Thermal and Reflectance Based Personal Identification Methodology in Challenging Variable Illuminations

Ognjen Arandjelović[†] Riad Hammoud[‡] Roberto Cipolla[†]

[†]Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK
{oa214,cipolla}@eng.cam.ac.uk

[‡]Delphi Corporation,
Delphi Electronics and Safety
Kokomo, IN 46901-9005, USA
riad.hammoud@delphi.com

Abstract

The objective of this work is to authenticate individuals based on the appearance of their faces. This is a difficult pattern recognition problem because facial appearance is generally greatly affected by the changes in the way a face is illuminated, by the camera viewpoint and partial occlusions, for example due to eye-wear. We describe a full automatic algorithm that systematically addresses each of these challenges. The main novelty is an algorithm for decision-level fusion of two types of imagery: one acquired in the visual and one acquired in infrared electromagnetic spectrum. Specifically: we examine (i) the effects of preprocessing of data in each domain, (ii) the fusion of holistic and local facial appearance, and (iii) propose an algorithm for combining the similarity scores in visual and thermal spectra in the presence of prescription glasses and significant pose variations, using a small number of training images (5-7). Our system achieved a high correct identification rate of 97% on a freely available test set of 30 individuals and extreme illumination changes.

Index Terms

Face, recognition, thermal, infrared, fusion, illumination, invariance.

I. Introduction

For decades, the personal identification task had shown progress by employing technological means like *secret knowledge*, such as Personal Identification Numbers, and by using *personal possessions*, such as Identity Cards and Radio Frequency Identification chips. As opposed to these means which are generally easy targets for fraud, biometric modalities like facial geometry, ear form and iris are universal and relatively consistent over time. As the cost of biometric sensors, visible and thermal imagers, microphones, and motion sensors continues to sink due to higher demand, biometric systems have the tendency to employ more than a single sensor to identify an individual upon as many as non-redundant biometric modalities. In recent years there have been largely three categories of approaches to face recognition for obtaining complementary information that can be loosely termed multi-sample, multi-sensor and multi-modal. This paper presents a multi-sensory multi-sample face biometric fusion methodology for personal identification.

A. Problem challenges and proposed methodology

Variations in head pose and illumination are the most challenging aspects of face recognition. In practice, the effects of changing pose are usually less problematic and can oftentimes be overcome by acquiring data over a time period e.g. by tracking a face. Consequently, image sequence or image set matching has recently gained a lot of attention in the literature [2], [19], [53], as a practical way of obtaining multi-view appearance information. In this paper we too consider the image set matching paradigm. In contrast to pose, illumination changes are much more difficult to deal with for face recognition algorithms: the illumination setup in which a face is imaged is in most cases not possible to control, its physics difficult to accurately model and training data containing typical appearance variability is usually not available.

Thermal spectrum imagery is useful in this regard as it is virtually insensitive to illumination changes, see Fig. 1. On the other hand, it lacks much of the individual, discriminating facial detail contained in visual images. In this sense, the two modalities can be seen as complementing each other. The key idea behind the system presented in this paper is that robustness to extreme illumination changes can be achieved by *fusing* the two. This paradigm will further prove useful when which we consider the difficulty of recognition in the presence of prescription glasses, which are a major problem for recognition using thermal spectrum imagery.

B. Paper organization

The remainder of this paper is organized as follows. In the next section we review relevant previous work on face recognition in visual and thermal spectra, as well as

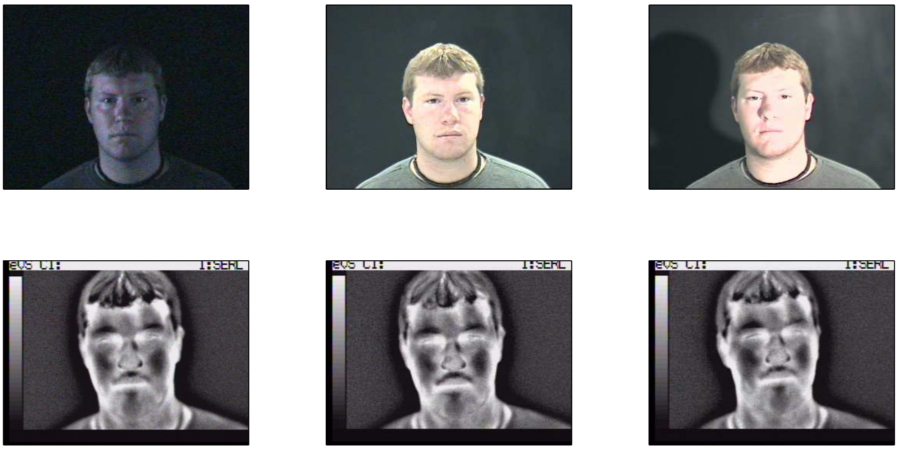


Fig. 1. *Illumination changes have a dramatic effects on images acquired in the visible light spectrum (top row). In contrast, thermal imagery (bottom row) shows remarkable invariance.*

multi-sensory techniques. Section III describes each of the main components of the proposed system in detail: baseline face set matching is covered first (Section III-A), followed by sections on data preprocessing and extraction (Section III-C), recognition using only a single modality (Section III-D), the proposed modality fusion (Section III-E) and, finally, dealing with occlusions caused by prescription glasses (Section III-F). Our empirical evaluation methodology, data sets used and the performance of the proposed algorithm are reported and discussed in Section IV. The paper is concluded with a summary and an outline of promising directions for future research.

II. Previous work

A. Mono-sensor based techniques

a) Optical sensors: 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. In order to overcome the lighting, pose and facial expression changes, a flurry of face recognition algorithms, from the two well-known broad categories, statistical appearance-based and model-based methods, has been proposed [43].

Appearance-based methods, as their name suggests, perform recognition directly from the way faces appear in images, interpreting them as ordered collections of pixels. Faces are typically represented as vectors in the image space; discrimination between individuals is then performed by employing statistical models that explain inter- and/or intra- personal appearance changes.

The main limitation of purely appearance-based recognition is that of limited generalization ability, especially in the presence of greatly varying illumination conditions. On the other hand, an appealing property of this group of algorithms is that they can be typically extended to recognize using sets or sequences of faces. Eigenfaces, for example, can be used on a per-image basis, with recognition decision cast using majority vote [3]. A similar voting approach was also successfully used with local features in [22], which were extracted by tracking a face using a Gabor Wavelet Network [22], [36], [37]. In [59] video information is used only in the training stage to construct person-specific PCA spaces, *self-eigenfaces*, while verification was performed from a single image using the Distance from Feature Space criterion. Classifiers using different eigenfeature spaces were used in [45] and combined using the sum rule [34]. Better use of training data is made with various discriminative methods such as Fisherfaces, which can be used to estimate *database-specific* optimal projection [23]. An interesting extension of appearance correlation-based recognition to matching sets of faces was proposed by Yamaguchi *et al.* [69]. The so-called Mutual Subspace Method (MSM) has since gained considerable attention in the literature. In MSM, linear subspaces describing appearance variations within sets or sequences are matched using canonical correlations [27], [31].

However, all of the aforementioned appearance-based methods poorly generalize in the presence of large illumination changes. Thus, model-based methods have been proposed to deal with this problem. These formulate models of image formation with the intention of recovering (i) mainly person-specific (e.g. face albedo or shape) and (ii) extrinsic, nuisance variables (e.g. illumination direction, or head yaw). The key challenge lies in coming up with models for which the parameter estimation problem is not ambiguous or ill-conditioned.

The Active Appearance Model [20], for example, was proposed for describing objects that vary both in shape and appearance. A deformable, triangular mesh is fitted to an image of a face, guided by combined statistical models of shape and shape-free appearance, so as to best explain the observed image. The 3D Morphable Model can be seen as a 3D extension of this approach. This model consists of albedo values at the nodes of a 3-dimensional triangular mesh describing face geometry. Model fitting is performed by combining a Gaussian prior on the shape and texture of human faces with photometric information from an image [62].

