

Performance Evaluation of Face Recognition using Visual and Thermal Imagery with Advanced Correlation Filters

Jingu Heo, Marios Savvides, and B.V.K. Vijayakumar

Department of Electrical and Computer Engineering
Carnegie Mellon University, U.S.A

jheo@cmu.edu, msavvid@ri.cmu.edu, kumar@ece.cmu.edu

Abstract

This paper presents the face recognition performance evaluation using visual and thermal infrared (IR) face images with correlation filter methods. New correlation filter designs have shown to be distortion invariant and the advantages of using thermal IR images are due to their invariance to visible illumination variations. A combined use of thermal IR image data and correlation filters makes a viable means for improving the performance of face recognition techniques, especially beyond visual spectrum. Subset of Equinox databases are used for the performance evaluation. Among various advanced correlation filters, minimum average correlation energy (MACE) filters and optimum trade-off synthetic discriminant function (OTSDF) filters are used in our experiments. We show that correlation filters perform well when the size of face is of significantly low resolution (e.g. 20x20 pixels). Performing robust face recognition using low resolution images has many applications including performing human identification at a distance (HID). The eyeglass detection and removal in thermal images are processed to increase the performance in thermal face recognition. We show that we can outperform commercial face recognition algorithms such as FaceIt[®] based on Local Feature Analysis (LFA).

1. Introduction

Despite a significant level of maturity with a few practical successes, face recognition is still a highly challenging task in pattern recognition and computer vision [1][2]. The performance of face recognition systems varies significantly according to the environments where face images are captured and according to the way user-defined parameters are adjusted in different applications [3][4][5][6]. Since a face is essentially a 3D object, lighting sources from different directions may dramatically change visual appearances (e.g. due to self-shadowing and specular reflections). Face recognition accuracy degrades quickly when the lighting is dim or does not uniformly illuminate the face [7].

Illumination tolerant face recognition is an essential part of any face recognition[8][9] system as even when we consider the best case scenarios of operation where the user is co-operative and provides the recognition system with a suitable face pose; in most cases the user does not have control of the lighting conditions in the surrounding environment. Currently there are many algorithms that have been developed with the aim of handling visual based face recognition in the presence of illumination variations [10][11][12][13] are a just few.

Unlike using the visible spectrum, recognition of faces using different multi-spectral imaging modalities, in particular infrared (IR) imaging sensors [14][15][16][17][18][19] has become an area of growing interest. Thermal IR and particularly Long Wave Infra-Red (LWIR) imagery is independent of illumination since thermal IR sensors operating at particular wavelength bands measure heat energy emitted and not the light reflected from the objects. More importantly IR energy can be viewed in any light conditions and is less subject to scattering and absorption by smoke or dust than visible light. Hence thermal imaging has great advantages in face recognition under low illumination conditions and even in total darkness (without the need for IR illumination), where visual face recognition techniques fail. However, thermal imaging needs to solve several challenging problems. Thermal signatures can be changed significantly according to different body temperatures caused by physical exercise or ambient temperatures. Thermal images of a subject wearing eyeglasses may lose information around the eyes since glass blocks a large portion of thermal energy. Thermal imaging has difficulty in recognizing people inside a moving vehicle. A combined use of visual and thermal face recognition system can alleviate the problems caused by only one modality face recognition system.

The use of correlation filters as a biometric verification is expanding due to their efficiency and robustness. However, relatively less effort has been made to demonstrate their performance. In this paper, in an attempt to show the effectiveness of face recognition with correlation filters, we evaluate the performance of visual and thermal imagery over several different face recognition algorithms including Principal Component Analysis (PCA) [20], normalized correlation, and Local

Feature Algorithm (LFA). Figure 1 show the framework experimented in this paper. Co-registered visual and long-wave infrared (8-12 μm) images acquired from the Equinox databases are used for the performance evaluation.

