Gender Classification in Video

Aman Agarwal (aaman@wharton.upenn.edu), Syed Ali Raza (syedar@seas.upenn.edu)

Faculty Advisor: Dr. Jainbo Shi
Abstract:

A relatively new and interesting area in computer science and robotics is computer vision; the power to extract useful information automatically from videos and images. This project aims to assign a gender to the humans detected by video. The input is a video taken from a moving car on the streets of Philadelphia. As a result, the human objects are full body and the resolution of facial expressions is low. Even the resolution of the person is very bleak.

To solve this problem, instead of concentrating on facial features, we tried to garner more information from the surrounding to determine the gender of the individual. Features used were length of the skin showing in the neck, length of clothes and presence of pink among others to arrive at a hypothesis on gender. We maintained a variable that stored the constant count of the gender factor and once all classifiers were used, we outputted the gender based on the factor. The success rate achieved was ~64% depending on the data set.

This information could be useful for media and advertising companies who by using such software could change electronic ads on billboards to suit the gender around the board. They could do this on the fly by taking constant readings of the people in the surrounding. It could also prove invaluable in study of geography and sociology to study changing gender trends in a specific region. Survey companies would also find this useful to track shifts or survey gender in a specific area.

Related Work:

*High Resolution, Frontal Face:*

Most work in Gender Identification has been done using high quality frontal faces. The projects have used pictures of faces and analyzed them using various calculus and math based techniques to arrive at conclusions regarding those people. Examples of such research are:

*Age and Gender Estimation Based on Facial Image Analysis* by the Faculty of Engineering, Kagawa University in Japan focuses on extracting age and gender in portraits. It is known to have a 68% success rate and uses wrinkles primarily to determine the age and gender of a high resolution human face. This technique only works with frontal, high resolution portraits.

*An analysis of automatic gender classification* - Castrillon-Santana, M. (Univ. de Las Palmas de Gran Canaria, Las Palmas, Spain); Vuong, Q.C. *Source: Progress in Pattern Recognition, Image Analysis and Applications. Proceedings 12th Iberoamerican Congress on Pattern Recognition, CIARP 2007* (Lecture Notes in Computer Science vol. 4756), 2007, 271-80; ISBN-10: 3-540-76724-X. This article claims a 96% accuracy rate on gender classification of images from a 500 male and 500 female database. It again uses full frontal, high resolution pictures and a mix of differential equations with face masks to determine the correct gender. This kind of analysis would never work with the video input we have since we don’t get a high facial resolution.
Another website ‘celebrity.myheritage.com’ compares frontal face pictures with celebrity masks and though doesn’t explicitly identify the gender but matches the face to a celebrity. This implicitly calculates gender and is a method that can be used given we have high resolution frontal image of only the face.

**Low Resolution, Frontal Face:**

Other work being done in this field relates to low resolution pictures but still apply front facial recognition. They have lower quality picture, so they cannot determine each feature of the face perfectly, but they still have enough resolution to use the face as the primary factor in determining the gender. The most prominent paper in that regard is **Paper Identity and Gender Recognition Using the ENCARA Real-Time Face Detector** by IUSIANI - Universidad de Las Palmas de Gran Canaria Spain. This group of researchers used a real-time face detector, ENCARA, to detect frontal faces in video streams, providing fast performance. The project used visual information provided by a single camera which wasn’t very high resolution (a high quality web cam) and employed explicit and implicit knowledge to arrive at a conclusion. Although this system used lower resolution pictures, it still primarily worked by extracting frontal images from videos and performing analysis on them.

Another paper attempting such a problem is **Face identification in the near-absence of focal attention** - Reddy, L. (Computer & Neural Syst., California Inst. of Technol., Pasadena, CA, USA); Koch, C. Source: Vision Research, v 46, n 15, July 2006, 2336-43; ISSN: 0042-6989 CODEN: VISRAM; Publisher: Elsevier, UK. The key point for this paper is the ‘near-absence’ factor. The paper still uses mostly focal recognition with some attention to spatial recognition. The paper is slightly more relevant to our project in terms of the challenges faced but the outcome of the paper is to achieve identification of the person rather than classification. They don’t attempt to define the gender but rather match it to relevant people’s masks and/or compare with a set of pictures and indentify the human being in the picture. The project is relevant for pictures and videos of celebrities to automatically classify which celebrities have appeared in the video or picture.