An attractive feature of model-based approaches is that they explicitly model both intrinsic and extrinsic variables. On the other hand, they suffer from convergence problems in the presence of background clutter or facial occlusions such as glasses, unlike the method proposed in this paper. Furthermore and importantly for the our work they often require high quality image input and struggle with non-Lambertian effects or multiple light sources.

b) Thermal infrared-sensor based techniques: A number of recent studies suggest that face recognition in the thermal spectrum offers a few distinct advantages over the visible spectrum, including invariance to ambient illumination changes [46], [55], [56],

[67]. 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. In outdoor environments, and particularly in direct sunlight, illumination invariance only holds true to good approximation for the Long-Wave Infrared (LWIR: 8–$14\mu m$) spectrum, which is fortunately measured by the less expensive uncooled thermal infrared camera technology. Human skin has high emissivity in the Long-Wave Infrared (MWIR: 3–$5\mu m$) spectrum and even higher emissivity in the LWIR spectrum making face imagery by and large invariant to illumination variations in these spectra.

Appearance-based face recognition algorithms applied to thermal infrared imaging consistently performed better than when applied to visible imagery, under various lighting conditions and facial expressions [35], [51], [54], [56]. Further performance improvements were achieved using decision-based fusion [56]. In contrast to other techniques, Srivastana and Liu [57], performed face recognition in the space of Bessel function parameters. First, they decompose each infrared face image using Gabor filters. Then, they represent the face by modelling the marginal density of the Gabor filter coefficients using Bessel functions. This approach has further been improved by Buddharaju *et al.* [14]. Recently, Friedrich and Yeshurun [26] showed that IR-based recognition is less sensitive to changes in 3D head pose and facial expression.

A thermal sensor generates imaging features that uncover thermal characteristics of the face pattern. Another advantage of thermal infrared imaging in face recognition is the existence of a direct relationship to underlying physical anatomy such as vasculature. Indeed, thermal face recognition algorithms attempt to take advantage of such anatomical information of the human face as unique signatures. The use of vessel structure for human identification has been studied during recent years using traits such as hand vessel patterns [33], [39], finger vessel patterns [41], [52] and vascular networks from thermal facial images [47]. In [15] a novel methodology that consists of a statistical face segmentation and a physiological feature extraction algorithm, and a matching procedure of the vascular network from thermal facial imagery has been proposed.

The down side of employing near infrared and thermal infrared sensors is that glare reflections and opaque regions appear in presence of subjects wearing prescription glasses, plastic and sun glasses. For a large proportion of individuals the regions around the eyes – that is an area of high interest to face recognition systems – become occluded and therefore less discriminant [5], [38].

B. Multi-sensor based techniques

In the biometric literature several classifiers have been used to concatenate and consolidate the match scores of multiple independent matchers of biometric traits [17] [7], [10], [42], [65]. In [13] a HyperBF network is used to combine matchers based on voice and face features. Ross and Jain [49] use decision tree and linear discriminant classifiers for classifying the match scores pertaining to the face, fingerprint and hand geometry modalities. In [48] three different colour channels of a face image are independently subjected to LDA and then combined.

Recently, several successful attempts have been made to fuse the visual and thermal infrared modalities to increase the performance of face recognition [13], [18], [28]–[30], [35], [49], [55], [64], [68]. Visible and thermal sensors are well-matched candidates for image fusion as limitations of imaging in one spectrum seem to be precisely the strengths of imaging in the other. Indeed, as the surface of the face and its temperature have nothing in common, it would be beneficial to extract and fuse cues from both sensors that are not redundant and yet complementary.

In [30] two types of visible and thermal fusion technique have been proposed. The first fuses low-level data while the second fuses matching distance scores. Data fusion was implemented by applying pixel-based weighted averaging of co-registered visual and thermal images. Decision fusion was implemented by combining the matching scores of individual recognition modules.

The fusion at the score level is the most commonly considered approach in the biometric literature [50]. Cappelli et al. [16] use a *double sigmoid function* for score normalization in a multi-biometric system that combines different fingerprint matchers. Once the match scores output by multiple matchers are transformed into a common domain they can be combined using simple fusion operators such as the sum of scores, product of scores or order statistics (e.g., maximum/minimum of scores or median score). Our proposed method falls into this category of multi-sensor fusion at the score level. To deal with occlusions caused by eyeglasses in thermal imagery, Heo et al. [30] used a simple ellipse fitting technique to detect the circle-like eyeglass regions in the IR image and replaced them with an average eye template. Using a commercial face recognition system, FaceIt [32], they demonstrated improvements in face recognition accuracy. Our method differs both in the glasses detection stage, which uses a principled statistical model of appearance variation, and in the manner it handles detected occlusions. Instead of using the average eye template, which carries no discriminative information, we segment out the eye region from the infrared data, effectively placing more weight on the discriminative power of the same region extracted from the filtered, visual imagery.

III. Method Details

In the sections that follow we explain our system in detail, the main components of which are conceptually depicted in Fig. 2.

A. Matching image sets

In this paper we are dealing with face recognition from sets of images, both in the visual and thermal spectrum. We will demonstrate that illumination invariance and recognition robustness in the presence of prescription glasses can be achieved using a

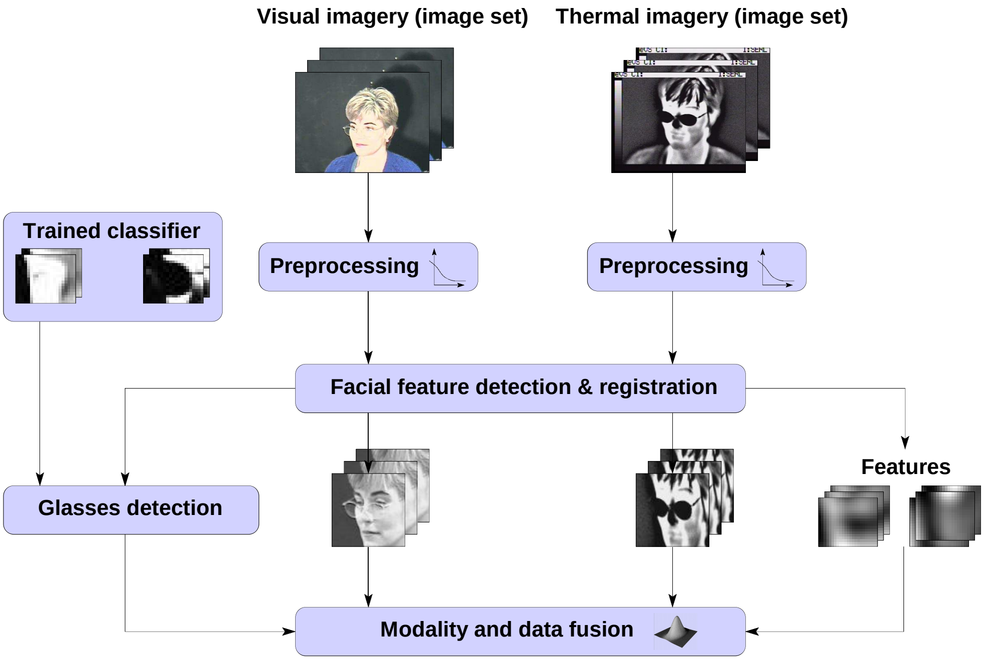


Fig. 2. *Our system consists of three main modules performing (i) data preprocessing and registration, (ii) glasses detection and (iii) fusion of holistic and local face representations using visual and thermal modalities.*

combination of simple data preprocessing (Section III-C), a combination of holistic and local features (Section III-D) and the fusion of the two modalities (see Section III-E). These stages normalize for the bulk of appearance changes caused by extrinsic (non person-specific) factors. Hence, the requirements for our basic set-matching algorithm are those of (i) some pose generalization and (ii) robustness to noise. We will show that the two criteria are successfully met by modelling appearance variations within an image set using a low-dimensional linear subspace and then by comparing two such subspaces by finding the most similar modes of variation within them.

