is the older woman on the right, while Lisa is the younger woman on the left. This recognition is possible for humans because of their extensive cultural training.

We describe a new approach for recognizing people and estimating their ages and genders using a contextual prior derived from first names. We demonstrate that first names contain much useful information, and in many cases allow us to recognize people in images for which no training examples were ever labeled. In Section 2, we review the related work. We describe a first name database (Section 3) that allows us to understand the relationship between first names, birth years, and gender. We introduce the descriptors of age and gender (Section 4) and our model for inferring explanations for the ages, genders, and first names of people in an image (Section 5), and our image-based gender and age classifiers (Section 6). Finally, we describe the performance of our model in Section 7.

2. Related Work

A recent thrust in computer vision concerns the use of context in object detection and recognition. For example, Boutell and Luo use image capture metadata and timestamp to classify images as either indoor or outdoor [15]. Hoiem *et al*. [11], and Torralba and Sinha [21] describe the context (in 3D and 2D, respectively) of a scene and the relationship between context and object detection. We observe that the context in which an image is captured extends far beyond the pixel values in the image itself. Geographic location, cultural influences, and time all affect the likelihood that specific objects will appear in images. For a few simple examples, we would not expect to see images of airplanes captured prior to 1900, images of ocean beaches in Kansas, or images of snow in Panama.

Certainly, the most popular method for recognizing images of people is face recognition. There are many techniques for recognizing faces, or for comparing the similarity of two faces [26], and under controlled environments, recognition rates exceed 90% [19]. However, there are significant differences between the problem of face recognition in general and the problem we are addressing. Often, a face of unknown identity is compared against a gallery of face images with known identity, where each gallery image is captured with similar pose, illumination and expression [10, 18]. For individual consumers, developing such a gallery is inconvienient at best and impossible at worst. Reseachers have incorporated face recognition techniques to aid searching, retrieving, and labeling of consumer images [9, 25, 23, 1]. All of these systems rely on the user to label example faces for each individual to be recognized.

Both the Satoh and Kanade [20] "Name-It" system and Berg *et al*. [5] associate names from captions with faces from images or video. The main focus in these papers is to use a large number of images to aid in the unsupervised clustering. Similarly, in Zhang *et al*. [25], a user indicates a set of images that contain a certain person. The algorithm selects one face from each image, maximizing the similarity, and concludes these faces must be the certain person. In these papers, names are treated merely as labels that contribute no infomation to the problem solution. The desire is always to assign the same label to similar faces from different images. As a result, none of these papers could resolve the problem of associating multiple names and images in single image (as readily noted in [25]).

Several researchers have attempted to recognize people from contextual information that extends beyond pixel data. In an extreme example, Naaman *et al*. [17] describe an interactive labeling application that uses only context (e.g. popularity, co-occurance, and geographic re-occurance) to create a short drop-down list for labeling the identities of people in the image. This method uses no image features, although the authors note that the combination of context- and content-based techniques would be desirable. Gallagher and Chen use a group prior [7] to learn social groups that well-explain the observed image facial features of groups of people in consumer image collections.

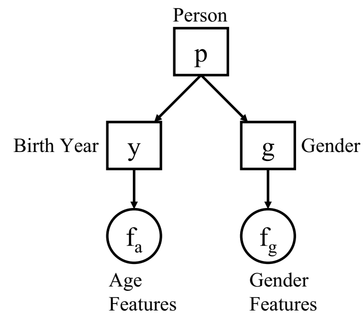

Figure 2. A graphical model that represents the relationship between a person p having a first name n, the descriptors of birth year y and gender g, and the image-based features f_y and f_g.

It has long been known that the appearance of a person is not independent of the name. Parents spend extraordinary time and effort to select the perfect name for an expected child. This decision is influenced by many factors, such as the gender of the child, the local culture, and family heritage. Researchers [14] have shown that humans associate certain names with certain stereotypical facial features, for example the name "Bob" is associated with a rounder face than the name "Tim."

We show the value of using names as not just labels, but also as a rich source of contextual information about individuals. We use a data-driven approach to the problem, using a large first name database for learning priors. We model the relationship between first names, age, gender, and appearance. Similar to the approach the reader might take to solve the problem shown in Figure 1, our model (Figure 2) considers age and gender estimates from image features, as well as first names. The model then uses all the available information to find a plausible explanation for the ages, genders, and names of the people in the image.

