Dissimilarity representations for thermal signature recognition at a distance

Yenisel Plasencia*, Edel García*, Robert P. W. Duin†, Heydi Méndez*, César San Martín‡and Claudio Soto‡

*Advanced Technologies Application Center. 7a # 21812 e/ 218 y 222, Rpto. Siboney, Playa, C.P. 12200, La Habana, Cuba

†Faculty of Electrical Engineering, Mathematics and Computer Sciences, Delft University of Technology, The Netherlands

‡Dep. of Electrical Eng. University of La Frontera. Casilla 54-D, Temuco, Chile

yplasencia@cenatav.co.cu

Keywords: dissimilarity representation, thermal signature recognition, noise robust

Abstract

The recognition of faces at a distance has several challenges. One is the uncontrolled illumination, another is the low resolution of the images. One approach to tackle the first limitation is to use longwave infrared (LWIR) face images since they are invariant to illumination changes. In this paper we studied the application of dissimilarity representations for LWIR face images to overcome the low resolution problem. A comparison is made between feature representations and dissimilarity representations for various image resolutions. Our experiments using the Equinox benchmark database showed that classification results using the nearest neighbour (1-NN) classifier on a dissimilarity space were stable when the image resolution was decreased. On the other hand, classification accuracy decreased when using the 1-NN on features, supporting the idea that a dissimilarity representation can be a more proper solution for the recognition of faces at a distance.

1 Introduction

The use of infrared (IR) imagery in order to develop face recognition systems, has started to receive a special attention in the last years because of its robustness to illumination changes [17, 18, 5]. Moreover, this kind of images allows us to develop face recognition systems in a complete dark environment. In this way, a number of researches have been done using active near infrared (NIR) imagery ($0.7 - 2.4\mu$m), due to its property of being reflected by objects, and to be invisible and unobtrusive. However, the use of LWIR ($8 - 12\mu$m) has received little attention in the literature, in spite of the fact that the thermal infrared presents several advantages, some of them are:

i) LWIR sensors collect the heat energy emitted by a body instead the light reflected.

ii) LWIR sensors have an invariant behavior under changes in illumination, being able to operate even in complete darkness.

iii) The human skin has a high emissivity representing a thermal signature own to each individual.

In this paper, only LWIR face images are considered. The opaqueness to glass is one of the principal limitations of thermal infrared face recognition. It is the equivalent to occlusion in visible face recognition. This implies that the presence of glasses degrades significantly the recognition performance. Many of the works in the literature on LWIR face recognition are devoted to solve this, they show that there is a preference for fusion strategies to tackle this problem [8, 3, 9, 12]. Another disadvantage of LWIR face representation is that it is sensitive to body temperature changes. Such changes can be provoked by external temperature like cold or warm air, by body exercising, or simply by consuming alcoholic beverages.

A number of algorithms have been proposed for the classification of faces from LWIR images, but in the representation level, only feature representations have been studied. In general in pattern recognition, the two most studied approaches for representation have been the feature or vector space representation and the structural representation. The vector space representation is the most frequently used due to the availability of statistical techniques that have been showing a good performance on the different pattern recognition problems. One disadvantage of this approach is that, in the process of encoding objects in vectors, we can loose discriminative information, and the representations of the classes in a vector space may superpose. This class overlap [7] implies that dissimilar objects may be close, carrying classification errors. The structural approach seems to be a more robust representation, but there isn't as much tools available in structural pattern recognition as in statistical pattern recognition.

Recently Duin and Pekalska [15] proposed a new approach for pattern representation, named Dissimilarity Representation. It arises from the idea that the notion of proximity is more fundamental than that of a feature or a class. This dissimilarity approach offers the possibility of unifying both representations, the statistical and the structural [4]. In case of lacking knowledge on good features it may be preferred the dissimilarity approach over arbitrarily selected features.

In this paper we formulated the hypothesis that the dissimilarity representation is more robust than the feature representation to identify persons in low spatial resolution LWIR imagery. To verify that the hypothesis holds, classification on a dissimilarity space is compared with classification on features, for thermal infrared face recognition. The main contributions of this work are:

i) the proposal of a new dissimilarity representation of the face thermal signature, suitable for the classification of faces at a distance;

ii) a comparative study of the proposed dissimilarity representation and feature representations for various image resolutions.

