Complexion as a Soft Biometric in Human-Robot Interaction

Wallace Lawson, J. Gregory Trafton
Naval Center for Applied Research in Artificial Intelligence
Washington, DC
{ed.lawson, greg.trafton}@nrl.navy.mil

Eric Martinson
NRC Postdoctoral Fellow
Naval Research Laboratory *
Washington, DC

Abstract

Complexion plays a remarkably important role in recognition. Experiments with human subjects have shown that complexion provides as much distinctiveness as other well-known features such as the shape of the face. From the perspective an autonomous robot, changes in lighting (e.g., intensity, orientation) and camera parameters (e.g., white balance) can make capturing complexion challenging. In this paper, we evaluate complexion as a soft biometric using color (histograms) and texture (local binary patterns). We train a linear SVM to distinguish between the individual and impostors. We demonstrate the performance of this approach on a database of over 200 individuals collected to study biometrics in human-robot interaction. In our experiment, we identify 9 individuals that interact with the robot on a regular basis, rejecting all others as unknown.

1. Introduction

An autonomous robot must be able to learn the identities of individuals with whom it interacts. Identity is important for a number of reasons. For example, a robot should be able to remember preferences, interaction histories, etc. When identity is first learned, we anticipate the person to be generally cooperative, but may be speaking, changing head pose and facial expression. A robot may be expected to operate in uncontrolled environments, where the lighting during evaluation may not be the same as it was during enrollment.

Our goal is for autonomous systems to be able to meet and remember new people, in a way that is much like humans. For this to be possible, all biometric signatures must have a short enrollment time, preferably real-time. Enrollment must occur completely autonomously. Biometric signatures should be invariant to changes in illumination (e.g., outdoor, indoor, etc.). Finally, we expect both a high level of accuracy and permanence.

*Prior affiliation

In this paper, we examine the use of complexion as a partial solution to this problem. Complexion is a *soft biometric*[4]. Soft biometrics lack the distinctiveness and permanence of traditional modalities (e.g., face recognition, fingerprint recognition, etc.). However, they can be a powerful way of identifying people when combined together [12, 13, 8, 2]. In complexion, we measure the color appearance of the head, which can include skin color, hair color, facial hair color, etc. Complexion has nice properties that make it ideal for an autonomous robot. Complexion does not require high resolution, is not greatly affected by changes in pose, and can be both learned and evaluated quickly.

Although it has not been studied extensively as a soft biometric, complexion appears to be one of the more important cues that humans use for recognition. Sinha et al. [14] demonstrated that pigmentation cues provide as much distinctiveness as shape cues. This effect becomes even more pronounced when dealing with faces that have reduced resolution, which can obscure shape. Pascalis and Schonen [10] demonstrated that the contrast between hair and skin is one of the more important feature used by neonates to recognize their mothers. By obscuring the hair, they effectively reduced the neonates ability to distinguish between their mother and a stranger. For neonates, this cue is even more important than intrinsic facial features.

The major challenge is providing invariance to changes in illumination and camera parameters, which can greatly affect the accuracy of identification. Changes in illumination are caused by differences in the position and orientation of the light source(s) in the environment. For example, shadows cast on the face can make the complexion seem darker than it actually is. Prior work in skin detection has demonstrated that the Hue / Saturation (HS) color space provides a measure of invariance to shadow. Changes in the camera parameters can affect the white balance of the camera. This can greatly affect the measured color from the cameras. We can compensate for this by measuring the observed texture on the face, using local binary patterns.

In this paper, we capture complexion using both Hue-

Saturation color histograms and center-symmetric local binary patterns (CS-LBP) [11]. We compare the performance of each approach, showing the benefits of each in different situations.

Figure 1. The MDS robot used in this experiment,

Our robotics platform is an Mobile-Dextrous-Social (MDS) robotics named Octavia. Octavia is equipped with an RGB-D sensor using a pair of color cameras in her eyes and a SwissRanger SR-4000 camera in the forehead. The SR-4000 is used to detect individuals in the environment, so that Octavia's eyes (i.e. RGB cameras) look directly people when she is interacting with them.

