The Emperor's New Masks: On Demographic Differences and Disguises

Katherine L. Gibson and Jonathan M. Smith
CIS Department
University of Pennsylvania
[gibsonk,jms]@cis.upenn.edu

Abstract

Remaining unrecognized in an era of ubiquitous camera surveillance remains desirable to some, but advances in face recognition technology make it increasingly difficult to do so. A large database of high-quality imagery was used to explore the effectiveness of disguise as an approach to avoiding recognition. A commercial system that was highly rated in NIST's Face Recognition Vendor Test[13] was used to evaluate a variety of disguises worn by each member of a study population that was diverse in age, gender, and race. Analysis of the recognition results for subsets extracted from the population shows that disguise can be remarkably effective. However, the efficacy of the disguises against face recognition varies so significantly with demographics that, for some, the disguises are not worth wearing.

1. Introduction

Biometric systems are useful for a variety of purposes, such as authentication [4], recognition [27] and forensics [10]. The ability to perform face recognition (FR) from a distance without the subject's cooperation or knowledge [7] makes FR an especially suitable biometric for use by security and law enforcement. As FR system performance has steadily increased [5, 21, 23, 13], FR use as a surveillance tool has become more common. While the problem of temporary occlusion (*e.g.*, self-occlusion and occlusion by objects) can be overcome via multiple cameras [14], deliberate occlusion is considerably more difficult to address. Even absent malicious intent, dark sunglasses, heavy bangs, and surgical masks[1] obscure parts of the face, hiding them from the FR system.

As biometric systems, particularly FR, have become more common, examining the effects of deliberate occlusion on system accuracy becomes more important. To that end, we used a unique new database of occluded faces, the DISGUISED FACE DATABASE (DFACED), in combination a high-performance commercial FR system to explore the effects of disguise [12].

We make three novel contributions. First, we sketch the capabilities of a new database of facial imagery, unique both in the diversity of the imaged population, and in the diversity and consistency of disguises applied to each subject. Second, we provide an analysis of disguise performance based on a highly rated commercial FR system. Third, while differences in the effectiveness of disguises have been observed for individuals [29], we used demographic data associated with our subject population to show that certain disguises perform poorly for entire racial and gender subpopulations in our study, giving rise to the paper's title.

2. Related Work

The negative effect of occlusion on face recognition has been heavily studied, with many algorithms developed to address the problem, and numerous databases created for testing the effects.

2.1. Algorithms

Algorithms tackling the problem of occlusion fall into one of two general areas: those that detect disguises, and ignore the affected area; and those that split the face into multiple smaller parts, expecting that only a fraction of the face will be affected by disguise.

An algorithm presented by Ramanathan, *et al*. [24] detects disguises by breaking the image into left and right halves, relying on the symmetry of the human face, before analyzing edge densities to choose the "better" half with which to continue. Min, *et al*.'s [19] approach begins similarly, but instead breaks the face into upper and lower halves before running local binary patterns (LBP) on each non-occluded portion, producing superior results compared to LBP alone. Dhamecha, *et al.* [9] divide the face image into *biometric* and *non-biometric* "patch" classes, before using LBP on the biometric patches.

Pavlidis and Symosek [20] have demonstrated that disguises can be easily detected in images captured in near-

[1] Surgical masks are often worn in Asia to inhibit the spread of illnesses; airline travelers have also worn masks as a precaution against Ebola [15].

infrared, as they reduce the face's natural thermal signal. Ahonen, *et al*. [1] do not attempt disguise detection; instead, they break the face into small regions, and use LBP to compute a description of each, before recombining them into a spatially enhanced histogram or feature vector. Wright, *et al*. [28] exploit sparsity to robustly handle occlusion (and image corruption), taking advantage of occlusions "corrupting" only a fraction of an image's pixels. An approach put forth by Martinez [17] compensates for possible distortions (including occlusion) by applying both localization and warping to each image before attempting recognition. Finally, Yang and Zhang [30] employ sparse representation-based classification (SRC), using it on the image Gabor-features; this not only reduces the computational time needed to code the face images, but also greatly improves the SRC accuracy.