The appearance modelling step is a simple application of Principal Component Analysis (PCA) without mean subtraction. Formally, given a data matrix $\mathbf{d} \in \mathbb{R}^{D \times N}$ (each column representing a rasterized image), the corresponding subspace is spanned by the eigenvectors of the matrix $\mathbf{C} = \mathbf{d}\mathbf{d}^T$ corresponding to the largest eigenvalues. We used 5D subspaces, as sufficiently expressive to on average explain over 90% of data variation within intrinsically low-dimensional face appearance changes within a set.

B. Principal angles

Principal, or canonical, angles $0 \leq \theta_1 \leq \ldots \leq \theta_D \leq (\pi/2)$ between two D-dimensional linear subspaces U_1 and U_2 are recursively uniquely defined as the minimal angles between any two vectors of the subspaces:

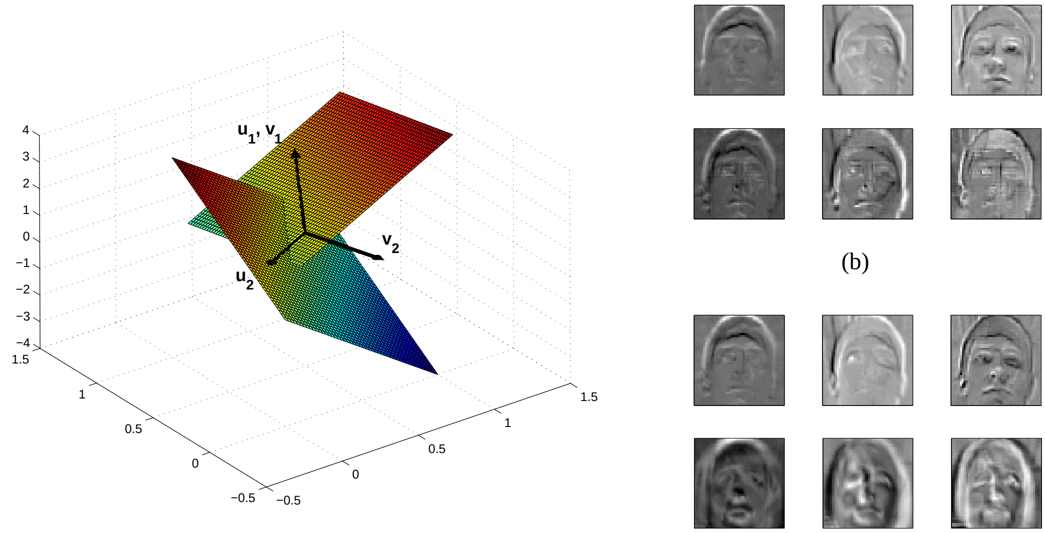
$$\rho_i = \cos \theta_i = \max_{\mathbf{u}_i \in U_1} \max_{\mathbf{v}_i \in U_2} \mathbf{u}_i^T \mathbf{v}_i \tag{1}$$

subject to the orthonormality condition:

$$\mathbf{u}_i^T \mathbf{u}_i = \mathbf{v}_i^T \mathbf{v}_i = 1, \;\; \mathbf{u}_i^T \mathbf{u}_j = \mathbf{v}_i^T \mathbf{v}_j = 0, \;\; j = 1, ..., i-1 \tag{2}$$

We will refer to $\mathbf{u}_i$ and $\mathbf{v}_i$ as the i-th pair of *principal vectors*, see Figure 3 (a). The quantity ρ_i is also known as the i-th canonical correlation [31]. Intuitively, the first pair of principal vectors corresponds to the most similar modes of variation within two linear subspaces; every next pair to the most similar modes orthogonal to all previous ones. We quantify the similarity of subspaces U_1 and U_2, corresponding to two face sets, by the cosine of the smallest angle between two vectors confined to them i.e. ρ_1.

This interpretation of principal vectors motivates the suitability of canonical correlations as a similarity measure when subspaces U_1 and U_2 correspond to face images. First, the empirical observation that face appearance varies smoothly as a function of camera viewpoint [2], [9] is implicitly exploited: since the computation of the most similar



(a) (b) (c)

Fig. 3. *An illustration of the concept of principal angles and principal vectors in the case of two 2D subspaces embedded in a 3D space. As two such subspaces necessarily intersect, the first pair of principal vectors is the same (i.e.* $\mathbf{u}_1 = \mathbf{u}_2$*). However, the second pair is not, and in this case forms the second principal angle of* $\cos^{-1}\rho_2 = \cos^{-1}(0.8084) \approx 36°$*. The top three pairs of principal vectors, displayed as images, when the subspaces correspond to image sets of the same and different individuals are displayed in (b) and (c) (top rows corresponds to* $\mathbf{u}_i$*, bottom rows to* $\mathbf{v}_i$*). In (b), the most similar modes of pattern variation, represented by principal vectors, are very much alike* in spite of different illumination conditions *used in data acquisition.*

modes of appearance variation between sets can be seen as an efficient“search” over entire subspaces, *generalization* by means of linear pose interpolation and extrapolation is inherently achieved. This concept is further illustrated in Figure 3 (b,c). Furthermore, by being dependent on only a single (linear) direction within a subspace, by employing the proposed similarity measure the bulk of data in each set, deemed not useful in a specific set-to-set comparison, is thrown away. In this manner *robustness to missing data* is achieved.

An additional appealing feature of comparing two subspaces in this manner is contained in its computational efficiency. If $\mathbf{B}_1$ and $\mathbf{B}_2$ are orthonormal basis matrices corresponding to U_1 and U_2, then writing the Singular Value Decomposition (SVD) of the matrix $\mathbf{B}_1^T\mathbf{B}_2$:

$$\mathbf{M} = \mathbf{B}_1^T\mathbf{B}_2 = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{3}$$

The i-th canonical correlation ρ_i is then given by the i-th singular value of $\mathbf{M}$ i.e. $\mathbf{\Sigma}_{i,i}$, and the i-th pair of principal vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ by, respectively, $\mathbf{B}_1\mathbf{U}$ and $\mathbf{B}_2\mathbf{V}$ [11]. Seeing that in our case $\mathbf{M}$ is a 5×5 matrix and that we only use the largest canonical correlation, ρ_1 can be rapidly computed as the largest eigenvalue of $\mathbf{M}\mathbf{M}^T$ [44].

C. Data preprocessing & feature extraction

The first stage of our system involves coarse normalization of pose and illumination. Pose changes are accounted for by in-plane registration of images, which are then passed through quasi illumination-invariant image filters.

We register all faces, both in the visual and thermal domain, to have the salient facial features aligned. Specifically, we propose to explicitly align the eyes and the mouth. This choice of features is motivated by their characteristic appearance in both thermal and visual spectra, which makes them most reliably detected (e.g. see [6], [8], [21], [24], [60]). The 3 point correspondences, between the detected and the canonical features' locations, uniquely define an affine transformation which is applied to the original image. Finally, faces are then cropped to 80×80 pixels, as shown in Figure 4.

Coarse brightness normalization is performed by band-pass filtering the images [6], [25]. The aim is to reduce the amount of high-frequency noise as well as extrinsic appearance variations confined to a low-frequency band containing little discriminating information. Most obviously, in visual imagery, the latter are caused by illumination changes, owing to the smoothness of the surface and albedo of faces [1].

