(44)

A Pattern Recognition System for Recognizing Gender from Silhouettes

by

Yuying Chen

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science in Computer Science and Electrical Engineering

and Master of Engineering in Computer Science and Electrical Engineering

at the Massachusetts Institute of Technology

May 26, 1998

Author .....
Department of Electrical Engineering and Computer Science
May 22, 1998

Certified by .....
Áaron F. Bobick
Assistant Professor Of Computational Vision
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

A Pattern Recognition System for Recognizing Gender from Silhouettes

by

Yuying Chen

Submitted to the
Department of Electrical Engineering and Computer Science

May 22, 1998

In Partial Fulfillment of the Requirements for the Degrees of
Bachelor of Science in Computer Science and Electrical Engineering
and Master of Engineering in Computer Science and Electrical Engineering

Abstract

We developed a system that allows the computer to decide on the gender of the person. This system takes as input black and white silhouette images of the full body and looks for 'ideal' images, images where the gender indications are the clearest. From those images, gender-discriminating features are calculated and compared against preset threshold values to decide the gender of the subject. Images are classified as male, female or unsure based on the feature values. The system performs with 90% accuracy for identification of both males and females.

Thesis Supervisor: Aaron F. Bobick
Title: Assistant Professor Of Computational Vision

Acknowledgments

First and foremost, I would like to thank Jim Davis. His generous help and advice were essential to the success of this project. Thank you very much Jim. Next I would like to thank my advisor, Professor Aaron Bobick, for believing in me enough to take me on and guiding me through this project in a field that I am not familiar with. Everyone who helped me out by volunteering to be one of the subjects for the training and testing data: Amanda Gruhl, Andrew Wright, Andy Wheeler, Chris Bentzel, Erwin Lau, Jim Davis, Josh Weaver, Haixia Lin, Kaiyuh Hsiao, Kate Mahoney, Kathy Paur, Lajos Molnar, Mark Herschberg, Owen Johnson, Paul Covell, Stephen Intille; thank you very much. Without you, I would have no data to analyze. I want to thank Fernando Padilla(the master of latex) for teaching me Latex in one day and Lajos Molnar for answering all my random questions. Lastly, I want to thank Han B. Chou for believing in me and hanging in there with me till the end, without whom, I would have never worked on Athena for so long.

Contents

Chapter 1

Introduction/Motivation

The majority of research done in personal identification has been concentrated in face recognition. Systems based on the Karhunen-Loeve feature extraction method have obtained recognition rates as high as 99% on the ARPA FERET database of nearly 2000 facial photographs.[10] One system uses eigen spaces as a compact representation of the face, then searches through the database to find the closest match. This process parallels our own recognition process of photographs. In life however, we see people from far away and can recognize them before their facial features are even distinguishable. How does this type of recognition occur? This recognition is done through the identification of physical features in the body. For this project, we attempt one step in enabling the computer to do the same recognition, identifying the person's gender.

1.1 Why Identify Gender?

We chose gender because it is a major classifier of people. With one decision, 50% of the population can be eliminated from the scope of identification. Also, many sociological, cultural, psychological, behavioral and economic traits divide along the gender lines. Once the gender identity is known, a great deal of other information can be deduced. To facilitate accurate identification of gender specific characteristics, we are using black and white binary silhouette images as input to the system.

1.2 Why Silhouette?

What is gender but a physiological difference, a different composition/construction of the body? Because it is a visible physical difference, we need the input image to be as clear and precise as possible with regard to the body shape and dimensions. Silhouettes can provide us with this clarity and precision. Being black and white, they remove the noise that clothing and lighting variations can create and indicate the shape of the body (with clothing) in its purest form. Shape from shading might be beneficial for face recognition, but the body is clothed and the shadows are not direct results of the body. We use a binary silhouette image as opposed to grey scale for the same reasoning. The grey scale does not provide much additional information about the body shape. Therefore, we simplify everything to black or white and minimize ambiguity for the decision process.

Chapter 2

Previous Work

As simplistic as the identification of gender may sound to a person trying to make the classification, it is difficult to know how to begin to teach a computer the same feat without some guidance. In related areas, there has been work done in gender identification from faces and perception research as to exactly what we perceive as gender-specific in the way people walk. Unfortunately, there has been very little work done in this specific area of gender identification from full body images. The most similar project is the non-published work done in a Japanese laboratory that attempted identification of gender using heat-vision cameras.

2.1 Face Recognition Extensions

The most prolific work done in gender recognition have been extensions on face recognition systems such as Sexnet[6] and EMPATH[3]. The main research in a face recognition system involves finding a representation that is compact yet detailed enough to reconstruct the face from. The number of parameters range anywhere from 40-100. A fully automated face recognition system reads in an image, normalize and reposition it based on the locations of the main facial features (i.e. eyes, nose, mouth), extract the parameter values for that particular representation, and compare these values with those of faces in the system to find the match. For these systems to do gender identification, they simply train the system to produce a 1 for male and a

0 for female subjects from the available parameter values. For the testing set, the system will then identify any image with a value > 0.5 to be male and any image with a value < 0.5 to be female. System identification rates all range from 97%-99% for theses systems. Unfortunately, because none of the projects disclose exactly what parameters distinguish the genders from each other, these systems do not help us in identifying gender discriminating features.

2.2 Perception Research

The second project was a part of some perception research done in 1970's by James Cutting[4] to try to quantify how people perceived repetitive motion, among which was walking. In one of the experiments he conducted, Cutting affixed lights on the joints of his subjects (a moving light display), then asked observers to describe what they saw as subjects walked around a dark field. Observers familiar with the display mechanism could recognize specific people as well as the motion; even those unfamiliar with the display or the experiment could identify the movement and the gender of the person. This final result intrigued Cutting greatly, so he and his colleagues set out to find exactly what characteristics indicated the gender of the walking person. After many false starts with walking speed, step size, and arm swing which indicated gender but not necessarily, they began a biomechanical study of a person. With this, they discovered that roughly 75% of the variance in gender judgments were due to the fact that the locus of the centers of movement were different for males and females. The center of movement was best approximated by intersecting diagonals that connected the shoulder and hip joints. The male center of movement tends to be lower than that of the female. Because of this difference, there were also other systematic differences in the way men and women walked: "females swing their arms more, males their shoulders more; women rotate their hips more and walk more smoothly across a surface; males, in contrast, tend to bob up and down more when they walk." With these results, Cutting was able to simulate computer characters walking with gender specific characteristics by simply adjusting the center of movement of the animated

characters. Based on these results, we began feature extraction with calculations of center of gravity in the silhouette images. This is based on the theory that the difference in the center of movement caused by differences in shoulder and hip width and positioning would also cause gender discriminating differences in the center of gravity.

2.3 Heat-vision Recognition

Lastly, the most similar project was a system developed in Japan that identified gender through using heat vision cameras[7]. The heat vision camera outputs a heat-print of everything in its field of view. After extracting the person from the image, the heat distribution over different areas of the body was analyzed for gender differences. From this, they discovered that the heat distribution in the chest area differed in men and women: a woman's heat signature in the torso contained two circular areas of heat that did not exist in a man's heat signature. This fact may or may not be verified by human physiology, but this result is unreliable because of a circumstantial dependence: clothing. Disregarding the body temperature and heat distribution of the person, thicker cloth retains more heat inside the clothing than thinner fabrics. For this project, all test subjects were personnel of the lab. This fact biases thickness of clothing along gender lines simply because it is a technology lab in Japan. It turns out, the majority of the men subjects were research staff of the lab, whose dress code was formal, consisting mostly of suits and lab coats. The majority of the female subjects however, were support staff of the lab, secretaries and such. Their business attire usually consisted of a blouse and skirt. A blouse is made of much thinner material than suits or lab coats. This predisposes the women subjects to have a heat signature of larger areas in the torso than men subjects without actual heat distribution difference. Therefore, the result of this project does not indicate a physical difference along gender lines. Rather, because of this cultural/social factor, the camera was basically identifying the difference in clothing fabric thickness. Because of this, the only thing valuable we learned for was: be careful that the gender

discriminating features choosen are not biased by non-physiological factors.

Chapter 3

Project Specifications

3.1 Proposed Problem

The problem we are trying to solve is the identification of a person's gender from full body silhouettes. Because of the time constraints on this project, some assumptions/constraints must be made about the input to the overall system. The input to the computer is RGB images from a video camera. We assume there is only one or less person in the field of view at any time. Without this limit, we would need to solve the problem of extracting individuals from a crowd image. We also constrain the movements of the person to movements such as walking, standing and turning around, movements where the body is upright and contain no large movements of the limbs. Since we are interested only in the construction of the body for this project, we are trying to eliminate major distortions of the body shape through large movements. For the most common place application of this system (i.e. access control, monitoring), walking, standing, turning around encompass the majority of the movements observed, thereby making these constraints plausible. Non-restrained movements would also require the system to perform pose estimation and outline labeling. Finally, we assume that the full body is visible in the image and the camera placement to be about hip level to the person. Placing the camera at about hip level gives us the flattest image of the person and minimizes distortion caused by the subject not being parallel to the camera lens. Also assuming that we see the full

body in the image insures that the calculations we do based on body dimensions are proportionally correct as well as eliminating the need for camera calibration. These constraints/assumptions were made to narrow the actual problem solved to specifically that of gender identification. Without them, we would need to solve other major problems in the field of computer vision such as people extraction, pose estimation, and camera calibration. These constitute the constraints on the input to the system. For the system itself, there is a design criteria of correctness. Because gender is a very sensitive issue, correctness when an identification is made is essential. We tolerate misses(not identifying someone's gender) but have no tolerance for false alarms(misidentifying someone's gender). Finally, as with all systems based on bodily dimensions, we do not expect optimal behavior from this system for people of non-average proportions.

3.2 Proposed Solution

Having narrowed down the problem as much as possible to simply that of gender identification, here is how I propose to solve the problem. The proposed solution to this problem is divided into 3 parts: silhouette data retrieval, feature extraction and differentiation. Silhouette data retrieval is done through background subtraction. An image of the background is captured as the reference image. For each new image, it is compared per pixel with the reference image and the difference gives us the black/white silhouette image through thresholding. Pixels with an absolute difference value greater than the threshold are set to white, and the rest is set to black. From these silhouette images, we can indicate by hand the images that give the clearest indication of gender. A range of possible features that will reflect inter-gender differences are calculated on these images alone and plotted versus gender to extract the most discriminating features. Knowing the specific range and thresholding values for these features, we can then code the correct differentiators into the system. (Because all of the features are extracted from 'ideal' images only, the system must also detect when 'ideal' images occur in the input stream of video.) Since we expect the

probabilistic densities of male and female to overlap, there are three classifications possible for each feature value: male, female or unsure. A selected image indicates gender only when all discriminating features indicate the same gender for this image. Otherwise, the image is classified as unsure.

Chapter 4

The Project

4.1 Obtaining Silhouette Images

This system takes full body silhouette images as input. So as a first step, we need to obtain clean silhouette images from the video data. Many methods were attempted before achieving the cleanest silhouette for the least amount of processing time. The key to our method of obtaining clear silhouettes was the physical setup of the project.

4.1.1 Physical Setup

This system was developed to run in a setup called SIDEshow: A Silhouette-based Interactive Dual-screen Environment created by James W. Davis[5]. The space contains two large video screens 10 feet apart, one in front of the user and one in back of the user. In front of the user, there is a 5x4 foot back-projection screen elevated 3 feet off the floor, and a black-and-white video camera attached to the bottom of it. A 10x8 foot back-projection screen is used as the back wall and a small carpeted area in the center restricts the subject's movements to 3 feet in front of the back screen. Behind the back screen, there are 6 invisible infrared light emitters (consumer 840nm emitters) backlighting it: 3 lighting the top half and 3 lighting the bottom half of the screen. By using an infrared-pass/visible-block filter tuned for this wavelength of infrared light on the video camera, we can retrieve clean and reliable silhouettes

without more cameras[11], extra processing time, or restrictions on clothing color of the subject[2]. The camera is restricted only to see this infrared light "since a person standing in front of the screen physically blocks the infrared light coming out of the screen, the video camera in front of the subject therefore sees a bright image(of the infrared light passing through the screen) with a black silhouette of the person(from where the person is blocking the infrared light to the camera)."(Figure 4-1)[5]

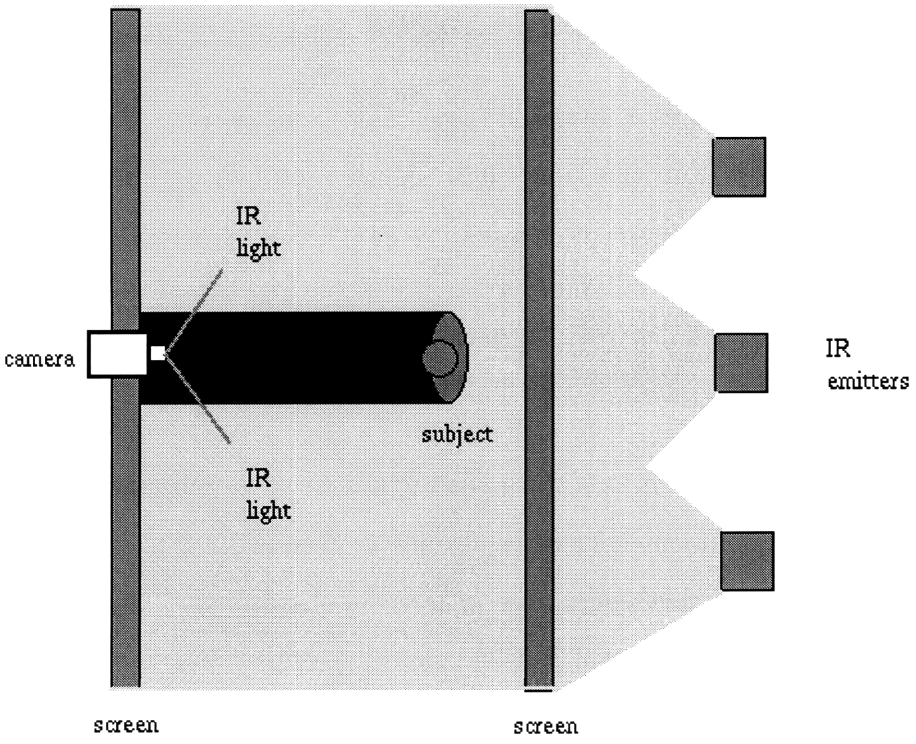

Figure 4-1: Conceptual drawing of blocking infrared light from the camera to generate a silhouette of the person.

