

Identifying People with Soft-Biometrics at Fleet Week

Eric Martinson
NRC Post-Doctoral Fellow
US Naval Research Laboratory
Washington, DC USA
eric.martinson.ctr@nrl.navy.mil

Wallace Lawson, J. Gregory Trafton
US Naval Research Laboratory
Washington, DC USA
{ed.lawson, greg.trafton}@nrl.navy.mil

Abstract— Person identification is a fundamental robotic capability for long-term interactions with people. It is important to know with whom the robot is interacting for social reasons, as well as to remember user preferences and interaction histories. There exist, however, a number of different features by which people can be identified. This work describes three alternative, soft biometrics (clothing, complexion, and height) that can be learned in real-time and utilized by a humanoid robot in a social setting for person identification. The use of these biometrics is then evaluated as part of a novel experiment in robotic person identification carried out at Fleet Week, New York City in May, 2012. In this experiment, Octavia employed soft biometrics to discriminate between groups of 3 people. 202 volunteers interacted with Octavia as part of the study, interacting with the robot from multiple locations in a challenging environment.

Index Terms— I.2.9 Robotics, Person Identification, Soft Biometrics

I. INTRODUCTION

Most robots do not know whom they are interacting with. Even in relatively sophisticated interaction scenarios (e.g., museum guides [1] or game scenarios [2]), the robot does not typically know the identity of the person. Knowing someone's identity allows the robot to gather user preferences, use appropriate forms of address and pronouns, keep track of what that person knows (or does not [3]), and many others. Probably the most commonly used method of identifying people is to use registered RFID tags (or similar); this approach has the advantage of being simple and robust in a large number of environments. Unfortunately, it is only appropriate for people who already have the technology and allow access or requires the robots to give out RFID tags so they can interact.

Our approach is to attempt to uniquely identify people on a mobile robotics platform without providing unique tags, emitters, or clothing: to be able to identify people using some of the same cues that people do. Typically, this would be accomplished by building individual models of people using traditional biometrics like face recognition and/or speaker recognition. To enroll in these systems, an individual will approach the robot and have a picture taken and/or speak to the robot for a few seconds. Later, the robot can identify people who have been enrolled. These biometrics have been shown to

be extremely robust across large datasets [4]. Unfortunately, on a mobile platform, uniquely identifying an individual becomes very difficult, primarily because of dynamic changes in the environment or sensing situations. For example, as the robot moves to new locations or shifts its head elevation and orientation, lighting conditions or ambient noise levels may change, people may alter their pose, or they may not be looking at the camera in an optimal manner. It is neither practical nor good interaction practice to ask an individual to re-enroll every time lighting, noise, pose, or location changes.

Our solution to this problem is to use soft biometrics. In contrast to traditional biometrics, which can uniquely identify people, “soft” biometrics, such as gender, height, clothing, hair color, etc, lack enough distinctiveness and permanence [5] for unique identification. The precision of soft biometrics can be affected by the same issues as traditional biometrics, but because each soft biometric uses different information, they can be overall quite robust. In fact, Reid and Nixon [6] suggest that by fusing enough soft biometric modalities, a computer could still uniquely identify people.

This work demonstrates that even small (three) numbers of soft biometrics can achieve quite good person identification results in a perceptually challenging environment. Biometrics for describing clothing, complexion, and height were successfully utilized in a human-study conducted at Fleet Week in New York (Figure 1), with our humanoid robot, Octavia. Interacting with over 200 volunteers as part of more than 100 “identification” games, Octavia averaged a 90% correct identification rate with just soft biometrics.



Fig 1. Octavia at Fleet Week.

II. RELATED WORK

Soft biometric research has generally focused on fusion with other soft modalities or traditional modalities to recognize individuals. Park [7] combined gender and other facial marks to improve the precision of identification when combined with face recognition. Lawson and Martinson [8] demonstrated increased precision using a Markov Logic Network to fuse soft modalities (complexion, clothing) with traditional biometrics (speaker recognition, face recognition). The advantage of this type of fusion is that it allows contextual information (i.e. environmental conditions that might impact precision, like distance to the person) to be integrated into the identification decision. Jain [5] used gender, ethnicity and height to improve the precision of a fingerprint recognition system. Fingerprint is a “contact biometric”, which means that the types of soft biometrics that can be used in this domain can be extended to include those that might include some measurement that requires physical contact with the person. Towards that end, Ailisto et al [9] explored the use of body weight and fat to enhance the performance of fingerprint recognition.