2.2. Databases

The 1998 AR Face Database [18] was the first significant database of occluded facial images, although its disguises were limited to sunglasses and a scarf around the lower half of the face. Other databases have been created since, such as the National Geographic dataset [24], Singh, *et al*.'s synthetic data base [25], and the MASKS data base gathered by Alexander [3]. However, all of these databases have limitations, such as the number of disguises used, the small size of the subject pool, and/or the genuineness of the disguises.

Recently, Singh, *et al*. [8, 9] created the IIIT-Delhi Disguise Version 1 (ID V1) face database, which contains a wide variety of disguises, as well as images captured in both visible and thermal spectra. ID V1 is limited in some ways due to the size of its subject pool (75 individuals), and by the lack of a uniform disguise set across all subjects (ID V1 subjects were asked to disguise themselves, which provides more variation across individuals, but makes it challenging to evaluate a single disguise's performance across all subjects). The DISGUISED FACE DATABASE [12] overcomes these limitations, allowing more thorough testing of the effects of disguise against FR systems.

2.3. Demographic Variation in Automated Face Recognition

There is a considerable scientific literature on the "Other Race Effect" [6, 16], an observable bias in human recognition performance favoring members of the same racial group. Furl, *et al*. [11] raised the question of whether such biases affect computer algorithms for face recognition, a question answered affirmatively in a thorough experimental study by Phillips, *et al*. [22]. The experiments suggest that state-of-the-art machine recognition algorithms are affected by the demographics of the test population used to train the algorithms. These papers did not, however, address cross-couplings between disguises and demographics and their effects on automated face recognition system performance.

3. The DISGUISED FACE DATABASE

The images used in this paper came from the DISGUISED FACE DATABASE (DFACED), a database gathered at Penn. DFACED is notable for its incorporation of genuine disguises (as opposed to disguises added via digital post-processing of undisguised facial imagery), the wide variety of disguises, the consistent application of those disguises to every subject, and the diversity of its subject population across age, gender, and race.

DFACED is comprised of 325 individual subjects captured with high-quality cameras. The interpupillary distance of all images is in excess of 250 pixels. Subjects were also captured from five different angles, with the cameras spaced at 45 degree intervals. In this paper we restrict our analyses to images from the center, head-on camera.

3.1. Subject Diversity

The subjects are demographically diverse. Race, gender, and age are supplied by the subjects themselves to ensure accuracy, along with their individual ethnicity. The race options provided were Black (or African American), East Asian (Chinese, Japanese, Korean, *etc*.), South Asian (Indian), White (Caucasian), and Other. The "Other" race option was selected by 20 subjects, and included subjects who identified as Middle Eastern, Southeast Asian, or multiracial. Due to this inconsistent ethnic makeup, those subjects who identified as "Other" were excluded from our analysis. The remaining race and gender breakdowns can be seen in Table 1, along with the total number of subjects in each group.

One unusual characteristic of DFACED is the age distribution, which skews significantly younger than the general population: 78% of the subjects were between the ages of 18 and 30 at the time of capture, a consequence of drawing subjects from a population within easy reach of Penn. Although we do not believe this has had a major effect on our findings, it should be considered when evaluating results.

3.2. Disguises

DFACED incorporates six disguise components: an eyepatch, a surgical face mask, a baseball hat, an adhesive mustache, a domino mask, and sunglasses. The six disguises can be seen in Figure 1, along with examples of "As Arrived" and "Clean" images that were gathered for each subject.

We chose these disguise components to fit into two categories: isolation disguises (to isolate key facial features and areas of the face) and realistic disguises (to simulate attire commonly worn in public, though not necessarily as a disguise).

The **eyepatch** was chosen as an isolation disguise for the eyebrows and eyes (although eyepatches are sometimes worn in public due to eye injuries).

The **surgical face mask** was also chosen as an isolation disguise, although it too is sometimes worn in public [15]. The second and third variations (seen in Figure 1 as FM2 and FM3) were included to test the effect of major facial features (*i.e*., nose, chin) being available for recognition, while still covering the mouth and cheeks.