4.1.2 Background Subtraction

With nearly ideal black/white silhouette images resulting from the infrared backlighting, the task of background subtraction becomes almost trivial. Before anyone steps into the camera's field of view, an image of the background is stored as the reference image(image resolution: 160x120). For each subsequent image, each pixel is compared to that of the reference image. If the difference between the two pixel values is greater than 50 out of 255, then the pixel is set to white. Else, the pixel is set to black. This results in a white silhouette on a black background image with the white blobs indicating the new elements of change. To filter out noise caused by lighting variations, after the pixel by pixel comparison, any blobs of area less than 500 pixels is erased from the image. Then, the code eliminates all but the largest blob in the image bigger than the minimum size as a result of the one person or less simplification. The final resulting image is input to the gender identification system.

4.2 Dataset

4.2.1 Subjects

The training and testing dataset used to determine the discriminating features were images of 12 men and 6 women from a video taken on two separate days in the SIDEshow setup.

First Day's Subjects

The first set of video data was taken December 17, 1997, Wednesday 4pm-6pm with the SIDEshow setup in the cube(real name) space. 8 males and 1 female were videotaped on the first day: M1(5'10, 180lbs), M2(6', 160lbs), M3(6'1, 175lbs), M4(5'11, 160lbs), M5(6', 180lbs), M6(5'8, 140lbs), M7(5'11, 250lbs), M8(5'7, 150lbs), F1(5'5, 140lbs). Each subject was videotaped doing the following sequence of movements: The subject had to walk in from screen left to the center of the screen. Once in the center, he turns slowly for a full circle. After completing the circle, he faces the cam-

era with both feet together and raises his arms with hands open to above his head and then back down again. This motion is repeated with the feet shoulder width apart. Still facing the camera, he then bends over to touch his toes with straight legs and with arms outstretched. After returning to the upright position, the toe touch motion is repeated facing sideways to the camera, usually facing the left of the screen. Still facing sideways after the toe touch, he then crouches as low as balance and flexibility allows. This motion is repeated with the subject facing the camera. From here, he simply walks off screen left to retrieve a chair that is used for the following movements. The chair is placed center screen facing screen left. When the chair is properly placed, the subject stands sideways to the camera in front of the chair with his back to the chair. He then sits down and gets up twice in this orientation before turning around to face the chair. From this point on, the specified movements simply conclude with the subject picking up the chair with both hands and walking off screen left.

Second Day's Subjects

The second set of video data was taken December 18, 1997 Thursday 5pm-6pm in the same setup. On this day, 4 males and 5 females were videotaped: M9(6'1, 180lbs), M10(5'7, 135lbs), M11(6'2, 195lbs), M12(5'4, 120lbs), F2(5'5, 140lbs), F3(5'7, 120lbs), F4(5'4, 115lbs), F5(5'8, 130lbs), F6(5'10, 130lbs). They were asked to do a similar set of movements:

*note: The subjects include personnel of the MIT Media Lab and the author's friends. All but two are undergraduates at MIT. All are under age 30.

4.2.2 Obtaining the Images for Training Data

For Gender Identification

The images that were judged to give the clearest indications of gender were extracted by hand from the above video data for the gender identification training dataset. The background subtraction code produced silhouette images while the input video

of training data subjects played. A script wrote to file the PPMs(portable pixmap file format) of the silhouette images whenever indicated by a key stroke. Using this process, I extracted 113 images for the training data set: 50 female images and 63 male images. From my point of view, the silhouette images that most clearly indicated gender were side profiles of the subject(Figure 4-2). This is mainly a result of a woman's breasts making the most visible difference attributed to gender in profile. The human eye can also detect shape differences in the overall body shape, but the differences were very difficult to quantify.

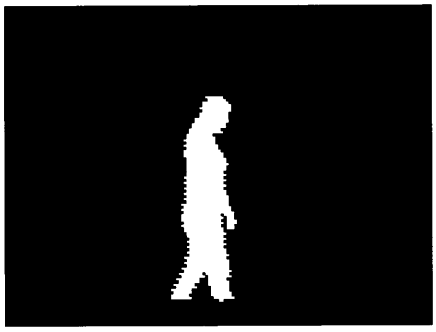

Figure 4-2: Sample training data image for gender identification.

For Profile Detection

Since the gender identification training set consisted of side profile images, it was apparent that the system must be able to detect when the body is facing sideways from the camera. For the training dataset for profile detection, we needed profile images and non-profile images. For the profile images, we simply reused the gender identification training data set. For the non-profile images, a script outputed every tenth frame of the video to file. From this set of images, I selected by hand 180 frames of non-profile images(10 per subject). In the selection of the non-profile images, I favored frontal images(Figure 4-3) because the difficulty in profile detection lay in distinguishing between the orientations of the body. Other variations in the body positioning (ex. when the body is not upright) produce a lot more variation in the shape and can be more easily detected.

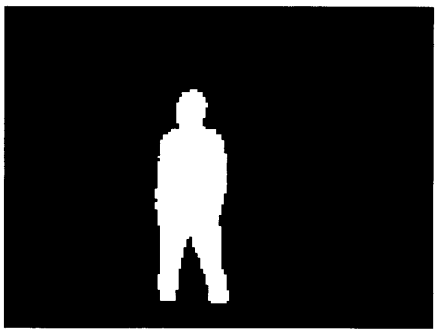

Figure 4-3: Sample training data non-side-profile image for profile detection.

4.2.3 Testing Data

For profile detection as well as overall system performance, the testing data set included all of the above generated frames, a total of 5806 frames, consisting of every 10th frame grabbed from the video.

For gender identification alone, the testing data consisted of all the side-profile images in the profile detection testing set. The profile images were classified by hand for a total of 1475 frames with 539 female frames and 936 male frames.

4.3 Feature Extraction

For both gender identification and profile detection, we followed the same procedure to find the discriminating features.

4.3.1 Features Calculated

The features we calculated on the silhouette images were inspired by Cutting's perception research results. His research had indicated that 75% of the gender difference perceived were contributed to the difference in center of movement[Cut78]. Cutting approximates the center of movement by cross connecting the shoulder and hip joint positions. We could not accurately track shoulder and hip joint positions in our video

data. Therefore, we tried to imitate his result by testing the theory that the difference in center of movement is also visible in the moments of the body shape by calculating and analyzing the following features: average value, median value, Hu moments (all 7 orders), spatial moments, and centroids on the silhouette images. Average_y, median_y and ycentroid were normalized by height to remove the height dependence of the features.

F(j, k) =the discrete image function that represents all white pixels in the silhouette image;

J=width of image function;

K=height of image function;

Image is (0,0) at top left corner;

Height

$$Height = y_1 - y_k = K;$$

Average Value, normalized

$$average_x = \sum_{j=1}^{J} \sum_{k=1}^{K} x_j / TotalPixels;$$

$$average_y = \sum_{j=1}^{J} \sum_{k=1}^{K} y_k / TotalPixels;$$

$$normalized\_average_y = \frac{average_y}{height};$$

Median Value, normalized

Sort all pixels by x, MED_x is the x of the median pixel coordinate.

Sort all pixels by y, MED_y is the y of the median pixel coordinate.

$$Norm\_MED_y = \frac{MED_y}{height};$$

Hu Moments

$$V(m,n) = \frac{U_u(m,n)}{[M(0,0)]^{\alpha}};$$

$$M(0,0) = M_u(0,0);$$

$$\alpha = \frac{m+n}{2} + 1;$$

$$h_1 = V(2,0) + V(0,2);$$

$$h_2 = [V(2,0) - V(0,2)]^2 + 4[V(0,3) - 3V(2,1)]^2;$$

$$h_3 = [V(3,0) - 3V(1,2)]^2 + [V(0,3) - 3V(2,1)]^2;$$

$$h_4 = [V(3,0) + V(1,2)]^2 + [V(0,3) - V(2,1)]^2;$$

$$\begin{aligned} h_5 = {} & [V(3,0) - 3V(1,2)][V(3,0) + V(1,2)] \\ & \{[V(3,0) + V(1,2)]^2 - 3[V(0,3) + V(2,1)]^2\} \\ & +[3V(2,1) - V(0,3)][V(0,3) + V(2,1)] \\ & 3[V(3,0) + V(1,2)^2 - [V(0,3) + V(2,1)]^2; \end{aligned}$$

$$\begin{aligned} h_6 = {} & [V(2,0) - V(0,2)]\{V(3,0) + V(1,2)^2 \\ & -[V(0,3) + V(2,1)]^2\} + 4V(1,1) \\ & [V(3,0) + V(1,2)][V(0,3) + V(1,2)]; \end{aligned}$$

$$\begin{aligned} h_7 = {} & [3V(2,1) - V(0,3)][V(3,0) + V(1,2)] \\ & \{[V(3,0) + V(1,2)]^2 - 3[V(0,3) + V(2,1)]^2\} \\ & +[3V(1,2) - V(3,0)][V(0,3) + V(2,1)] \\ & \{3[V(3,0) + V(1,2)]^2 - [V(0,3) + V(2,1)]^2\}; \end{aligned}$$

Moments

M_u (m, n): the (m,n)th unscaled spatial moment;

$$M_u(m,n) = \sum_{j=1}^{J}\sum_{k=1}^{K}(xj)^m(y_n)^n F(j,k);$$

M(m,n): the (m,n)th scaled spatial moment;

$$M(m,n) = \frac{M_u(m,n)}{J^m K^n};$$

Spatial Central Moments

U_u: unscaled spatial central moment

$$U_u(m,n) = \sum_{j=1}^{J}\sum_{k=1}^{K}[x_k - x_k]^m[y_j - y_j]^n F(j,k);$$

$$x_k = \frac{M_u(1,0)}{M_u(0,0)};$$

$$y_k = \frac{M_u(0,1)}{M_u(0,0)};$$

Centroids

$$xcentroid = \frac{M(1,0)}{M(0,0)} = \frac{M_u(1,0)}{M_u(0,0)J};$$

$$ycentroid = \frac{M(0,1)}{M(0,0)} = \frac{M_u(0,1)}{M_u(0,0)K};$$

We also attempted similar calculations on the 1-D profile signal alone.

Profile center of gravity(COG) and Moments

$$Profile(k) = \sum_{j=1}^{J} F(j,k);$$

Profile(k): is a 1D signal of the silhouette image, where Profile(k) is total number of pixels in row k;

$$M_u(n) = \sum_{k=1}^{K} y_k^n Profile(k);$$

$$COG = \frac{M_u(1)}{M_u(0)K};$$

In order to detect the side profile image, we also calculated area, maximum width, average width, and circularity:

Area(A_o)

$$area = \sum_{j=1}^{J} \sum_{k=1}^{K} F(j,k);$$

Average Width(avgw)

$$avgw = \frac{\sum_{k=1}^{K} J_k}{K};$$

Maximum Width(maxw) = J;

Perimeter(P_o)

$$P_o = \sum_{k=1}^{K} F(1,k) + F(J,k)_{J \neq 1};$$

Circularity(C_o)

$$C_o = \frac{4\pi A_o}{P_o^2};$$

$$A_o = area;$$

$$P_o = perimeter;$$

4.3.2 Discriminating Features

On both training datasets, we calculated area, average width(avgw), circularity(circ), perimeter(peri), height(ht), all seven order Hu moments(hu1, hu2, hu3, hu4, hu5, hu6, hu7), maximum width(maxw), the 1st cross row column moment($M_u(1,1)$), average value normalized by height(navg), median value normalized by height(nmed), the first cross row column normalized spatial central moment(U(1,1)), center of gravity of the 1D profile signal normalized by height(nprfcog), 1st moment of the 1D profile signal(prfmom1), the 1st cross row column spatial central moment(scm11), xcentroid(xcd) and ycentroid(ycd) for all the images. Each feature was plotted by itself and analyzed for differentiation between side-profile image vs. non-side-profile image or male vs. female silhouette. For side-profile detection, we also calculated aspect ratios ht/maxw(maxasp), ht/avgw(avgasp), ht/maxw/area(nmaxasp) and ht/avgw/area(navgasp).

Features for Profile Detection

From these features we discovered that circ, hu1, hu2 and the various aspect ratios differentiated between side-profile and non-side-profile images. Because the feature values for the two classes of images overlapped, there was no clear choice for a threshold value. Possible threshold values were selected for maximized correct identifications for each increase in error tolerance for each feature. Because we were only concerned with discriminating between 2 classes of images, only 1 threshold value was needed. Every image with a feature value greater than the threshold is identified as a side-profile image. Figures 4-1 to 4-7 indicate the values for each of the discriminating features and possible threshold values:

For the aspect ratio feature, because there were so many options for how the value is calculated, all the indicated threshold values could not be tested. We wanted to test the performance of thresholds at different error levels, so we chose the one condition per each false alarm level that had the maximum hits across possible threshold values of all four methods of calculation. Table 4.1 lists those conditions that maximized