Similarly, multiple soft biometrics can be used to identify individuals uniquely. One advantage of this type of approach is that a textual description of an individual can be turned into a biometric signature to recognize individuals. Further, when a sufficient number of soft modalities are combined, it is quite possible to get very high levels of precision. Reid and Nixon [6] combined multiple manually marked semantic traits of the individual to recognize people. Demirkus et al. [10] use a number of facial soft biometrics in addition to other soft modalities to track individuals in handoff regions between different video surveillance cameras.

III. METHODOLOGY

This section details the person identification system used by Octavia during interactions with people. This includes descriptions of the person tracking system, the three soft biometric modalities being investigated (clothing, complexion, and height), and our method of data fusion.

A. Person Detection and Tracking

A time-of-flight (ToF) camera (SR4000) is used for detection and tracking. The advantage of using the ToF camera is the ease with which foreground can be isolated. Large depth discontinuities will exist along the boundary between the foreground and the background. Therefore, objects in the foreground will have similar depth information and will have significant depth discontinuities along the edges of the object. We use connected components analysis to find these regions (Figure 2). In this step, each pixel in the depth image is compared to its 8-neighbors. If the difference in distance is less than a threshold they are included together in the same object.

Large connected components are processed to determine if they are a person. While there are a number of ways to do this, one conceptually simple yet effective approach is to look for large connected components with a face. If a face is detected with a sufficiently high confidence, the region is determined to be a person and it is added to the list of active tracks. Figure 5



Fig 2. (Left) Intensity image from the SR-4000. (Right) Tracked and segmented human observer.

shows an image of this process. In the figure, two large connected components have been found. Both have plausible human-like shape, but only one connected component has a face. For face detection, we use the Pittsburgh Pattern Recognition SDK (PittPatt) [11].

Tracking takes advantage of the notion that the shape of a person should not change significantly from one frame to the next. We compare connected components found in consecutive frames using Tanimoto similarity. Tanimoto measures the ratio of the number of overlapping pixels over all of the pixels. Higher Tanimoto similarity scores (Eq 1) between components (C_1, C_2) indicate that the two connected components belong to the same individual.

$$T(C_1, C_2) = \frac{C_1 \cap C_2}{C_1 \cup C_2} \quad \text{Eq 1}$$

B. Soft Biometrics:

Octavia uses 3 soft biometrics to identify people: clothing, complexion, and height. All three currently assume an initial training session, during which a model of the individual can be constructed. During Fleet Week, models were created as part of a game played by the participants before the robot guesses identities. More details are provided in Section IV.

1) Clothing and Complexion

Clothing and complexion depend on face detection. When a face is detected in the RGB image using PittPatt, regions relative to the detected face are extracted from the image for creating color histograms (Figure 3). Part of the detected face is used to create the complexion model (Figure 3), and an area below is used to create a clothing model. Selection of regions in this manner provides relatively well lit and pose tolerant areas that can be used for sampling. Although alternative

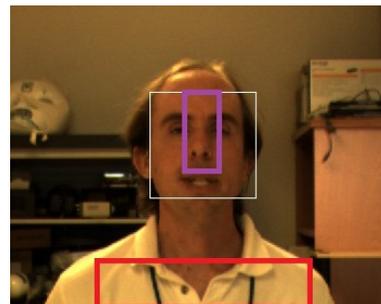


Fig 3. Boxes showing the regions used for complexion (about nose and forehead) and clothing are shown relative to the detected face position.

methods exist (e.g. GMM [12]), color histograms have been used successfully in other fused biometrics systems [8].

Color histogram models are created to measure the frequency of each color in the identified regions. To create a color histogram, H , each channel (r, g, b) is quantized using the 4 most significant bits. The resulting data are then concatenated, producing a 4096 bin histogram.

Histograms are compared using a modified χ^2 similarity measure (Eq. 2). During enrollment, a new histogram is compared against existing models. If it is sufficiently different from known histograms it is retained, otherwise, its support for an existing histogram is noted. Those histograms that have not received a sufficient amount of support during training are discarded [8]. This provides tolerance to situations when landmarks are not correctly detected, or motion blur affects the appearance of the individual.

$$\chi^2(H_a, H_b) = 2 \sum_{i=0}^{4096} \frac{(H_a(i) - H_b(i))^2}{(H_a(i) + H_b(i))} \quad \text{Eq 2}$$

2) Height

An individual's height is determined by converting detected face positions (x, y, z) in the SR4000 intensity image into a real-world coordinate frame. By comparing pixel locations to the corresponding depth image, Octavia determines both a known direction and distance. Then, forward kinematics are applied to identify the transformation matrix from the base to the SR4000.