The **baseball hat** and **adhesive mustache** were chosen as realistic disguises; hats are often worn in public, and many men grow out (or remove) their facial hair during their lifetime. The adhesive mustache was also worn by female subjects, both for the sake of completeness and to allow future exploration of its effect on determined gender.

The **domino mask** is an isolation disguise: it allows testing on the importance of the area of the face surrounding the eyes, while leaving the eyes themselves available for recognition.

Finally, the **sunglasses** were chosen as a realistic disguise; not only are they commonly worn in public, they are also often used as an example when discussing FR, and are included in many other databases [8, 9, 18, 24]. We included four varieties (black/white frames and dark/clear lenses) in order to allow testing of both sunglasses and eyeglasses (*i.e*., clear lenses), and of high-contrast frames on all skin tones.

Where possible, especially for the isolation disguises, we used disguises with a bright green coloration. We knew the color would not blend in with a subject's hair or skin, and would also be relatively easy to remove in post-processing if necessary.

It was not feasible to image all possible combinations of the disguise components and their variations on the subjects, each of whom was asked to repeatedly change disguises and position themselves for a conistent image set (a complete image set required approximately one hour to collect). Therefore we chose what we believe to be a representative sample, which includes the disguise components and their variations (six individual components and 10 variations), as well as 12 disguises that use multiple components. In addition to these disguised images, a single "As Arrived" image was captured at the start of the session, and six "Clean" images were captured throughout the session, for a total of 35 images per subject.

Figure 1: Individual disguises and their abbreviations:
Top row:
AA – subject "As Arrived" with facial accessories (glasses, bangs, etc.) intact
Er – *subject's* right eye covered by eyepatch
El – left eye covered
EBl – left eyebrow covered using eyepatch half
EBb – both eyebrows covered
EBr – subject's right eyebrow covered
Middle row:
C – subject's face "Clean" of disguises
FM1 – surgical face mask covering from nose to chin
FM2 – face mask pulled down to below the nose
FM3 – face mask additionally pulled up to show chin
H – baseball hat
Hb – baseball hat worn backwards
Bottom row:
M – mustache
DM – domino mask
Sbd – sunglasses with black frames and dark lenses
Sbc – black frames and clear lenses
Swd – white frames and dark lenses
Swc – white frames and clear lenses
When used in combination with other disguises, FM1 (full face mask) and Sbd (black frames and dark lenses) are referred to as FM and S, respectively. The other face mask and sunglasses variations do not appear in combination with other disguises.

4. Evaluation of Disguise Imagery using Commercial Recognition Technology

Due to the nature of the DISGUISED FACE DATABASE, preprocessing was not performed on the images prior to evaluation. The disguises worn by the subjects caused automatic facial feature detection to be ineffective, as they covered the eyes and affected the perceived skin color. Manual marking of facial features would have been tedious, with over a quarter of the images involving one or both of the eyes in occlusion. Especially for those images in which both eyes were occluded, even manual feature marking would have been difficult in addition to producing inaccurate results.

We contacted multiple vendors of highly-rated commercial face recognition systems, in attempts to acquire a wide range of FR systems on which to test; however, only one vendor could offer us access with minimal publication restricitions. We also sought to compare with non-commercial FR systems, but initial test runs of non-commercial systems produced results so poor in relation to those of the commercial system that comparison of the results would have little value.

As stated earlier, the commercial system we used performed well in analyses conducted as part of the 2013 FRVT Evaluation [13]. Using the commercial FR system, we attempted face detection on each image from the 325 subjects. A face was successfully detected in 70.6% of the available images. Failure to detect a face was more common on images with a high amount of disguise coverage, such as those seen in Figure 2. Images with low disguise coverage (*e.g.*, mustache, eyebrows, baseball hat) or none at all ("Clean" and "As Arrived") had very high rates of face detection.

For each of the images with a successfully detected face, we next performed verification, using that image as a probe; the gallery was populated with all of the probe subject's "Clean" images, and the test was run with a False Acceptance Rate (FAR) of .01%. (The commercial FR system we used employs a "matching threshold" to control the FAR, and returns all results for which its confidence is above the set threshold.)