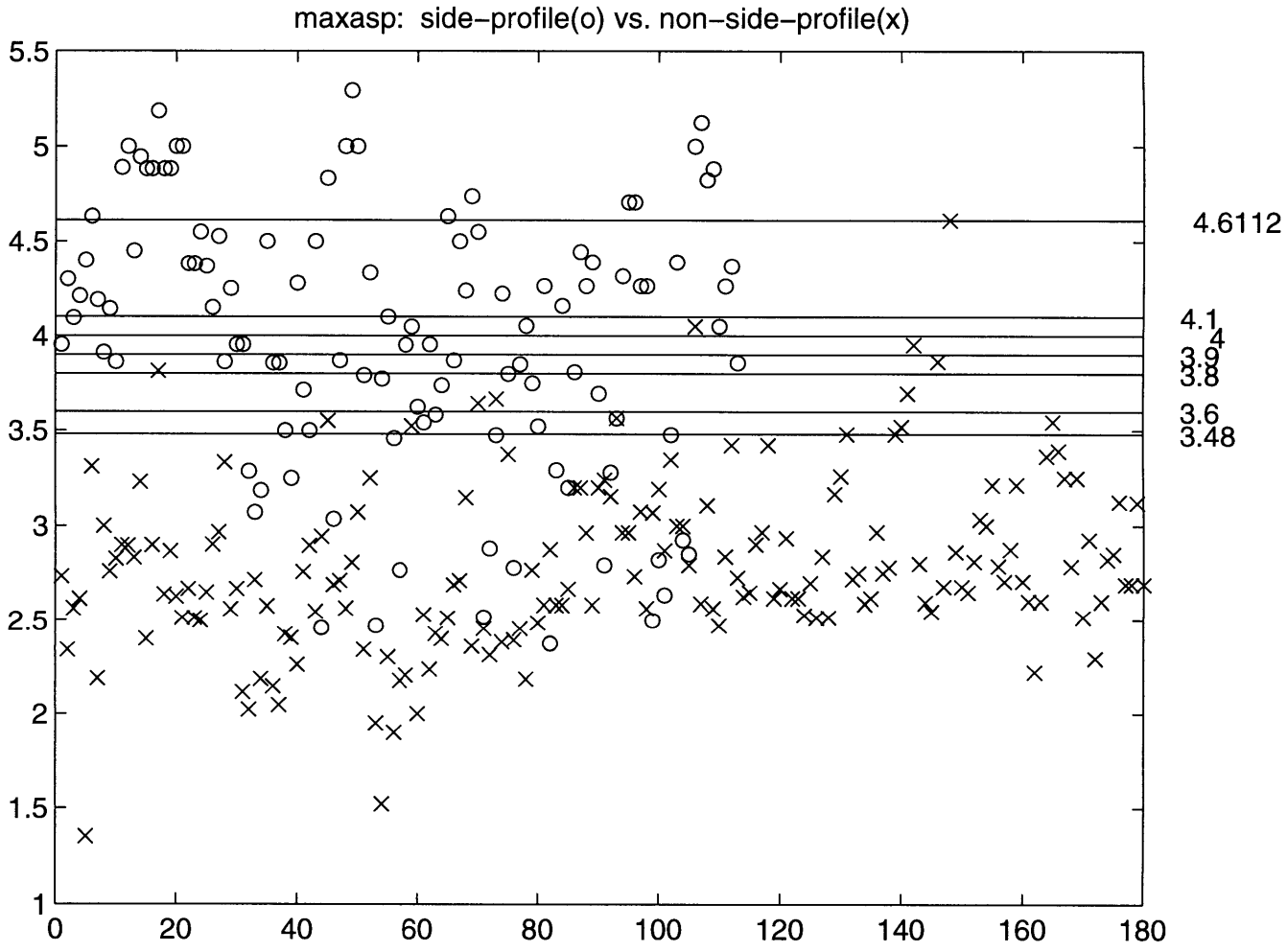

Figure 4-4: Height/maxw aspect ratio values and possible threshold values

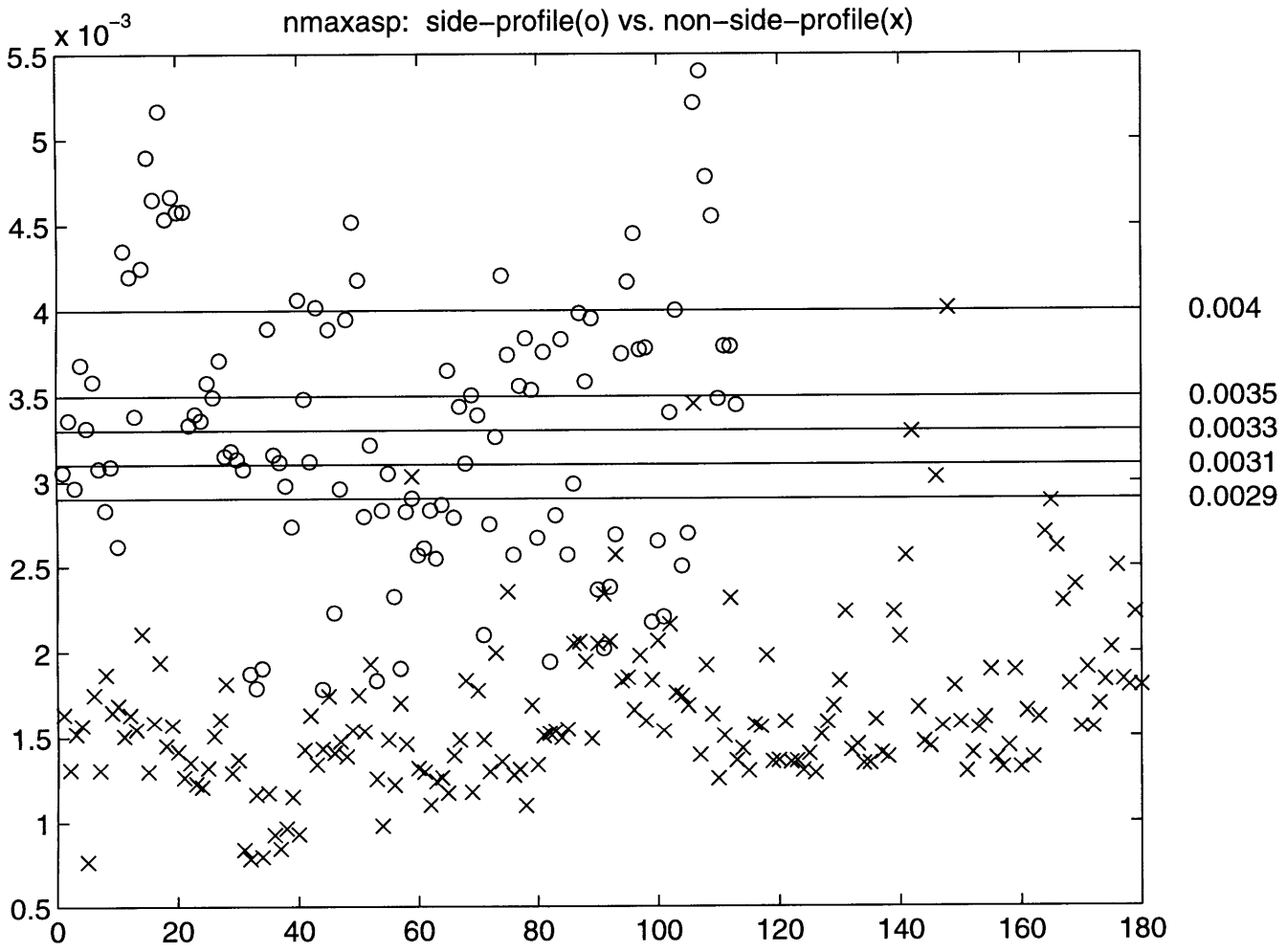

Figure 4-5: Height/maxw/area aspect ratio values and possible threshold values

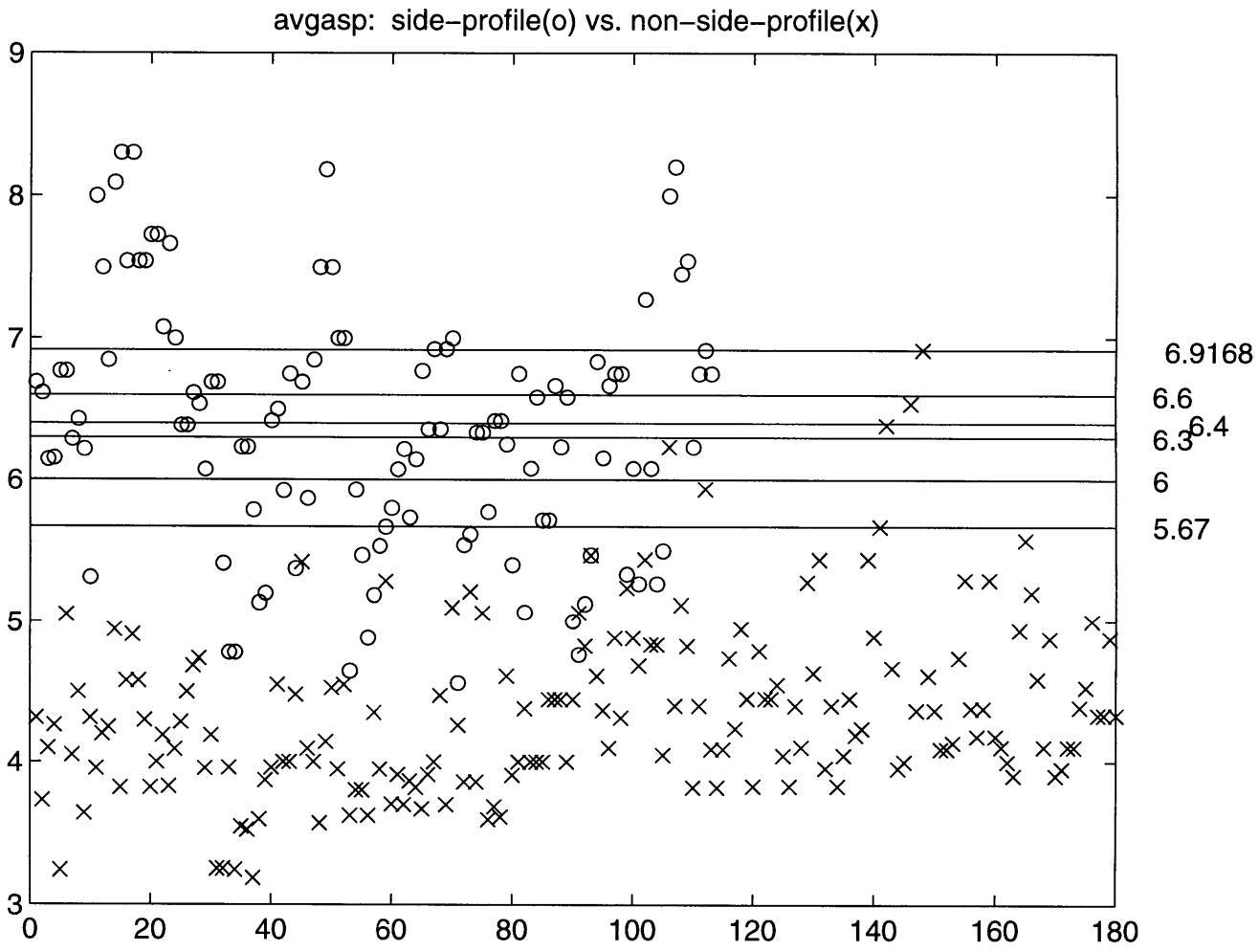

Figure 4-6: Height/avgw aspect ratio values and possible threshold values

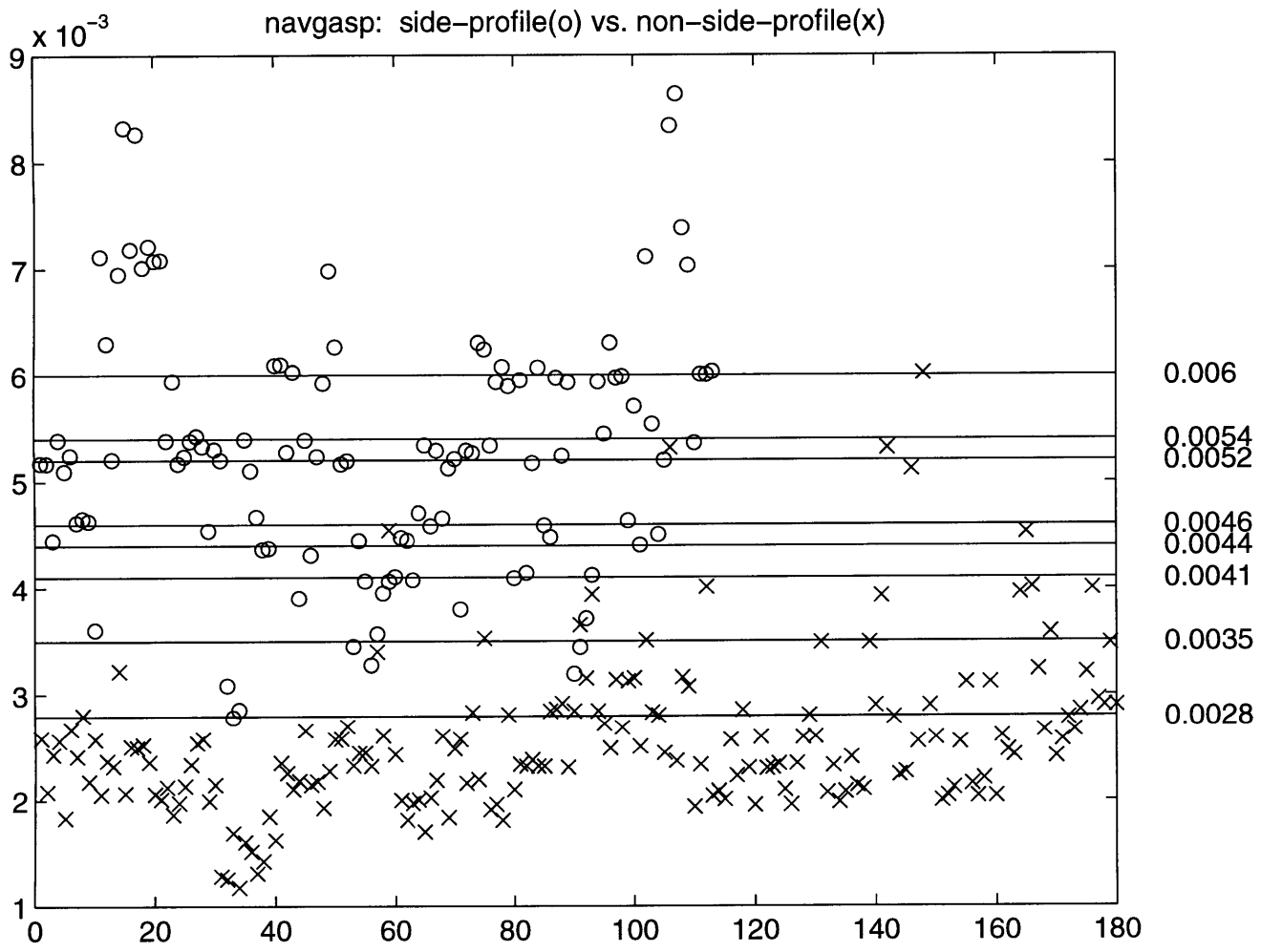

Figure 4-7: Height/avgw/area aspect ratio values and possible threshold values

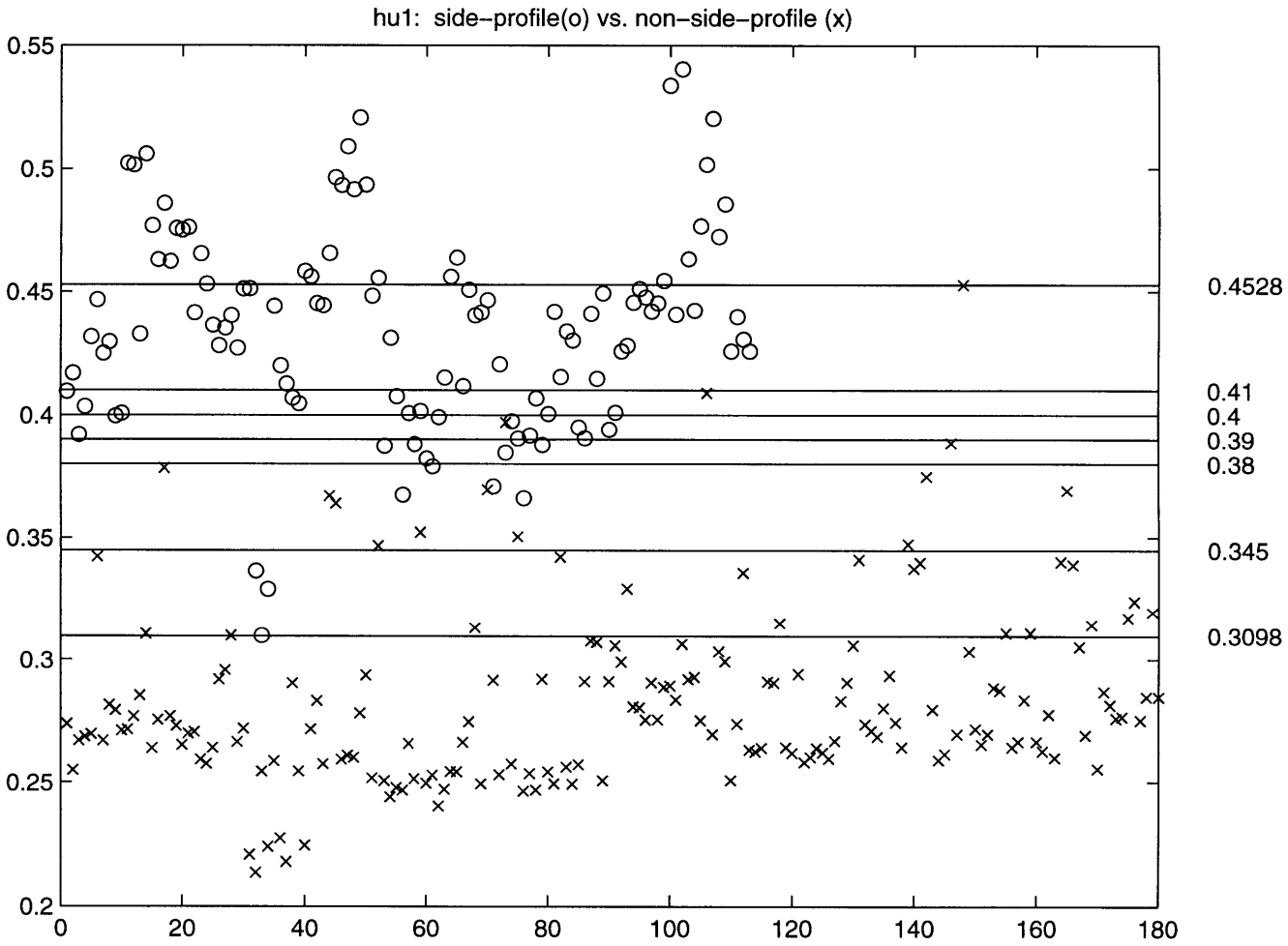

Figure 4-8: 1st order Hu moment values and possible threshold values

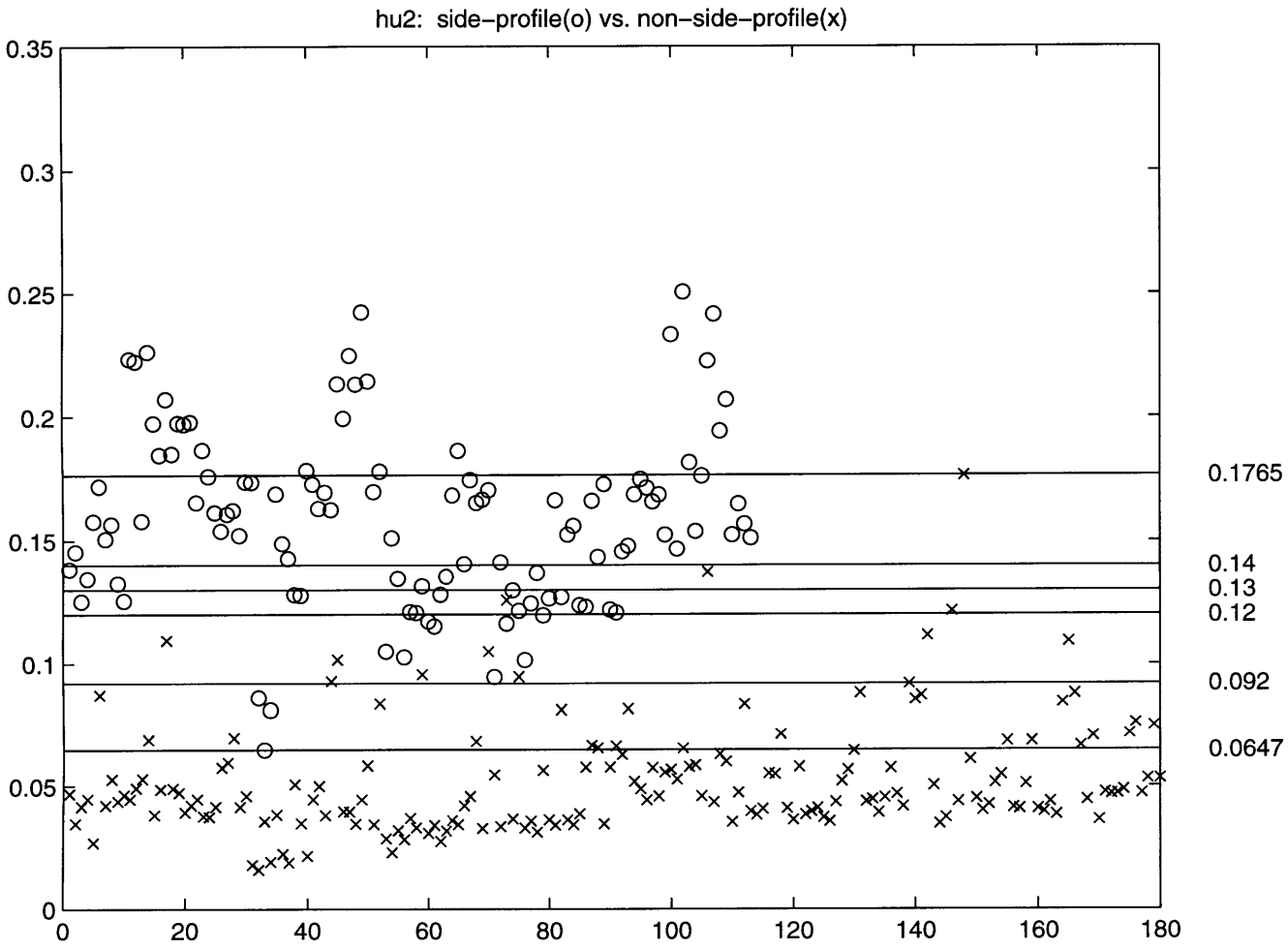

Figure 4-9: 2nd order Hu moment values and possible threshold values

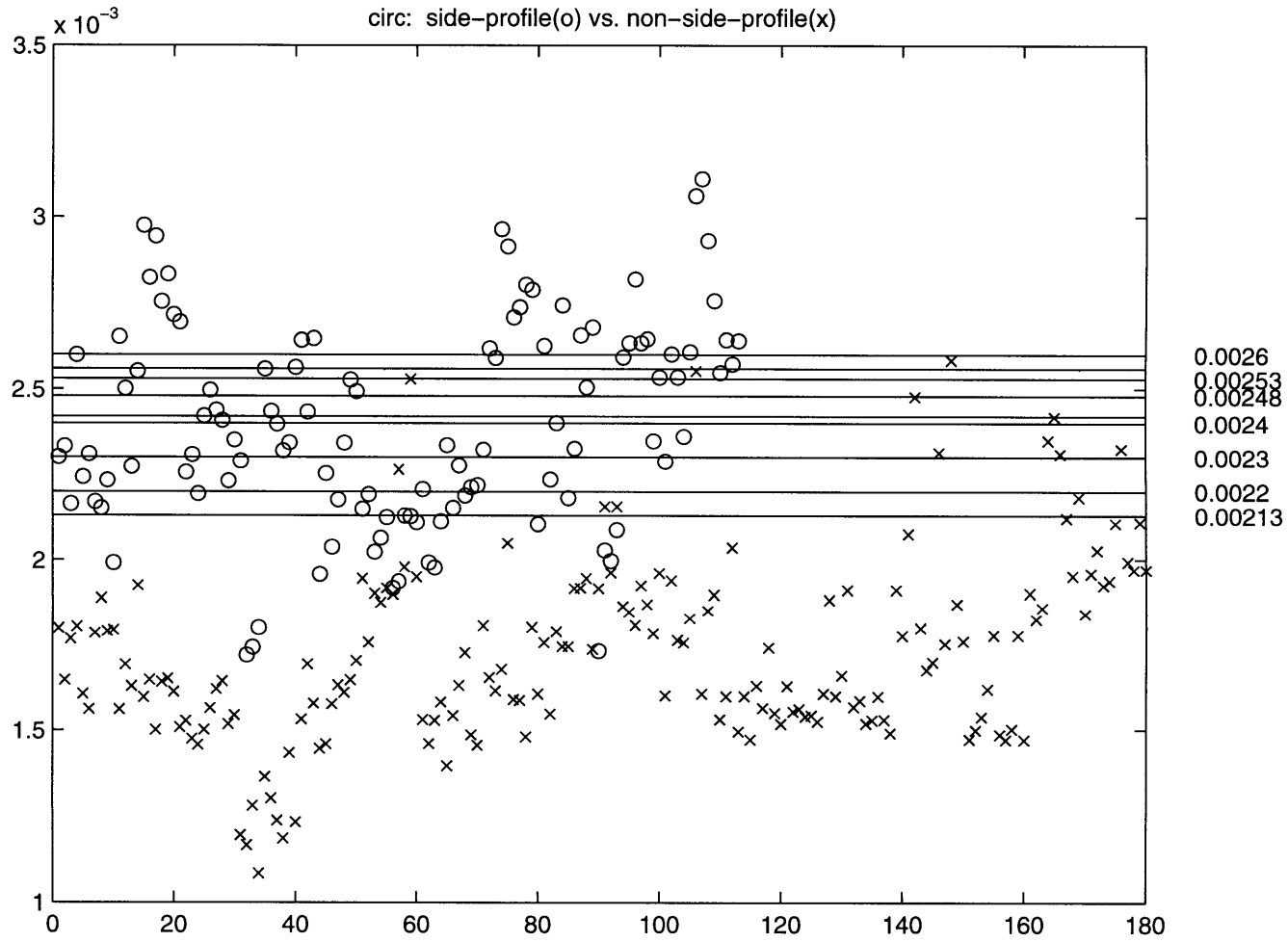

Figure 4-10: Circularity values and possible threshold values

the hits for each increasing false alarm level and the next best condition.

FalseAlarms	Max Hits	Feature	Threshold	Next Best (Hit/Condition)
0	28	navgasp	0.006	26/avgasp@4.6112
1	55	maxasp	4.1	47/avgasp@6.6
2	60	maxasp	4	58/nmaxasp@0.0033
3	67	nmaxasp	0.0031	66/maxasp@3.9
4	80	navgasp	0.0046	78/avgasp@6
5	87	avgasp	5.67	77/nmaxasp@0.0029
>=6	96	navgasp	0.0041	90/navgasp@0.0044

Table 4.1: Aspect ratio threshold values sorted by number of false alarms.

The rest of the aspect ratio threshold values tested were chosen for even distribution in the range of values. Some values for circ, hu1, and hu2 were also eliminated for even distribution and to lessen the length of each test run. Table 4.2 shows a complete list of individual profile-detection conditions that were tested on the test dataset.

Feature	Threshold	Feature	Threshold
maxasp	4.1	maxasp	4
avgasp	5.67	nmaxasp	0.0031
navgasp	0.006	navgasp	0.0046
navgasp	0.0041	navgasp	0.0028
navgasp	0.0035		