To test identification in a human-robot interaction scenario, we designed an experiment where the public was asked to interact with Octavia over a 5 day period. We had over 200 individuals participate, 9 of which were experimenters that interacted with the robot on a regular basis. We use complexion to identify the experimenters that interact with the robot regularly, while rejecting everybody else as unfamiliar.

The remainder of the paper is organized as follows. We report on related work in section 2. We introduce methodology in section 3, then show experimental results in section 4. We conclude in section 5 with a discussion of our results and future directions.

2. Related Work

Complexion is closely related to the problem of detection skin in images. Zarit et al. [16] studied skin detection in five different color spaces: normalized RGB, YCrCb, HSV, Fleck HS, and CIE LAB. As has been consistent with other more recent works (e.g., [5]), Zarit found that Hue/Saturation/Value (HSV) tends to perform slightly better than or comparably to many other metrics.

One important question is invariance to lighting. Some [9] have proposed that the illuminance (or Value) channel is most affected by shadow, and that removing this component can increase performance. However, Khan et al. [5] demonstrated that while this may be true, removing the illuminance channel can also decrease skin detection performance. From the perspective of complexion as a soft biometric, this introduces an interesting trade-off. Without this component (i.e., using Hue/Saturation - HS), we may be invariant to illumination conditions, however, with this component accuracy may increase. In our case, lighting invariance is of critical importance, so we choose to use the HS color space. When the lighting is controlled, HSV may be more appropriate.

Complexion as a soft biometric is also closely related to color based object recognition [6], and detection [15]. Recent works by Zhu et al [17] have suggested the use of local binary patterns to handle changes in lighting. LBPs are computed using a difference in values from the center pixel, which implicitly handles a situation of a linear change in color intensity [17]. However, CS-LBPs alone do not describe the color of the object, they simple describe the relative difference between color pixels. While this makes it an effective way of describing texture, its not appropriate to distinguish between different colored objects that may, perhaps have a similar texture.

Local binary patterns can also be used to recognize faces [1]. In this work, the authors explored the tradeoff between the radius and sampling size of the LBP as well as the size of the extracted patches. As noted by the authors, as the sizes of the patches increase, the spatial information becomes increasingly lost. While this is a negative for face recognition, it may be a positive for the use of LBP as a soft biometric. By using the face as a whole, we may be able to capture more holistic features capturing information such as the general shape of the face and the contrast between features.

Some prior works have addressed the question of complexion as a soft biometric from the point of view of ethnicity [7, 3] or using categorical labels (e.g., light or dark skinned) [2]. While these works have shown promise, they remove a great deal of distinctive information that is present due t o the ranges of skin color that are possible within ethnic groups.

3. Methodology

In this section, we discuss our approach to modeling complexion using both color (section 3.3) and texture (section 3.2). However, we begin with a brief overview of the problem of modeling color and the challenges of color in the presence of illumination and shadow.

3.1. Color Invariance

The images in figure 2 illustrate a typical problem: variation in illumination conditions. In some cases, parts of the

face are cast in shadow, in other cases specular reflections are caused by different lighting conditions. In this section, we review color invariance as described by Van De Sande et al. [15].

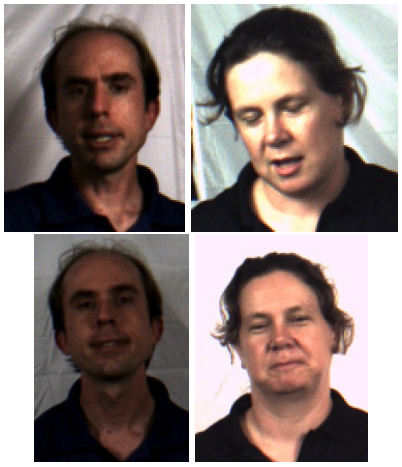

Figure 2. Complexion can vary greatly due to different illumination conditions. The figure above shows the same individuals in different lighting conditions. Some lighting changes were due to specular reflections (right); others were due to shadow (left).

Changes in color values can be classified into 6 different categories. The first category (*light intensity change*) can model the effect of shadows by providing a linear scaling of each color channel.

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \quad (1)$$

The second category (*light intensity shift*) models the intensity at which light appears at a pixel, which can model specular reflections.