3. First Name Data

In our work, we use the U.S. Social Security baby name database [22]. This database contains the 1000 most popular male and female baby names (among applicants for a U.S. Social Security Number) for each year between 1880 and 2006 (representing over 280 million named babies.) The results described here could be extended to other countries and cultures given the appropriate training data. Using this data, we can compute statistics related to distributions over birth year, gender, and first name.

The influence of popular culture on selected names in evident in the database. For example, between 1936 and 1937, the popularity of the female name "Deanna" increased by 2000%, the largest percentage increase in the database, coinciding with the first feature length film starring popular actress Deanna Durbin in 1936. Likewise, the largest decline in name popularity occurred between 1977 and 1978, when "Farrah" fell by 78% coinciding with actress Farrah Fawcett leaving the popular show "Charlie's Angels" in 1977.

The database contains a total of 6693 unique names, with 3401 names associated with male babies, 3960 associated with female babies, and 668 shared between both genders. There is nearly twice the diversity in the names selected for females (entropy of first names, given female $H(p|g = \text{female}) = 9.20$ bits) than for males ($H(p|g = \text{male}) = 8.22$ bits). The majority of first names are strongly associated with one gender or the other. The entropy of gender is nearly one bit (0.998) but the conditional entropy of gender given first name is only $H(g|p = n) = 0.055$. However, some names are surprisingly gender-neutral. For example, the names "Peyton", "Finley", "Kris", "Kerry" and "Avery" all have nearly equal probability of being assigned to either a boy or girl.

First names also names convey a great deal of information about year of birth. Names such as "Aiden", "Caden", "Camryn", "Jaiden", "Nevaeh", "Serenity", and "Zoey" all have expected birth years more recent than 2001. Therefore, we expect recent images of people with these names to be small children. Other names experience cyclical or level popularity, and consequently do not reveal much about the age of the individual. For example, of all the first names, the name "Nora" leaves us with the greatest uncertainty regarding the year of birth. Figure 3 shows the distribution over birth year for a selection of first names, assuming that the person is alive in 2007. We consider life expectancy in our calculations, using a standard actuarial table [2]. We estimate there are approximately 3.9 million men named "James" and 2.6 million women named "Mary" alive today in the U.S.; the most popular names for each gender.

4. Modeling the Appearance of a Name

We would like to model the relationship between the appearance f_i of a person p in an image with the first name n. Ideally, this relationship $P(f_i, p = n)$, a statistical model of appearance for each possible first name, could be learned given a huge number of training images of people and associated names. For example, by collecting hundreds or thousands of portraits of people for each possible first name, a model of the appearance of that first name could be learned. This could be an attractive approach, but it is not yet feasible for a number of reasons. First, while there are billions of images of people on the internet and websites such as Flickr (www.flickr.com), it is still not easy to find images of people that have been labeled with accuracy, and a manual human review might still be necessary. Second, celebrities generally are labeled with greater accuracy but in far greater numbers than are non-celebrities. For example, a search for "Angelina" returns an inordinate number of pictures of actress Angelina Jolie, creating a sampling bias that is difficult to address. Third, this appearance model changes over time as a particular first name decreases or increases in popularity, and those already with a given first name change in appearance as they age. Managing this evolution is a difficult task in itself.

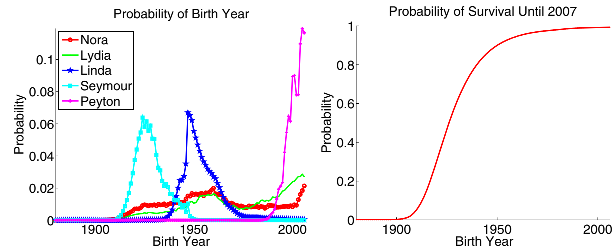

Figure 3. (Left) The distribution over birth year for a selection of first names, given the person is alive in 2007. (Right) Considering life expectancy, the probability that a person from a given birth year is alive in 2007.