hu1	0.4528	hu1	0.41
hu1	0.38	hu1	0.345
hu1	0.3098		
hu2	0.1765	hu2	0.14
hu2	0.12	hu2	0.092
hu2	0.0647		
circ	0.0026	circ	0.0024
circ	0.0023	circ	0.0022
circ	0.00213		

Table 4.2: Profile detection conditions that were tested on the test dataset

Features for Gender Discrimination

For gender identification, we did not find any of the calculated features on the full silhouette images to discriminate between genders. Following the theory that the

most visible body difference as a result of gender is in the torso area, we recalculated all the above features for only the top half of the silhouette (from top of the head to half of the silhouette height). Then we discovered that halfhu3(3rd order Hu moment of the half silhouette), halfhu4(4th order Hu moment of the half silhouette), halfhu5(5th order Hu moment of the half silhouette), halfhu6(6th order Hu moment of the half silhouette), halfnmedy(y value of normalized median of half silhouette), and halfnprfcog(center of gravity of profile signature normalized of half silhouette) differentiated between male and female images. Because we are separating the images into Male(M), Female(F), and Unsure(U) to keep the error rate at a minimum, we needed 2 threshold values per feature. The 5 threshold values closest to the majority of male values decides all values greater than it to be male; the 5 threshold values closest to the majority of female values decides all values smaller than it to be female. What values not indicated as M or F by the two thresholds are classified as U. Figures 4-8 to 4-14 are training data values for each feature with the threshold values for M and F indicated and table 4.3 is a complete list of individual gender identification conditions that tested on the test dataset.