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix} \quad (2)$$

The third category (*light intensity change + light intensity shift*) models both together:

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix} \quad (3)$$

Note that in Eq. 1 - 3 each channel is scaled by the same factor, a or adjusted by the same amount o_1. Scaling each channel by different factors a, b, c and/or adjusting by different amounts o_1, o_2, o_3 can adjust different channels in different ways. For example, the red channel can be scaled by $a = 1.2$, whereas the other channels are not scaled at all ($b = c = 1.0$). Such situations are common in white balancing, where the color space can be slightly different in time 1 and time 2.

From figure 2 it is clear that both shadow and specular reflection are common problems , but can be adequately described by eq. 1-3. When images are taken at different points of time, it is possible that the camera will be white balanced differently, which will cause the color channels to vary indepedently.

The RGB color space is not invariant to any scaling or adjusting factor, which makes it unsuitable for modeling complexion. Prior work ([15]) has shown that the HS color space is tolerant to light intensity change, light intensity sift, and a combination of the three (i.e., eq. 1-3).

Invariance to adjusting and scaling by different factors is much more difficult. On potential solution is the center symmetric local binary pattern (CS-LBP) [17, 11]. LBPs are computed by comparing the difference between a pixel values. Scaling by the same factor does not impact this decision.

3.2. Local Binary Patterns

Local binary patterns [11] are a flexible and effective way of describing the texture. The flexibility of LBPs are due to the various ways in they can be formulated.

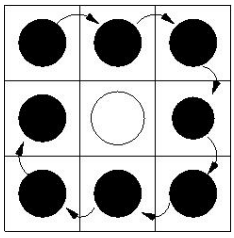

Figure 3. Local binary patterns are computed by comparing the center pixel (white circle) to surrounding pixels (black circle).

The canonical local binary pattern (see Figure 3) is computed by comparing the value of the center pixel to the surrounding pixels. This comparison results in an 8 bit number, where each comparison results in a 1 if the surrounding (black) pixel is greater than the center (white) pixel. Intuitively, each number represents a different kind of texture. LBP=0 when there is a flat region of no texture, LBP=255 when there is a spot, other numbers represent corners, edges, etc.

The basic LBP used a radius of 1, it is possible to extend this out by comparing values at other radii. For example, when radius=2, we look at neighbors that are 2 pixels away. This can result in either a 16-bit or an 8-bit number. However, we use an 8-bit number by interpolating between values when a comparison falls between several possible pixels.

To summarize the texture of a region, we accumulate values into a histogram, where each bin describes a different

kind of texture. By looking at smaller regions, we can find a more local description of texture. In our experiments, we divide the face into the top and bottom, summarizing the textures present in each separately.

Hue-Saturation local binary patterns (HS-LBP) are computed by finding the local binary pattern on each channel independently, then concatenating the resulting histograms. In [1] a similar approach was used for face recognition, where it was shown experimentally that radius=2 sampling along 8 different bits provides a great deal of effectiveness.

3.3. Histograms

We detect faces using the Pittsburgh Pattern Recognition SDK. We resize the detected faces to a normalized size of 10 pixels interocular. Typical examples are shown in figure 4. We compute the color histograms for the Hue, Saturation channels separately, using 10 bins for each channel. We concatenate the two histograms resulting in 20 different color features for each image.

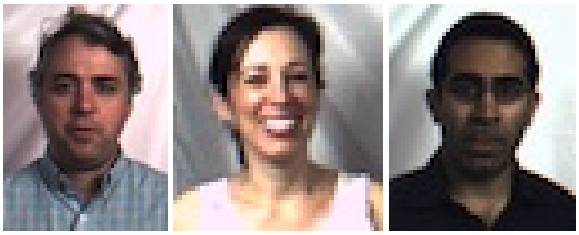

Figure 4. Low resolution images are used to learn the complexion of participants.

3.4. Classification

For both the purposes of efficiency and classification accuracy, we use a linear support vector machine to identify individuals using the features obtained either through color histograms or through color local binary patterns.