There are five motors involved in estimating the transformation matrix: torso pan, neck pitch, head pan, head pitch, and head roll. Although head roll is not generally used in tracking, in order to keep people vertically aligned in the image, it must still be included when calculating the transformation matrix from camera to robot local space. The transformation matrix for each individual joint is derived from the associated Denavit-Hartenberg parameters.

$$T_{cam2robot} = T_{torsopan} * T_{neckpitch} * T_{headpan} * \dots * T_{headpitch} * T_{headroll} \quad \text{Eq 3}$$

$$Face(x, y, z) = T_{cam2robot} * [x, y, z, 1]^T$$

The height of an individual was then created from the average height of the last 20 detected face positions (~1.2 sec). A height model for an individual was simply the last average measured face height of the training data.

Given two height measurements, the similarity between them (e.g. likelihood of being the same individual) was then a linear function up to 15 cm in difference.

$$S(m_1, m_2) = \begin{cases} 1 - \frac{|m_1 - m_2|}{0.15}, & |m_1 - m_2| < 0.15 \\ 0, & |m_1 - m_2| \geq 0.15 \end{cases} \quad \text{Eq 4}$$

C. Fusing Results

The fusion rule was that of weighted summation. All scores were normalized to 1.0, then summed together and divided by the number of available modalities. The weights $\{W_{cloth}, W_{compl}, W_{height}\}$ were assigned to be $\{1.0, 0.1, 1.0\}$ during Fleet Week. The complexion weight was set low because of poor

demonstrated precision during initial testing. However, the nonzero weight meant complexion was still the tie breaker when combined clothing and height scores were not enough.

When data are missing for a given modality, the weight for that modality was assigned to be zero. This rule was necessary to handle partially missing modalities. For example, a clothing model could not be created for individual X because too little clothing data were acquired during training due to the person moving around. However, during testing, X stood still and a color histogram was created. That histogram can still help suppress other identification scores, but should not contribute to X's combined score since he does not have a clothing model.

IV. FLEET WEEK EXPERIMENT

Fleet Week in New York City is a once a year event in which the US Navy and Coast Guard make active duty ships available for public viewing. It is also a time for demonstrating other Navy hardware and research endeavors. This year, 2012, the robot Octavia was invited back to meet the public. Octavia made her first appearance at Fleet Week in 2010.

The person identification experiment was structured as a game. Three volunteers were instructed to stand in specific locations (~1.5-m away) in the region facing the robot. Then the robot would look at each individual and build models for three soft biometrics: clothing, appearance, and height. Models were constructed from up to a maximum of 20 sec of data. Usually, however, models were from much less, as individuals did not always look at the robot. After learning models, the robot would "close its eyes", and people would change places. Then the robot would look at each of them and try to identify them using the soft biometrics models.

Octavia played a total of 105 games. 67% of the game participants were volunteers from the audience, while the remaining 33% were with the experimenters. The purpose of this experiment was three-fold. First, we are establishing the applicability and performance of soft biometrics in human-robot interaction. Second, it is a rejection experiment, evaluating the false-positive rate of watch-list based hard biometrics (e.g. face and speaker). Finally, it is data collection for learning person identification models over longer periods of time. The end goal is to develop a system from this data that can improve hard-biometrics models over time without requiring repeated training type interactions.

A. Robot Data Collection

The Xitome Mobile-Dextrous-Social (MDS) humanoid robot, Octavia (Figure 4) is our platform for exploring lifelong learning in biometrics. Designed for human robot interaction, the MDS robot has 8 DoF for the head, neck and eyes allowing the robot to pan or tilt the camera as needed to look at arbitrary locations in 3D space. An additional 33 degrees of freedom enable gestures and facial expressions. This humanoid body sits atop a two-wheeled Segway base.

Throughout all of the games, Octavia continuously recorded data from her onboard sensors. This included:

- Color Camera: Located in the right eye, Octavia stored 640x480 RGB images at ~4 Hz from a Point Grey Firefly camera. A sample image is shown in Figure 5 (top).



Fig 4. Octavia has a color camera and an SR-4000 time of flight sensor mounted in the head. For audition, it uses a 4-element mic array spread across body and backpack.

- SR4000: Located in Octavia’s forehead, the SR4000 produced a 176x144 intensity and depth image. These were stored at ~17 Hz. A sample image is shown in Figure 5 (bottom).
- 4 Microphones: Located on the body and backpack, these were Shure MX100 lavalier microphones with the cardioid package (sampled at 32768 Hz).
- Joint Angles: stored the current state of the entire kinematic chain from torso to eye (torso pan, neck pitch, head pan, head pitch, head roll, eye pitch, eye pan) at ~4 Hz.

Games were indexed by time, and the boundaries of both training and testing session clearly marked in the database. The exhibit itself measured 20x15’ and was roped off to prevent bystanders to get too close to the robot.