We take an alternate approach. Rather than directly learning the appearance for each name, we instead propose a set of descriptors that have an easy-to-learn relationship with both first names and the visual appearance of person images. The descriptors we select are birth year y and gender g, as we can learn $P(a|f_a)$, an image-based estimate of the person's age given age-relevant appearance features f_a and $P(g|f_g)$, the gender of the person given gender-relevant appearance features f_g. When the image has the associated image capture time stored in the EXIF header, the relationship between $P(a|f_g)$ and $P(y|f_g)$ is simply:

$$P(y|f_a) = P(a = c - y|f_a) \quad (1)$$

where c represents the image capture year, y represents a possible birth year and a is the age of the person. We use the terms "age" and "birth year" synonymously because each conveys the same information, given that the age is known with respect to a reference year.

Likewise, given the first name database, the distributions over these same descriptors ($P(y|p = n)$ and $P(g|p = n)$), the distribution of birth years for a given first name and the distribution over gender for a given first name, are learned

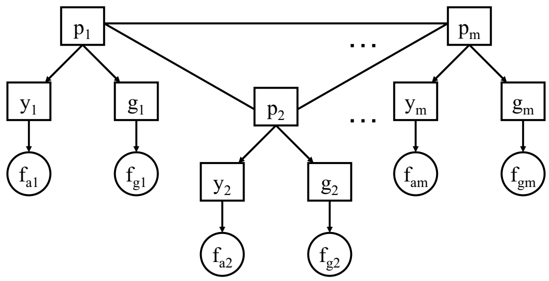

Figure 4. A graphical model that represents the relationship between a person p, the descriptors of birth year y and gender g, and the associated features f_y and f_g.

with maximum likelihood estimation.

In essence, our approach amounts to the following: A first name provides a description of attributes associated with an individual. By extracting descriptors from a person image that describe these same attributes, we can compute distributions over name, age, and gender.

5. Recognition from First Names

For a person in an image, we extract features related to each of the descriptors (gender and age). The name of the person and the values of the descriptors are represented as random variables. We make the simplifying assumption that given a first name, birth year and gender are independent, as in the graph model of Figure 2. Features related to age f_a and gender f_g are observed in the image, and we want to find the likelihood of a particular first name given these descriptor-specific features. The joint distribution can be written:

$$P(p, y, g|f_a, f_g) = P(p)P(y|p)\frac{P(y|f_a)}{P(y)}P(g|p)\frac{P(g|f_g)}{P(g)} \tag{2}$$

The term $P(g|p = n)$ is the probability that person with first name n has a particular gender. The term $P(y|p = n)$ is the probability that person with first name n was born in a particular year. This distribution is estimated from the name data, while considering the life expectancy as follows:

$$P(y = i|p = n, c) \propto \text{count}(y = i, p = n)_{c-i}p_0 \tag{3}$$

where the notation ${}_{c-i}p_0$ is used in actuarial science to indicate the probability of survival from birth (age 0) to age $c - i$, where c is the image capture year (since we know the person is alive in this year).

Finding the likelihood $P(p = n|f)$ of a particular name assignment $p = n$ given all the features $f = \{f_a, f_g\}$ is accomplished by marginalizing the joint distribution over all possible assignments of birth year and gender.

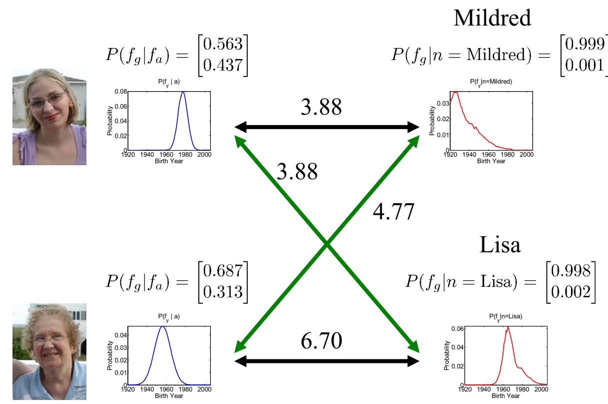

Figure 5. Assigning names to people can be represented as a bipartite graph. The estimates of gender and birth year given the names Mildred and Lisa as well as the appearance features are shown.The cost of each assignment is shown on each edge, and Munkres algorithm correctly assigns the names to faces (green edges).