Fig. 4. *Shown is the original image in the visual spectrum with detected facial features marked by yellow circles (left), the result of affine warping the image to the canonical frame (centre) and the final registered and cropped facial image.*

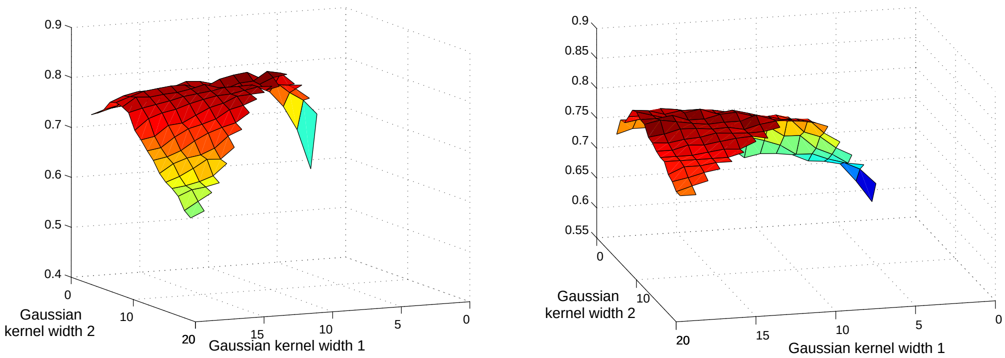
We consider the following type of a spatial-domain band-pass filter:

$$\mathbf{I}_F = \mathbf{I} * \mathbf{G}_{\sigma=W_1} - \mathbf{I} * \mathbf{G}_{\sigma=W_2}, \tag{4}$$

which has two parameters - the widths W_1 and W_2 of isotropic Gaussian kernels. The optimal values of these are estimated from a small training corpus of individuals in varying illumination. Figure 5 shows the recognition rate across the corpus as the values of the two parameters are varied. The optimal values were found to be 2.3 and 6.2 for visual data; the optimal filter for thermal data was found to be a *low-pass* filter with $W_2 = 2.8$ (i.e. W_1 was found to be very large). This is indeed expected, as thermal imagery is not affected by any predominantly low-frequency extrinsic factors, but can be seen to contain a rather large amount of high frequency noise.

Examples of optimally filtered, registered and cropped images are shown in Figure 6. It is important to note from Figure 5 that the recognition rate varied smoothly with changes in kernel widths, showing that the method is not very sensitive to their exact values, which is suggestive of good generalization to unseen data.

Finally, we also propose to further scale the result of filtering of visual data by the



(a) Visual (b) Thermal

Fig. 5. *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 W_1 and W_2 of the two Gaussian kernels in* (4)*. 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.*

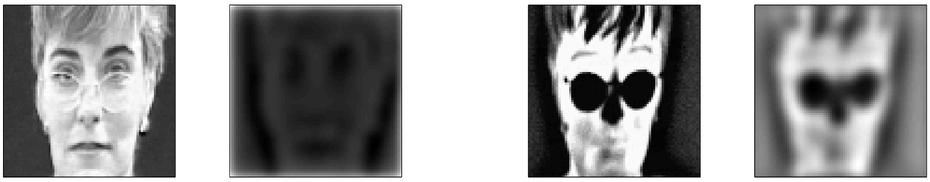
smooth version of its original image:

$$\hat{\mathbf{I}}_F(x, y) = \mathbf{I}_F(x, y)./(\mathbf{I} * \mathbf{G}_{\sigma=W_2}), \tag{5}$$

where $./$ represents element-wise division. The purpose of local scaling is to equalize edge strengths in dark (weak edges) and bright (strong edges) regions of the face; this is similar to the Self Quotient Image of Wang *et al.* [63]. This step further improves the robustness of the representation to illumination changes, as we demonstrate in Section IV.

D. Single modality-based recognition

We compute the similarity of two individuals using only a single modality (visual or thermal) by combining the holistic face representation described in Section III-C and a representation based on local image patches. The idea behind using patches is to gain further viewpoint robustness. Since smaller surface areas of the face are closer to planar



(a) Visual (b) Thermal

Fig. 6. *Effects of the optimal band-pass filters on registered and cropped faces in (a) visual and (b) thermal spectra.*

than the entire face, when affine registered they exhibit a greater degree of appearance invariance. This is supported by previous empirical accounts in the literature, e.g. see [12], [53], [66].

As before, we use the eyes and the mouth as the most discriminative regions by extracting rectangular patches centred at the detections, see Figure 7. We treat holistic and local patch-based similarities as independent, the overall similarity score thus being obtained by weighted summation:

$$\rho_{v/t} = \omega_h \cdot \rho_h + \overbrace{\omega_m \cdot \rho_m + (1 - \omega_h - \omega_m) \cdot \rho_e}^{\text{Local patches contribution}}, \tag{6}$$

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.

The optimal values of the weights were estimated from the offline training corpus. As expected, eyes were shown to carry a significant amount of discriminative information, as for the visual spectrum we obtained $\omega_e = 0.3$. On the other hand, the mouth region, highly variable in appearance in the presence of facial expression changes, was found not to improve recognition (i.e. $\omega_m \approx 0.0$).

The relative magnitudes of the weights were found to be different in the thermal

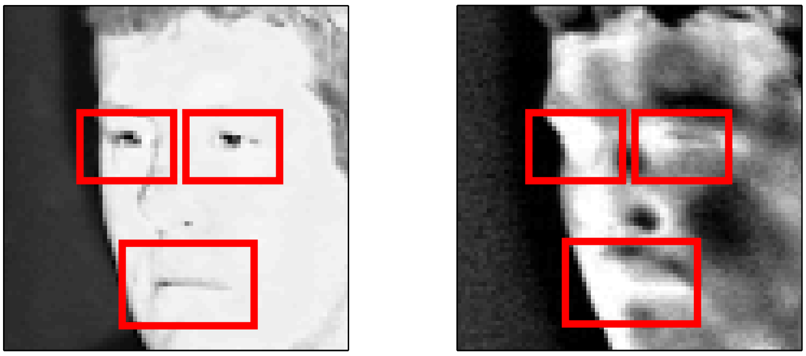


Fig. 7. *In both the visual and the thermal spectrum our algorithm combines the similarities obtained by matching the holistic face appearance and the appearance of three salient local features – the eyes and the mouth.*

spectrum, both the eye and the mouth region contributing equally to the overall score: $\omega_m = 0.1, \ \omega_h = 0.8$. Notice the rather insignificant contribution of individual facial features. This is most likely due to inherently spatially slowly varying nature of heat radiated by the human body.

E. Fusing modalities

Until now we have focused on deriving a similarity score between two individuals given sets of images in either thermal or visual spectrum. A combination of holistic and local features was employed in the computation of both. However, the greatest power of our system comes from the fusion of the two modalities.

Given ρ_v and ρ_t, the similarity scores corresponding to visual and thermal data, we compute the joint similarity as:

$$\rho_f = \omega_v(\rho_v) \cdot \rho_v + \left[1 - \omega_v(\rho_v)\right] \cdot \rho_t. \tag{7}$$

Notice that 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), we can be confident that illumination difference between the two images sets compared is mild and well compensated for by the visual spectrum preprocessing of Section III-C. 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 value of $\omega_v(\rho_v)$ can then be interpreted as statistically the optimal choice of the mixing coefficient ω_v given the visual domain similarity ρ_v. Formalizing this we can write

$$\omega_v(\rho_v) = \arg\max_{\omega} p(\omega|\rho_v), \tag{8}$$

or, equivalently

$$\omega_v(\rho_v) = \arg\max_{\omega} \frac{p(\omega, \rho_v)}{p(\mu)}. \tag{9}$$

Under the assumption of a uniform prior on the confusion margin, $p(\mu)$

$$p(\alpha|\mu) \propto p(\alpha, \rho_v), \tag{10}$$

and

$$\omega_v(\rho_v) = \arg\max_{\omega} p(\omega, \rho_v). \tag{11}$$

1) Learning the α-function: The function $\omega_v \equiv \omega_v(\rho_v)$ is estimated in three stages: first (i) we estimate $p(\omega_v, \rho_v)$, then (ii) compute $\omega(\rho_v)$ using (11) and finally (iii) make an analytic fit to the obtained marginal distribution. Step (i) is challenging and we describe it next.

a) Iterative density estimate: The principal difficulty of estimating $p(\omega_v, \rho_v)$ is of practical nature: in order to obtain an accurate estimate (i.e. a well-sampled distribution), a prohibitively large training database is needed. Instead, we employ a heuristic alternative. Much like before, the estimation is performed using the offline training corpus.