Given the set of features x_i and the associated label y_i, a two class problem is solved using a linear discriminant a to project a pattern x

$$g(y) = a^t x \tag{4}$$

An unknown pattern will be classified using a, where the resulting prediction is based on the the sign of g. Data is usually situated in a manner where there are a great many possible hyperplanes that will solve the problem in the manner outlined above. The most obvious solution may be a pair of hyperplanes, one of which lies on the positive decision boundary, the other on the negative decision boundary. The *margin* is the distance between these two hyperplanes. The most general solution will be the one that maximizes this margin

$$\frac{z_k g(y_k)}{||a||} \geq b \tag{5}$$

maximizing b, subject to $b||a|| = 1$. *Support vectors* will be those points that lie on the boundary with the maximum margin. These represent the patterns that are most difficult to classify, the removal of these points from the training set will result in a different margin.

It is common to make use of a *kernel function* in cases where the data is not linearly separable. In this case, rather than comparing two patterns x_i and x_j directly, we compare $k(x_i, x_j)$, where the kernel k generally is a polynomial or an RBF kernel. This may result in an improved classification accuracy. However, the desire for efficient classification, we limit experimentation to a linear kernel.

4. Experimental Results

To evaluate biometrics for human-robot interaction, we designed a study in the form of a "game" where three individuals have a short interaction with the robot. The game is divided into two separate phases: a training phase, and an evaluation phase.

In the first phase, the robot needed to acquire a model and an identifier for each individual. For each of three prespecified 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, the robot readjusted the necessary motors 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 at least 0.1 rad.

After detecting the individual and adjusting pose, the robot first asked for an identifier. This was done as a question about the volunteers favorite flavor of ice cream, color, animal, planet, or sport. The answer provided by the volunteer was 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, their names were used to preserve consistency.

With an established identifier, the robot asked the volunteer to speak on a subject for up to 30-sec. Subjects were randomly selected from one of 5 topics. The robot 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.

Once the training phase was completed for all three volunteers, the robot 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, the robot said, Ready or not, here I come, and opened her eyes.

In the final phase, the robot again turned to face each of the locations in turn and tried to recognize the individual. As during training, the robot had to find the individuals 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, she 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. The purpose of asking for a hint was to encourage the subjects to provide facial expression and/or speech data, and was not used during the identification process.

After asking for a hint, the robot again looked at the subject for 20-sec, using all images where a face was detected to match the biometric signatures. At the end, she guessed the participants identity from the 3 possible subjects based on maximum likelihood. We have reported the results from this "shell game" in [8], where we made use of multiple soft biometrics to identify individuals. In this work, we limit the analysis to complexion information.

We evaluate open-world identification of 9 experimenters using complexion information alone. For the purposes of evaluation, we split those sessions that do not belong to the experimenters into 5 different folds. At each step, one fold is used to train the classifier, the remaining four are used for evaluation.

Experimental results are shown the ROC curve in figure 5. It's interesting to note the overall performance of complexion, as well as the fact that neither approach dominates the other.

4.1. Results in Transformed Color Space

To illustrate the problems introduced by different color balancing algorithms, we synthetically create a second data set, where the red channel in the test images are scaled by a factor of 1.2, while keeping the blue and green channels the same. The results of this experiment are in figure 6.

This produces some dramatically different results, but shows the effectiveness of LBPs.

We can draw several conclusions from these two experiments. First, in cases where color camera calibrations do not change, HS color histograms are an effective tool. However, over the long term, CS-LBPs are likely to be the more robust approach.

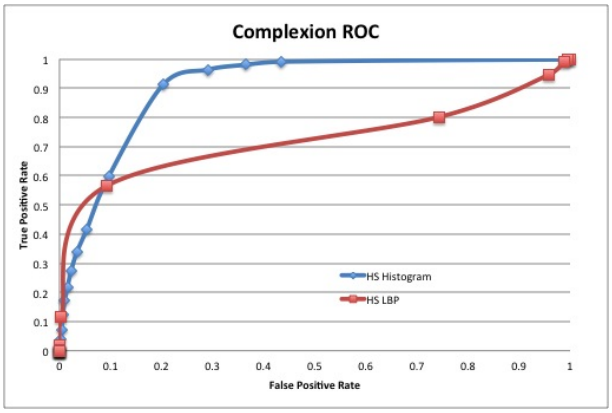

Figure 5. ROC showing performance of complexion using soft biometrics. In this experiment, we evaluate complexion by comparing the subject against 191 non-subjects using 5-fold cross-validation.