When multiple people are in the image, the interactions between the name-person assignments are represented with the graph model shown in Figure 4. Our model incorporates the independence assumption that once the first name of a person p_m is known, the age and gender of this person are independent of the attributes of names of other people in the image. A particular person in the image is p_m, the associated features are f_{am} and f_{gm}, and the name assigned to person p_m is n_m. We seek to map a set of K first names $\mathbf{N}$ to the set of M people $\mathbf{p}$ in a single image with associated features $\mathbf{f}$ where there are no labeled training faces from which to directly estimate $P(\mathbf{f}|\mathbf{p} = \mathbf{n})$, where $\mathbf{n}$ is a particular assignment of names to people $\mathbf{p}$ in the image.

Using the independence assumptions from the graph model, we write $P(\mathbf{p} = \mathbf{n}|\mathbf{f})$:

$$P(\mathbf{p} = \mathbf{n}|\mathbf{f}) = \frac{P(\mathbf{f}|\mathbf{p} = \mathbf{n})P(\mathbf{p} = \mathbf{n})}{P(\mathbf{f})} \tag{4}$$

$$\propto P(\mathbf{p} = \mathbf{n})\prod_m P(f_m|p_m = n_m) \tag{5}$$

The maximum likelihood assignment of names to people is the one that maximizes $P(\mathbf{p} = \mathbf{n}|\mathbf{f})$. $P(\mathbf{p} = \mathbf{n})$ is the group prior [7] for a particular set of individuals appearing together in an image. In our case, we assume this term is a non-zero constant only for valid assignments of names to people. Then, the log likelihood we desire to minimize is:

$$\mathcal{L} = -\log P(\mathbf{p} = \mathbf{n}) - \sum_m \log P(f_m|p_m = n_m) \tag{6}$$

The term $\log P(\mathbf{p} = \mathbf{n})$ enforces that the name assignments are valid (no more than one person for each name, and no more than one name for each person). Name assignments $\mathbf{p} = \mathbf{n}$ with zero probability incur an infinite penalty. Assuming K first names and M people in the image, there are at most $\max(M, K)!$ possible combinations of names to people to consider. However, the complexity is reduced by recognizing that equation 6 exactly describes the classic assignment problem. The assignment problem is represented as a bipartite graph where one set of nodes represents people in the image, and the other set represents first names, as illustrated in Figure 5. The cost between each vertex is $-\log(P(f_m|p_m = n_m))$. This problem can be solved in $O(\max(M, K)^3)$ using Munkres algorithm [16].

According to our model, age is influenced by both the first name and the age-specific features extracted from the image of the person. Likewise, gender is affected by both the first name and gender-specific features. Our model is used to select the most likely name to person assignment, and also to refine the image-based estimates of age and gender. When finding distributions over age and gender for a given person p, the marginal probability of the name assigned to that person is used. For example, for finding the distribution $P(g|f_g, p = n)$ with our model:

$$P(g|f_g, p = n) \propto \sum_n P(p = n)P(g|p = n)\frac{P(g|f_g)}{P(g)} \quad (7)$$

A similar calculation is performed for using the model to find the distribution over age $P(y|f_a, p = n)$. In the inference step, we first find the maximum likelihood name assignments given the initial age and gender estimates, then update the age and gender estimates. Further iteration did not significantly affect the results, so only one iteration is performed.

6. Image-Based Gender and Age Classifiers

Our model requires estimates of $P(y|f_a)$, age given age-specific features and $P(g|f_g)$, gender given gender-specific features extracted from an image.

We implemented age and gender classifiers following the examples of [13, 8] and [24, 3]. For age classification, we acquired the image collections from three consumers, and labeled the individuals in each image, for a total of 117 unique individuals. The birth year of each individual is known or estimated by the collection owner. Using the image capture date from the EXIF information and the individual birthdates, the age of each person in each image is computed. This results in an independent training set of 2855 faces with corresponding ground truth ages. Each face is normalized in scale (49×61 pixels) and projected onto a set of Fisherfaces [4] created from an independent set of faces from 31 individuals. The age of a query face is found by normalizing its scale, projecting onto the set of Fisherfaces, and finding the nearest neighbors (we use 25) in the projection space using a Euclidean distance measure. The estimated age of the query face is the median of the ages of these nearest neighbors. Given this estimate for the age, we then model $P(a|f_a)$ as a Gaussian having a mean value of the estimated age, a standard deviation of one-third the estimated age (the accuracy of our age classifier decreases with age), and truncated so that ages less than zero have no density. Figure 6 shows several age classification results.