B. Environment Setup

The Octavia exhibit was located on the interior of pier 90, where the USS Wasp aircraft carrier was moored. People lined up in the middle of the pier to visit the carrier, and then returned along the edge past the exhibits. Octavia was the last exhibit on the pier, but not the last exhibit at Fleet Week.

Normally used as a parking lot, lighting on the pier was a significant problem for perception. Large, metal rolling doors let in natural light behind the Octavia exhibit. These were supplemented by small numbers of overhead lights. To make the environment friendlier for color perception, we left the metal doors mostly closed and setup lights and a backdrop.

Acoustic perception was also very challenging in this environment. Ambient noise due to weather (rain, thunder, and wind) was significant and irregular. This was usually dominated, however, by speech noise from the crowd. Average noise levels regularly topped 70 dBA with peak noise being much higher. Even using 2 speakers mounted approximately 3 ft to either side of the robot to amplify Octavia’s speech output (Cepstral TTS), volunteers regularly asked for help in understanding what the robot said due to ambient noise levels.

C. Game Procedures

The game itself involved 3 stages: training, moving around, and recognition. Volunteers were separated from the crowd, and brought behind the ropes to stand against the white backdrop. They were told by the experimenter to stand on one of 3 locations where the robot would look for people during the game. Then, once all three volunteers were situated, the experimenter described the expected game progress from beginning to end. For additional assistance, an experimenter stood nearby, ready to answer questions about what the robot said and to maintain participant safety. Any questions regarding the details of the system or otherwise not related to the game in progress were deferred until after the game had finished.

1) Training

In the first phase, Octavia needed to acquire a model and an identifier for each individual. For each of three pre-specified locations, she rotated her head and torso to center that location in the image. Then she looked straight, up, and then down, trying each position until a face was found in the SR4000 intensity image. After detecting the individual, Octavia readjusted torso pan, neck pitch, head pan, head pitch, eye pitch, and eye pan to place the individual at the top of the image in both the rgb and depth cameras. Additional adjustments to pose were made throughout the entire game whenever pose varied by >0.1 rad.

After detecting the individual and adjusting pose, Octavia first asked for an identifier. This was done as a question about the volunteer’s favorite flavor of ice cream, color, animal, planet, or sport. The answer provided by the volunteer was

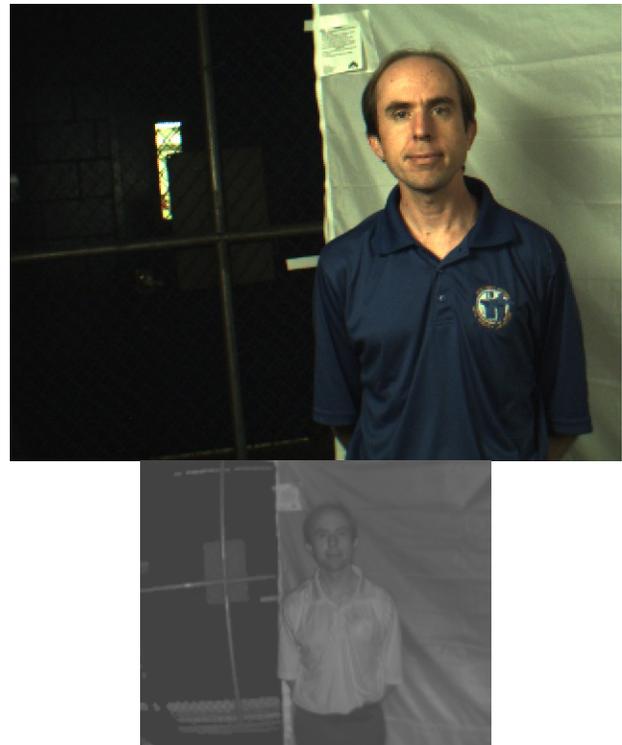


Fig 5. (Top) An rgb image collected during a game with Octavia. (Bottom) An SR4000 intensity image collected at the same time.

manually entered by the experimenter into the person-ID system, and was attached to all subsequently collected data and biometric models. When the experimenters participated in the shell game, their names were used to preserve consistency across games.

With an established identifier, Octavia next asked the volunteer to speak on a subject for up to 30-sec. Subjects were randomly selected from one of 5 topics. No subject topic or identifier question was repeated during the game to avoid confusion. The topics initiated by Octavia were as follows:

- “Tell me about your favorite part of the U.S.S. Wasp.”
- “Tell me about your favorite movie robot and why.”
- “I give up. What would YOU like to talk to a robot about?”
- “You are just about to escape fleet week. What are you going to do next?”
- “After fleet week, do you think I could play for the Mets?”