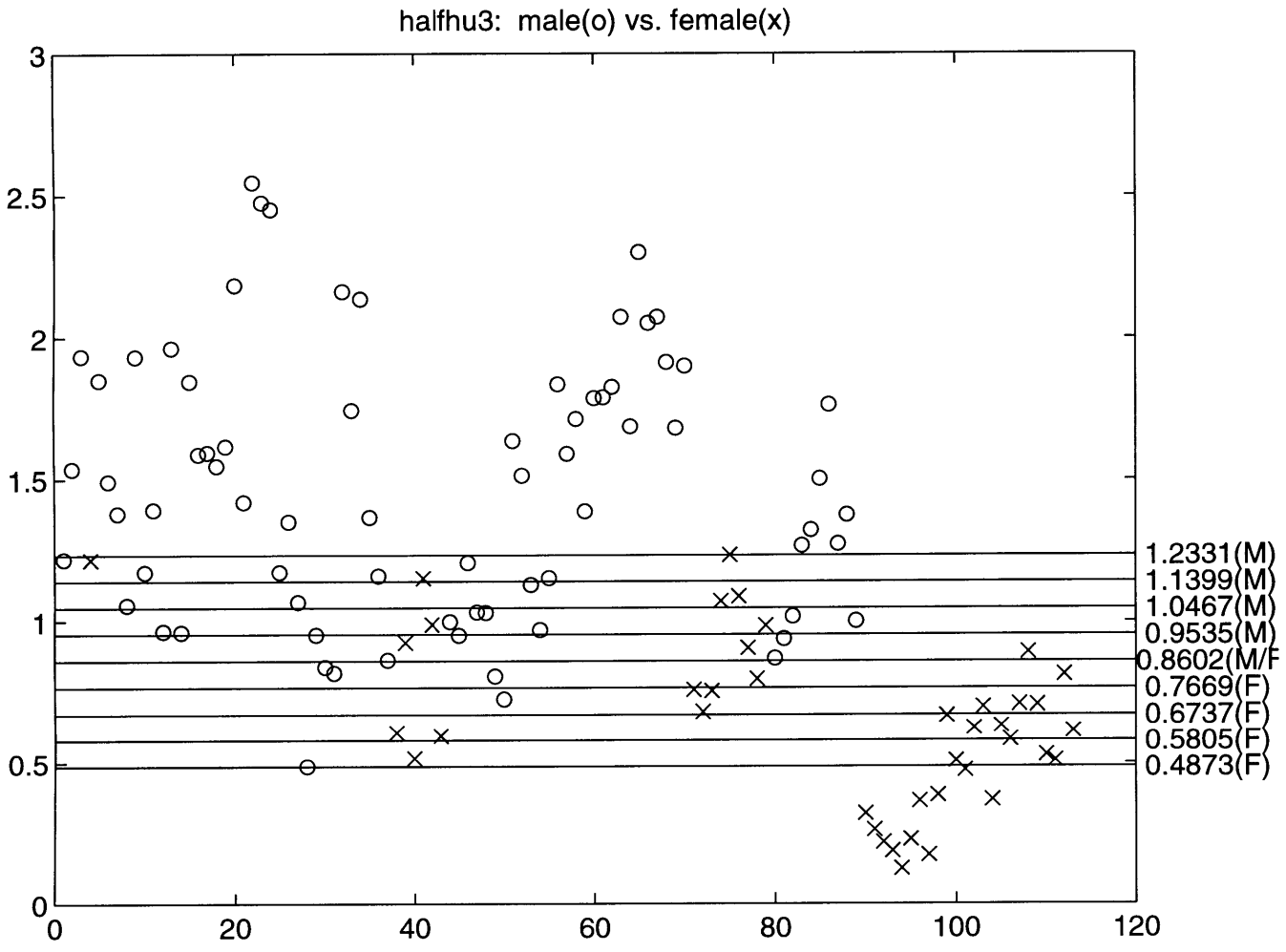

Figure 4-11: 3rd order Hu moment values and possible threshold values

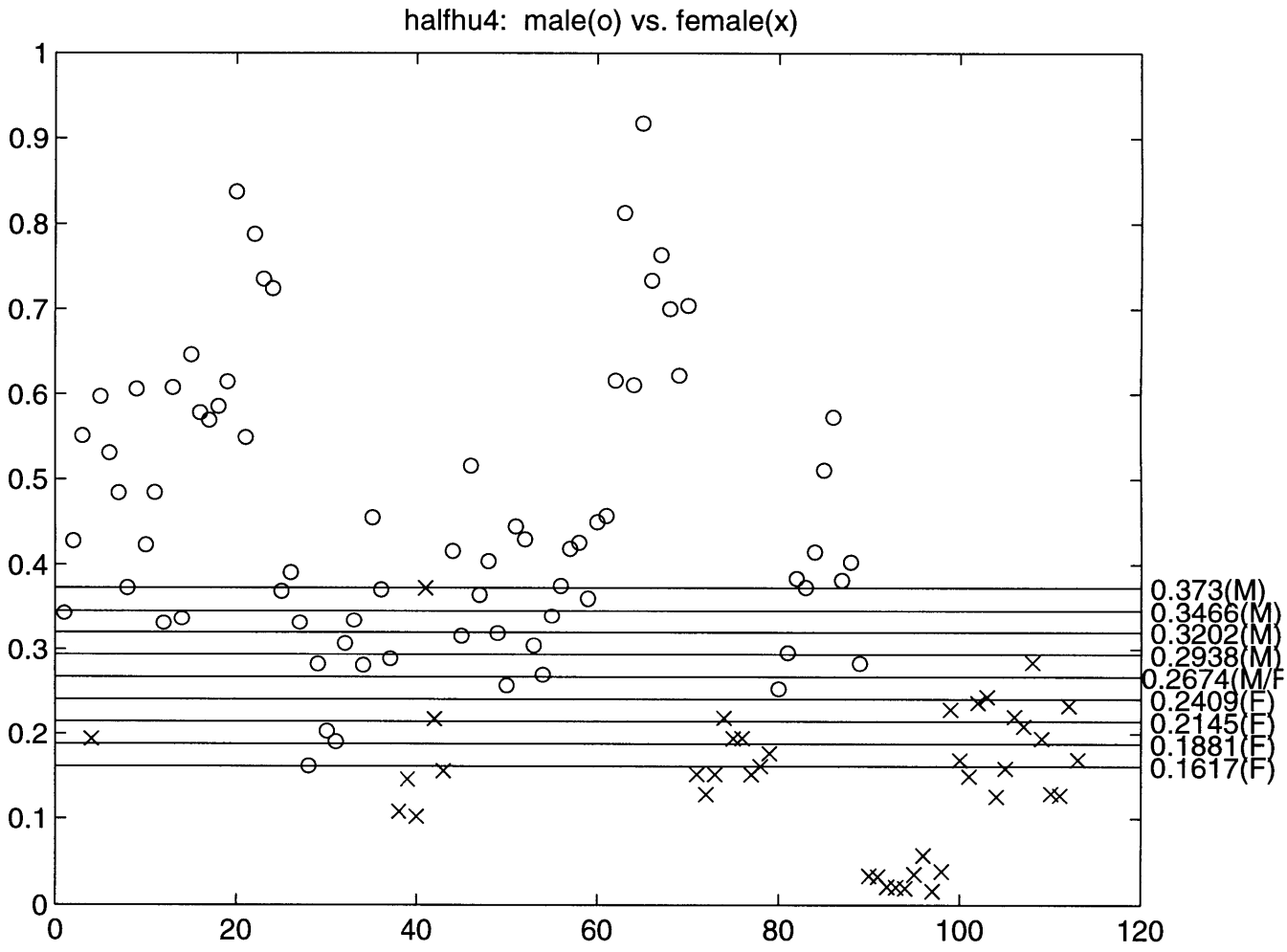

Figure 4-12: 4th order Hu moment values and possible threshold values

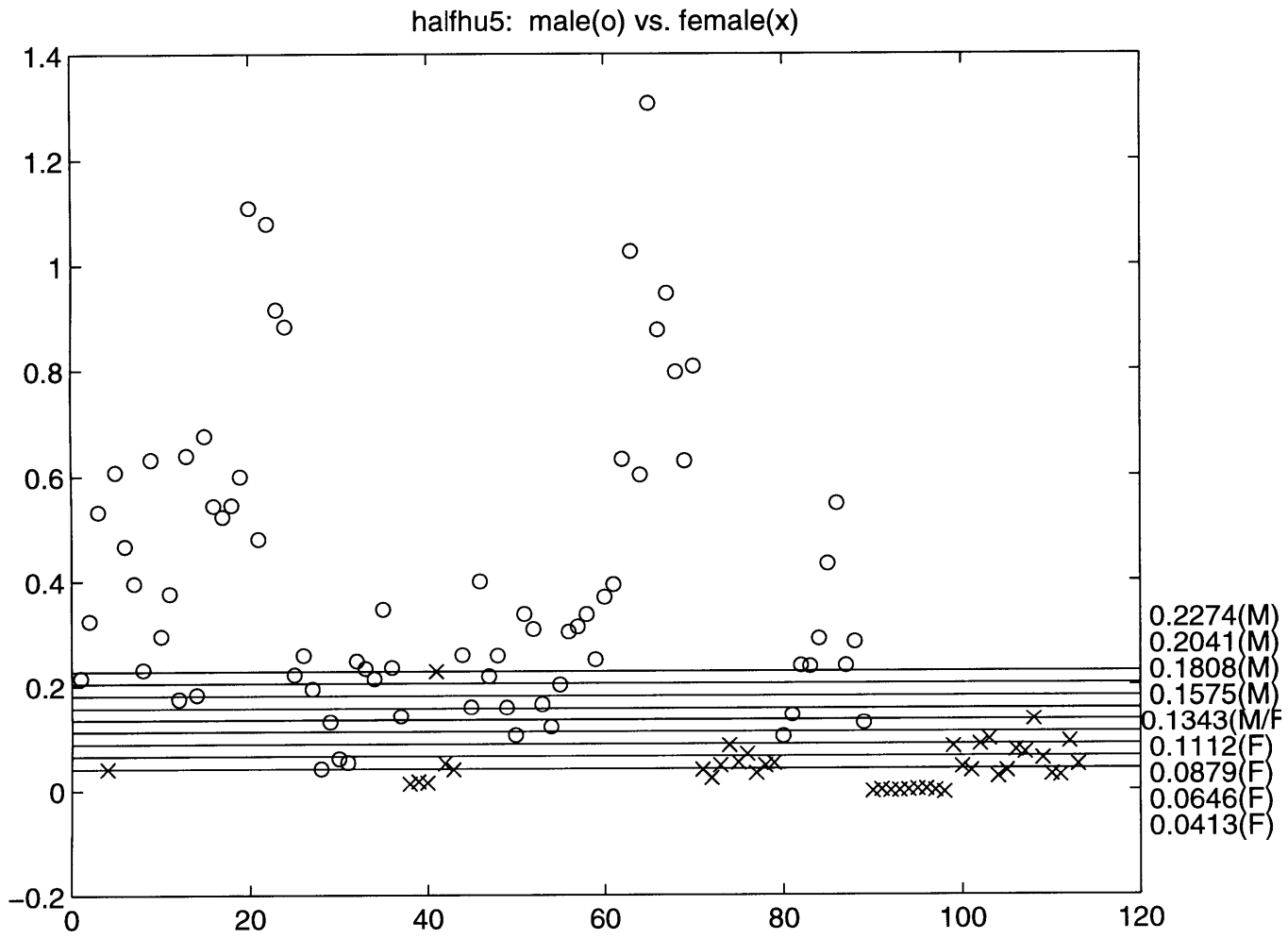

Figure 4-13: 5th order Hu moment values and possible threshold values

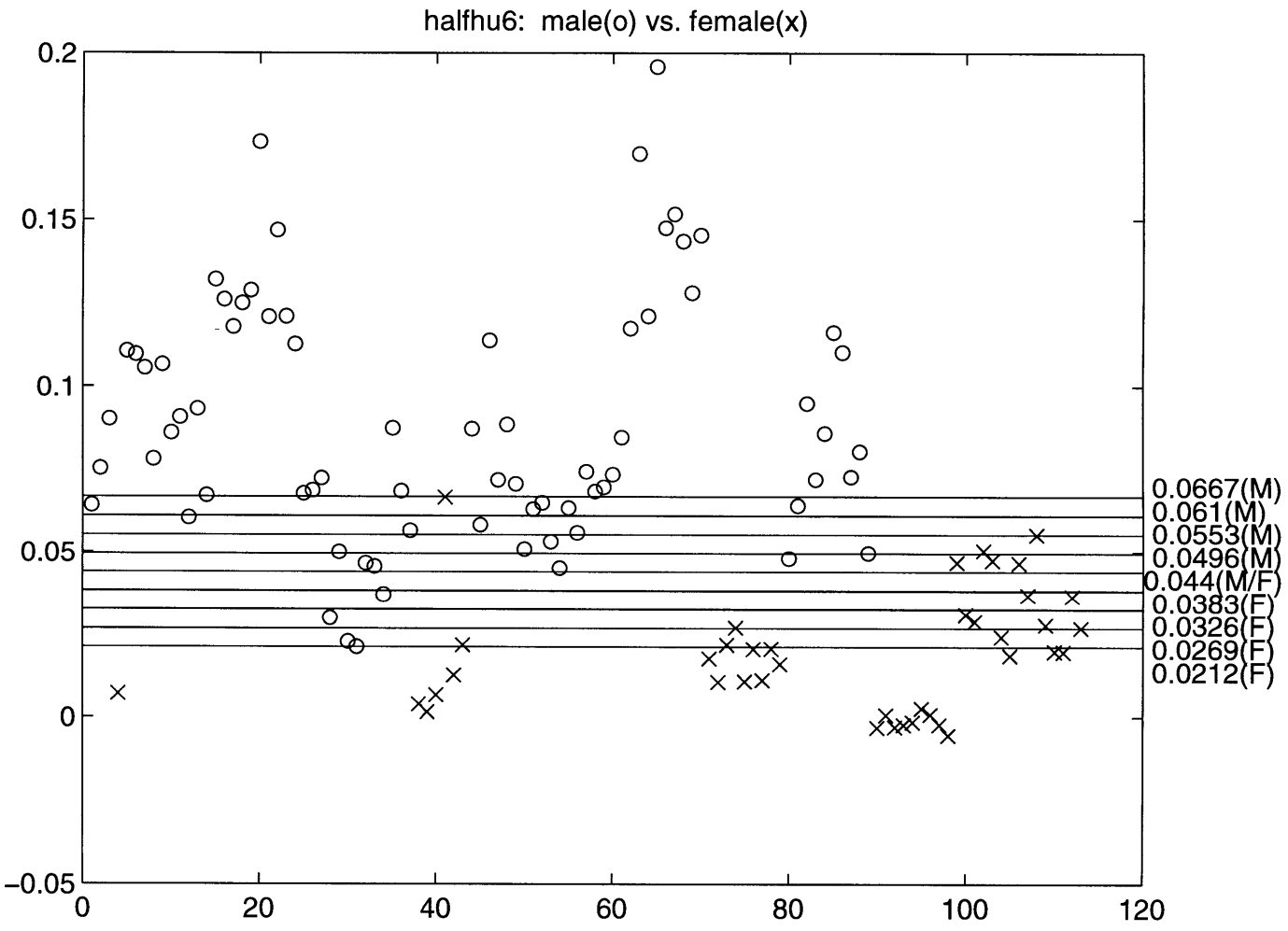

Figure 4-14: 6th order Hu moment values and possible threshold values

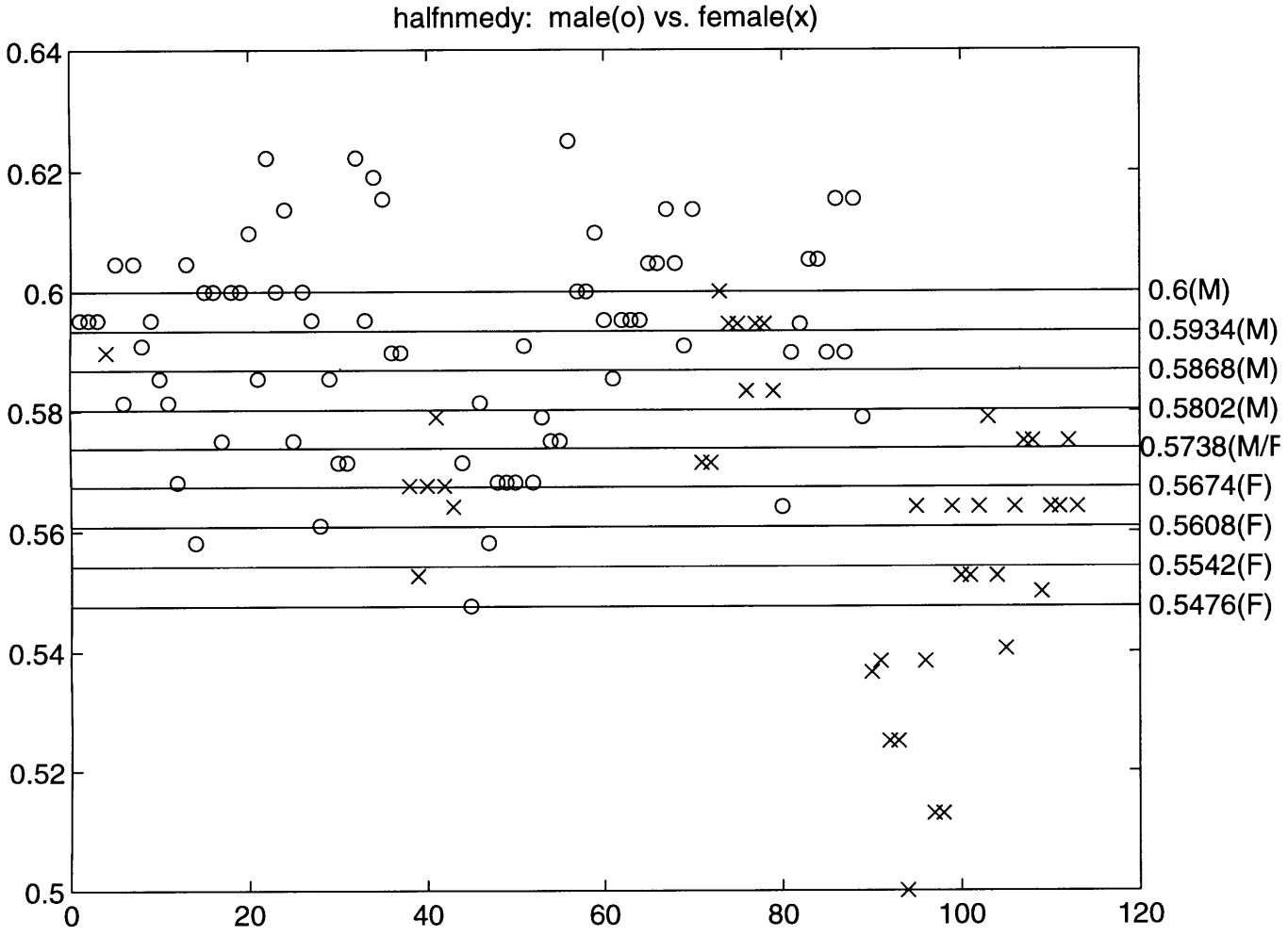

Figure 4-15: normalized y median values and possible threshold values

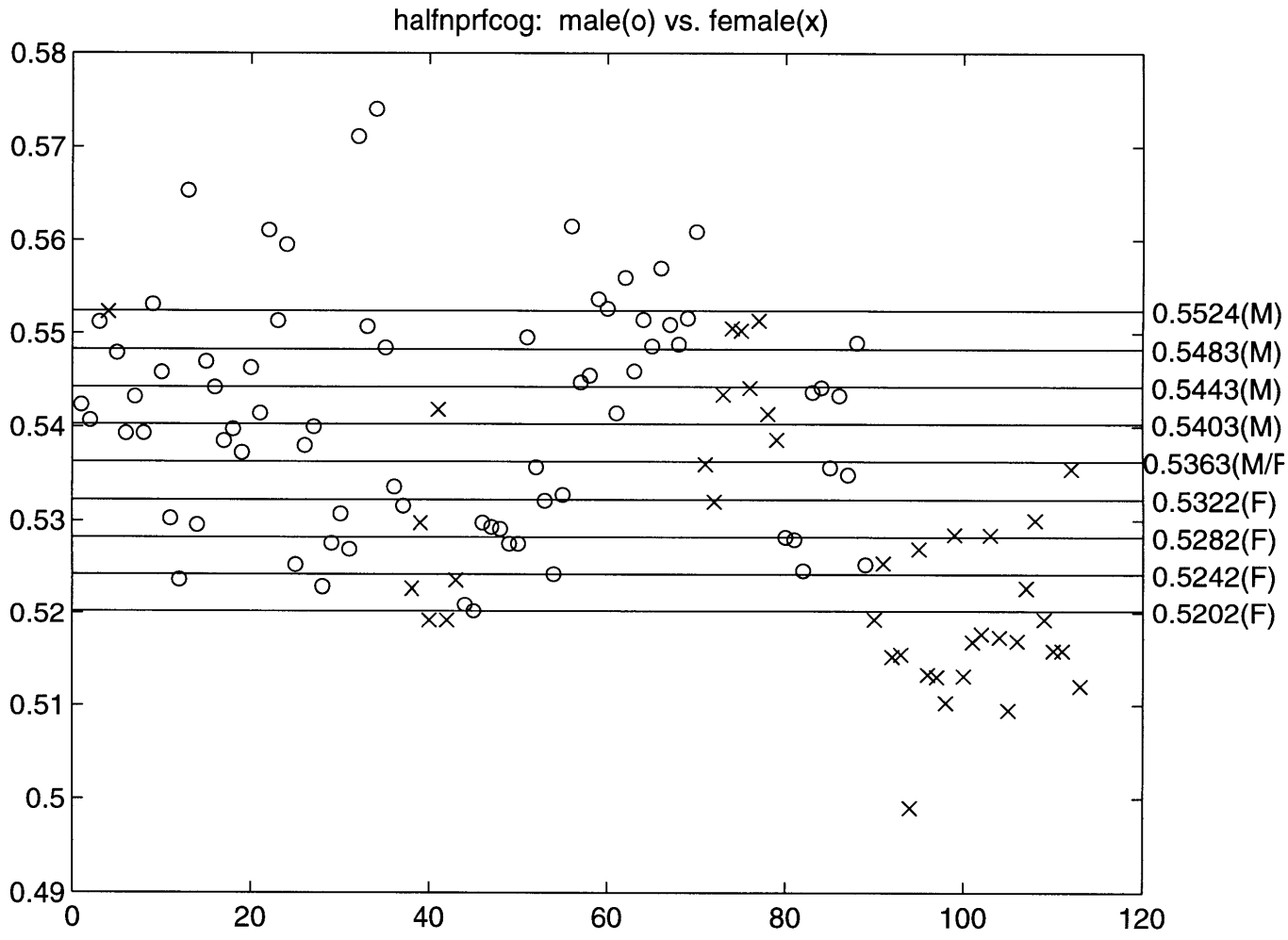

Figure 4-16: normalized center of gravity values and possible threshold values

Feature	Threshold	Feature	Threshold
halfhu3 (M>)	1.2331	halfhu3(F<)	0.4873
halfhu3 (M>)	1.1399	halfhu3(F<)	0.5805
halfhu3 (M>)	1.0467	halfhu3(F<)	0.6737
halfhu3 (M>)	0.9535	halfhu3(F<)	0.7669
halfhu3 (M>)	0.8602	halfhu3(F<)	0.8602
halfhu4 (M>)	0.373	halfhu4(F<)	0.1617
halfhu4 (M>)	0.3466	halfhu4(F<)	0.1881
halfhu4 (M>)	0.3202	halfhu4(F<)	0.2145
halfhu4 (M>)	0.2938	halfhu4(F<)	0.2409
halfhu4 (M>)	0.2674	halfhu4(F<)	0.2674
halfhu5 (M>)	0.2274	halfhu5(F<)	0.0413
halfhu5 (M>)	0.2041	halfhu5(F<)	0.0646
halfhu5 (M>)	0.1808	halfhu5(F<)	0.0879
halfhu5 (M>)	0.1575	halfhu5(F<)	0.1112
halfhu5 (M>)	0.1343	halfhu5(F<)	0.1343
halfhu6 (M>)	0.667	halfhu6(F<)	0.0212
halfhu6 (M>)	0.061	halfhu6(F<)	0.0269
halfhu6 (M>)	0.0553	halfhu6(F<)	0.0326
halfhu6 (M>)	0.0496	halfhu6(F<)	0.0383
halfhu6 (M>)	0.044	halfhu6(F<)	0.044
halfnmedy (M>)	0.6	halfnmedy(F<)	0.5476
halfnmedy (M>)	0.5934	halfnmedy(F<)	0.5542
halfnmedy (M>)	0.5868	halfnmedy(F<)	0.5608
halfnmedy (M>)	0.5802	halfnmedy(F<)	0.5674
halfnmedy (M>)	0.5738	halfnmedy(F<)	0.5738
halfnprfcog (M>)	0.5524	halfnprfcog(F<)	0.5202
halfnprfcog (M>)	0.5483	halfnprfcog(F<)	0.5242
halfnprfcog (M>)	0.5443	halfnprfcog(F<)	0.5282
halfnprfcog (M>)	0.5403	halfnprfcog(F<)	0.5322
halfnprfcog (M>)	0.5363	halfnprfcog(F<)	0.5363

Table 4.3: Gender identification conditions that will be tested

Chapter 5

Experiments Run and Results

5.1 Experiments on Individual Features

For each of the conditions* listed in Table 4.2 and 4.3, an experiment was run on the appropriate test dataset to ascertain the performance of each selected threshold values for individual features.

For the profile detection conditions(Table 4.2), the system indicated whether or not the image's value for that particular feature was greater than the specified threshold value. If it was, the frame number was added to the list of frames identified as side-profile. This list is compared to the list of hand-picked profile images to find the hits(correct identification), misses(incorrect rejection), false alarms(incorrect identification). From these statistics, we calculate the correctness percentage(C) and detection percentage(D).

(*note: a condition is a feature with a specific threshold value. Later on, it is a combination of several different features at different threshold values.)

$$C = \frac{Hits}{Hits + FalseAlarms} = \frac{\#frames\ IDed\ correctly}{\#frames\ IDed\ as\ profile} * 100;$$

$$D = \frac{Hits}{Hits + Misses} = \frac{\#frames\ IDed\ correctly}{total\#profile\ frames} * 100;$$

In order to be consistent in our evaluation of the performance of each condition,

we also calculate a weighted sum of C and D as overall evaluation. C is weighted more than D here because gender-discrimination makes decisions based on the side profiler's results: it is more important for the side-profiler to be correct in its decision.