Section 2 presents related work in face recognition using the dissimilarity space classification. Section 3 introduces the dissimilarity space classification and its advantages over the 1-NN classification. Since a dissimilarity measure or matrix is needed in order to create a dissimilarity representation, image histograms and the Chi square distance computed between the histograms are briefly described in Section 3. Section 4 presents the experimental analysis, including data description, experimental setup, results and discussion. The conclusions are drawn in Section 5.

2 Related work

There are few works in the literature where dissimilarity representations are introduced for face recognition, and none of them make use of thermal infrared imagery. In [13] after reducing dimensionality with the Principal Component Analysis (PCA) approach, the authors used the Euclidean distances to conform the dissimilarity matrix that characterizes the face data. They built a dissimilarity space from the Euclidean distances derived from a feature space. Then they compared linear and quadratic classifiers in that space with the 1-NN classifier applied directly to the dissimilarity matrix, as a function of the amount of prototypes selected per class. In their experiments they showed that the dissimilarity space classifiers outperformed the 1-NN rule.

In [10], the author proposed the use of dissimilarity representations to solve the Small Sample Size problem that affects the direct application of the Linear Discriminant Analysis (LDA) method for face recognition. This is an alternative to the conventional use of methods like PCA as a previous step before the application of LDA. The method joined to a classifier fusion strategy was proved in face recognition and the results were comparable to the state of the art results.

In [11], the authors proposed the use of dimensionality reduction methods on the dissimilarity matrix constructed with all the training objects, instead of prototype selection methods. In this way they maintained more useful information for discrimination than with prototype selection methods. This was evidenced in the experimental results where the classification rates were higher using dimension reduction. Nevertheless, this approach has the disadvantage of the high computational cost compared to prototype selection, since for classifying a new object, the dissimilarity has to be measured to the complete training set first. In the case of prototype selection techniques, when a new object comes, we only need to compute the dissimilarities to some prototypes and not to the whole training set.

3 Dissimilarity space classification

A suitable and well accepted technique for classification from dissimilarities is the 1-NN rule [6]. For classifying new objects, the rule assigns the class with the smallest dissimilarity. It can work well for large training sets due to its theoretical properties. For robust dissimilarity measures, 1-NN is expected to be the best classifier [15]. One of its disadvantages is the high computational cost since it needs to compute distances to the whole training set. It is also sensitive to noisy objects or outliers because its decision relies on nearness and not on densities. Its performance worsens for small training sets and for not robust dissimilarities. One way to overcome the 1-NN limitations is the use of spaces where we can construct classifiers that base their decisions not only on the nearest object, but on the proximities to a set of objects that was properly chosen. This is the case of the dissimilarity space proposed by Pekalska and Duin [15].

The dissimilarity space is a Euclidean vector space. For its construction a representation set
$R = \{r_1, r_2, ..., r_n\}$ is needed, where the objects belonging to this set (also called prototypes) are chosen adequately based on some criterion that can be dependant of the problem at hand. Let X be the training

set, R and X can have the following relationships: $R \cap X = \varnothing$, or $R \subseteq X$. Once we have R, the dissimilarities of the objects in X to the objects in R are computed. When a new object r comes, it is also represented by a vector of dissimilarities d_r to the objects in R (1).

$$d_r = [d(r, r_1)d(r, r_2)...d(r, r_n)]. \quad (1)$$

The dissimilarity space is defined by the set R so each coordinate of a point in that space corresponds to a dissimilarity to some prototype and the dimension of this space is determined by the amount of prototypes selected. This allows us to control the computational cost and to guarantee the trade off between classification accuracy and computational efficiency. Compared to the nearest neighbour rule, classifiers in this space use more information that only the neighbourhood of each point, so they are less sensitive to outliers. Constructing classifiers in this space can also be less computational expensive, since only the dissimilarities to the representative objects are computed. In this space we can use a diversity of classifiers [14]. Previous to the construction of the space, a dissimilarity measure or a dissimilarity matrix should be provided.

3.1 Histograms

In our approach, before computing the dissimilarity values for the creation of the dissimilarity space, pixel intensity histograms of the whole image were used as an intermediate representation. Essentially, image histograms encode the intensity frequencies. A histogram of the whole image is invariant to rotation and translation of objects and pixels inside the image. This can be an advantage in some applications and a disadvantage in other applications, because we may want to maintain the structural information. The use of local image histograms is a suitable solution to address the last problem, but this also has drawbacks such as the optimization of the size of the regions.
In our approach, the use of histograms has the advantage of allowing horizontal shifts and rotations of the face in the image. As it is shown in Fig. 1, in the face images selected for our experiments the background is almost constant with some exceptions like the non uniformity noise. Also the majority of the background pixel intensities are different from the face pixel intensities, implying that the background information is not supposed to interfere with the face information.