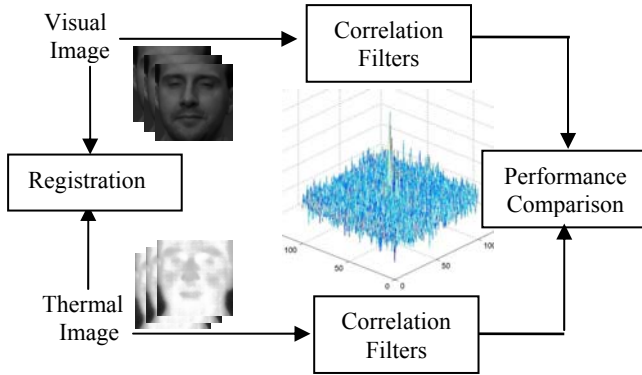


Figure 1: A framework for the performance evaluation

2. Background

One of the first correlation filters is the known synthetic discriminant function (SDF) [21] filter which is simply a weighted sum of matched filters corresponding to the training images. These weights are chosen so that the correlation peak resulting from cross-correlating this filter with any of the training images for this class is equal to a nonzero constant (e.g., 1) and the correlation peak due to other classes is zero. However, these filters only control a single point in the correlation plane (namely, the origin), leading to correlation outputs where sidelobes would be larger than the values at the origin, thus yielding false detections. This large sidelobe problem was addressed by the minimum average correlation energy (MACE) filter [22][23].

2.1 Minimum average correlation energy (MACE) filter

The MACE filter is designed to minimize the average correlation plane energy resulting from the training images, while constraining the value at the origin to certain pre-specified values. Correlation outputs from MACE filters typically exhibit sharp correlation peaks making the peak detection and location relatively easy and robust. The minimization of energy is done while satisfying the linear constraints that the correlation values at the origin resulting from the training images take on pre-specified values (stored in row vector \mathbf{u}), i.e.,

$$\mathbf{X}^+ \mathbf{h} = \mathbf{u} \quad (1)$$

Where \mathbf{X} is a $d^2 \times N$ complex matrix, where the i th column contains the 2-D Fourier transform of the i th training image lexicographically re-ordered into a column vector. Minimizing the average correlation energy while

satisfying Eq. (1) leads to the following closed form solution for the MACE filter \mathbf{h} .

$$\mathbf{h} = \mathbf{D}^{-1} \mathbf{X} (\mathbf{X}^+ \mathbf{D}^{-1} \mathbf{X})^{-1} \mathbf{u} \quad (2)$$

where \mathbf{D} is a $d^2 \times d^2$ diagonal matrix containing the power spectrum of training images along its diagonal and \mathbf{h} is a $d^2 \times 1$ column vector containing the 2-D correlation filter. The MACE filter in Eq. (2) yields sharp correlation peaks in response to training images and near-training images from the desired class and small output values in response to images from all the other classes.

2.2 Noise-tolerant correlation filters

In their attempt to produce sharp correlation peaks, MACE filters tend to amplify high spatial frequencies and any noise in the input. Thus, it is important to reduce the noise sensitivity of these correlation filters. Noise-tolerant correlation filters were first introduced in [24].

If the input images are corrupted by zero mean, additive, and stationary noise, then the output variance σ^2 is as follows:

$$\sigma^2 = \mathbf{h}^T \mathbf{C} \mathbf{h} \quad (3)$$

where \mathbf{C} is a $d^2 \times d^2$ diagonal matrix whose diagonal elements $\mathbf{C}(k,k)$ represent the noise power spectral density at frequency k . Minimizing the output noise variance in Eq. (1) subject the usual linear constraints leads to the following closed form solution.

$$\mathbf{h} = \mathbf{C}^{-1} \mathbf{X} (\mathbf{X}^+ \mathbf{C}^{-1} \mathbf{X})^{-1} \mathbf{u} \quad (4)$$

In many applications where the noise power spectral density is unknown, a good model is white noise which assumes $\mathbf{C} = \mathbf{I}$, the identity matrix.

2.3. Optimal tradeoff synthetic discriminant function (OTSDF) filter

A way to optimally trade off [25] between noise tolerance (achieved through low frequency emphasis) and peak sharpness (through high frequency emphasis) producing the optimal trade-off filter:

$$\mathbf{h} = \mathbf{T}^{-1} \mathbf{X} (\mathbf{X}^+ \mathbf{T}^{-1} \mathbf{X})^{-1} \mathbf{u} \quad (5)$$

where $\mathbf{T} = (\alpha \mathbf{D} + \sqrt{1 - \alpha^2} \mathbf{C})$, and $0 \leq \alpha \leq 1$. It is important to

note that when $\alpha=1$, the optimal tradeoff filter reduces to the MACE filter in Eq. (2) and when $\alpha=0$, it simplifies to the noise-tolerant filter in Eq. (4).

Varying α allows us to produce filters with optimal tradeoff between noise tolerance and discrimination. Typically using α values close to, but not equal to 1 (e.g., 0.99) improves the robustness of MACE filters.