Our algorithm is based on an iterative incremental update of the density, initialized as uniform over the domain $\omega_v, \rho_v \in [0, 1]$. We iteratively simulate matching of an unknown person against a set gallery individuals. In each iteration of the algorithm, these are randomly drawn from the offline training database. Since the ground truth identities of all persons in the offline database is known, for each $\omega_v = k\Delta\omega_v$ we can compute (i) the initial visual spectrum similarity $\rho_v^{p,p}$ of the novel and the corresponding gallery sequences, and (ii) the resulting separation $\delta(k\Delta\omega_v)$ i.e. the difference between the similarities of the test set and the set corresponding to it in identity, and that between the test set and the most similar set that does *not* correspond to it in identity. This gives us information about the usefulness of a particular value of ω_v for observed $\rho_v^{p,p}$. Hence, the density estimate $\hat{p}(\omega_v, \rho_v)$ is then updated at $(k\Delta\omega_v, \rho^{p,p}), k = 1 \ldots$. We increment it proportionally to $\delta(k\Delta\omega_v)$ after passing through a y-axis shifted sigmoid function:

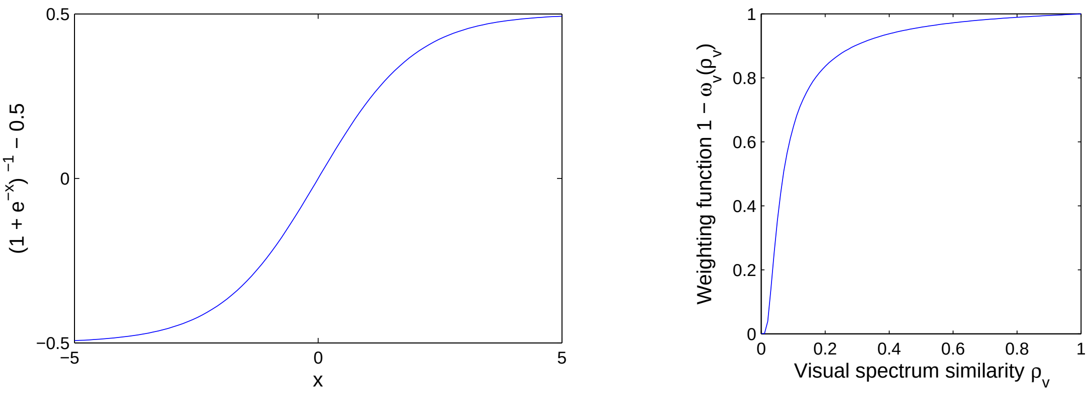
$$\hat{p}(k\Delta\omega_v, \rho_v^{p,p})_{[n+1]} = \hat{p}(k\Delta\omega_v, \rho_v^{p,p})_{[n]} + \Big[\text{sig}(C \cdot \delta(k\Delta\omega_v)) - 0.5\Big], \tag{12}$$

where subscript $[n]$ signifies the n-th iteration step and

$$\text{sig}(x) = \frac{1}{1 + e^{-x}}, \tag{13}$$

as shown in Figure 8 (a). The sigmoid function has the effect of reducing the overly confident weight updates for the values of ω_v that result in extremely good or bad separations $\delta(k\Delta\omega_v)$. The purpose of this can be seen by noting that we are using separation as a proxy for the statistical goodness of ω_v, while in fact attempting to maximize the average recognition rate (i.e. the average number of cases for which $\delta(k\Delta\omega_v) > 0$). This is one of the main differences of the described algorithm to that of Arandjelović and Cipolla proposed in [4].

Figure 9 summarizes the proposed offline learning algorithm. An analytic fit to $\omega_v(\rho_v)$ in the form $(1 + e^a)/(1 + e^{a/\rho_v})$ is shown in Figure 8 (b).



(a) y-axis shifted sigmoid function

(b) Weighting function

Fig. 8. *The contribution of visual matching, as a function of the similarity of visual imagery. A low similarity score between image sets in the visual domain is indicative of large illumination changes and consequently our algorithm leant that more weight should be placed on the illumination-invariant thermal spectrum.*

F. Prescription glasses

The appeal of using the thermal 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, see Figure 6. The practical importance of achieving recognition robustness in the presence of corrective eyewear can be seen by noting that in the US in 2000 roughly 96 million people, or 34% of the total population, wore prescription glasses [61].

In our system, the otherwise undesired, gross 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 used in recognition; otherwise the eye regions in thermal images are ignored as they contain no useful recognition (discriminative) information.

Input: visual data $d_v(person, illumination)$,
thermal data $d_t(person, illumination)$.
Output: density estimate $\hat{p}(\omega_v, \rho_v)$.

1: Initialization
$\hat{p}(\omega_v, \rho_v) = 0$,

2: Iteration
for all illuminations $i,\ j$ **and** persons p

3: Iteration
for all $k = 0, \ldots, 1/\Delta\omega_v,\ \omega_v = k\Delta\omega_v$

5: Separation given ω

$$\delta(k\Delta\omega_v) = \min_{q \neq p}[\omega_v \rho_v^{p,p} + (1-\omega_v)\rho_t^{p,p} \\ -\omega_v \rho_v^{p,q} + (1-\omega_v)\rho_t^{p,q}]$$

6: Update density estimate

$$\hat{p}(k\Delta\omega_v, \rho_v^{p,p}) = \hat{p}(k\Delta\omega_v, \rho_v^{p,p}) \\ + \Big[\text{sig}(C \cdot \delta(k\Delta\omega_v)) - 0.5\Big]$$

7: Smooth the output
$\hat{p}(\omega_v, \rho_v) = \hat{p}(\omega_v, \rho_v) * \mathbf{G}_{\sigma=0.05}$

8: Normalize to unit integral
$\hat{p}(\omega_v, \rho_v) = \hat{p}(\omega_v, \rho_v) / \int_{\omega_v} \int_{\rho_v} \hat{p}(\omega_v, \rho_v) d\rho_v d\omega_v$

Fig. 9. *The proposed fusion learning algorithm, used offline.*

b) Glasses detection: We detect the presence of glasses by building representations for the left eye region (due to the symmetry of faces, a detector for only one side is needed) with and without glasses, in the thermal spectrum. The foundations of our classifier are laid out in §III-A. Appearance variations of the eye region with out without