Figure 6. A sampling of our age estimation results. Each row shows a random selection of people for which the age classification result was within a specific range. (Top) Babies and children under the age of five. (Middle) Adults between the ages of 18 and 41. (Bottom) Adults older than 42. The colored bar indicates whether that classification agreed with the human-estimated age for the person, where green indicates agreement.

Following the example of [24], we implement a face gender classifier using a support vector machine. We reduce the feature dimensionality by first extracting facial features using an Active Shape Model [6]. The ASM locates 82 key points including the eyes, eyebrows, nose, mouth, and face border. Following the method of [7], PCA is futher used to reduce the dimensonality to five features. A training set of 3546 gender-labeled faces from our consumer image database is used to learn a support vector machine that outputs probabilistic density estimates for gender. Figure 7 shows the gender estimation results for a selection of face images.

7. Experiment

Tags are often used to indicate objects within an image without providing the spatial location of the objects. For example Flickr and Adobe Albums software both allow users to tag images with keywords. Our experiments address the scenario where images contain multiple people, and are tagged to indicate the first names of the people in the image. Our goal is to disambiguate the tags by assigning names to people based on a single image and to estimate the age and gender of each person. This name-person assignement could be a useful first step for an application that then searches for these same individuals in other images.

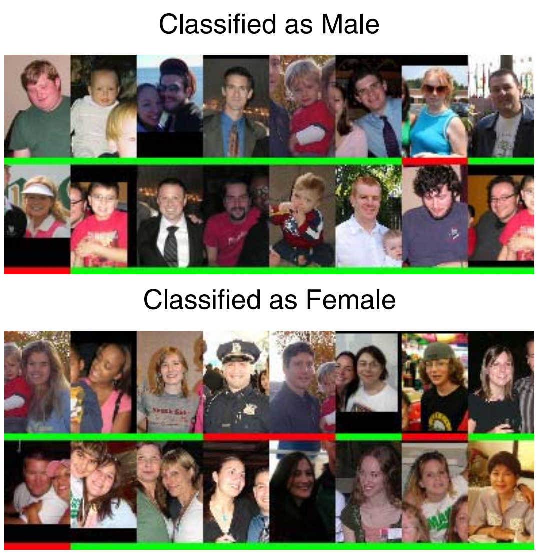

Figure 7. Gender classification results. (Top) A random selection of people classified as male. (Bottom) A random selection of people classified as female. The colored bar beneath each image is green if the classification is correct.

	Set A	Set B	Set C	Overall
Total images	134	14	48	148
Total people	307	32	105	339
Total males	132	8	26	140
Total females	175	24	79	199
Total children under 10	36	8	12	44
Images with >2 people	31	3	8	34
Uniform gender images	40	8	48	48

Table 1. A summary of our test sets. Set C is comprised of all images from Sets A and B where the people all have the same gender. The Overall Set is the union of sets A and B.

	Set A	Set B	Set C	Overall
Random	43.7%	43.8%	45.7%	43.7%
Age	47.9%	**59.4**%	58.1%	49.0%
Gender	59.3%	56.3%	51.4%	59.0%
Age+Gender	**62.2%**	56.3%	**61.9%**	**61.7%**

Table 2. Using image-based age and gender classifiers for recognizing people in a single image. The percentage of correct name assignments is reported. The "Random" row values are expectations rather than an actual experiment. The other three rows show the performance of first name assignment using the image-based age classifier, gender classifier, or both.

We used the following method to collect test sets of names and faces. For Set A, the U.S. baby name database is used to generate random first names. We produce 100 independent pairs of random names. A search is performed on Flickr to find images containing the pairs of people with those first names. The images from the search were painstakingly examined to manually assign names to faces (using captions and other tagged images from the same user's collection). Most of the images are 500×375 pixels, and contain people with challenging poses and expressions, difficult lighting, sunglasses, and occlusion. For most of the name pairs at least one image was located, resulting in a test set of 134 images with 307 people.