3.2 Chi square distance

For the comparison of the LWIR histograms, the Chi square distance measure [2] was used. This distance has been proving to be effective for histogram

Figure 1: Examples of LWIR images from the Equinox database.

comparison. Let S and M be two histograms. The Chi square distance is defined as follows:

$$\chi^2(S, M) = \sum_{i=1}^{n} \frac{(S_i - M_i)^2}{(S_i + M_i)}, \quad (2)$$

where n is the number of bins in the histogram. This measure is often used as input for the 1-NN classifier.

4 Experimental analysis

In this section we investigate the adequacy of dissimilarity representations for thermal IR face recognition in the presence of low resolution images.

4.1 Data

For the experiments we used the Equinox face database, which is a benchmark database for thermal IR face recognition. It was collected by Equinox Corporation under DARPAs HumanID program [1]. The images of the database have a size of 320X240 pixels. The LWIR images were stored as grayscale images with 12 bits per pixel. Image sequences of each subject with three illumination conditions, frontal, left and right illumination, were acquired. For each illumination, image sequences of 40 frames were taken with the subjects pronouncing the vowels. Three static shots of each subject with the expressions smile, frown, and surprise were also taken. The complete process was repeated for those subjects wearing glasses. In total the database has 89 subjects.
The images are affected by a non uniformity effect manifested in a fixed pattern noise, that is a pixel to pixel variation in the sensor array caused by the difference in the semiconductors. Examples of images from the database as well as the described effect can be seen in Fig. 1.

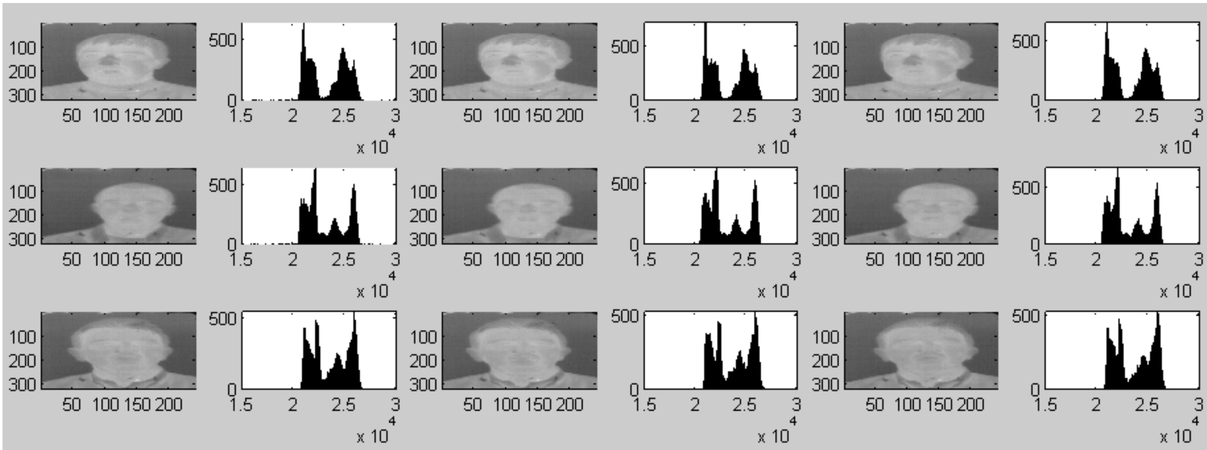

Figure 2: Examples of images with the highest resolution and their related histograms. The rows contain images of different subjects, the columns contain examples of images of the same subject.