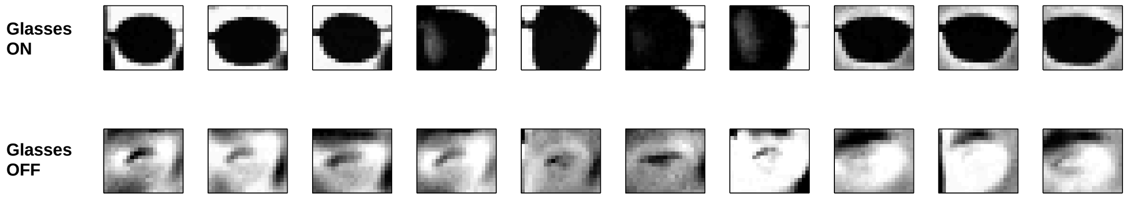
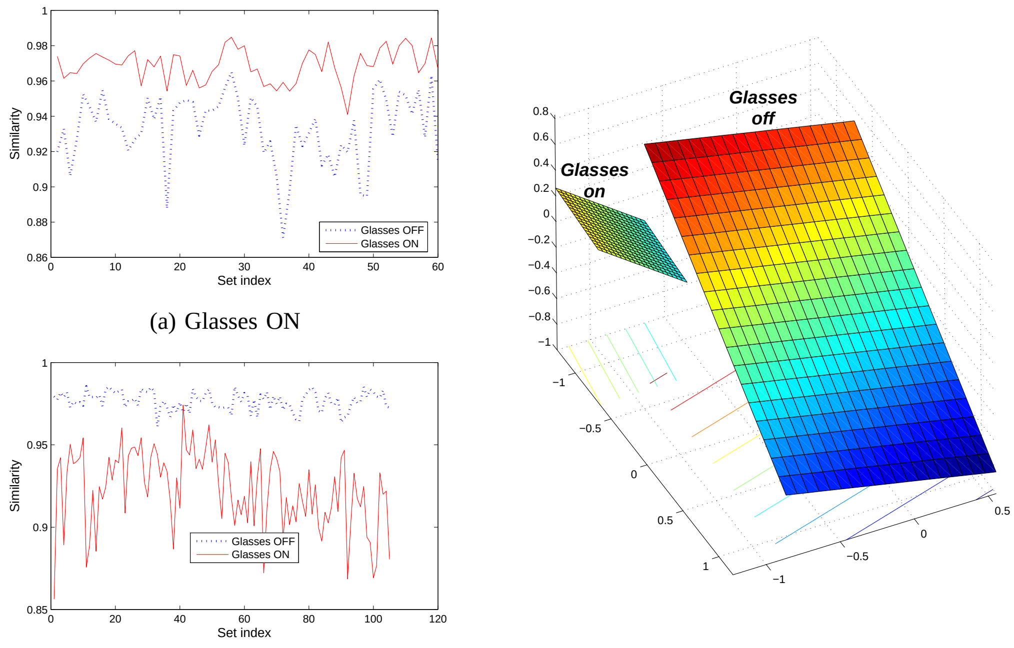


Fig. 10. *Shown are examples of glasses-on (top) and glasses-off (bottom) thermal data used to construct the corresponding appearance models for our glasses detector.*

glasses are represented by two 6D linear subspaces estimated from the training data corpus, see Fig. 10 for examples of training data used for subspace estimations. The linear subspace corresponding to eye region patches extracted from a set of thermal imagery of a novel person is then compared with “glasses on” and “glasses off” subspaces using principal angles. The presence of glasses is deduced when the corresponding subspace results in a higher similarity score. We obtain close to flawless performance on our data set (also see §IV for description), as shown in Fig. 11 (a,b). Good discriminative ability of principal angles in this case is also supported by visual inspection of the “glasses on” and “glasses off” subspaces; this is illustrated in Fig. 11 (c) which shows the first two dominant modes of each, embedded in the 3D principal subspace.

The presence of glasses severely limits what can be achieved with thermal imagery, the occlusion heavily affecting both the holistic face appearance as well as that of the eye regions. This is the point at which our method heavily relies on decision fusion with visual data, limiting the contribution of the thermal spectrum to matching using mouth appearance only i.e. setting $\omega_m = 1.0$ in (6).



(a) Glasses ON

(b) Glasses OFF

(c) Model subspaces

Fig. 11. *(a) Inter- and (b) intra- class (glasses on and off) similarities across our data set. (c) Good discrimination by principal angles is also motivated qualitatively as the subspaces modelling appearance variations of the eye region with and without glasses on show very different orientations even when projected to the 3D principal subspace. As expected, the "glasses off" subspace describes more appearance variation, as illustrated by the larger extent of the linear patch representing it in the plot.*

IV. 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], freely available for download at `http://www.cse.ohio-state.edu/OTCBVS-BENCH/`. this database contains 30 individuals, 11 roughly matching poses

[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.

in visual and thermal spectra and large illumination variations (some of these are exemplified in Figure 12). Images were acquired using the Raytheon Palm-IR-Pro camera in the thermal and Panasonic WV-CP234 camera in the visual spectrum, in the resolution of 240×320 pixels.

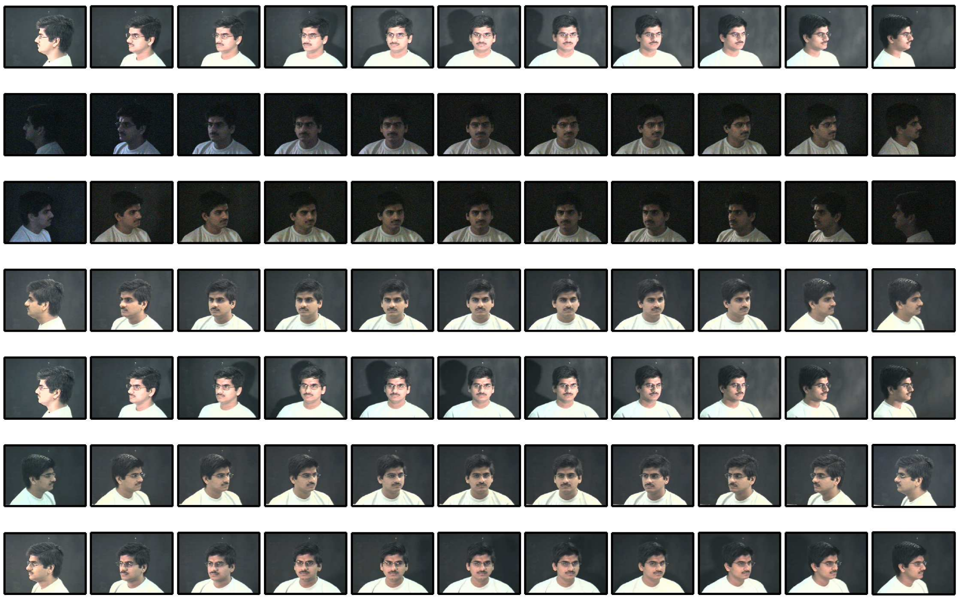
Our algorithm was trained using all images in a single illumination in which all 3 salient facial features could be detected. This typically resulted in 7-8 images in the visual and 6-7 in the thermal spectrum, see Figure 13, and roughly $\pm 45°$ yaw range, as measured from the frontal face orientation.

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.

A. Results

A summary of 1-to-N matching results is shown in Table I. 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 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 Figure 15 and 1-to-1 matching results in Table II.

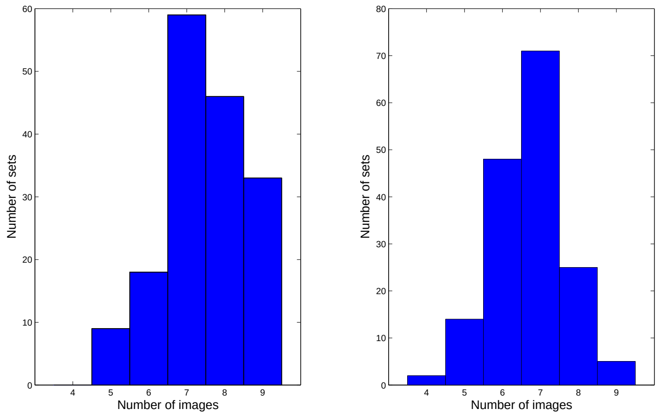
On the other hand, the reason for low recognition rate of raw thermal imagery is twofold: it was previously argued that the two main limitations of this modality are the inherently lower discriminative power and occlusions caused by prescription glasses.