Octavia recorded 20-sec of data after asking the volunteer about a subject. Although participants were told beforehand that they could speak for as long or as little as they would like, most participants spoke only a sentence in response. Despite the limited utterance length, 20-sec was still necessary to guarantee participants looked at the robot long enough to build a model. Often, they would look around while the robot was building models, limiting the number of full-on frontal views of the face.

After 20-sec, Octavia said, “Thank you {ID}. I think I can recognize you now,” where {ID} was the identifier provided previously. Then she would rotate to face the next individual and repeat.

2) *Moving Around*

Once the training phase was completed for all three volunteers, Octavia said, “Now I will close my eyes and people should move around. Then I will try to recognize each of you.” She then closed her eyes and slowly counted to five. People were told previously that they could either move or stay where they were to try and fool the robot, but they overwhelmingly choose to change positions. On reaching the number five, Octavia said, “Ready or not, here I come,” and opened her eyes.

3) *Recognition*

In the final phase, Octavia again turned to face each of the locations in turn and tried to recognize the individual based on soft biometrics models. As during training, Octavia had to find the individual’s face first by looking up and down, and then continuously updated her viewing angle to keep the face towards the top of the image in both cameras. Once they were detected, Octavia said, “I’m sure I can do this, but maybe you could give me a little hint about your identity?” Because participants were warned ahead of time that this was optional, most simply stated, “No” or shook their heads in the negative.

After asking for a hint, Octavia again looked at the subject for 20-sec, using all images where a face was detected to match clothing and appearance, and using tracking data to

estimate height. At the end, Octavia guessed the participants identity from the 3 possible subjects based on maximum likelihood. No memory was employed to improve performance, so Octavia could make the mistake of guessing the same individual twice in a single game.

V. RESULTS

The experiment at Fleet Week was structured as a game to encourage participation by the volunteers and attract additional volunteers from the audience. As part of the game, Octavia would announce her guess after each testing session. Over all 6-days, including games with both volunteers and experimenters, Octavia averaged 90% correct identification rates using just the three soft biometrics. This performance is very encouraging as precision would have been even higher, except that this includes hardware failures and software problems (e.g. failing to retrieve models from the database). Such problems are not usually presented in biometrics software, as they are evaluating datasets instead of systems. Unfortunately, we do not have the actual numbers to exclude hardware concerns, as not all system issues were immediately evident. A post-analysis of the results (described in Section V.C) suggests that precision should have been closer to 94%.

In the remainder of this section, we analyze in depth each of the individual biometrics that Octavia used for person identification. These post analyses focus on only a single day of data containing 21 games of 3 people, with 46 volunteers and 6 experimenters participating. Over each 20-sec period, ~80 color and ~340 intensity/depth images were collected. Each participant was described by a minimum of two such sessions: one training, and one testing. Of the color images, 95% of those images have faces in them, and ~80% are images with angles of incidence <15 degrees. These measurements were estimated using the PittPatt SDK [11]. For the SR4000, used by the height biometric, the average testing session contained 285 measurements/person.

A. *Post Analysis Description*

Performance of individual soft-biometrics was measured using both closed-world and open-world analyses. In the closed-world analysis, each datapoint was compared to K models, where K was the number of players. The model with the highest similarity was the winner. Games were remixed to create matchups between game participants and increase the number of overall possible comparisons. Given N different models (this is not the number of participants, because experimenters were present in multiple games), all possible combinations of 2, 3, 4, or 5 players were investigated. Two models for the same experimenter were never evaluated as part of the same game. Furthermore, when an experimenter was one of the subjects in the matchup, the closed world analysis for that matchup also included all test-sets in which the experimenter in question was present.

In the open-world analysis, each model was compared to every datapoint in the database, across all subjects, and raw similarity scores were calculated. All scores greater than some threshold were accepted as belonging to that model. By varying

that threshold, we examined how the correct accept vs false accept rate changes for each biometric.

B. Clothing

Clothing was the most reliable of the individual biometrics. The closed-world analysis for each is given in Table 1. Note that precision is estimated across all testing images with detected faces, and is defined as the number of correctly identified individuals over the total number of faces. Also note that this analysis excludes those games where an individual did not have a model.

Table 1. Soft-biometrics precision as it varies with the number of players.

Number of Players	Clothing	Complexion	Height
2	0.932	0.841	0.884
3	0.884	0.742	0.796
4	0.848	0.674	0.725
5	0.818	0.624	0.664

When using the entire 20-sec training session, only one individual was not modeled. This happens when too few faces are detected during the training session, and/or there is too much variation in the recorded clothing data. In this particular case, the robot failed to center the camera on the individual's head correctly.

An open question regarding clothing is the need for the full 20-sec of training. To explore the effects of the training set size, models for each person were also created using only the first 5-sec of training data. For clothing, the effect on precision was <0.1% for all game sizes.