Figure 2: From left to right: *Hat-Sunglasses-Face Mask* combination, with 0% successful face detection; *Sunglasses-Face Mask* combination, with successful face detection for only 1.2% of subjects; and *Hat-Domino Mask-Face Mask* combination, with faces detected for only 1.7% of subjects.

5. Results

From the recognition results, we calculated the percentage of successful matches for each disguise tested. For example, when subject 285 wore a face mask (FM1), it correctly matched against two of her six clean images, giving a correct match percentage of 33%. If a face was not detected for a disguise image, the result was automatically 0%. The inability to detect a face by default meant that image could not be successfully matched to any of the subject's "Clean" images.

The relationship between correct match percentage and disguise effectiveness is inversely correlated: the higher the match percentage, the more poorly the disguise performed. In other words, disguises with *low* match percentages are *more* effective at disguising the wearer – they are preventing the FR system from making a correct identification. We sorted the disguises from worst performance (100% successful matches) to best (0% successful matches), based on their effectiveness as averaged across all subjects.

For brevity's sake, we discarded from analysis those disguises that performed very poorly, with greater than 99% recognition rate: the baseball hat, eyebrow coverings, and mustache (Hb, EBb, EBl, H, EBr, and M). We also discarded those disguises that performed incredibly well, with a less than 2% recognition rate: S-FM, H-S-FM, DM-FM, and H-DM-FM. The latter three (*Hat-Sunglasses-Face Mask*, *Domino Mask-Face Mask*, and *Hat-Domino Mask-Face Mask*) were "perfect" disguises, with a 0% recognition rate.

We then grouped the subjects' results by race and gender, and calculated each demographic group's average performance for each of the 18 remaining disguises, discarding the "Other" racial group as previously mentioned. The raw numbers for these results can be seen in Table 1, broken down by race/gender and disguise.

6. Discussion

In this work, we presented a new DISGUISED FACE DATABASE for use in evaluating the effects of disguise on face recognition, and performed analysis on this database using a commercial system. The results of the disguise performance were then averaged for each demographic group (broken down by race and gender) as presented in Table 1.

6.1. Full Data Set Analysis

We first analyze the performance of the disguises generally, over the full data set. We can clearly see that as overall disguise coverage increases, recognition rate falls. This follows conventional wisdom: as the amount of the face available for recognition decreases, so does recognition rate.

Supporting this, eyebrow coverage alone has essentially no impact on recognizability, with recognition rates of 99.3% or higher for every demographic group. However, eyebrows have been found to be one of the most important features for face recognition by humans. [26] Given this vast difference in the feature's importance to human and computer face recognition, it is evident that disguise elements which confound a human observer may not be similarly effective against an FR system.

It is also observable that two or more disguises in combination can perform far better than either of the disguise components alone, or even better than the sum of the two. For example, sunglasses with black frames and dark lenses

(Sbd) have a 72.3% overall correct recognition rate, while the mustache (M) has a recognition rate of 99.7% (high enough that it is excluded from the results in Table 1). However, the combination of the two (S-M) has a recognition rate of only 36.1%, much lower than either of the components alone would suggest. Similar findings can be seen for the combinations of H-FM, H-S, S-M, and DM-M.

Finally, recognition rates vary widely between subjects, including those belonging to the same demographic group, and even subjects of the same age. As an example, subject 156 has a recognition rate of 81.5% over all disguises, but subject 193 has an overall recognition rate of only 53.6%, despite both subjects being "identical" demographically. This may be due to inherent disguisability, as discussed by Yager and Dunstone [29], although it may also be caused by the imprecise demographic grouping of "race," which we intend to address in future research.

6.2. Demographic Analysis

In addition to these overall trends, the performance of certain disguises varies widely between demographic groups as seen in Table 1. We will examine in detail a few of these differences in performance, some examples of which can be seen in Figure 3.

We calculate the statistical significance of these differences using the Mann-Whitney-Wilcoxon test, as the observations are independent, and we expect that the underlying distribution of results is non-normal. Sample sizes can be seen in parentheses next to each group's label in Table 1.

