On Person Authentication by Fusing Visual and Thermal Face Biometrics

Ognjen Arandjelović[†] Riad Hammoud[‡] Roberto Cipolla[†]
[†]University of Cambridge, UK [‡]Delphi Electronics and Safety, USA
{oa214,cipolla}@eng.cam.ac.uk riad.hammoud@delphi.com

Abstract

Recognition algorithms that use data obtained by imaging faces in the thermal spectrum are promising in achieving invariance to extreme illumination changes that are often present in practice. In this paper we analyze the performance of a recently proposed face recognition algorithm that combines visual and thermal modalities by decision level fusion. We examine (i) the effects of the proposed data preprocessing in each domain, (ii) the contribution to improved recognition of different types of features, (iii) the importance of prescription glasses detection, in the context of both 1-to-N and 1-to-1 matching (recognition vs. verification performance). Finally, we discuss the significance of our results and, in particular, identify a number of limitations of the current state-of-the-art and propose promising directions for future research.

1 Introduction

Among the most sensors used in face biometric systems is the optical imager. This is driven by its availability and low-cost. An optical imager captures the light reflectance of the face surface in the visible spectrum. The visible spectrum provides features that depend only on surface reflectance. Thus, it is obvious that the face appearance changes according to the ambient light. Over the years a number of methods have been proposed to solve this problem [1] [4] [7] [19], but it still remains challenging in most practical circumstances (e.g. [3]).

Recent studies have proved that face recognition in the thermal spectrum offers a few distinct advantages over the visible spectrum, including invariance to ambient illumination changes [20] [16] [11] [15], see Fig. 1. This is due to the fact that a thermal infrared sensor measures the heat energy radiation emitted by the face rather than the light reflectance. Appearance-based face recognition algorithms applied to thermal IR imaging consistently performed better than when applied to visible imagery, under various lighting conditions and facial expressions [9] [14] [16] [13]. Further performance improvements were achieved using fusion of different modalities [5] [12] [6] [8] [2].

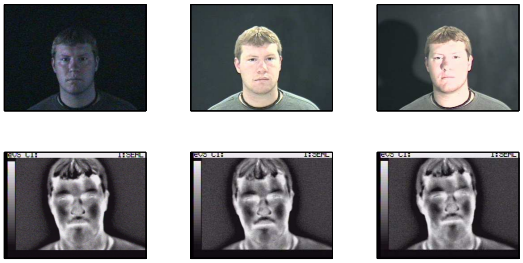


Figure 1: ***Invariance:*** *Illumination changes have dramatic effects on data acquired in the visible spectrum (top row). In contrast, thermal imagery (bottom row) shows remarkable invariance.*

2 Method Overview

In this section we briefly explain the system of Arandjelović *et al.* [2]. Its main components are conceptually shown in Fig. 2.

At the lowest level, image sets – rasterized images in either the visual or thermal spectrum – are matched using the cosine of first principal angle θ between the corresponding Principal Component Analysis (PCA) subspaces U_1 and U_2 [2] [21]:

$$\rho = \cos\theta = \max_{\mathbf{u}\in U_1} \max_{\mathbf{v}\in U_2} \mathbf{u}^T\mathbf{v}. \quad (1)$$

Feature fusion. Two types of features are used in set matching: holistic and local. Holistic face appearance is matched after affine warping using the locations of the two detected eyes and the mouth, see Fig. 3. Local patches used correspond to the same three detected regions, see Fig. 4.

The similarity of two individuals using only a single modality (visual or thermal) is computed by combining the holistic face representation and a representation based on local image patches. The similarity score is obtained by a constant-weighted summation:

$$\rho_{v/t} = \omega_h \cdot \rho_h + \omega_m \cdot \rho_m + (1 - \omega_h - \omega_m) \cdot \rho_e, \quad (2)$$

where ρ_m, ρ_e and ρ_h are the scores of separately matching, respectively, the mouth, the eyes and the entire face regions, and ω_h and ω_m the weighting constants. We followed the recommendation of the original paper's authors