Figure 6 plots the ROC curve for all modalities. Again, clothing is highest with a correct accept rate of 79.0% for a false accept rate of 10%.

C. Complexion

Complexion was the least precise soft-biometric. With 3 people, precision was 74% in the closed-world analysis. This was evident during initial testing at Fleet Well as well, hence the reason why complexion was assigned a very low weight in the final fusion system.

Complexion also had significant effects from changing the training session size. When using only the first 5-sec of the training session to build a model, 6 models failed to gather enough data. Even when models were constructed, they were less accurate. In Figure 6, the curve representing the smaller training session size (Appearance – 5sec) reaches only 49.9% correct accept rate when at 10% false accept rate. Using the full training set raises the correct accept rate to 52.7%. In the closed world analysis, the smaller training session demonstrated in a decreased precision of ~2% for each player count condition.

D. Height

The final soft-biometric, height, appears to fall somewhere in the middle in terms of precision. It is 4-5% better than complexion, but falls off much faster than clothing with the number of players in the game. The ROC curve shows a similar story, reaching only a 62.5% correct accept rate at a 10% false accept rate.

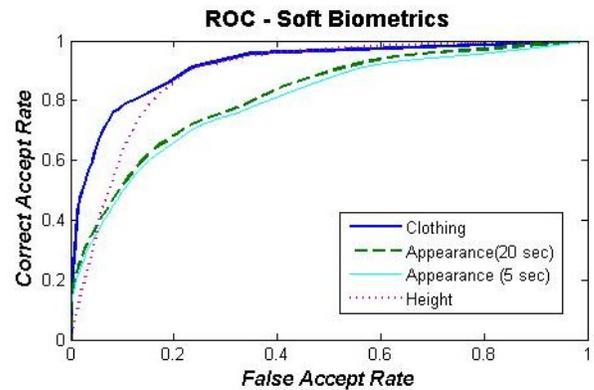


Figure 6. ROC curve comparing correct accept rate vs false accept rate for each of the individual biometrics.

Unlike clothing and complexion, the height model is calculated from a running average of 20 frames. Although more data could be incorporated from the rest of the session, that would lower precision for two reasons. First, people do not always stand straight when talking. As they change their pose, their height will appear less than normal to the robot. The second problem is a robot problem. Because the robot needed to center people in its field of view, it moved during data collection. This resulted in erroneous height estimates when the recorded robot joint angle positions could not be updated as fast as a rate as the SR4000.

To evaluate height without these movement related issues, we scored just the end points of each session. By the end of testing, people were done moving around, and the robot had long since stopped moving to finish centering subjects in the image. Using only these scores, which were the ones actually used by the fusion system, increased precision by 4.8% for games with 3 individuals. The full results are in Table 2.

Table 2. Precision of the height biometric when examining only the end measurements of each session.

Player Count	2	3	4	5
Precision	0.918	0.844	0.779	0.720

Note that similar analyses of clothing and complexion failed to demonstrate any such difference between overall and endpoint performance.

E. Fusion

Combining the three biometrics together provided the best overall precision for all player counts. For 3 players, the 94.3% precision across all hypothetical games is close to the actual 95% measured at Fleet Week for that day. The combined results are summarized in Table 3:

Player Count	Clothing*	Clothing + Height	Complete Fusion System
2	0.932 (0.922)	0.969	0.970
3	0.884 (0.873)	0.941	0.943
4	0.848 (0.835)	0.914	0.917
5	0.818 (0.804)	0.887	0.891

*The numbers in () indicate precision over all games, including those where an individual was not modeled in training.

Note that Table 3 examines precision over all games, unlike Tables 1 and 2, which excluded games where no models were present. As such, the clothing precision is decreased due to a single missing model. The addition of height to the clothing biometric provided a substantial precision increase over clothing alone, the best single biometric. Appearance further increased precision, but by a more modest amount (see the ROC curve, Figure 7).

One of the most interesting issues with soft biometrics concerns which specific measures play a significant role in increasing the accuracy of the system. To investigate this question, we used logistic regression. A simple description of logistic regression is that it is a multiple linear regression model with a dichotomous variable as an outcome variable; a more detailed description can be found in [13]. The outcome of the system (correct or incorrect) was the dichotomous dependent variable. The match score of each of the three soft biometrics were the predictors¹.

The overall logistic regression was statistically significant, $\chi^2(3) = 463.5, p < 0.0001$. All three variables were statistically significant, suggesting that all three made an important contribution to the overall success of the model. Another way to examine whether the complexion biometric had an impact on the overall model was to compare a model that had only clothing and height with a model that had clothing, height, and complexion. Consistent with the earlier analysis, the full model was statistically better than the model with only clothing and height, $\chi^2(1) = 67.4, p < 0.0001$. Both these results suggest that, even though the gain in performance was relatively modest, it is an important biometric to include in future models.