In constructing Set B, we selected name pairs that might be difficult for humans to perform the name assignment task. For example, the names "Chris" and "Dana" can each be male or female but each lean towards a specific gender. Also, we used name pairs that have a large disparity in expected birth year, but are perhaps less well known, for example "Tammy" (most popular in the 1960s) and "Paige" (popular in the past decade). This small but challenging set contains 14 images and the associated first name tags. Set C contains all those images from Sets A and B where all people have a common gender. Name assignment is difficult in this subset since recognizing gender alone is not sufficient to ensure good performance. Table 1 summaries the characteristics of the test images for our experiments.

For detecting faces, we use a cascade face detector similar to [12]. As our focus is not on face detection, we manually add faces that are missed by our face detector by clicking on the eyes of the missed face. Faces range in size from 12 to 74 pixels between the pupils. We compute image-based estimates of the age and gender of each person using the classifiers described in Section 6. Finally, our model (Section 5) is used to find the most likely assignment of first names to faces and estimates of age and gender that incorporate evidence from both image features and first names.

Name Assignment Accuracy: Table 2 reports the accuracy of our algorithm, considering different subsets of the test set and the model. We show a considerable improvement over random guessing for all subsets of test images. Using the image-based age classifier provides improved name assignment with images of constant gender (Set C), and in the challenging Set B. Using both the age and the gender descriptors provides the greatest improvement in name assignment accuracy, from 43.7% to 61.7% overall, reducing the error rate by 32%. Figure 8 discusses several image examples, the name assignments, and the age and gender classifications from the image-based classifiers and from the model.

Age and Gender: Our model improves the age and gender estimates over the estimates from the image-based classifiers. For each person image, we manually labeled the age and the gender of the person (without looking an any name infomation or tags associated with the image). Our image-based age classifier has a mean absolute error of 10.0 years, and 28.6% of the genders are misclassified by the image-based gender classifier. Our model is used to assign names to people, and then the age and the gender are re-estimated based on the name assignments as described in Section 5.

	Age	Gender
	Image-Based → Model	Image-Based → Model
Age	10.0 →9.33	→35.4%
Gender	→13.7	28.6% →18.3%
Age+Gend	10.0 →9.38	28.6% →19.5%

Table 3. Our model provides improvement over the image-based age and gender classifiers. This table shows the error reduction achieved by estimating age and gender with our model. For the age column, we show that the mean absolute difference between an age estimate and a manually labeled age. For the gender column, the percent is the gender classification error rate. The rows show the error reduction using the image-based age classifier, the image-based gender classifier, or both. The model predicts age even when no image-based age classifier is used, and gender even when no image-based gender classifier is used.

	Set A	Set B	Set C	Overall
Subject 1	79.2%	81.3%	**65.7**%	79.4%
Subject 2	78.2%	68.8%	61.0%	77.3%
Subject 3	79.5%	43.8%	54.3%	76.1%
Subject 4	69.1%	53.1%	41.9%	67.6%
Human Age+Gender	**80.8**%	**93.8**%	63.8%	**82.0**%

Table 4. Results for Humans. Four subjects perform the same name assignment task as does our algorithm, and this table reports each subject's accuracy for assigning names to faces. The last row ("Human Age+Gender") reports the results of using our model for name assignment, but using manually labeled values for age and gender rather than image-based classifiers.

Both the age estimation and the gender classification are improved through this process, as shown in Table 3. The gender classification error is reduced by 32% compared to using only image-based classifiers. The age classification error reduction is a more modest 6%, likely due to the fact that most names vary only slowly in popularity over time.

Human Performance: It is interesting to compare the results of our algorithm with the accuracy of a human attempting the same task. A user interface was created to allow a human subject to easily assign each tagged name to the person that the subject felt was most plausible. A total of four subjects repeated this exercise for each of the 146 images in the test set. The results of this human experiment are reported in Table 4. The values in this table can be compared directly with those for our model, shown in Table 2.