Looking first at recognition rates for the full-coverage surgical face mask disguise (FM1), we observe that recognition is lower for Black women and men (23.8% and 41.7%, respectively) than for any other group, making it an effective disguise for them. In comparison, FM1 fails as a disguise for White women and men, with high recognition rates (79.5% and 80.7%); this difference in performance is statistically significant, with $p < 0.001$.

We can observe similar effects with the EBb-FM (full cover surgical mask with both eyebrows covered) disguise, which performs exceptionally well (0.8% and 0%) on Black women and men, in contrast with its less satisfactory performance on White women and men (37.1% and 38.8%), with $p < 0.001$.

Conversely, the domino mask and mustache disguise (DM-M) performs relatively well across all demographic groups (20% overall), but is particularly effective on White women and men (1.3% and 11.8%), while failing to sufficiently disguise Black women and men (64.3% and 60%, respectively), with $p < 0.001$.

Likewise, the Sbd (black framed, dark lensed sunglasses) disguise leads to a decline in recognition rate across all demographic groups, but this decline is especially pronounced for White women and men (46.2%, 60.1%), with $p < 0.001$.

Figure 3: Effective and Ineffective Disguises
First row: (Effective) The *sunglasses* (Sbd) and *Domino Mask-Mustache* (DM-M) disguises work well for White women and men.
Second row: (Ineffective) The *Face Mask* (FM1) disguise is ineffective on White women and men, just as the *Sunglasses-Mustache* combination is for South Asian women and men. The *Eyebrows-Face Mask* combination (EBb-FM) also performs poorly as a disguise for White women and men.

Although the effects of contrast are most conspicuous when examining the Black and White demographic groups, differences can also be seen for certain disguises in the South Asian group, as well as between the two genders. For example, the S-M (sunglasses and mustache) disguise generally works well, with a correct recognition rate of only 28% across all demographic groups. However, South Asian women and men are far more recognizable while wearing it (71.4% and 60.4%), with $p < 0.001$. Finally, within every racial group, the H-S (baseball hat and sunglasses) disguise is less effective when worn by men, increasing successful recognition by a minimum of 10%. However, the significance of these results varies widely: for Black women and men, $p = 1.873$; for East Asians, $p = 0.0369$; for South Asians, $p = 0.4309$; and for Whites, $p < 0.001$.

We hypothesize that many of these divergences in performance are caused by differences in the level of contrast between a disguise and the subject's skin tone. As put forth by Alexander [2], in addition to a decline in accuracy due to a disguise occluding the features, recognition is further degraded by a disguise that is in high contrast with the surrounding area of the face.

Demographics (# of subjects)	Swc	Er	H-M	FM2	Sbc	FM3	El	FM1	Sbd	Swd	DM	H-DM	EBb-FM	H-S	S-M	DM-M	H-FM	H-S-M
Black ♀ (21)	100	89.7	100	79.4	100	84.9	88.9	23.8	81.0	76.2	79.4	57.9	0.8	37.3	47.6	64.3	0.0	19.8
Black ♂ (10)	100	86.7	88.3	80	100	70.0	76.7	41.7	73.3	66.7	95.0	61.7	0.0	65.0	23.3	60.0	20.0	6.7
S Asian ♀ (14)	95.2	92.9	92.9	100	100	100	100	92.9	78.6	71.4	54.8	33.3	44.0	28.6	71.4	31.0	13.1	0.0
S Asian ♂ (24)	95.8	98.6	95.8	95.8	99.3	83.3	84.7	71.5	70.8	86.8	62.5	53.5	2.1	40.3	60.4	46.5	3.5	16.7
E Asian ♀ (37)	100	98.2	89.2	97.3	96.8	97.3	97.3	83.8	86.5	71.6	92.3	61.7	44.4	27.0	18.9	30.4	20.3	2.7
E Asian ♂ (22)	100	100	96.2	100	87.9	95.5	95.5	54.5	82.6	76.5	89.4	46.2	28.8	54.5	27.3	26.5	4.5	11.4
White ♀ (88)	95.8	94.1	93.4	98.9	85.6	90.9	88.6	79.5	46.2	28.2	28.8	21.6	37.1	12.9	14.4	1.3	25.0	0.9
White ♂ (89)	95.3	95.5	97.6	92.1	89.1	87.6	86.5	80.7	60.1	51.5	36.5	41.0	38.8	36.0	25.7	11.8	27.6	26.2
Overall (305)	96.9	95.2	94.8	94.6	91.4	89.8	89.4	73.5	64.5	54.9	52.6	40.8	31.7	30.6	28.0	20.9	19.6	12.0