2.4 Peak-to-Sidelobe Ratio (PSR) and Peak-to-Correlation Energy (PCE)

There are many performance measures for correlation filters. PSR and PCE will be one of the commonly used

methods to measure the correlation output. The peak-to-sidelobe ratio (PSR) can be formed by

$$PSR = \frac{peak - mean}{\sigma} \quad (6)$$

where peak indicates the correlation output and the mean and standard deviation (σ) comes from the sidelobe region surrounding the peak region.

Another peak-sharpness fitness metric is the peak-to-correlation energy (PCE) which is defined as the ratio of the correlation peak and the energy in the correlation plane as shown in Eq. 7.

$$PCE = \frac{|c(0,0)|^2}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |c(x,y)|^2 dx dy} \quad (7)$$

where $c(0,0)$ is the correlation peak and the denominator term is the energy in the correlation plane. More literature on advanced correlation filters can be found in [23][24][25].

3 Experiments

Equinox Corporation built an extensive database of face images using co-registered broadband-visible/IR camera sensors for experimentation and statistical performance evaluations [26]. A subset of Equinox databases are used for the evaluation of visual and thermal face recognition performances as shown in Table 1. The database consists of visual and thermal IR images of 3,244 (1,622 per modality) face images from 90 individuals. Images taken with frontal lighting conditions are used as the gallery images. Probe images were also divided into different conditions. Probe 1 to probe 3 indicate where individuals do not wear eyeglasses whereas probe 4 to probe 6 indicate individuals with eyeglasses. Since eyeglasses in thermal images block the thermal emission from the eye section and thus do not provide any information on the thermal images around the eyes, we separate the performance evaluation depending on the presence of the eyeglasses. The performance is measured via the first success rates (rank-1 recognition).

Table 1: The Equinox database of visual and thermal IR face images

Dataset	Visual (Thermal)	Eyeglasses	Lighting	Expression
Gallery	90(90)	Off	Frontal	Neutral
Probe 1	283(283)	Off	Frontal	Vary
Probe 2	370(370)	Off	Left	Vary
Probe 3	365(365)	Off	Right	Vary
Probe 4	177(177)	On	Frontal	Vary
Probe 5	172(172)	On	Left	Vary
Probe 6	165(165)	On	Right	Vary

Figure 2 shows different pairs of visual and its corresponding thermal image used in the experiments. Manually selected eye positions from the visual images are used for the normalization of faces not only visual images and but also thermal images. This is one of the challenges of using only thermal images as it is very hard to determine where the eyes are and thus the use of a co-registered camera which acquires both visual and thermal images is necessary to proper face normalization for scale, rotation using the eye locations. After the face normalization, histogram equalization is performed in visual images to compensate the effects of different lighting variations. Figure 3 shows an example of normalized faces acquired from both modalities.

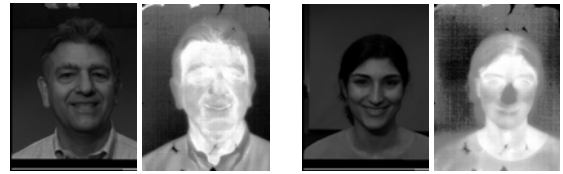


Figure 2: Examples of visual and thermal images



Figure 3: The normalized faces: visual (top row) and thermal (bottom row) images of the same individuals taken from different conditions

The performance is evaluated while varying the face size to very low resolutions. Robust low resolution face recognition is very important in any human identification at a distance system as in such scenarios the person is far away and only a low resolution face is available for recognition. Low resolutions also allow for using low-memory computing platforms and face recognition speed is faster with smaller image resolutions. As shown in Figure 4, different scales are applied on the face images varying from 12x12 pixels to 128x128 pixels.

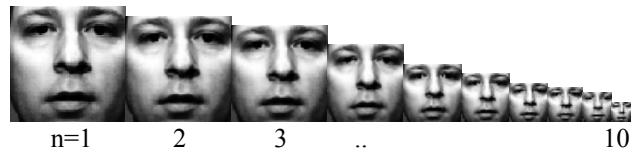


Figure 4: Different resolutions of the normalized face images; Image size from the left $s[n] \times s[n]$ pixels, where $s[n] = [128, 80, 64, 48, 32, 28, 24, 20, 16, 12]$

3.1 No eyeglasses

Figure 5 shows the performance of correlation filters (MACE and OTSDF). The first three probes in Table 1 are evaluated where individuals are not wearing eyeglasses. We also changed the size of gallery images for each person (ranging from 2 to 3 images acquired from probe 1 to probe 3), since the composition of correlation filters using different number of training images generates only one correlation filter per person and shows better performance than using only a single training image. The best performance is acquired from the thermal imagery using the OTSDF filter with 96.5% recognition rate when 3 training images are used while the visual face recognition performance shows at best 85.0%. The result shows that under variations such as illumination and expression, thermal face recognition can perform better than visual face recognition which is significantly affected by variations.