Subjects 1 and 2 have the overall best performances and are U.S. born, while subjects 3 and 4 have each lived in the United States for about five years and have lower classification accuracy. This supports our assertion that this image understanding problem requires an understanding of cultural context. Subjects with less time in the U.S. had less time to form this contextual prior, and therefore find the name assignment task more challenging. In fact, by virtue of having a more complete contextual prior, our model outperforms subjects 3 and 4 on the difficult Sets B and C.

We did an additional experiment to verify our model. Rather than relying on age and gender estimates from the image-based classifiers, we manually labeled each person's age and gender, without any knowledge of the names associated with the image. Then the model is used to produce name assignments using these manually derived estimates for $p(y|f_a)$ and $p(g|f_g)$. The accuracy of this approach is reported in the "Human Age+Gender" row of Table 4. This method produces the highest overall name assignment accuracy compared to the four test subjects, beating the best human subject by 2.6%. This success can be explained by considering that the model has complete domain knowledge regarding first names in the United States, but each human's contextual knowledge of first names is incomplete to some degree. When the model is given a human-level ability to classify gender and age, then it is difficult for a human to achieve greater accuracy. From this experiment, we draw several conclusions. First, we expect that improved gender and age predictors will improve the performance of our model. Second, because the performances of the human subjects and the "Human Age+Gender" method are similar, our model is validated and the independence assuptions that we made are shown to be reasonable.

8. Conclusion

In this paper we introduced a model for the relationship between first names, age, gender, and appearance in images. With this model, we infer likely name assignments for images tagged with the first names of the people in a single image. Further, we show that the model's estimates of age and gender are superior to those from an image-based classifier. Our model could be extended by understanding the context of nicknames (for example, baby "Timothy" might be called "Timmy"), and familial titles such as "Mom" or "Uncle" that likely also provide strong priors for the appearance of people in the image.

In a broader scope, our work is a case study emphasizing that images must be interpreted in the context of the culture in which they are captured. A good understanding of the cultural context provides a strong prior for image understanding.

References

[1] D. Anguelov, K.-C. Lee, S. Burak, Gokturk, and B. Sumengen. Contextual identity recognition in personal photo albums. In *Proc. CVPR*, 2007.

[2] E. Arias. United States life tables, 2003. Technical report, National Center for Health Statistics, 2006.

[3] S. Baluja and H. Rowley. Boosting sex identification performance. In *IJCV*, 2007.

[4] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *PAMI*, 1997.