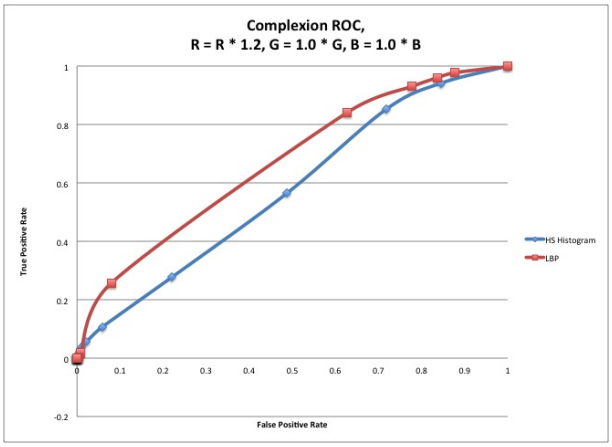

Figure 6. ROC showing performance of complexion using soft biometrics. In this experiment, we evaluate complexion by comparing the subject against 191 non-subjects using 5-fold cross-validation.

5. Discussion

We've shown in this paper the effectiveness of color for complexion analysis. The strength of the HS-color histogram lies in it's ability to distinguish between people in the same color space. When the color parameters are adjusted, it loses this ability. The CS-LBPs in this case perform quite well.

Complexion has a number of advantages, which we intend to fully investigate in future works. Among these advantages are the ability to recognize individuals in low resolution images. This provides the ability to recognize subjects at a distance in combination with other modalities (e.g., gait).

The biggest issue that we have observed is related to the color space changes produced by white balancing. This step has to be performed carefully, and ad-hoc approaches to

white balancing may not be sufficient.

Acknowledgment

This work was partially supported by the Office of Naval Research under job order number N0001407WX20452 and N0001408WX30007 awarded to Greg Trafton. Wallace Lawson was also funded by the Naval Research Laboratory under a Karles Fellowship.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[2] A. Dantcheva, C. Velardo, A. D'Angelo, and J. Dugelay. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51(2):739–777, January 2011.

[3] S. Hosoi, E. Takikawa, and M. Kawade. Ethnicity estimations with facial images. *IEEE International Conference on Automated Face and Gesture Recognition*, pages pp. 195–200, 2004.

[4] A. Jain, S. C. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition. *Proc. of the SPIE*, 5404:561–572, 2004.

[5] R. Khan, A. Hanbury, J. StÃttinger, and A. Bais. Color based skin classification. *Pattern Recognition Letters*, 33(2):157 – 163, 2012.

[6] F. Liu and M. Gleicher. Learning color and locality cues for moving object detection and segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[7] X. Lu and A. Jain. Ethnicity identification from face images. *Proceedings of SPIE*, 5404:pp. 114–123, 2004.

[8] E. Martinson, W. Lawson, and J. Trafton. Identifying people with soft biometrics at fleet week. *ACM / IEEE Intl Conference on Human-Robot Interaction*, 2013.

[9] S. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler. Tracking groups of people. *Computer Vision and Image Understanding (CVIU)*, 80(1):42–56, 2000.

[10] O. Pascalis and S. Schonen. Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, 16(1):pp. 79–85, January-March 1995.

[11] M. Pietikainen, G. Zhao, A. Hadid, and T. Ahonen. *Computer Vision Using Local Binary Patterns*. Springer London, 2011.

[12] D. Reid and M. Nixon. Inputting human descriptions in semantic biometrics. *Multimedia in Forensics, Security and Intelligence*, 2010.

[13] S. Samangooei, B. Guo, and M. S. Nixon. The use of semantic human description as a soft biometric. In *Proceedings of IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2008.

[14] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.

[15] K. van de Sande, T. Gevers, and C. Snoek. Evaluation of color descriptors for object and scene recognition. In *Computer Vision and Pattern Recognition, IEEE Conference on*, pages 1–8. IEEE, 2008.

[16] B. Zarit, B. Super, and F. Quek. Comparison of five color models in skin pixel classification. In *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on* [16].

[17] C. Zhu, C. Bichot, and L. Chen. Multi-scale color local binary patterns for visual object classes recognition. *International Conference on Pattern Recognition*, 2010.