Table 1: Percentage of Correct Matches Against "Clean" Images (Broken Down by Race/Gender and Disguise): The darker the shading of individual cells, the lower the correct match percentage (*i.e.*, the more effective the disguise). The final row gives the overall performance across the entire database (minus subjects whose chosen race was "Other"). Explanations of the disguise abbrevations can be found in the description of Figure 1.

However, the last two results we examined (Sbd's performance on South Asian women and men, and the intra-racial gender-based performance differences of H-S) cannot be easily explained through contrast, and warrant further and more nuanced study. These results may be due to anomalies in the data, but it is also possible that the FR system used may weigh certain facial features more heavily than others, or that the intra-racial differences in skin tone are affecting the outcome.

7. Conclusion

We described a substantial new database of facial imagery, the DISGUISED FACE DATABASE (DFACED), characterized both by the diversity of the imaged population and the number of example disguises applied consistently to each and every member of the subject population. DFACED was used to perform a data-driven analysis of the performance of disguises using matches by a highly rated commercial FR system as a benchmark.

The analysis showed substantial differences in the comparative effectiveness of the disguises across the entire subject population, but when the population was broken into subgroups based on the subjects' demographic data, a substantial skew associated with disguise effectiveness began to emerge, such that some disguises perform poorly for entire subpopulations (such as those who indicated specific racial and gender identities) in our study. The data suggest that for the members of these groups, the effect of some disguises is minimal, to the point of ineffectiveness.

8. Future Work

The demographic results were surprising, and while many can be explained by the contrasts between skin tones and disguises, other results lack such a clear and simple explanation. These results may stem from biases in training recognizers (see Section 2.3) or from other subtle interactions between demographics and recognizers, and are worth further investigation.

Rather than rely on the racial labels self-designated by each subject, we intend to investigate a potentially more accurate strategy, that of organizing the subjects by skin tone. This strategy has several advantages, amongst the more important of which are that quantitative models can emerge, intra-subgroup variance can be accommodated, and individual variance (*e.g.*, tanned in the summer, versus pale in the winter) could also be modeled as part of the skintone palette.

Given the importance of the recognizer in our analysis, we chose a very capable commercial FR system for evaluating disguise effectiveness. However, this system, as with most commercial products, is not designed to contend with adversarial behaviors, such as disguises. We are now investigating algorithms specifically designed to be robust against occlusions, including algorithms created by academics [9, 17, 20, 30].

Finally, while DFACED uses real disguises on real people, there remains the possibility of extracting disguises and digitally applying them to subjects not present in the database. This could have many applications, such as using the images with digitally applied disguises as training data to improve recognition when real disguised images are not readily available.

9. Ackowledgments

This work has been partially supported by the Olga and Alberico Pompa Professorship, by the National Science Foundation under Award CNS-07-16552, by the Office of Naval Research under Award Number N000141210757, and by the Defense Advanced Research Project Agency (DARPA) and Space and Naval Warfare Systems Center Pacific under Contract No. N66001-11-C-4020. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Office of Naval Research, the Defense Advanced Research Project Agency or Space and Naval Warfare Systems Center Pacific.

The authors thank the anonymous reviewers for their valuable comments, insights, and suggestions for improvements; the paper is better as a result of your efforts. We are also grateful to João Sedoc for his advice regarding appropriate statistical measures and analysis. Finally, a heartfelt "Thank You" to the subjects imaged for the DISGUISED FACE DATABASE, without whom this work would not have been possible, particularly those subjects who allowed unrestricted use of their images in publications.