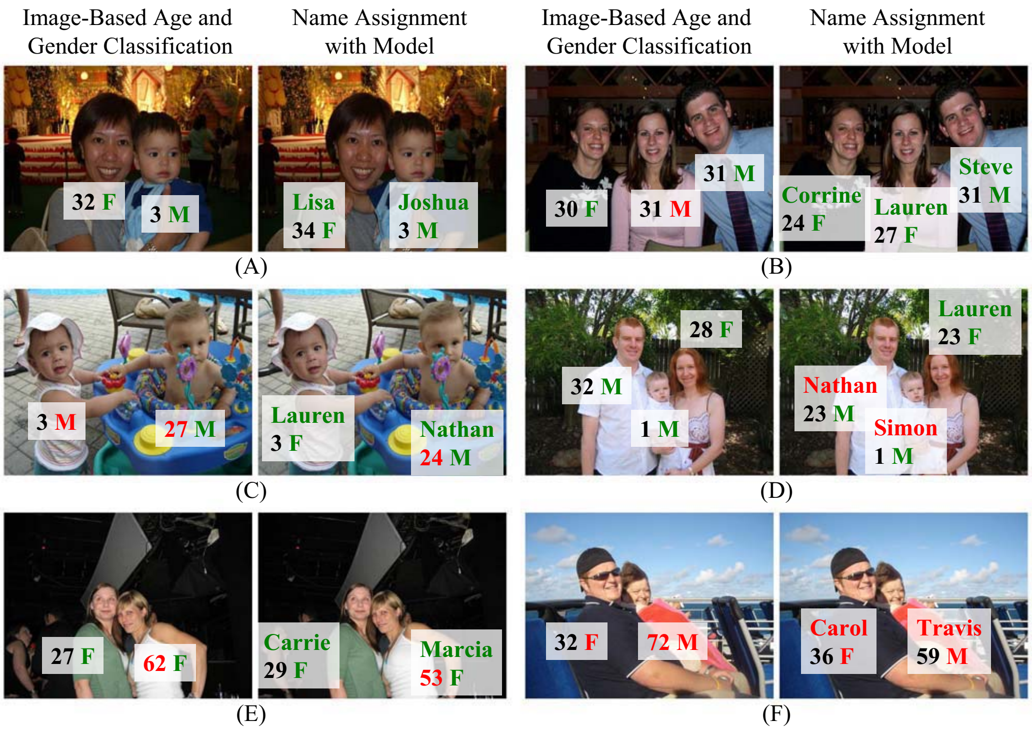

Figure 8. Our model assigns names to people in images, and improves the performance of gender and age classifiers. The left image in each pair shows the estimated age and gender directly from our image-based classifiers. (Green indicates correct, and red indicates incorrect. Age classification results are only marked incorrect when they are grossly wrong.) The right image shows the assigned names for each person, and the estimated age and gender from the model. Poor age classifications due to pose and occlusion are generally mitigated (though still not perfect), in (C), (E), and (F). Gender misclassifications are often corrected by the influence of the first names as in (B) and (C). Image (D) illustrates that reasonable gender and age classifications do not guarantee a good name to person assignment, due to the ambiguity of the problem. In (F), multiple image-based misclassifications can place the model in a position where recovery is unlikely.

[5] T. Berg, A. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D. Forsyth. Names and faces in the news. In *Proc. CVPR*, 2004.

[6] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *CVIU*, 1995.

[7] A. Gallagher and T. Chen. Using group prior to identify people in consumer images. In *Proc. CVPR SLAM workshop*, 2007.

[8] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai. Learning from facial aging patterns for automatic age estimation. In *ACM MULTIMEDIA*, 2006.

[9] A. Girgensohn, J. Adcock, and L. Wilcox. Leveraging face recognition technology to find and organize photos. In *Proc. MIR*, 2004.

[10] R. Gross, I. Matthews, and S. Baker. Eigen light-fields and face recognition across pose. In *FGR*, 2002.

[11] D. Hoiem, A. Efros, and M. Hebert. Putting objects in perspective. In *Proc. CVPR*, 2006.

[12] M. Jones and P. Viola. Fast multiview face detector. In *Proc. CVPR*, 2003.

[13] A. Lanitis, C. Taylor, and T. Cootes. Toward automatic simulation of aging effects on face images. *PAMI*, 2002.

[14] M. Lea, R. Thomas, N. Lamkin, and A. Bell. Who do you look like? evidence of facial stereotypes for male names. *Psychonomic Bulletin and Review*, 2007.

[15] J. Luo, M. Boutell, and C. Brown. Pictures are not taken in a vacuum - an overview of exploiting context for semantic scene content understanding. *IEEE Signal Processing Magazine*, 2006.

[16] J. Munkres. Algorithms for the assignment and transportation problems. *SIAM*, 1957.

[17] M. Naaman, R. Yeh, H. Garcia-Molina, and A. Paepcke. Leveraging context to resolve identity in photo albums. In *Proc. JCDL*, 2005.

[18] K. Nishino, P. Belhumeur, and S. Nayar. Using eye reflections for face recognition under varying illumination. In *Proc. ICCV*, 2005.

[19] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, K. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale results, 2007.

[20] S. Satoh and T. Kanade. Name-it: Association of face and name in video. In *Proc. CVPR*, 1997.

[21] A. Torralba and P. Sinha. Statistical context priming for object detection. In *Proc. ICCV*, 2001.

[22] U.S. Social Security Administration. Baby name database. http://www.socialsecurity.gov/OACT/babynames.

[23] R. X. Y. Tian, W. Liu, F. Wen, and X. Tang. A face annotation framework with partial clustering and interactive labeling. In *Proc. CVPR*, 2007.

[24] M.-H. Yang and B. Moghaddam. Support vector machines for visual gender classification. *Proc. ICPR*, 2000.

[25] L. Zhang, Y. Hu, M. Li, and H. Zhang. Efficient propagation for face annotation in family albums. In *Proc. MM*, 2004.

[26] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 2003.