**Very Low Resolution, No Defined Face:**

Till date, no paper has been published on classifying gender of humans from a low resolution moving camera. Most papers have been utilizing high resolution facial features. What makes our project vastly different from the approaches tried before is the concentration on aspects of context and surrounding rather than facial features to arrive at a hypothesis. In our input, it was impossible to extract any useful information of the face. In fact the resolution is often so bad that the eyes cannot be discerned from the face. The face looks like a mere blob of skin color.

Similar work in the past or even in the present relies on extracting common feature points from a huge set of male and female frames and store them. After that they see if there is a match between the set of feature points they have stored and with the given frame whose gender is to be determined. They also might compare to masks of certain genders. We, on the other hand, use no such approach. First, we
find the alignment of the body relying mostly on color whereas the aforementioned feature points rely on edges and corners. After this we rely on facts such as length of skirt, relative height, color of clothes and length of neck among others to ascertain the gender which is a completely different approach than what has been done before and what’s happening right now.

**Technical Approach:**

We have been provided with a ‘cut out’ of human figures in the input video. To facilitate our identification of gender, we decided to find the body alignment of the human in the figure. In last year’s project on human identification, there were many cases where even though the human was identified correctly, the head or the legs or other parts of the body were cut off. The data set we used was of hand cropped images in which the human has been cut out (in a rectangular shape with certain elements in the background present). Since working with false positive images to begin with could be a deterrent to quick progress, we worked on this data set which differed from the old data set in the sense that the images were properly cut and within the image the human was whole and complete. Our thought here was that if we cannot find an appropriate alignment of the body, then it is probably a false positive. Despite our want to identify false positives, we haven’t achieved that in our project. We also assumed that finding the alignment would be the hardest part but on the contrary, relative to finding the gender, it was actually easier.

**Alignment:**

Finding the alignment wasn’t a major part of our projection so it wasn’t attempted perfectly and only the very necessary information was extracted. For our purposes we only needed the location of the head and feet. For identifying feet, we start by looking for two clusters of points within the lower portion of the image. First, we look at clusters of similar color to the face, if not, for similar color per se so that we provide for people wearing shoes. Often women wear sandals and as such no two small distinct blobs are visible. Then we trace the skin color (which happens to be legs) as far down as possible and locate the last two points which we mark as feet. Though this approach does not provide the alignment with a hundred percent accuracy, it gives a good approximation that can be worked with. Problems that we have faced include a bright day where a light-colored person’s face melds in with the background due to the brightness, or the presence of shadows which made color analysis tough. Another problem we had was when people were standing (in side profile) in such a way that the feet appeared as a single color blob. We tackled it by assuming that this is the case when other probable cases fail. The location of the head for all images was roughly at the same point. We start at that point, and then in the surrounding, we look for a cluster of points with the same color as the sample facial colors we have stored. Once found, that color is stored to be used for further processing. In the case we are unable to find an alignment; an estimate of the location of the head and feet is made. The alignment allows us to find the person’s height ratio with the image. This feature will be explained further later. Other uses of determining the body alignment are that we can recognize features such as color of clothes and other similar gender distinguishing features. A simple color test which was implemented
was that if there is a certain amount of pink or purple in the body of the individual, then the chances of that individual being a female is higher.

**Diagram**

![Diagram showing process of identification](image)

**Process of Identification:**

Once we found the alignment, we applied classifications on the basis of various identifiers on these key points. The factors used to classify are mentioned in detail below. With the results from these factors, we extracted information and arrived at our conclusion on the gender of the human being. We had earlier hoped that our end result would be the same inputted video edited to include the classification by presenting colored bars to create a distinction between the male and female humans. This could not work out since our work was restricted to hand cropped images from the video.

Our algorithm uses a double variable called ‘factor’ which is initialized to zero. The relevant picture is then subject to alignment. After that, it goes through 5 classification sub routines which changes the ‘factor’ value. The 5 classifications used are: height ratio of the person, neck ratio, cloth ratio, leg ratio and color ratio. Each classification characteristic has a percentage of accuracy and a threshold value (based on our test sample). If the male/female threshold is met, then we increase/decrease the factor by the percentage accuracy of the test performed. Once all the five classifications have been gone through, our final factor value gives us the gender of the individual. We have assumed that a positive factor implies a male and a negative factor implies a female. Therefore, when we find a classification that meets the threshold for a female, we decrease the factor by the percentage accuracy of the test and conversely, if the threshold has been met for a male, we would increase the factor by the relevant percentage of accuracy. The only factor that has a common threshold is height. This is the deciding factor for those images who don’t satisfy any of the thresholds. In that case...
every image will have its ‘factor’ value modified by the height feature and as such will fall into one category or another. All the classification functions have been explained below with the thresholds and accuracies mentioned.