$$Overall = \frac{3}{4} * C + \frac{D}{4};$$

Tables 5.1-5.4 lists results for individual condition's performance on the test dataset as indicated by C, D and Overall. The best overall performing conditions for each feature are noted with an * in the tables. The correctness percentages(C) range from 52.34% to 90.10% while the detection percentages(D) range all the way from 6.33% to 99.18% with Overall performance ranging from 45.34 to 83.05. At this point, the only clear result is that circularity as a feature, performs a lot worse than the other 3 features of hu1, hu2 and aspect ratio in differentiating between side-profile and non-side-profile images.

Feature	Threshold	C	D	Overall
maxasp (>)	4.1	71.54%	6.33%	55.24
maxasp (>)	4	71.54%	6.33%	55.24
avgasp (>)	5.67	90.10%	61.90%	83.05*
nmaxasp (>)	0.0031	80.37%	35.92%	69.26
navgasp (>)	0.006	61.07%	11.63%	48.71
navgasp (>)	0.0046	80.70%	50.07%	73.04
navgasp (>)	0.0041	82.50%	69.59%	79.27
navgasp (>)	0.0028	73.79%	95.78%	79.29
navgasp (>)	0.0035	81.23%	83.33%	81.76*

Table 5.1: Results for individual aspect ratio values

Feature	Threshold	C	D	Overall
hu1 (>)	0.4528	70.24%	20.07%	57.65
hu1 (>)	0.41	75.83%	57.41%	71.22
hu1 (>)	0.38	72.00%	82.59%	74.65*
hu1 (>)	0.345	66.30%	95.03%	73.48
hu1 (>)	0.3098	56.83%	99.12%	67.4

Table 5.2: Results for individual hu1 values

For each of the gender discrimination conditions listed in Table 4.3, the system indicated a frame to be male if its feature value was greater than that of the male

Feature	Threshold	C	D	Overall
hu2 (>)	0.1765	85.32%	18.98%	68.73
hu2 (>)	0.14	89.56%	54.29%	80.74
hu2 (>)	0.12	85.54%	75.24%	82.97*
hu2 (>)	0.092	77.89%	93.95%	81.9
hu2 (>)	0.0647	65.00%	99.18%	73.55

Table 5.3: Results for individual hu2 values

Feature	Threshold	C	D	Overall
circ (>)	0.0026	52.34%	24.35%	45.34
circ (>)	0.0024	57.81%	44.08%	54.38
circ (>)	0.0023	58.55%	54.97%	57.66
circ (>)	0.0022	59.72%	70.20%	62.34
circ (>)	0.00213	59.01%	78.44%	63.87*

Table 5.4: Results for individual circ values

threshold or female if it was less than that of the female threshold. The lists of identified male and female frames are compared to the known lists of male and female frames to again calculate the correctness percentage(C), the detection percentage(D) and Overall performance.

$$C = \frac{Hits}{Hits + FalseAlarms} = \frac{\#\ frames\ correctly\ IDed\ as\ (F/M)}{\#frames\ IDed};$$

$$D = \frac{Hits}{Hits + Misses} = \frac{\#frames\ correctly\ IDedas\ (F/M)}{total\#frames\ that\ are\ (F/M)};$$

$$Overall = \frac{3}{4} * C + \frac{D}{4};$$

No percentages of Unsure are calculated at this point because one, it adds no new information about the performance of the condition, and two, it requires both the male and the female threshold to calculate and the 2 threshold values are not yet paired up.

Tables 5.5-5.10 lists the C, D, and Overall for each gender- discrimination experiment. For these experiments, C range from 63.4% to 100%; D range from as low as 3.17% to as high as 96.04%. On average, female identification conditions have much lower detection percentages(D) and slightly higher correctness percentages(C). The

Overall performance range from 57.81 to 90.98, and attained better performance for male identifications than for female identifications. The best performing condition for each feature is indicated by *'s in the mentioned tables.