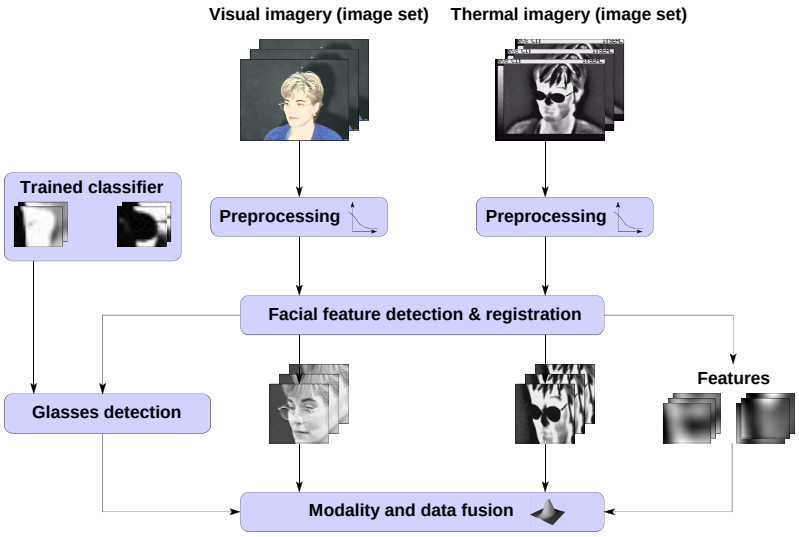


Figure 2: ***System overview:*** *The evaluated system consists of three main modules performing (i) data preprocessing and registration, (ii) glasses detection and (iii) fusion of holistic and local face representations using visual and thermal modalities.*

Figure 3: ***Registration:*** *Original image with detected facial features marked by yellow circles (left), result of affine warping the image to canonical frame (centre) and the cropped result.*

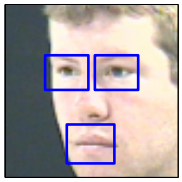


Figure 4: ***Features:*** *In both spectra, matching is done using holistic appearance and appearance of 3 salient facial features.*

in the choice of values of the weights:

$$\omega_h = 0.7 \qquad \omega_m = 0.0, \tag{3}$$

for the visual spectrum and

$$\omega_h = 0.8 \qquad \omega_m = 0.1, \tag{4}$$

for the thermal.

Modality fusion. Given ρ_v and ρ_t, the similarity scores corresponding to visual and thermal data, Arandjelović *et al.* compute joint similarity as:

$$\rho_f = \omega_v(\rho_v) \cdot \rho_v + (1 - \omega_v(\rho_v)) \cdot \rho_t, \tag{5}$$

making the weighting factors are no longer constants, but *functions*. The key idea is that if the visual spectrum match is very good (i.e. ρ_v is close to 1.0), one can be confident that illumination difference between the two images sets compared is mild. In this case, visual spectrum should be given relatively more weight than when the match is bad and the illumination change is likely more drastic.

The function $\omega_v \equiv \omega_v(\rho_v)$ is estimated in three stages: first (i) we estimate $\hat{p}(\omega_v, \rho_v)$, the probability that ω_v is the optimal weighting given the estimated similarity ρ_v, then (ii) compute $\omega(\rho_v)$ in the maximum a posteriori sense and finally (iii) make an analytic fit to the obtained marginal distribution. To solve step (i) a heuristic, iterative algorithm is proposed. In short, matching of an unknown person against a set gallery individuals is simulated using the offline gallery. Since the ground truth identities of all persons in the offline database is known, an estimate of the likelihood that each $\omega = k\Delta\omega$ is optimal, is readily computed. Density $\hat{p}(\omega, \rho)$ is then incremented proportionally after being passed through the sigmoid function, see Fig. 5 and the original paper for details.

Prescription glasses. The appeal of using the thermal

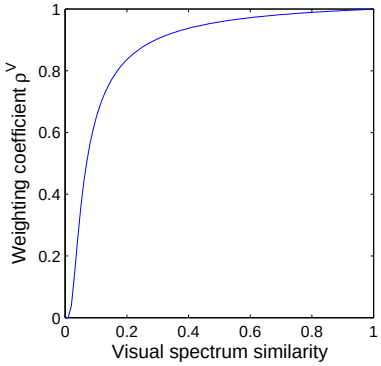
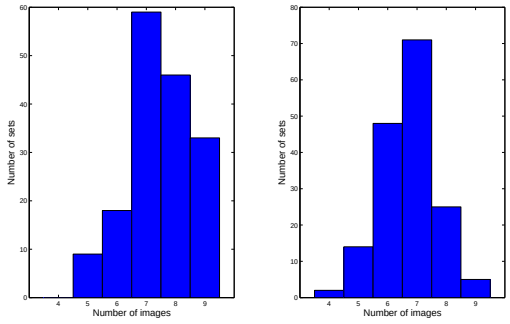