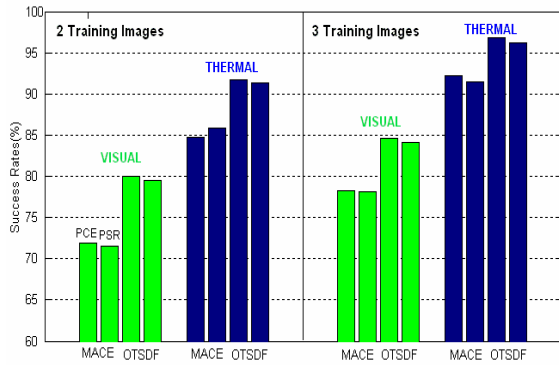


Figure 5 : Performance comparisons with correlation filters using visual and thermal face recognition (No eyeglasses)

Figure 6 shows the performance of the OTSDF filter with different scales. The larger face size does not lead to better performance due to redundancy information associated with the faces. The best performance acquired with the face size of between 20*20 and 32*32 pixels in thermal images whereas between 32*32 and 48*48 pixels required in visual images. Typically, face recognition systems require at least 100*100 pixels of face size to extract features from the face. However, our results show that OTSDF filter performs well when face size is extremely small making this method attractive for applications where low-resolution face images are available. Using lower resolution also means lower memory requirements and faster recognition speed. In an effort to show the effectiveness of correlation filters, we compared commonly used face recognition algorithms, Normalized Correlation, PCA and a well-known commercial face recognition software FaceIt[®], which has been highly ranked by the face recognition vendor test

(FRVT) [3][4][5] based on Local Feature Analysis (LFA) are also included in this experiment.

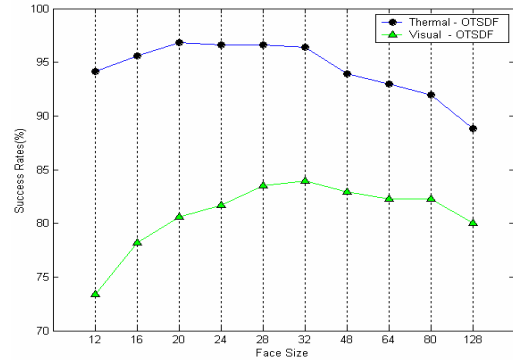


Figure 6: Performance comparison by changing face size.

Exactly the same training and testing conditions have been used throughout the performance comparison for all the algorithms. Table 2 shows the comparison results using a low-resolution face size (32*32 pixels) and a higher resolution face size of 128*128 pixels. FaceIt[®] gave robust performance results with visual and thermal images when face size is large enough, however for low resolution images (such as 32x32), OTSDF filter outperforms FaceIt[®] recognition algorithm. Throughout the performance comparison, thermal face recognition gives better performance over visual face recognition regardless of the different algorithms. Therefore, we can assume the thermal images provided by Equinox [26] (note that thermal noise is reduced by the use of radiometric calibration) show less variations under different conditions (expression, illumination and slight pose) than visual images, thus leading to better performance than visual face recognition.

Table 2 : Performance comparison using different face recognition algorithms; Recognition Rates (%), No eyeglasses

	Small face size (32*32) – 3 gallery		Big face size (128*128) – 3 gallery	
	Visual	Thermal	Visual	Thermal
OTSDF	84.10	96.84	79.98	88.83
PCA	60.85	90.38	55.53	91.38
Normalized Correlation	46.16	56.28	51.54	60.91
LFA (FaceIt [®])	82.33	85.66	94.54	96.22

3.2 Eyeglasses

Thermal images of a subject wearing eyeglasses may lose information around the eyes since the glass blocks a large portion of thermal energy emitted by the eyes as discussed before. Therefore, the eyeglasses region should

detected and removed for better performance in thermal based face recognition. We measured statistical temperatures of the facial components as shown in Figure 7. Figure 7 (a) and (b) indicate the thermal images averaged over 60 image samples of with or without eyeglasses from the Equinox database. Figure 7(c) shows the histograms of eyeglasses, the eyes, and the nose obtained from the averaged thermal face images. The eyeglass regions are the coldest areas in thermal images. The nose area is often associated with lowest temperature of a face when a subject does not wear glasses. The eyes are relatively warm regions, while the eyebrows, the mouth, and the cheeks are often all in similar temperature ranges. The sizes of the face components used in the experiment are 35×25 pixels for the eye and glasses, 24×35 pixels for the nose, and 150×120 for the face.