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*, 2006.

[2] J. Alexander and J. M. Smith. Engineering Privacy in Public: Confounding Face Recognition. In *Privacy Enhancing Technologies: Third International Workshop*, 2003.

[3] J. M. Alexander. *MASKS: Maintaining Anonymity by Sequestering Key Statistics*. PhD thesis, University of Pennsylvania, 2009.

[4] Authentication Technologies and Their Privacy Implications, Committee On. *Who Goes There?: Authentication Through the Lens of Privacy*. The National Academies Press, 2003.

[5] D. M. Blackburn, J. M. Bone, and P. J. Phillips. Face Recognition Vendor Test 2000 Evaluation Report, February 2001.

[6] R. K. Bothwell, J. Brigham, and R. Malpass. Cross-racial Identification. *Personality & Social Psychology Bulletin*, 15:19–25, 1989.

[7] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-Invariant 3D Face Recognition. *Lecture Notes in Computer Science*, 2003.

[8] T. Dhamecha, A. Nigam, R. Singh, and M. Vatsa. Disguise detection and face recognition in visible and thermal spectrums. In *Biometrics (ICB), 2013 International Conference on*, 2013.

[9] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar. Recognizing Disguised Faces: Human and Machine Evaluation. *PLoS ONE*, 2014.

[10] Forensic Science Community, Committee on Identifying the Needs Of. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, 2009.

[11] N. Furl, P. J. Phillips, and A. J. O'Toole. Face Recognition Algorithms and the Other-Race Effect: Computational Mechanisms for a Developmental Contact Hypothesis. *Cognitive Science*, 26:797–815, 2002.

[12] K. L. Gibson, V. Shankar, B. Dolhansky, J. M. Smith, and B. Taskar. DFaceD: A Disguised Face Database. Technical Report #MS-CIS-15-03, University of Pennsylvania, 2015.

[13] P. Grother and M. Ngan. Face Recognition Vendor Test (FRVT), Performance of Face Identification Algorithms, 2014.

[14] J. Harguess, C. Hu, , and J. K. Aggarwal. Fusing Face Recognition from Multiple Cameras. 2009.

[15] P. LeBeau. Worried fliers wear masks as precaution against Ebola, Oct. 2014.

[16] R. Malpass and J. Kravitz. Recognition for Faces of Own and Other Race Faces. *Journal of Personality and Social Psychology*, 13:330–334, 1969.

[17] A. M. Martinez. Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2002.

[18] A. M. Martinez and R. Benavente. The AR Face Database. Technical Report #24, CVC, 1998.

[19] R. Min, A. Hadid, and J. Dugelay. Improving the recognition of faces occluded by facial accessories. In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, 2011.

[20] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system. In *Computer Vision Beyond the Visible Spectrum: Methods and Applications, 2000. Proceedings. IEEE Workshop on*, 2000.

[21] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone. Face Recognition Vendor Test 2002, March 2003.

[22] P. J. Phillips, F. Jiang, A. Narvekar, J. Ayyad, and A. J. O'Toole. An Other Race Effect for Face Recognition Algorithms. Technical Report NISTIR 7666, NIST, February 2010.

[23] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 Large-Scale Results, 2007.

[24] N. Ramanathan, R. Chellappa, and A. Roy-Chowdhury. Facial similarity across age, disguise, illumination and pose. In

Image Processing, 2004. ICIP '04. 2004 International Conference on, 2004.

[25] R. Singh, M. Vatsa, and A. Noore. Face recognition with disguise and single gallery images. *Image and Vision Computing*, 2009.

[26] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. In *Proceedings of the IEEE*, 2006.

[27] Whither Biometrics Committee. *Biometric Recognition: Challenges and Opportunities*. The National Academies Press, 2010.

[28] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.

[29] N. Yager and T. Dunstone. The Biometric Menagerie. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.

[30] M. Yang and L. Zhang. Gabor Feature Based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision - ECCV 2010*, Lecture Notes in Computer Science, pages 448–461. 2010.