**Leg Ratio:**

The data sample we worked with comprised of videos taken in the summer. We made use of that in our estimate of their gender. Females often wore shorts, dresses, and skirts. Guys were seen wearing shorts too but their shorts were relatively long and came up to their knees. From the alignment, we calculated the pixel height of the person, and then we traced the facial color downwards to the legs and found the pixel height of the legs. The fraction of this height to the person’s pixel height was classified as leg ratio. We tested this on multiple images and found our hypothesis correct. We played with different thresholds and chose one which was both precise and had a high recall ratio. This value was 0.25 and had an accuracy of 71.43%.

While carrying out repeated trials of this, we noticed that often the leg ratio came out to be zero. After further investigation we realized that most of the time when this happened, it was usually a male wearing formal clothing. We put this to the test and found improved results. Indeed, the accuracy for this test was 85.71% but the recall rate was comparatively smaller.

**Neck Ratio:**

This also popped up during repeated trials of leg ratio. We were marking on our image all the points in the image which matched the extracted facial color of the person. We noticed that females tended to wear low cut shirts with more skin showing than men and this was captured in the aforementioned trial. We started measuring the width and length of the region below the face which matched the facial color and found that the length of the neck showing was an effective test for determining the gender. We realized that often females would wear shirts too so we eliminated those test pictures where the human had no or barely any neck showing. After that, we set the threshold of males at .055 and for females at .085 and achieved an accuracy of 67% and 70% respectively. The thresholds here represent the length of the neck extended downwards divided by the person’s height.

**Height Ratio:**

Originally, we were told that we would be provided with the distance from the camera to the object. Given the distance from the camera, we wrote a function which would give us the height of the object. Unfortunately, we couldn’t do this in the end because to calculate distance we needed left and right calibrated pictures. This was supposed to be the crux of our gender identification and a major setback but then we came up with another kind of height ratio. While pouring over our data sample we realized that regardless of a person’s height, the amount of pixels present above the head was more or less constant. We decided to put it to the test and calculated the fraction of human body in the frame. This figure was surprisingly different for males and females and as such became one of our
distinguishing features. Keeping a threshold at the 0.6 fraction mark we achieved an accuracy of 83.33% for females and 56.25% for males.

**Cloth Ratio:**

This works in the opposite sense to the leg ratio. Here we first find the color of the cloth approximately around the chest area and then we trace it downwards and upwards until the color changes. If the person in question is a female wearing a long dress then the length traced would be longer. As expected, high values tended to correspond with females wearing long dresses but it wasn’t as effective as our early estimates. The recall rate was low and the accuracy was 66.7%. Also this feature could only be used to place weight on females since males could not benefit from this feature in any way.

**Color:**

We hypothesized that the presence of colors like pink and bright purple should be attributed to the presence of females present in the picture. We originally thought that this would be a strong feature but many guys would wear similar colored clothing and this did not prove out to be as accurate as other features. The recall rate was not high since many females did not wear pink or purple. This is also a female only feature with an accuracy of 66.7%.

**Major Challenges:**

Our first task was to find the alignment of the body. Finding the exact alignment of every image turned out to be a tough task. With the low resolution and different objects in the background blurring with the human, it was hard to get a distinguishing line between them. Often the hands were hidden behind a person’s back or there would be multiple people in the picture. Sometimes it was hard to tell if the person was facing backwards or forwards. Even telling the gender for our own eyes was a guess at best at times. In the end, we realized that we did not need the whole body for our purposes so we left it with a basic alignment output which was satisfactory for our purposes.

When we started the project we had lots of ideas. Once we started coding, they started failing us and we were left with a handful of them. We first thought of calculating the length of hair but regardless of our countless different approaches, it was futile. We tried using edge detection, tracing it from the top of the head downwards and many other similar approaches including color density of hair color. Since all the papers we read gender identification but they worked with pictures having a high resolution facial view, we did not find any help there.