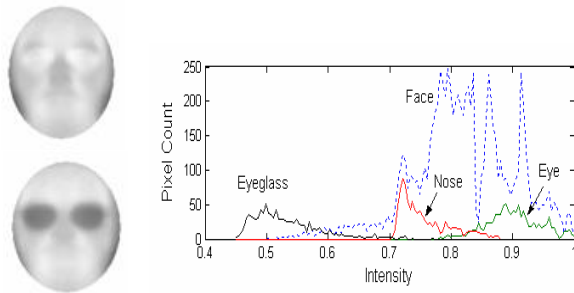


Figure 7: Average thermal images (no eyeglasses- upper left, eyeglasses- lower left) and histograms of different facial components.

A thermal image is binarized using a threshold which can be acquired from the statistical variations, provides data points for fitting with ellipses. Ellipse fitting can be used for face detection since the biggest ellipse is considered being associated with the face. Similarities of all the ellipses within the face region, or inside the biggest ellipse, are tested for possible eyeglass candidates. Among all the ellipses available, two ellipsoidal shapes of similar shape, size, and location are considered as candidate eyeglasses in thermal images. Detailed analysis and similarity measure can be found in [19].

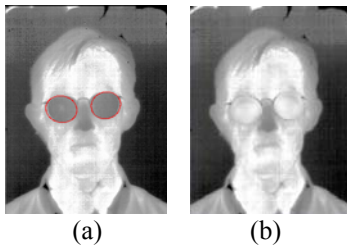


Figure 8: Replacement of eyeglass regions with template. (a) Result of eyeglass detection, (b) A face image replaced with an eyeglass template acquired from the average thermal image

We compared the performance using Probe 4-6 (Total 514 images) from the database described in Table 1. Table 3 shows the performance comparison results where individuals are wearing eyeglasses. We consider the best performance of different algorithms regardless of the size of the face. Visual face recognition is only slightly affected by the presence of the eyeglasses. FaceIt[®] shows almost similar performance regardless of the presence of the eyeglasses in visual images. Without the eyeglass removal, thermal face recognition shows a very low performance rates. After the eyeglass removal, overall performance is significantly improved in thermal face recognition regardless of the different algorithms.

Table 3 : Performance comparison using different face recognition algorithms; Recognition Rates (%), Eyeglass

	Visual	Thermal	
	N/A	Before Removal	After Removal
OTSDF	78.10	45.76	85.77
Normalized Correlation	50.11	44.72	50.68
LFA (FaceIt [®])	90.11	40.56	80.34

4. Summary and Discussion

This paper presents the performance evaluation of visual and thermal infrared (IR) images using correlation filters and also includes comparison with other face recognition algorithms. From the experimental results, correlation filters show the best performance at low resolutions (as low as 32×32 pixels) for both visual and thermal images. Low-resolution face recognition is important for any human-identification at a distance system where we typically only have low-resolution images to work with. Throughout the performance comparison, thermal face recognition showed higher performance than visual face recognition under various lighting conditions and facial expressions when no eyeglasses are present regardless of face recognition algorithms. Therefore, we can assume the thermal images provided by Equinox (thermal noise is reduced by the radiometric calibration) show less variations under different conditions than visual images, leading to better performance than visual face recognition. We show that eyeglass can be detected in thermal images using the ellipse fitting method and replaced with template eye patterns to minimize the effect of eyeglass. Eyeglass replacement significantly improved the recognition accuracy in thermal images. In contrast, eyeglasses only slightly affected the visual based face recognition algorithms.

The advantage of using advanced correlation filter methods is that we did not use any images from other

classes to build the filters. Each filter only uses one or two images from each person to build his/her filter. Thus this approach is better suited for scaling of the system as it does not require re-training with the addition of a new person in the system.