Figure 5: ***Modality fusion:*** *Contribution of visual score, as a function of similarity of visual imagery. Low similarity is indicative of large illumination changes and more weight placed on thermal data.*



(a) Visual (b) Thermal

Figure 6: ***Training sets:*** *Histograms of the number of images per person used to train the algorithm.*

spectrum for face recognition stems mainly from its invariance to illumination changes, in sharp contrast to visual spectrum data. The exact opposite is true in the case of prescription glasses, which appear as dark patches in thermal imagery. The practical importance of this can be seen by noting that in the US in 2000 roughly 96 million people, or 34% of the total population, wore prescription glasses [18].

In the evaluated system, the appearance distortion that glasses cause in thermal imagery is used to help recognition by detecting their presence. If the subject is not wearing glasses, then both holistic and all local patches-based face representations can be be used in recognition; otherwise the eye regions in thermal images are ignored (i.e. $\omega_h = \omega_e = 0.0$ is used in (2)).

3 Empirical Evaluation

We evaluated the described system on the *"Dataset 02: IRIS Thermal/Visible Face Database"* subset of the *Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS)* database[1], available for download at `http://www.cse.ohio-state.edu/OTCBVS-BENCH/`. Briefly, this database contains 29 individuals, 11 roughly matching poses in visual and thermal spectra and large illumination variations, some of which are shown in Fig. 1).

The algorithm was trained using all images in a single illumination in which all 3 salient facial features could be detected. This typically resulted in 7-8 images in the visual and 6-7 in the thermal spectrum, see Fig. 6.

[1]IEEE OTCBVS WS Series Bench; DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968; DOD/TACOM/NAC/ARC Program under grant R01-1344-18; FAA/NSSA grant R01-1344-48/49; Office of Naval Research under grant #N000143010022.

We evaluated:

- Performance improvement with band-pass and Self-Quotient image filters, respectively:

$$\mathbf{I}_{BP} = \mathbf{I} * \mathbf{G}_{\sigma=W_1} - \mathbf{I} * \mathbf{G}_{\sigma=W_2} \quad (6)$$

$$\mathbf{I}_{SQI} = \mathbf{I}_{BP}./(\mathbf{I} * \mathbf{G}_{\sigma=W_2}), \quad (7)$$

- Recognition performance using only individual local features and fusion with holistic face appearance, and
- Importance of dealing with prescription glasses.

The performance of the algorithm was evaluated both in 1-to-N and 1-to-1 matching scenarios. In the former case, we assumed that test data corresponded to one of people in the training set and recognition was performed by associating it with the closest match. Verification (or 1-to-1 matching, "is this the same person?") performance was quantified by looking at the true positive admittance rate for a threshold that corresponds to 1 admitted intruder in 100.

3.1 Results

A summary of 1-to-N matching results in shown in Tab. 1.

Firstly, note the poor performance achieved using both raw visual as well as raw thermal data. The former is suggestive of challenging illumination changes present in the OTCBVS data set. This is further confirmed by significant improvements gained with both band-pass filtering, see Fig. 8, and the Self-Quotient Image which increased the average recognition rate for, respectively, 35% and 47%. The same is corroborated by the Receiver-Operator Characteristic curves in Fig. 7 and 1-to-1 matching results in Tab. 2.

On the other hand, the reason for low recognition rate of raw thermal imagery is twofold: it was previously argued

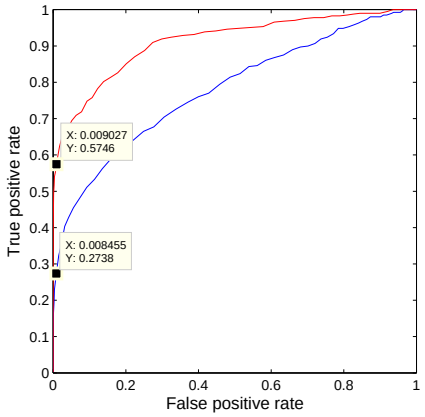
Representation		Rec.
Visual	Holistic raw data	0.58
	Holistic, band-pass	0.78
	Holistic, SQI	0.85
	Mouth+eyes+holistic data fusion, SQI	0.87
Thermal	Holistic raw data	0.74
	Holistic raw w/ glasses detection	0.77
	Holistic, low-pass	0.80
	Mouth+eyes+holistic data fusion, low-pass	0.82
Proposed thermal + visual fusion	w/o glasses detection	0.90
	w/ glasses detection	*0.97*