(a) Visual

(b) Thermal

Fig. 12. *Each row corresponds to an example of a single training (or test) set of images used for our algorithm in (a) the visual and (b) the thermal spectrum. Note the extreme changes in illumination, as well as that in some sets the user is wearing glasses and in some not.*



(a) Visual (b) Thermal

Fig. 13. *Shown are histograms of the number of images per person used to train our algorithm. Depending on the exact head poses assumed by the user we typically obtained 7-8 visual spectrum images and typically a slightly lower number for the thermal spectrum. The range of yaw angles covered is roughly* $\pm 45^\circ$ *measured from the frontal face orientation.*

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 Table 15.

Neither the eyes or the mouth regions, in either the visual or thermal spectrum, proved very discriminative when used in isolation, see Figure 14. Only 10-12% true positive admittance was achieved, as shown in Table III. 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

TABLE I

Shown is the average rank-1 recognition rate using different representations across all combinations of illuminations. Note the performance increase with each of the main features of our system: image filtering, combination of holistic and local features, modality fusion and prescription glasses detection.

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

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 Table I. 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.

V. SUMMARY AND CONCLUSIONS

In this paper we described a system for personal identification based on a face biometric that uses cues from visual and thermal imagery. The two modalities are shown to complement each other, their fusion providing good illumination invariance

TABLE II

A summary of the comparison of different image processing filters for 1 in 100 intruder acceptance rate. Both the simple band-pass filter, and even further its locally-scaled variant, greatly improve performance. This is most significant in the visual spectrum, in which image intensity in the low spatial frequency is most affected by illumination changes.

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 III

A summary of the results for 1 in 100 intruder acceptance rate. Local features in isolation perform very poorly.

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

TABLE IV

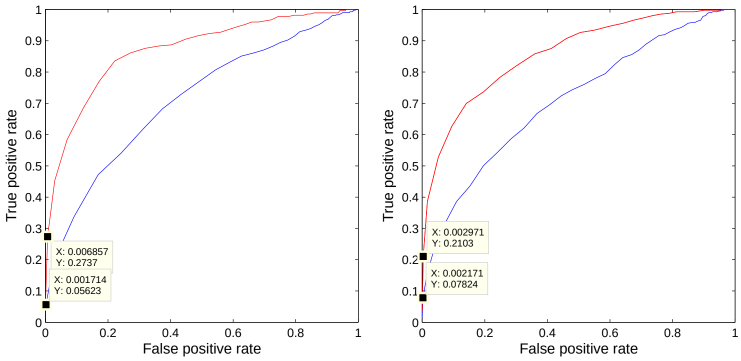
Holistic & local features – a summary of 1-to-1 matching (verification) results.

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 V

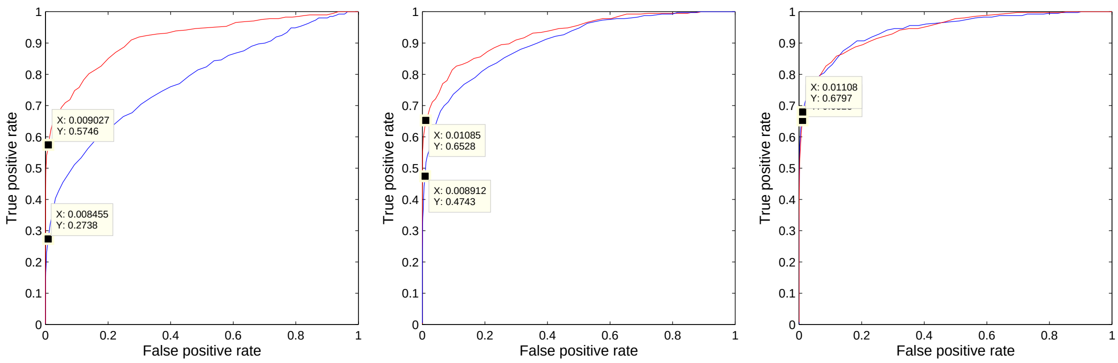
Feature and modality fusion – a summary of the 1-to-1 matching (verification) results.

Representation	**True admission rate**
1% intruder acceptance	
W/o glasses detection	0.7435
W/ glasses detection	0.8014



(a) Eyes (b) Mouth

Fig. 14. *Isolated local features Receiver-Operator Characteristics (ROC): for visual (blue) and thermal (red) spectra.*



(a) Unprocessed (b) Band-pass filtered (c) Self-Quotient Image filtered

Fig. 15. *Holistic representations Receiver-Operator Characteristics (ROC) for visual (blue) and thermal (red) spectra.*

and discriminative power between individuals. Prescription glasses, a major difficulty in the thermal spectrum, are reliably detected by our method, restricting the matching to non-affected face regions. Finally, we examined how different preprocessing methods affect recognition in the two spectra, as well as holistic and local feature-based face representations. The proposed method was shown to achieve a high recognition rate

(97%) using only a small number of training images (5-7) in the presence of large illumination changes.

Our results suggest several possible avenues for improvement. We intend to make further use of the thermal spectrum, by not only detecting the glasses, but also by segmenting them out. This is challenging across large pose variations, such as those contained in our test set. Another research direction we would like to pursue is that of synthetically enriching the training corpus to achieve increased robustness to pose differences between image sets (c.f. [40] [58]). Finally, we note that empirical evaluation on a larger data set is needed to clearly identify the limitations and scalability of the proposed method.

ACKNOWLEDGEMENTS

We would like to thank Trinity College Cambridge, the Toshiba Corporation and Delphi Electronics Ltd. for their kind support for our research, the volunteers whose face videos were entered in our face database and Cambridge Commonwealth Trust.

REFERENCES

[1] Y Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(7):721–732, 1997.

[2] O. Arandjelović and R. Cipolla. Face recognition from video using the generic shape-illumination manifold. *In Proc. European Conference on Computer Vision (ECCV)*, 4:27–40, May 2006.

[3] O. Arandjelović and R. Cipolla. An information-theoretic approach to face recognition from face motion manifolds. *Image and Vision Computing (special issue on Face Processing in Video)*, 24(6):639–647, June 2006.

[4] O. Arandjelović and R. Cipolla. A new look at filtering techniques for illumination invariance in automatic face recogition. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 449–454, April 2006.

[5] O. Arandjelović, R. I. Hammoud, and R. Cipolla. Multi-sensory face biometric fusion (for personal identification). *In Proc. IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS)*, pages 128–135, June 2006.

[6] O. Arandjelović and A. Zisserman. *Interactive Video: Algorithms and Technologies.*, chapter On Film Character Retrieval in Feature-Length Films., pages 89–103. Springer-Verlag, May 2006.

[7] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz. Fusion of face and speech data for person identity verification. *IEEE Transactions on Neural Networks*, 10(5):1065–1075, September 1998.

[8] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y. W. Teh, E. Learned-Miller, and D. A. Forsyth. Names and faces in the news. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2:848–854, 2004.

[9] M. Bichsel and A. P. Pentland. Human face recognition and the face image set's topology. *Computer Vision, Graphics and Image Processing: Image Understanding*, 59(2):254–261, 1994.

[10] E. S. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert conciliation for multimodal person authentication systems using bayesian statistics. *In Proc. International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 291–300, 1997.

[11] Å. Björck and G. H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973.

[12] D. S. Bolme. Elastic bunch graph matching. Master's thesis, Colorado State University, 2003.

[13] Roberto Brunelli and Daniele Falavigna. Person recognition using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(10):955–966, October 1995.

[14] P. Buddharaju, I. Pavlidis, and I. Kakadiaris. Face recognition in the thermal infrared spectrum. *In Proc. IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS)*, page 133, 2004.

[15] P. Buddharaju, I. T. Pavlidis, and P. Tsiamyrtzis. Physiology-based face recognition. *In Proc. IEEE Conference on Advanced Video and Singal Based Surveillance (AVSS)*, September 2005.