Since we do not consider the passage of time between gallery images and probe images, it is questionable thermal face recognition always performs better than visual face recognition. A combination of visual and thermal face recognition may be more viable means for face recognition, since using only one modality shows limitations under particular situations, such as dramatic illumination variations in visual images and facial temperature changes in thermal images. This will be our ongoing work using both modalities to improve recognition performance. We will be also evaluating the effect of sunglasses in visual face recognition.

References

- [1] I. Craw, N. Costen, T. Kato, and S. Akamatsu, "How should we represent faces for automatic recognition?," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No. 8, pp.725-736, 1999.
- [2] A. M. Burton, V. Bruce, and P. J. B. Hancock, "From pixels to people: a model of familiar face recognition," *Cognitive Science*, Vol. 23, pp.1-31, 1999.
- [3] P. J. Phillips and P. Rauss, "The Face Recognition Technology (FERET) Program," *Proc. Office of National Drug Control Policy, CTAC Int. Technology Symposium*, pp.8-11, 1997.
- [4] D. M. Blackburn, J. M. Bone, and P. J. Phillips, "Face Recognition Vendor Test 2000," *Evaluation Report, National Institute of Standards and Technology*, pp.1-70, 2001.
- [5] M. Bone and D. Blackburn, "Face Recognition at a Chokepoint: Scenario Evaluation Results," *Evaluation Report, Department of Defense*, 2002.
- [6] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face Recognition Vendor Test 2002," *Evaluation Report, National Institute of Standards and Technology*, pp.1-56, 2003.
- [7] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.721-732, 1997.
- [8] R. Chellappa, C.L. Wilson, S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE*, Vol.83 (5), pp705-741 (1995).
- [9] R. Gross, J. Shi, and J. Cohn, "Quo Vadis Face Recognition?," *Third Workshop on Empirical Evaluation Methods in Computer Vision*, December, 2001
- [10] P.Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. PAMI*-19(7), (1997).
- [11] T. Sim and T. Kanade, "Combining Models and Exemplars for Face Recognition: An Illuminating Example," *Proceedings of the CVPR* December (2001).
- [12] P.Belhumeur and D.Kriegman, "What is the Set of Images of an Object under All Possible Illumination Conditions," *Int. J. Computer Vision*, Vol.28, No.3, pp.245-260, (1998)
- [13] A.S.Georghiadis, P.N.Belhumeur, D.J.Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans PAMI*, V23,(6),(2001)
- [14] Y. Yoshitomi, T. Miyaura, S. Tomita, and S. Kimura, "Face identification using thermal image processing," *Proc. IEEE Int. Workshop on Robot and Human Communication*, pp.374-379, 1997.
- [15] F. Prokoski, "History, Current Status, and Future of Infrared Identification," *Proc. IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pp.5-14, 2000.
- [16] A. Selinger and D. A. Socolinsky, "Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study," *Technical Report, Equinox Corporation*, 2001.
- [17] L. B. Wolff, D. A. Socolinsky, and C. K. Eveland, "Face Recognition in the Thermal Infrared," *Technical Report, Equinox Corporation*, 2002.
- [18] L. B. Wolff, D. A. Socolinsky, and C. K. Eveland, "Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery," *Proc. CVPR Workshop on Computer Vision Beyond the Visible Spectrum*, 2001.
- [19] J. Heo, S. Kong, B. Abidi, and M. Abidi, "Fusion of Visual and Thermal Signatures with Eyeglass Removal for Robust Face Recognition," *IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum in conjunction with CVPR 2004*, pp. 94-99, Washington, D.C., July 2004.
- [20] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol. 3, pp.72-86, 1991.
- [21] C.F. Hester and D. Casasent, "Multivariant technique for multi-class pattern recognition", *Applied Optics* 19, pp.1758-1761 (1980).
- [22] A. Mahalanobis, B.V.K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl. Opt.* 26, pp. 3633-3630 (1987).
- [23] M. Savvides, B.V.K. Vijaya Kumar and P. Khosla, "Face verification using correlation filters," *Proc. Of Third IEEE Automatic Identification Advanced Technologies*, Tarrytown, NY, pp.56-61 (2002).
- [24] B.V.K. Vijaya Kumar, "Minimum Variance synthetic discriminant functions", *J. Opt. Soc. Am A*, Vol 3, pp1579-1584.
- [25] P. Refregier, "Filter Design for optical pattern recognition: Multi-criteria optimization approach," *Optics Lett*,V.15, 854-856, 1990.
- [26] <http://www.equinoxsensors.com/products/HID.html>