Feature	Threshold	R	C	Overall
halfhu3 (M>)	1.2331	90.62%	73.45%	86.33
halfhu3 (M>)	1.1399	88.43%	80.19%	86.37
halfhu3 (M>)	1.0467	87.17%	85.12%	86.66*
halfhu3 (M>)	0.9535	85.17%	90.36%	86.47
halfhu3 (M>)	0.8602	82.68%	92.51%	85.14
halfhu3 (F<)	0.8602	80.78%	54.10%	74.11*
halfhu3 (F<)	0.7669	84.06%	43.28%	73.87
halfhu3 (F<)	0.6737	83.71%	34.51%	71.41
halfhu3 (F<)	0.5805	85.62%	25.56%	70.61
halfhu3 (F<)	0.4873	90.00%	18.47%	72.12

Table 5.5: Results for individual halfhu3 values

Feature	Threshold	C	D	Overall
halfhu4 (M>)	0.373	93.76%	70.77%	88.01
halfhu4 (M>)	0.3466	92.76%	78.16%	89.11
halfhu4 (M>)	0.3202	91.99%	83.26%	89.8
halfhu4 (M>)	0.2938	91.05%	88.22%	90.34*
halfhu4 (M>)	0.2674	89.38%	91.01%	89.79
halfhu4 (F<)	0.2674	87.64%	71.46%	83.6*
halfhu4 (F<)	0.2409	87.64%	71.46%	83.7*
halfhu4 (F<)	0.2145	89.67%	61.57%	82.65
halfhu4 (F<)	0.1881	90.59%	48.51%	80.07
halfhu4 (F<)	0.1617	92.31%	35.82%	78.19

Table 5.6: Results for individual halfhu4 values

Now that we know how individual features perform, the best condition for each feature is chosen to represent that feature for the remaining experiments. Table 5.11 lists all of the best conditions for each feature of profile detection and gender discrimination.

Feature	Threshold	C	D	Overall
halfhu5 (M>)	0.2274	93.74%	73.77%	88.75
halfhu5 (M>)	0.2041	92.71%	80.30%	89.61
halfhu5 (M>)	0.1808	92.44%	83.83%	90.29
halfhu5 (M>)	0.1575	91.53%	87.90%	90.62*
halfhu5 (M>)	0.1343	90.32%	90.90%	90.47
halfhu5 (F<)	0.1343	83.59%	79.85%	82.66
halfhu5 (F<)	0.1112	85.44%	74.44%	82.69
halfhu5 (F<)	0.0879	88.53%	66.23%	82.96*
halfhu5 (F<)	0.0646	90.03%	52.24%	80.58
halfhu5 (F<)	0.0413	92.56%	37.13%	78.7

Table 5.7: Results for individual halfhu5 values

Feature	Threshold	C	D	Overall
halfhu6 (M>)	0.0667	93.86%	68.74%	87.58
halfhu6 (M>)	0.061	93.53%	75.80%	89.1
halfhu6 (M>)	0.0553	92.38%	81.80%	89.73
halfhu6 (M>)	0.0496	91.28%	88.54%	90.6*
halfhu6 (M>)	0.044	88.95%	91.33%	89.54
halfhu6 (F<)	0.044	83.97%	78.17%	82.52
halfhu6 (F<)	0.0383	87.09%	74.25%	83.88
halfhu6 (F<)	0.0326	90.34%	64.55%	83.89*
halfhu6 (F<)	0.0269	93.89%	54.48%	84.04
halfhu6 (F<)	0.0212	93.85%	42.72%	81.07

Table 5.8: Results for individual halfhu6 values

Feature	Threshold	C	D	Overall
halfnmedy (M>)	0.6	79.86%	37.37%	69.24
halfnmedy (M>)	0.5934	80.66%	49.57%	72.89
halfnmedy (M>)	0.5868	81.20%	62.42%	76.50
halfnmedy (M>)	0.5802	80.20%	77.62%	79.56*
halfnmedy (M>)	0.5738	76.11%	86.30%	78.66
halfnmedy (F<)	0.5738	63.40%	41.04%	57.81
halfnmedy (F<)	0.5674	75.00%	37.50%	65.63
halfnmedy (F<)	0.5608	71.75%	23.69%	59.74
halfnmedy (F<)	0.5542	90.23%	22.39%	73.27
halfnmedy (F<)	0.5476	94.44%	12.69%	74*

Table 5.9: Results for individual halfnmedy values

Feature	Threshold	C	D	Overall
halfnprfcog (M>)	0.5524	78.73%	66.17%	75.59
halfnprfcog (M>)	0.5483	77.86%	77.19%	77.69
halfnprfcog (M>)	0.5443	76.89%	85.12%	78.95
halfnprfcog (M>)	0.5403	75.35%	91.33%	79.35*
halfnprfcog (M>)	0.5363	73.58%	96.04%	79.20
halfnprfcog (F<)	0.5363	76.32%	21.64%	62.65
halfnprfcog (F<)	0.5322	91.03%	13.25%	71.59
halfnprfcog (F<)	0.5282	97.78%	8.21%	75.39
halfnprfcog (F<)	0.5242	100.00%	4.48%	76.12*
halfnprfcog (F<)	0.5202	100.00%	3.17%	75.79

Table 5.10: Results for individual halfnprfcog values

Feature(>)	Threshold	Feature	Threshold
avgasp(>)	5.67	navgasp	0.0035
hu1(>)	0.38	hu2	0.12
circ (>)	0.00213		
halfhu3 (M>)	1.0467	halfhu3 (F<)	0.8602
halfhu4 (M>)	0.2938	halfhu4 (F<)	0.2409
halfhu5 (M>)	0.1575	halfhu5 (F<)	0.0879
halfhu6 (M>)	0.0496	halfhu6 (F<)	0.269
halfnmedy (M>)	0.5802	halfnmedy (F<)	0.5476
halfnprfcog (M>)	0.5403	halfnprfcog (F<)	0.5242

Table 5.11: Conditions of best Thresholds for each feature

5.2 Experiments on Combinations of Features

For the rest of the experiments, we experimented with how different combinations of the features affect the performance of the system. First, we look at the performance of the profile detector and the gender discriminator independent of each other.

Table 5.12 lists conditions of features combined for better performance tested for profile detection and their results. The combinations were chosen based on the theory that ANDing the features would increase C. The overall performance for these conditions range from 0 to 87.24. The 3 best performing conditions were *hu*1&*hu*2&*seg*2, *navgasp*&*hu*1&*hu*2, and *navgasp*&*hu*1&*hu*2&*seg*2. (seg2 is a condition that automatically eliminates any image with a maximum number of segments per row exceeding two. This condition was added because the other conditions for profile detection were unable to distinguish between side-profile images and images where the subjected is facing the camera with both arms above his head in earlier experiments.)

Condition	C	D	Overall
avgasp&*hu*1&*hu*2&*circ*&*seg*2	NaN%	0%	0
avgasp&*hu*1&*hu*2&*seg*2	91.63%	58.85%	83.43
avgasp&*hu*1&*hu*2	91.26%	59.66%	83.36
avgasp&*seg*2	90.60%	60.95%	83.19
navgasp&*hu*1&*hu*2&*circ*&*seg*2	NaN%	0.00	0
navgasp&*hu*1&*hu*2&*seg*2	92.31%	71.02%	86.99*
navgasp&*hu*1&*hu*2	91.95%	72.24%	87.02*
navgasp&*seg*2	84.60%	81.50%	83.88
*hu*1&*hu*2&*seg*2	91.72%	73.81%	87.24*

Table 5.12: Experiments with combinations of features for profile detection and their results.

The tables in 5.13 list all the results for the experiments on combinations of features for gender discrimination. These conditions were chosen mainly in an attempt to test all possible combinations of features although not every combination was tested. The overall performance of these conditions range from 74.91 to 90.98 with the male identification conditions performing better on average than female conditions. The best performing conditions are noted with an * in the tables.

Condition	C	D	Overall
hu3&hu4&hu5&hu6&nmedy&nprfcog M(>)	93.05%	68.84%	87
hu3&hu4&hu5&hu6&nmedy M(>)	93.05%	68.84%	87
hu3&hu4&hu5&hu6&nprfcog M(>)	93.81%	77.84%	89.81
hu3&hu4&hu5&hu6 M(>)	93.96%	79.98%	90.47
nmedy&nprfcog M(>)	80.27%	77.09%	79.47
hu3&hu4&hu5&hu6&nmedy&nprfcog F(<)	100.00%	4.10%	76.03
hu3&hu4&hu5&hu6&nmedy F(<)	98.53%	12.50%	77.02
hu3&hu4&hu5&hu6&nprfcog F(<)	100.00%	4.29%	76.07
hu3&hu4&hu5&hu6 F(<)	92.52%	36.94	78.63
nmedy&nprfcog F(<)	100.00%	4.29%	76.07

Condition	C	D	Overall
hu3&hu4&hu5 M(>)	92.22%	82.55%	89.8
hu4&hu5&hu6 M(>)	93.25%	84.37%	91.03*
hu3&hu5&hu6 M(>)	93.85%	80.09%	90.41
hu3&hu4&hu6 M(>)	93.98%	80.30%	90.56*
hu3&hu4&hu5 F(<)	85.39%	49.07%	76.31
hu4&hu5&hu6 F(<)	94.22%	51.68%	83.58*
hu3&hu5&hu6 F(<)	92.52%	36.94%	78.63
hu3&hu4&hu6 F(<)	92.52%	36.94%	78.63

Condition	C	D	Overall
hu3&hu4&hu5&nmedy M(>)	91.22%	71.20%	86.21
hu4&hu5&hu6&nmedy M(>)	92.61%	69.81%	86.91
hu3&hu5&hu6&nmedy M(>)	92.93%	68.95%	86.94
hu3&hu4&hu6&nmedy M(>)	93.08%	69.16%	87.11
hu3&hu4&hu5&nmedy F(<)	97.10%	12.50%	75.95
hu4&hu5&hu6&nmedy F(<)	97.14%	12.69%	76.03
hu3&hu5&hu6&nmedy F(<)	98.53%	12.50%	77.02
hu3&hu4&hu6&nmedy F(<)	98.53%	12.50%	77.02

Condition	C	D	Overall
hu3&hu4&hu5&nprfcog M(>)	92.03%	80.41%	89.13
hu4&hu5&hu6&nprfcog M(>)	93.09%	80.84%	90.03
hu3&hu5&hu6&nprfcog M(>)	93.69%	77.94%	89.76
hu3&hu4&hu6&nprfcog M(>)	93.83%	78.16%	89.91
hu3&hu4&hu5&nprfcog F(<)	100.00%	4.29%	76.07
hu4&hu5&hu6&nprfcog F(<)	100.00%	4.29%	76.07
hu3&hu5&hu6&nprfcog F(<)	100.00%	4.29%	76.07
hu3&hu4&hu6&nprfcog F(<)	100.00%	4.29%	76.07

Condition	C	D	Overall
$hu3\&hu4$ M(>)	92.16%	83.08%	89.89
$hu4\&hu5$ M(>)	91.65%	86.94%	90.47*
$hu5\&hu6$ M(>)	93.15%	84.48%	90.98*
$hu3\&hu5$ M(>)	92.09%	83.51%	89.95
$hu3\&hu6$ M(>)	93.89%	80.62%	90.57
$hu4\&hu6$ M(>)	92.67%	85.33%	90.84*
$hu3\&hu4$ F(<)	84.47%	50.75%	76.04
$hu4\&hu5$ F(<)	88.47%	65.86%	82.82*
$hu5\&hu6$ F(<)	94.26%	52.05%	83.71*
$hu3\&hu5$ F(<)	85.39%	49.07%	76.31
$hu3\&hu6$ F(<)	92.52%	36.94%	78.63
$hu4\&hu6$ F(<)	94.00%	52.61%	83.65*

Condition	C	D	Overall
$hu3\&hu4\&nemdy$ M(>)	91.13%	71.52%	86.23
$hu4\&hu5\&nmedy$ M(>)	90.84%	72.16%	86.17
$hu5\&hu6\&nmedy$ M(>)	92.49%	69.91%	86.85
$hu3\&hu5\&nemdy$ M(>)	91.08%	72.16%	86.35
$hu3\&hu6\&nmedy$ M(>)	92.98%	69.49%	87.11
$hu4\&hu6\&nmedy$ M(>)	92.39%	70.24%	86.86
$hu3\&hu4\&nemdy$ F(<)	97.10%	12.50%	75.95
$hu4\&hu5\&nmedy$ F(<)	95.77%	12.69%	75.00
$hu5\&hu6\&nmedy$ F(<)	97.14%	12.69%	76.03
$hu3\&hu5\&nemdy$ F(<)	97.10%	12.50%	75.95
$hu3\&hu6\&nmedy$ F(<)	98.53%	12.50%	77.02
$hu4\&hu6\&nmedy$ F(<)	97.14%	12.69%	76.03

Condition	C	D	Overall
$hu3\&hu4\&nprfcog$ M(>)	91.96%	80.84%	89.18
$hu4\&hu5\&nprfcog$ M(>)	91.43%	83.40%	89.53
$hu5\&hu6\&nprfcog$ M(>)	92.99%	80.94%	89.98
$hu3\&hu5\&nprfcog$ M(>)	91.90%	81.37%	89.27
$hu3\&hu6\&nprfcog$ M(>)	93.73%	78.48%	89.92
$hu4\&hu6\&nprfcog$ M(>)	92.81%	81.58%	90.01
$hu3\&hu4\&nprfcog$ F(<)	100.00%	4.29%	76.07
$hu4\&hu5\&nprfcog$ F(<)	100.00%	4.29%	76.07
$hu5\&hu6\&nprfcog$ F(<)	100.00%	4.29%	76.07
$hu3\&hu5\&nprfcog$ F(<)	100.00%	4.29%	76.07
$hu3\&hu6\&nprfcog$ F(<)	100.00%	4.29%	76.07
$hu4\&hu6\&nprfcog$ F(<)	100.00%	4.29%	76.07

Condition	C	D	Overall
hu3&nmedy M(>)	86.25%	73.23%	83
hu4&nmedy M(>)	90.52%	72.59%	86.04
hu5&nmedy M(>)	90.70%	73.13%	86.31
hu6&nmedy M(>)	91.94%	72.06%	86.97
hu3&nmedy F(<)	95.71%	12.50%	74.91
hu4&nmedy F(<)	95.77%	12.69%	75.00
hu5&nmedy F(<)	95.77%	12.69%	75.00
hu6&nmedy F(<)	97.14%	12.69%	76.03

Condition	C	D	Overall
hu3&nprfcog M(>)	86.82%	82.55%	85.76
hu4&nprfcog M(>)	91.10%	84.37%	89.41
hu5&nprfcog M(>)	91.31%	84.37%	89.58
hu6&nprfcog M(>)	92.15%	84.26%	90.18
hu3&nprfcog F(<)	100.00%	4.29%	76.07
hu4&nprfcog F(<)	100.00%	4.29%	76.07
hu5&nprfcog F(<)	100.00%	4.29%	76.07
hu6&nprfcog F(<)	100.00%	4.29%	76.07

Table 5.13: Combinations of conditions for gender detection run on test dataset

*note: since these conditions are ANDed together, the threshold values can be changed to those with a higher error percentage to raise the D while not decreasing the C.