Table 1: ***1-to-N matching (recognition) results.***

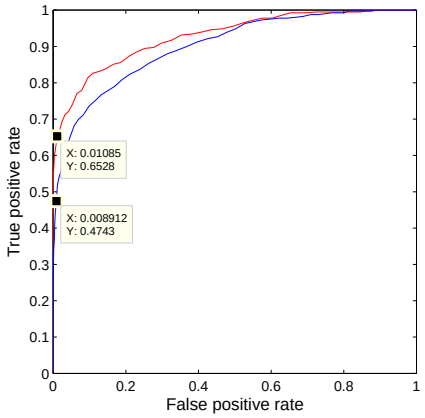
that the two main limitations of this modality are the inherently lower discriminative power and occlusions caused by prescription glasses. The addition of the glasses detection module is of little help at this point - some benefit is gained by steering away from misleadingly good matches between any two people wearing glasses, but it is limited in extent as a very discriminative region of the face is lost. Furthermore, the improvement achieved by optimal band-pass filtering in thermal imagery is much more modest than with visual data, increasing performance respectively by 35% and 8%. Similar increase was obtained in true admittance rate (42% vs. 8%), see Tab. 7.

Neither the eyes or the mouth regions, in either the visual or thermal spectrum, proved very discriminative when used in isolation, see Fig. 9. Only 10-12% true positive admittance was achieved, as shown in Tab. 3. However, the proposed fusion of holistic and local appearance offered a consistent and statistically significant improvement. In 1-to-1 matching the true positive admittance rated increased for 4-6%, while the average correct 1-to-N matching improved for roughly 2-3%.

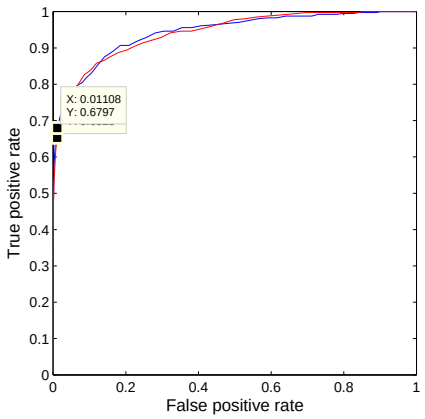
The greatest power of the method becomes apparent when the two modalities, visual and thermal, are fused. In this case the role of the glasses detection module is much more prominent, drastically decreasing the average error rate from 10% down to 3%, see Tab. 1. Similarly, the true admission rate increases to 74% when data is fused without special handling of glasses, and to 80% when glasses are taken into account. Finally, we note that the performance of the glasses detection module on this dataset was virtually perfect, incorrectly classifying an instance of a person without glasses only in a single case, see Fig. 10.