All the ideas we used in estimating the gender were the results of careful and long hours spent in analyzing the data. With our own features we had several problems. With every feature we had to make a decision between precision and recall. Except for the height distinguisher for males, nowhere else have we made a major comprise with the feature accuracy. The height distinguisher for males has an accuracy of only 54.6% to ensure that every picture has at least one identifier.
We had problems with finding out the face sometimes when the females wearing low clothes tended to lower the values of our face position estimate due to the excessive facial color present around the neck region. Funnily enough this is also how we came up with the neck ratio. Similarly the leg ratio tends to get skewed when in bright sunlight, most of the objects present in the background tend to shine and give off the same color as the face. Discrepancies like this were present in every feature and these erroneous results tended to skew our accuracy readings. The biggest challenge of this project was to keep trying without giving up hope, no matter how hopeless the situation looked.

Further Research:

Although not perfect, a lot of information regarding gender stereotypes and nuances can be obtained from an image by analyzing simple features such as those described here. There are a number of features that we either tried to implement unsuccessfully or just could not figure out a way to do so.

A classification function we were going to implement was to look at a circle around the hands to see if the person is carrying a bag which increases the chance of the person being a female. Our alignment did not accurately calculate hands and so this feature led to bad results. Another way we thought to test for bags was to look along the shoulders and see if there is a strip of color that is different than the shirt. If it is only on one side of the shirt then it’s more than likely the person is carrying a bag and there would be a higher chance that it would be classified as a female. We wrote code to try this out but did not manage to perfect it. The results were not improving our classification accuracy by much so we commented it out. If this algorithm was to be improved and made more accurate, our success rate would certainly rise.

The walk (gait) of the human can be used as a potential feature but was not possible to implement without a very effective use of videos and frames. We used cut out pictures like we mentioned before and did not use preceding frame data. If information from the video could be extracted to give the speed of the walk of an individual, it could help in classifying the gender. This would probably not give us the most accurate data since a lot of females walk fast and males tread along leisurely and vice versa.

If the distance to the person was known, then figuring out the person’s actual height would have helped. We have found a height ratio in our algorithm but that was an approximation made with the thought that a taller person would occupy more pixels longitudinally relative to the pixel height of the image. We had a function that could take the distance of the person and calculate the actual height of the individual. This would have certainly improved our success rate since we know that the average height of males is more than the average of females.

Body shape (using maybe canny edge detection) and structure would probably give the best indication of the gender. It would even help in identifying pictures where the back of the individual is seen. Though canny edge detection is easy, extracting information automatically from it was something that we could not do. Also, the edge detection was not always perfect and would gather noise.
An interesting aspect of such a project will be the responses of inanimate objects (fire hydrants, mailboxes, etc.) that get mistakenly identified as humans. One of our aims was to classify them as non-gender and thus, put them in an inanimate pool which we have not yet taken care of. Thus, our factor variable should remain at 0 for such an object or there should be a check that looks at alignment and eliminated these false positives that might come up. Since we didn’t run our code on any such pictures nor did we provide for such an error, we don’t know what our output would be. This is certainly something to look into.

Moreover, this data set was created using images in summer time only. For winter, a different algorithm would have to be created which uses some of these features and others that arise from winter clothing. Also, this analysis might not be generic enough to be used in varied geographical locations such as the eastern countries of the world since a lot of implicit knowledge regarding western cultures is used. Places where dressing styles and temperature are vastly different, different set of classification features could be used.

Conclusion:

The goal of this project was to classify gender and we achieved a ~64% accuracy rate. Though this is not a very accurate ratio compared to a probable 50-50 guess, it still suggests that a non-conventional approach to classify gender can be successful. Using simple information from the surrounding environment as classifiers, we can categorize humans into his/her gender fairly accurately. Given high resolution pictures, this would be much easier and more accurate. The project shows that a video taken from an ordinary camera can also undergo such analysis and bring about desirable results.

Resources Used:

Computers: We have mostly used our own laptops. We also used the Huntsman Hall computers for working on the non-coding aspects of our report.

Software Tools: MatLab.

Hardware Tools: Videos were taken from a car fitted with four cameras; two cameras aimed straight ahead while one was pointed towards the back and the last one towards the side. The two cameras pointed straight had 3D information. This was the hardware used to give us our input though we were not involved in the collection process. We also used data from only one camera rather than both.

References: The papers we read have been mentioned above. Once we went through those papers, we realized that our attempt was in a completely different manner than those mentioned in those papers. We tried to search for relevant papers but could not find much. We used a lot of what we learnt from the Computer Vision class and its slides in writing and implementing our code.