5.3 Experiments on Overall System

For our final experiments, we test how the system performs overall, with the gender discriminator only making decisions on images indicated as side-profiles by the profile detector. In theses experiments, we test how each of the best conditions for profile detection works with the best conditions for gender discrimination.

Tables 5.14-5.16 lists the results of these experiments. As indicated by the *'s in the tables, all the best performing conditions were with *hu1&hu2&seg2*, the best performing profile-detection condition as side-profiler. For the system that makes the most accurate decisions(C), condition *hu4&hu5&hu6* is best for male images and condition *hu5&hu6* is best for female images. For identification of both genders, condition

Condition(w/*navgasp&hu1&hu2&seg2*)	C	D	Overall
hu4&hu5&hu6 M(>)	90.86%	86.46%	89.76
hu4&hu6 M(>)	90.07%	87.71%	89.48
hu5&hu6 M(>)	90.72%	86.46%	89.66
hu4&hu5 M(>)	90.00%	82.02%	89.26
hu4&hu6orhu5&hu6orhu4&hu5 M(>)	89.12%	88.26%	88.90
hu4&hu6orhu4&hu5 M(>)	89.25%	88.26%	89.00
hu4&hu6orhu5&hu6 M(>)	89.94%	87.71%	89.38
hu5&hu6orhu4&hu5 M(>)	89.87%	87.02%	89.16
hu4&hu5&hu6 F(<)	92.83%	50.86%	82.34
hu4&hu6 F(<)	92.83%	50.86%	82.34
hu5&hu6 F(<)	92.89%	51.35%	82.50
hu4&hu5 F(<)	88.64%	67.08%	83.25
hu4&hu6||*hu5&hu6*||*hu4&hu5* F(<)	88.71%	67.57%	83.43
hu4&hu6||*hu4&hu5* F(<)	88.64%	67.08%	83.25
hu4&hu6||*hu5&hu6* F(<)	92.89%	51.35%	82.50
hu5&hu6||*hu4&hu5* F(<)	88.71%	67.57%	83.43

Table 5.14: Experiments and results on overall system with *navgasp&hu1&hu2&seg2* as the side-profiler condition.

Condition(w/*navgasp&hu1&hu2*)	C	D	Overall
hu4&hu5&hu6 M(>)	90.60%	85.71%	89.38
hu4&hu6 M(>)	89.85%	87.06%	89.15
hu5&hu6 M(>)	90.47%	85.71%	89.28
hu4&hu5 M(>)	89.76%	86.25%	88.88
hu4&hu6||*hu5&hu6*||*hu4&hu5* M(>)	88.92%	87.60%	88.59
hu4&hu6||*hu4&hu5* M(>)	89.04%	87.60%	88.68
hu4&hu6||*hu5&hu6* M(>)	89.72%	87.06%	89.06
hu5&hu6||*hu4&hu5* M(>)	89.64%	86.25%	88.79
hu4&hu5&hu6 F(<)	92.48%	50.61%	82.01
hu4&hu6 F(<)	92.48%	50.61%	82.01
hu5&hu6 F(<)	91.74%	51.09%	81.58
hu4&hu5 F(<)	88.14%	66.59%	82.75
hu4&hu6||*hu5&hu6*||*hu4&hu5* F(<)	87.66%	67.07%	82.51
hu4&hu6||*hu4&hu5* F(<)	88.14%	66.59%	82.75
hu4&hu6||*hu5&hu6* F(<)	91.74%	51.09%	81.58
hu5&hu6||*hu4&hu5* F(<)	87.66%	67.07%	82.51

Table 5.15: Experiments and results on overall system with *navgasp&hu1&hu2* as the side-profiler condition.

Condition(w/ *hu1&hu2&seg2*)	C	D	Overall
hu4&hu5&hu6 M(>)	*91.02%	90.16%	90.8
hu4&hu6 M(>)	90.28%	91.37%	90.55
hu5&hu6 M(>)	90.90%	90.16%	90.71
hu4&hu5 M(>)	90.21%	90.70%	90.33
hu4&hu6||*hu5&hu6*||*hu4&hu5* M(>)	89.38%	*91.91%	90.01
hu4&hu6||*hu4&hu5* M(>)	89.50%	*91.91%	90.10
hu4&hu6||*hu5&hu6* M(>)	90.16%	91.37%	90.46
hu5&hu6||*hu4&hu5* M(>)	90.09%	90.70%	90.24
hu4&hu5&hu6 F(<)	92.95%	51.09%	82.36
hu4&hu6 F(<)	92.95%	51.09%	82.49
hu5&hu6 F(<)	*93.01%	51.57%	82.65
hu4&hu5 F(<)	88.82%	67.31%	83.44
hu4&hu6||*hu5&hu6*||*hu4&hu5* F(<)	88.89%	*67.80%	83.62 *
hu4&hu6||*hu4&hu5* F(<)	88.82%	67.31%	83.44
hu4&hu6||*hu5&hu6* F(<)	93.01%	51.57%	82.65
hu5&hu6||*hu4&hu5* F(<)	88.89%	*67.80%	83.62 *

Table 5.16: Experiments and results on overall system with *hu1&hu2&seg2* as the side-profiler condition.

hu4&hu5&hu6 performs the best. For a system that does the best detection(D), conditions *hu4&hu6*||*hu5&hu6*||*hu4&hu5* and *hu4&hu6*||*hu4&hu5* perform equally well for male identification; conditions *hu5&hu6*||*hu4&hu5* and *hu4&hu6*||*hu5&hu6*||*hu4&hu5* perform equally well for female identification. For identification of both genders, the condition *hu4&hu6*||*hu5&hu6*||*hu4&hu5* performs the best. The Overall performance of the system based on both C and D range from 81.58 to 90.80. The best performing condition for male identification is *hu4&hu5&hu6* and for female identification is *hu4&hu6*||*hu5&hu6*||*hu4&hu5* or *hu5&hu6*||*hu4&hu5*. For both gender combined, the condition *hu5&hu6*||*hu4&hu5* performs the best. All of these gender discrimination conditions are identifying images indicated as side-profile with *hu1&hu2&seg2* as side-profiler. The system has an average correctness percentage(C) of 91.33%, detection percentage(D) of 79.8%, and Overall performance of 88.46 for gender identification on full body silhouette images. For an application that identifies gender from a video stream after 10 consecutive profile frames indicating the same gender, all but one subject is identified correctly within 15 seconds. That one subject is

classified as unsure and this is a result of not enough side-profile images detected.

Chapter 6

Discussion

6.1 Profile Detection

Overall this system performs very well and identifies the correct gender with a high accuracy while making very few misidentifications, classifying most ambiguous images as unsure. Even though the misidentification percentage is very small, due to the sensitivity of the issue of gender identification, we examined the images to identify the source of error.

Comparing the images that were misidentified as the opposite gender with misidentified side-profile images, we find that all mis-gendered images are also non-side-profile images. This indicates that if we increase the accuracy of the side-profile detector, the accuracy of the overall system will increase similarly. Looking through the misidentified profile images, it is apparent that one particular class of images dominate. The images that are consistently being misidentified as side-profile images are profile images occluded by the edge of the screen. The subject is usually in profile, but because of the occlusion, the gender is unclear even to the human eye. One most straight forward way to alleviate this problem is to put a minimum size restriction on the area of the silhouette. However, a potential problem with this solution is that more silhouettes will be eliminated for people of small stature than bigger/taller people; this solution is non-neutral across different subjects. Another solution would be to use a model based approach to estimate when parts of the body is occluded

by the field of view. However, this approach would most likely increase processing time dramatically without being able to lessen much of the misidentification. A final option would be to ignore images with more than a maximum number of white pixels near the vertical edges of the image.

As the mis-identification in overall gender identification is mainly a result of mis-identification in profile detection, the low percentage of identification(C) is also in large part due to the low identification rate for side-profile detection. Splitting all the side-profile detection down to individual subjects, we see that the detection rate is not evenly distributed among all the subjects: As we can see from Table 6.1, no side-profile

Subject	percent detection	Subject	percent detection
M1	81%	M7	0%
M2	82%	M8	76%
M3	93.5%	M9	86%
M4	95%	M10	46%
M5	73%	M11	80%
M6	86%	M12	66%
F1	59%	F4	33%
F2	81%	F5	75.5%
F3	85%	F6	88%

Table 6.1: Side-profile percentage of detection per subject

images were identified for M7 at all. Checking the list of the subjects for this project, we see that this is one subject who is overweight. The side-profiler identified no side-profiles for that one subject. This might be a result of the the training data being extremely biased towards lean people. Overall, a set of training data subjects that is more representative of the population, and has a wider distribution in subject height, weight and age, is needed to insure the performance of this system for the general population. A better training data set might fix this problem, but we did anticipate that the system would have trouble identifying the gender of overweight people from the start of this project. Earlier results did indicate that the gender discriminator correctly identified this subject when the gender-discriminator was independent of the side-profile detector. In attempting to address this problem, one possibility is to add an overweight flag to the profile detector. If a subject is detected as overweight,

a different set of thresholds can be used for profile-detection of that subject alone. Another solution is to use a model-based approach to do profile detection. A human model would be first fit to the the silhouette in an attempt to estimate limb positioning and body orientation before side-profile detection. However, it is also foreseeable that this method have difficulty with over-weight subjects as well.

6.2 Gender Discrimination

Even though we have just indicated that the majority of overall system inaccuracies is due to that of the profile-detector, there are also misidentifications in the gender discriminator independent of the side-profiler. In gender discrimination, on average, female identification have higher correctness percentages(C) but much lower detection percentages(D) than male identification. Looking at the training data for gender discrimination(Figures 4-8 to 4-11), it is apparent that these differences are inherent in the differentiating features themselves. For all of halfhu3, halfhu4, halfhu5, and halfhu6, a significant number of female values lie in the region occupied by that of the male values, while almost none of the male values stray into the female region. To improve both C and D, new gender discriminating features for body shape would be needed. Although it is fairly impossible to predict what those features might be if there are any, there is one method that these additional features maybe found by. This approach is similar to that used by face recognition systems. As a first step, we need to find representations of the silhouette image that are detailed enough to reconstruct the silhouettes from. Then, we can train a neural network with these parameters to find the features that discriminate between gender. If the representation is enough to identify individual people from profile-silhouette(as we can most of the time), we are certain to find additional features that will discriminate better between male and female body shapes. Another possible approach that may perform better is to make discriminations based on time-variant data analysis. This changes the entire approach to base gender difference on motion rather than physical shape.

6.3 Time-variant Analysis

We have the perception that movement is more difficult to alter than body shape, which suggests that a time-variant approach would result in more accurate identifications. However, a time variant approach would require many repetitions of the same movement before a decision can be made. Our approach requires as few as one image of the person in the correct orientation. For a real world application, there is much higher likelihood that a time-variant identification system will never make a decision. Another major difficulty in attempting time-variant analysis is the tracking of the same spot through a video sequence. It is very difficult for current computer vision systems to consistently tracking the same spot on the body through a sequence of images with the exception of body extremities. Time variant motion analysis in the past have been accomplished through either using moving light displays(attaching lights to the subject) or simply a lot of approximations[9][1]. Moving light displays would render real world applications impossible and to be able to distinguish gender, we can assume that the accuracy for tracking needs to be a lot higher than that needed to identify the movement or the positioning of the limbs. This also makes the time variant approach an undesirable alternative.

Overall, with a correctness percentage of 91.33% and detection percentage of 79.8%, it is very difficult to improve the performance by significant amount without a lot of extra work.

Chapter 7

Summary

This system identifies gender from silhouette images by first identifying the orientation of the body in the image. This decision is made by comparing the values of the 1st and 2nd order Hu moments and the maximum segments per row to preset values. If the body is in profile to the camera and upright, the gender discriminator makes a decision about the gender(M/F/U) using the 4th, 5th and 6th order Hu moments of the top half of the silhouette image. When a decision is made about the gender, there is a 91.3% likelihood that the indicated gender is correct. Given that it is the profile image of a person of a specific gender, there is a 79.8% likelihood that this system will identify the image as such.

Bibliography

[1] Ali Azarbayejani, Christopher Wren, and Alex Pentland. Real-time 3-d tracking of the human body. In *Proceedings of IMAGE'COM*, Bordeaux, France, May 1996.

[2] Trevor Darrell Christopher Wren, Ali Azarbayejani and Alex Pentland. Pfinder: Real-time tracking of the human body. In *IEEE Transactions on Pattern Analysis and Machine Intelligence Symposium on the Theory of Computing*, pages 780–785, July 1997.

[3] G.W. Cottrell and J. Metcalfe. Empath: Face, gender and emotion recognition using holons. In J. Moody R.P. Lippman and D.S. Touretzsky, editors, *Advances in Neural Information Processing Systems*, number 3, pages 564–571, San Mateo, 1991.

[4] James E. Cutting. Perceiving and recovering structure from events. In *Pattern Recognition Symposium on the Theory of Computing*, 1983.

[5] James W. Davis and Aaron F. Bobick. Sideshow: A silhouette-based interactive dual-screen environment. Technical Report 457, MIT Media Laboratory Perceptual Computing Section, 1998.

[6] B.A. Golomb, D.T. Lawrence, and T.J. Sejnowski. Sexnet: A neural network identifiies sex from human faces. In J. Moody R.P. Lippman and D.S. Touretzsky, editors, *Advances in Neural Information Processing Systems*, number 3, pages 572–577, San Mateo, 1991.

[7] Oral communications with Aaron Bobick.

[8] A. Lanitis, C.J. Taylor, and T.F. Cootes. A unified approach to coding and interpreting face images. In *International Conference of Computer Vision*, pages 368–373, 1995.

[9] Sourabh A. Niyogi and Edward H. Adelson. Analyzing and recognizing walking figures in xyt. Technical Report 223, MIT Media Laboratory Perceptual Computing Section.

[10] http://vismod.www.media.mit.edu/vismod/demos/facerec/index.html. VISMOD Face Recognition Home Page.

[11] Aaron Bobick Yuri Ivanov and John Liu. Fast lighting independent background subtraction. Technical Report 437, MIT Media Laboratory Perceptual Computing Section, 1998.