VI. DISCUSSION

The person identification experiment at Fleet Week demonstrated the utility of soft biometrics for identification in a social setting. Although the game was a closed-world system with only 3 subjects, groups of three or less are not uncommon social situations in which a robot would require rapid, accurate person identification, and soft biometrics can obviously contribute. By fusing multiple soft-biometrics together, a robot can even increase the interaction group size to 5 people and still maintain near 90% precision. That would handle all but the most challenging crowd situations, and allow people the freedom to move about during an interaction.

The obvious question, though, is how such soft biometrics compare to traditional biometrics like face recognition. The answer is that face recognition is a more mature technology with higher recognition rates, even from limiting training data. But it is not a solved problem, either. Face recognition systems are negatively impacted by different lighting conditions from training, high angles of incidence to a face, and even simple props like glasses. Under such circumstances, however, visual person identification by people degrades as well [14]. So people rely on multiple cues, like voice plus face [15] and likely many others. Soft biometrics can help robots do the

¹ For this analysis, the full training data set for each user was averaged to provide a single prediction point per biometric.

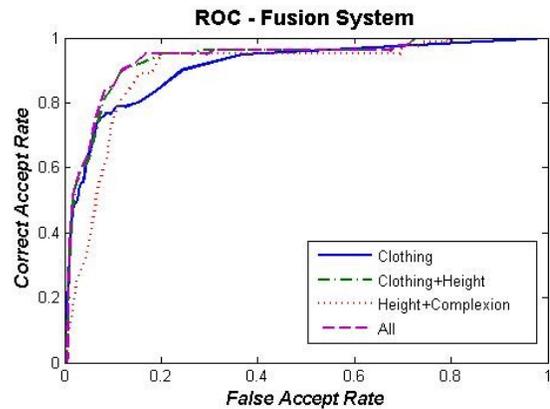


Fig 7. ROC curve comparing the performance of clothing alone to three fusion systems. The combined system is on top, often matched by clothing and height.

same, adding alternative recognition features to bolster the performance of other systems like face and/or speaker recognition.

Soft biometrics may also help improve communication with people, by adding relevant context to a scene. In contrast to face recognition, which often uses arbitrary machine generated features to describe a face, features such as clothing, height, gender, and even complexion have linguistic descriptors. A robot describing an unknown individual to a human partner could utilize soft biometrics as part of its descriptions of the environment (e.g. “He is tall, pale skinned, and wearing a white shirt with some red on the front”). Choosing the appropriate set of discriminators in a crowded setting is an interesting challenge in itself. Conversely, a human could tell a robot about unknown individuals the robot will meet in the future – letting the robot rapidly assign names to faces and build better models as it finally meets them.

In the remainder of this section, we will discuss the performance of the existing soft biometrics, as well as how to expand upon these results in the future.

A. Color Histogram Performance Differences

Color histograms were used as part of two different soft biometrics: clothing and complexion. Where clothing was the strongest of the three evaluated as part of the Fleet Week testing, complexion was the weakest. This discrepancy occurs because the color histogram model is good at picking up on major differences in color while being insensitive to small changes. When there are wide ranges of color differences between individuals, as with clothing, this is a strength. Small differences in those cases are most likely the effects of lighting or pose changes. Complexion, by contrast, consists of a much narrower range of colors. That same strength for clothing recognition means the current color histogram approach is not sensitive enough to pick up on small changes in color, which is necessary for robust complexion modeling.

B. Training Set Size

The post-analysis revealed differences between the biometrics in the size and shape of the training set required. Where clothing precision changed very little with a 5-sec

training set (as opposed to 20-sec), complexion improved significantly, with fewer missing models and a greater precision for those that were created. Height, by contrast, actually needed very little of the training data, but the timing of the data collection was important.

The minimum training duration is most likely related to the size of the feature space being modeled and the amount of noise in the feature space. As discussed previously, complexion is an impoverished feature space, and so benefitted from increased training time to eliminate the effects of noise. Clothing did not need the extra time because the variations were already so large. Height, by contrast, is also an impoverished space, but the noise is not random, so a longer training time does not help. Instead, significant height variations were due to posture changes that do not happen often, but which impact measurements for longer.

C. Removing Face Detection

The Fleet Week experiments demonstrated the potential of soft biometrics for recognition, but all of the existing algorithms currently rely on face detection. Clothing and complexion use the face position to identify where to sample the image, while height is estimated directly from the detected face location. At least two of these biometrics, however, should not need faces. Clothing is often recognizable from the back, or from far away, well before a face is detected. Height should be to the top of the head, and not require a face at all.