(a) Unprocessed

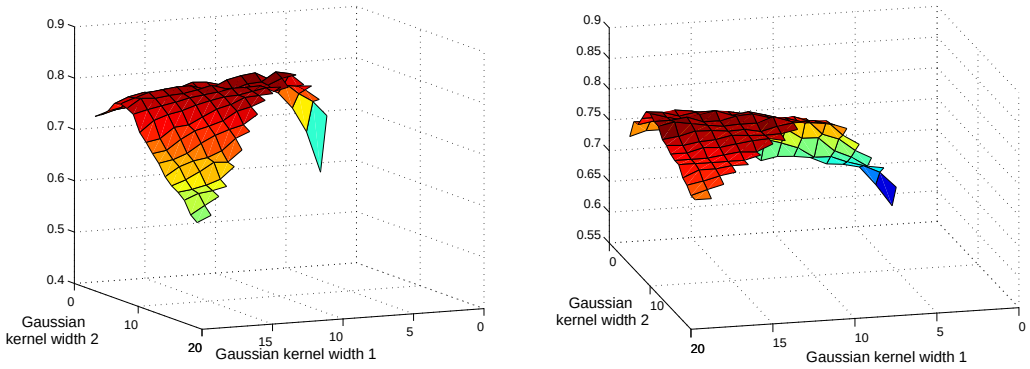


(b) Band-pass filtered



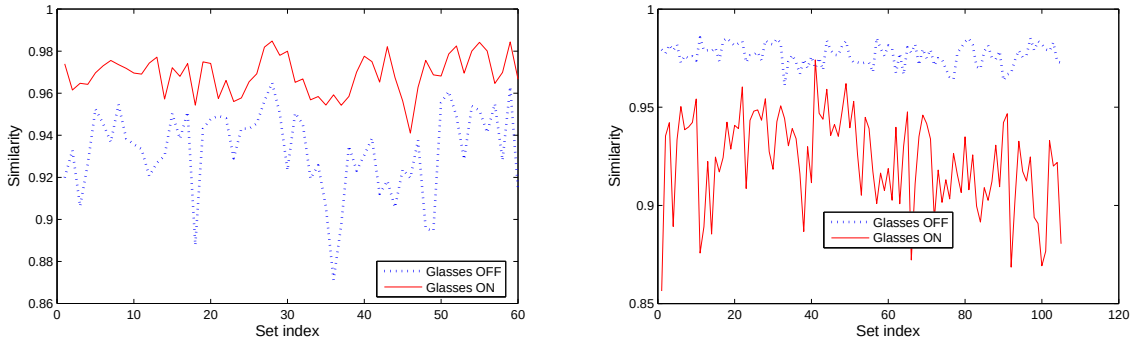
(c) Self-Quotient Image filtered

Figure 7: ***Holistic representations Receiver-Operator Characteristics (ROC):*** *Visual (blue) and thermal (red) spectra.*



(a) Visual (b) Thermal

Figure 8: ***Band-pass filter:*** *The optimal combination of the lower and upper band-pass filter thresholds is estimated from a small training corpus. The plots show the recognition rate using a single modality, (a) visual and (b) thermal, as a function of the widths of the two Gaussian kernels. It is interesting to note that the optimal band-pass filter for the visual spectrum passes a rather narrow, mid-frequency band, whereas the optimal filter for the thermal spectrum is in fact a* low-pass *filter.*



(a) Glasses ON (b) Glasses OFF

Figure 10: ***Glasses detection results:*** *Inter- and intra- class similarities.*

Representation	**Visual**	**Thermal**
1% intruder acceptance		
Unprocessed/raw	0.2850	0.5803
Band-pass filtered (BP)	0.4933	0.6287
Self-quotient image (SQI)	0.6410	0.6301

Table 2: ***Holistic, 1-to-1 matching (verification).***

Representation	**Visual (SQI)**	**Thermal (BP)**
1% intruder acceptance		
Eyes	0.1016	0.2984
Mouth	0.1223	0.3037

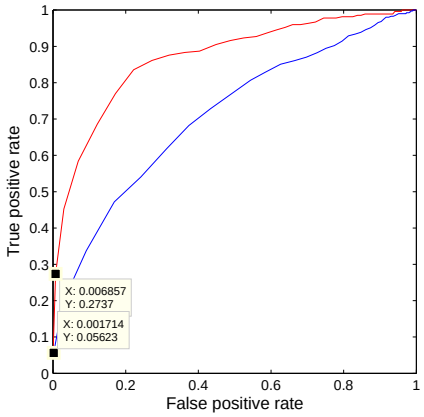
Table 3: ***Isolated local features, 1-to-1 matching (verification).***

Representation	**Visual (SQI)**	**Thermal (BP)**
1% intruder acceptance		
Holistic + Eyes	0.6782	0.6499
Holistic + Mouth	0.6410	0.6501
Holistic + Eyes + Mouth	0.6782	0.6558

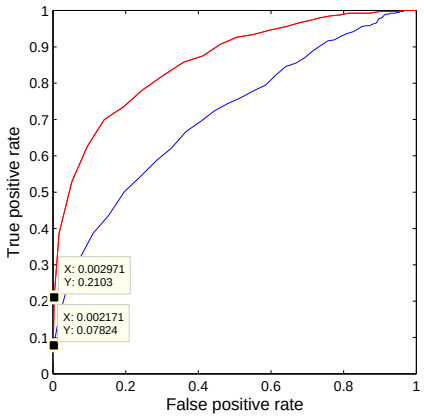
Table 4: ***Holistic & local features, 1-to-1 matching.***