[16] R. Cappelli, D. Maio, and D. Maltoni. A computational approach to edge detection. *In Proc. International Workshop on Multiple Classifier Systems*, pages 351–361, 2000.

[17] V. Chatzis, A. G. Bors, and I. Pitas. Multimodal decision-level fusion for person authentication. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 29(6):674–681, November 1999.

[18] X. Chen, P. Flynn, and K. Bowyer. Visible-light and infrared face recognition. pages 48–55, 2003.

[19] T-J. Chin and D. Suter. A new distance criterion for face recognition using image sets. *In Proc. Asian Conference on Computer Vision (ACCV)*, pages 549–558, 2006.

[20] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *In Proc. European Conference on Computer Vision (ECCV)*, 2:484–498, 1998.

[21] D. Cristinacce, Cootes T. F., and I Scott. A multistage approach to facial feature detection. *In Proc. IAPR British Machine Vision Conference (BMVC)*, 1:277–286, 2004.

[22] Teofilo Emidio de Campos, Rogerio Schmidt Feris, and Roberto Marcondes Cesar Junior. Eigenfaces versus

eigeneyes: First steps toward performance assessment of representations for face recognition. *In Proc. Mexican International Conference on Artificial Intelligence*, pages 193–201, 2000.

[23] G. J. Edwards, C. J. Taylor, and T. F. Cootes. Learning to identify and track faces in image sequences. *In Proc. IAPR British Machine Vision Conference (BMVC)*, pages 130–139, 1997.

[24] P. F. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision (IJCV)*, 61(1):55–79, 2005.

[25] A. Fitzgibbon and A. Zisserman. On affine invariant clustering and automatic cast listing in movies. *In Proc. European Conference on Computer Vision (ECCV)*, pages 304–320, 2002.

[26] G. Friedrich and Y. Yeshurun. Seeing people in the dark: face recognition in infrared images. *In Proc. IAPR British Machine Vision Conference (BMVC)*, pages 348–359, 2003.

[27] R. Gittins. *Canonical Analysis: A Review with Applications in Ecology.* Springer-Verlag, 1985.

[28] A. Gyaourova, G. Bebis, and I. Pavlidis. Fusion of infrared and visible images for face recognition. *In Proc. European Conference on Computer Vision (ECCV)*, 4:456–468, May 2004.

[29] J. Heo, B. Abidi, S. G. Kong, and M. Abidi. Performance comparison of visual and thermal signatures for face recognition. *Biometric Consortium Conference*, September 2003.

[30] J. Heo, S. Kong, B. Abidi, and M. A. Abidi. Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. *In Proc. IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS)*, pages 94–99, 2004.

[31] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–372, 1936.

[32] Identix Ltd. Faceit. `http://www.FaceIt.com/`.

[33] S. K. Im, H. S. Choi, and S. W. Kim. A direction-based vascular pattern extraction algorithm for hand vascular pattern verification. *ETRI Journal*, 25(2):101–108, April 2003.

[34] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[35] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent advances in visual and infrared face recognition – a review. *Computer Vision and Image Understanding (CVIU)*, 97(1):103–135, 2005.

[36] V. Krüger, A. Happe, and G. Sommer. Affine real-time face tracking using Gabor wavelet networks. *In Proc. IEEE International Conference on Pattern Recognition (ICPR)*, 1:127–130, 2000.

[37] V. Krüger and G. Sommer. Gabor wavelet networks for efficient head pose estimation. *Journal of the Optical Society of America*, 19(6):1112–1119, 2002.

[38] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(4):627–639, April 2007.

[39] C. L. Lin and K. C. Fan. Biometric verification using thermal images of palm-dorsa vein patterns. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(2):199–213, February 2004.

[40] A. M. Martinez. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(6):748–763, 2002.

[41] N. Miura, A. Nagasaka, and T. Miyatake. Feature extraction of finger vein patterns based on iterative line tracking and its application to personal identification. *Systems and Computers in Japan*, 35(7):61–71, April 2004.

[42] Verlinde P. and G. Cholet. Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application. *In Proc. International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, 5(2):188–193, 1999.

[43] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone. FRVT 2002: Overview and summary. *Technical report, National Institute of Justice*, March 2003.

[44] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C : the art of scientific computing.* Cambridge University Press, 2nd edition, 1992.

[45] J. R. Price and T. F. Gee. Towards robust face recognition from video. *In Proc. Applied Image Pattern Recognition Workshop, Analysis and Understanding of Time Varying Imagery*, pages 94–102, 2001.

[46] F. Prokoski. History, current status, and future of infrared identification. *In Proc. IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS)*, pages 5–14, 2000.

[47] F. J. Prokoski and R. Riedel. *BIOMETRICS: Personal Identification in Networked Society*, chapter 9 Infrared Identification of Faces and Body Parts, pages 191–211. Kluwer Academic Publishers, 1998.

[48] A. Ross and R. Govindarajan. Feature level fusion using hand and face biometrics. *In Proc. SPIE Conference on Biometric Technology for Human Identification II*, 5779:196–204, March 2005.

[49] A. Ross and A. K. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, September 2003.

[50] A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer, New York, USA, 2006.

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

[52] T. Shimooka and K. Shimizu. Artificial immune system for personal identification with finger vein pattern. *In Proc. International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, pages 511–518, September 2004.

[53] J. Sivic, M. Everingham, and A. Zisserman. Person spotting: video shot retrieval for face sets. *In Proc. IEEE International Conference on Image and Video Retrieval (CIVR)*, pages 226–236, 2005.

[54] D. Socolinsky and A. Selinger. Comparative study of face recognition performance with visible and thermal infrared imagery. *In Proc. IEEE International Conference on Pattern Recognition (ICPR)*, pages 217–222, 2002.

[55] D. Socolinsky and A. Selinger. Thermal face recognition in an operational scenario. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2:1012–1019, 2004.

[56] D. Socolinsky, A. Selinger, and J. Neuheisel. Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding (CVIU)*, 91(1–2):72–114, 2003.

[57] A. Srivastana and X. Liu. Statistical hypothesis pruning for recognizing faces from infrared images. *Image and Vision Computing*, 21(7):651–661, 2003.

[58] K. K. Sung and Tomaso Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 20(1):39–51, 1998.

[59] L. Torres, L. Lorente, and J. Vila. Automatic face recognition of video sequences using selfeigenfaces. *In Proc. IAPR International Symposium on Image/Video Communications over Fixed and Mobile Networks*, 2000.

[60] L. Trujillo, G. Olague, R. Hammoud, B. Hernandez, and E. Romero. Automatic feature localization in thermal images for facial expression recognition. *Computer Vision and Image Understanding (CVIU)*, 2007. (in press).

[61] T. C. Walker and R. K. Miller. *Health Care Business Market Research Handbook.* Norcross (GA): Richard K. Miller & Associates, Inc., 5th edition, 2001.

[62] Christian Wallraven, V. Blanz, and T. Vetter. 3D reconstruction of faces - combining stereo with class-based knowledge. *In Proc. Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM) Symposium*, pages 405–412, 1999.

[63] H. Wang, S. Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 819–824, 2004.

[64] J.G. Wang, E. Sung, and R. Venkateswarlu. Registration of infrared and visible-spectrum imagery for face recognition. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 638–644, May 2004.

[65] Y. Wang, T. Tan, and A. K. Jain. Combining face and iris biometrics for identity verification. *In Proc. International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 805–813, 2003.

[66] L. Wiskott, J-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, pages 355–396, 1999.

[67] L. B. Wolff, D. A. Socolinsky, and C. K. Eveland. Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery. *In Proc. IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS)*, 2001.

[68] Chen X., Flynn P., and Bowyer K. IR and visible light face recognition. *Computer Vision and Image Understanding (CVIU)*, 99(3):332–358, September 2005.

[69] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, (10):318–323, 1998.