The challenge is not so much in the biometrics, though, but in the person detection itself. Faces are the easiest way of separating a person from other objects in the background. Although tracking objects from depth images after a face has been detected alleviates the need for constant face detection, the robot cannot monitor every direction simultaneously. When it moves the head, it loses active tracks and needs to find faces all over again. People are not only separable from the background by faces, however. They also move, and have distinctive shapes [16]. Initiating tracks by more methods than just face detection, should enable the use of soft biometrics from a wider variety of relative positions.

VII. CONCLUSION

This work has demonstrated the utility of soft-biometrics for person identification as part of human-robot interaction. Three soft biometrics (clothing, complexion, and height) were evaluated individually and in a combined, fusion system as part of a human-study with a real robot at Fleet Week in New York City involving more than 200 volunteers. The combined person identification system correctly identified participants 90% of the time over a 6-day period, with the biometric describing clothing being the individual metric with the greatest utility.

Although one biometric was strongest, the contribution of each of the proposed biometrics was statistically significant. They differed in what they measured, and how much training was required, but each added precision to the combined system. The challenges now are in improving the interaction, reducing the structure required for training person identification, and expanding the role and generality of such soft biometrics in recognizing people.

ACKNOWLEDGEMENT

This research was supported by the Office of Naval Research under job order numbers N0001408WX30007 and N0001410WX20773.

REFERENCES

- [1] A. Yamazaki, K. Yamazaki, T. Ohyama, Y. Kobayashi, and Y. Kuno, "A techno-sociological solution for designing a museum guide robot: regarding choosing an appropriate visitor," in *Human Robot Interaction*, Boston, MA, 2012, pp. 309-316.
- [2] S. Chernova, N. Depalma, E. Morant, and C. Breazeal, "Crowdsourcing human-robot interaction: Application from virtual to physical worlds," in *RO-MAN*, Atlanta, GA, 2011, pp. 21-26.
- [3] L. Hiatt, A. Harrison, and G. Trafton, "Accommodating Human Variability in Human-Robot Teams through Theory of Mind," in *IJCAI*, Barcelona, Spain, 2011.
- [4] J. Phillips et al., "FRVT 2006 and ICE 2006 Large-Scale Results," NIST, Gaithersburg, MD, Technical Report NISTIR 7408, 2007.
- [5] A. Jain, S. Dass, and K. Nandakumar, "Can soft biometric traits assist user recognition?," *Proc. of SPIE*, vol. 5404, pp. 5601-572, 2004.
- [6] D. Reid and M. Nixon, "Imputing Human Descriptions in Semantic Biometrics," in *Multimedia in Forensics, Security and Intelligence*, Firenze, Italy, 2010, pp. 25-30.
- [7] U. Park and A. Jain, "Face matching and retrieval using soft biometrics," *IEEE Trans on Information, Forensics, and Security*, vol. 5, no. 3, pp. 406-415, 2010.
- [8] W. Lawson and E. Martinson, "Multimodal Identification using Markov Logic Networks," in *Proc. of Automatic Face & Gesture Recognition (FG)*, Santa Barbara, CA, 2011.
- [9] H. Ailisto, E. Vilkkijoune, M. Lindholm, S. Makela, and J. Peltola, "Soft biometrics - combining weight and fat measurements with fingerprint biometrics," *Pattern Recognition Letters*, vol. 27, no. 5, pp. 325-334, 2006.
- [10] M. Demirkus, K. Garg, and S. Guler, "Automated person categorization for video surveillance using soft biometrics," in *Proc. of SPIE, Biometric Technology for Human Identification VII*, Orlando, FL, 2010.
- [11] M. Nechyba, L. Brandy, and H. Schneiderman, "PittPatt Face Detection and Tracking for the CLEAR 2007 Evaluation," *Lecture Notes in Computer Science*, vol. 4625 / 2008, pp. 126-137, 2008.
- [12] J. Sivic, C. Zitnick, and Szeliski. R., "Finding people in repeated shots of the same scene," in *Proc of British Machine Vision Conference*, Edinburgh, UK, 2006, pp. 909-918.
- [13] C.-Y. Peng, K. Lee, and G. Ingersoll, "An introduction to logistic regression analysis and reporting," *The Journal of Educational Research*, vol. 96, pp. 3-14, 2002.
- [14] V. Bruce and A. Young, *In the Eye of the Beholder: The Science of Face Perception*. Oxford: Oxford University Press, 1998.
- [15] S. Campanella and P. Belin, "Integrating Face and Voice in Person Perception," *Trends in Cognitive Sciences*, vol. 11, no. 12, p. 535—543, 2007.
- [16] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Int. Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, 2005, pp. 886-893.