Representation	**True admission rate**
1% intruder acceptance	
Without glasses detection	0.7435
With glasses detection	0.8014

Table 5: ***Feature and modality fusion, 1-to-1 matching.***



(a) Eyes



(b) Mouth

Figure 9: ***Isolated local features Receiver-Operator Characteristics (ROC):*** *Visual (blue) and thermal (red) spectra.*

4 Conclusion

We analyzed and empirically evaluated a recently proposed system for personal identification based on face biometrics from visual and thermal imagery. Our results suggest that: (i) simple filters can be used to greatly improve recognition accuracy in both domains, (ii) little improvement is seen with the inclusion of local feature-based patches, (iii) the proposed algorithm for the detection of glasses is very reliable across individuals and different imaging conditions, (iv) the fusion of visual and thermal imagery is very promising in practical applications, the two modalities offering discriminative power in complementary ways.

In the verification setup the evaluated method achieved 80% correct admittance rate for 1% intruder admittance. Clearly, this is a much lower rate than that required in most practical applications, at this point making the system useful only in the pre-screening process. Much better performance was achieved in 1-to-N matching evaluation, in which a high correct identification rate of 97% was obtained using only a small number of training images (5-7) and in the presence of large illumination changes.

Our results suggest several possible avenues for improvement. Further use of the thermal spectrum should be attempted, by not only detecting the glasses, but also by segmenting them out so that holistic appearance can still be used in matching. This is challenging in the presence of extreme poses, as glasses in many cases "merge" with the background with more profile views. Another research direction that should be explored is of different representation of local appearance (e.g. as SIFT features), which too could possibly offer further benefit with large pose changes.

References

[1] O. Arandjelović and R. Cipolla. Face recognition from video using the global shape-illumination manifold. *ECCV*, May 2006.

[2] O. Arandjelović, R. Hammoud, and R. Cipolla. Multi-sensory face biometric fusion (for personal identification). *In Proc. IEEE Workshop on Biometics*, 2006. (to appear).

[3] O. Arandjelović and A. Zisserman. Automatic face recognition for film character retrieval in feature-length films. *CVPR*, 1, June 2005.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *PAMI*, 25(9), 2003.

[5] R. Brunelli and D. Falavigna. Personal identification using multiple cues. *PAMI*, 1995.

[6] X. Chen, P. Flynn, and K. Bowyer. Visible-light and infrared face recognition. 2003.

[7] A. S. Georghiades, D. J. Kriegman, and P. N Belhumeur. Illumination cones for recognition under variable lighting: Faces. *CVPR*, 1998.

[8] J. Heo, B. Abidi, S. G. Kong, and M. Abidi. Performance comparison of visual and thermal signatures for face recognition. September.

[9] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent advances in visual and infrared face recognition - a review. *CVIU*, 2004.

[10] A. M. Martinez. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *PAMI*, 24(6), 2002.

[11] F. Prokoski. History, current status, and future of infrared identification.

[12] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 2003.

[13] A. Selinger and D. Socolinsky. Appearance-based facial recognition using visible and thermal imagery: A comparative study. Technical Report 02-01, Equinox Corporation, 2002.

[14] D. Socolinsky and A. Selinger. Comparative study of face recognition performance with visible and thermal infrared imagery. 2002.

[15] D. Socolinsky and A. Selinger. Thermal face recognition in an operational scenario. 2004.

[16] D. Socolinsky, A. Selinger, and J. Neuheisel. Face recognition with visible and thermal infrared imagery. *CVIU*, 2003.

[17] K. K. Sung and Tomaso Poggio. Example-based learning for view-based human face detection. *PAMI*, 20(1), 1998.

[18] T. C. Walker and R. K. Miller. *Health Care Business Market Research Handbook.* Norcross (GA): Richard K. Miller & Associates, Inc., fifth edition, 2001.

[19] L. Wolf and A. Shashua. Learning over sets using kernel principal angles. *JMLR*, 4(10), 2003.

[20] L. B. Wolff, D. A. Socolinsky, and C. K. Eveland. Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery. 2001.

[21] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. *AFG*, (10), 1998.