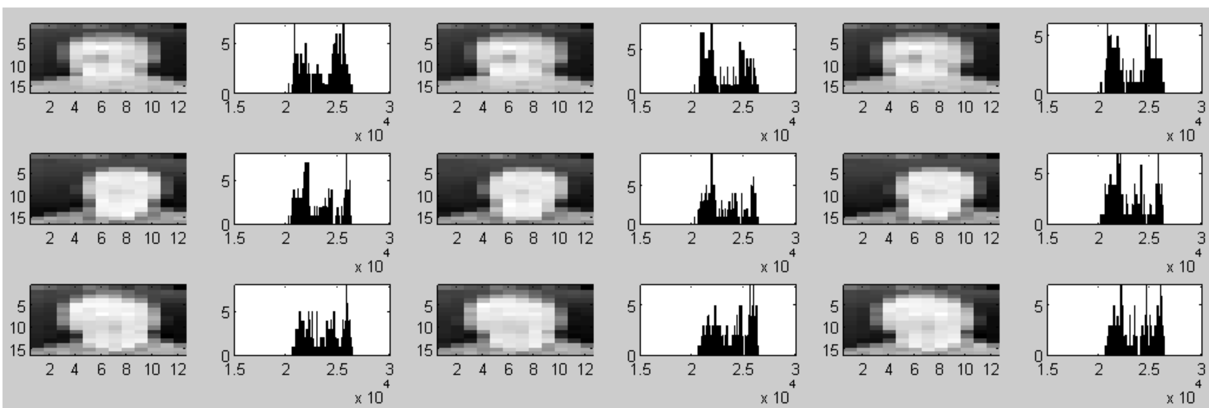

Figure 3: Examples of images with the lowest resolution and their related histograms. The rows contain images of different subjects, the columns contain examples of images of the same subject.

4.2 Experimental setup and preprocessing

The methodology described in [18] was followed for the experiments, but the subsets of the subjects wearing glasses were discarded. The description of the used subsets is the following:
VA: Vowel frames, all subjects, all illuminations.
EA: Expression frames, all subjects, all illuminations.
VF: Vowel frames, all subjects, frontal illumination.
EF: Expression frames, all subjects, frontal illumination.
VL: Vowel frames, all subjects, lateral illumination.
EL: Expression frames, all subjects, lateral illumination.

VA, VF, and VL vowel subsets were used as training sets, one at a time. VA, VF, VL, EA, EF and EL were used as test sets. In each training set we have 3 images per subject, amounting to 267 images. When the vowels subsets are taken as test sets, 9 images per subject are tested for classification (801 images in total), but only 3 images per subject at a time. Then the results of the 3 iterations are averaged. When the expression subsets are used as test sets, 6 images per subject are tested for classification (534 images in total). In this case also 3 iterations are made, and 2 images per subject are submitted for classification at a time. The final result is the average of the 3 iterations.
Training and test sets are independent except for some images that are repeated in the subsets with all illuminations because they contain images with frontal and lateral illuminations.
In LWIR face images there is a lack of accurate techniques for detecting face fiducial points. This points are needed for the geometric normalization of the face. Most of the feature representations will fail to describe the patterns properly if the face images are not aligned or registered before their feature extraction. We overcome this limitation using an histogram based representation that is robust to head rotations

		VF				VA				VL			
		320x240	80x60	32x24	16x12	320x240	80x60	32x24	16x12	320x240	80x60	32x24	16x12
VF	*1-NN*					99.87	100	99.50	98.00	100	100	99.25	95.38
	Diss					98.87	98.87	99.25	98.25	98.87	98.87	99.00	97.87
VA	*1-NN*	99.37	99.62	99.12	93.50					100	100	100	98.37
	Diss	99.62	99.75	99.75	97.78					100	100	100	99.75
VL	*1-NN*	99.75	99.87	99.12	90.13	100	100	99.75	97.12				
	Diss	99.00	99.25	99.37	95.88	99.87	100	99.87	98.62				
EF	*1-NN*	100	100	99.62	95.50	100	100	99.25	94.38	100	100	99.62	95.88
	Diss	99.62	99.62	99.62	97.75	100	99.62	99.62	98.12	100	100	99.62	97.37
EA	*1-NN*	100	99.62	99.62	88.38	100	100	100	96.25	100	100	100	99.62
	Diss	99.62	99.62	99.25	96.62	100	100	100	98.12	100	100	100	99.25
EL	*1-NN*	99.81	99.62	99.06	88.38	100	100	100	95.56	100	100	99.81	97.94
	Diss	96.81	98.87	99.25	96.81	98.68	100	99.43	96.62	99.25	100	99.43	98.68

Table 1: Correct classification rates (CCR) with different image sizes. *1-NN* rows refers to CCR using the 1-NN classifier on Chi square distance matrix, and *Diss* rows refers to CCR using a 1-NN classifier on the dissimilarity space.

image size	1-NN on distance matrix	1-NN on dissimilarity space
320x240 pixels	99.92	99.35
80x60 pixels	99.91	99.63
32x24 pixels	99.58	99.56
16x12 pixels	94.96	97.83

Table 2: Average correct classification rates with all the image sizes, 1-NN classifier on the Chi square distance matrix and 1-NN classifier on the dissimilarity space.

and horizontal translation of the body in the scene. This representation of the whole face image including the scene background allow us to skip the face preprocessing step.

For the experiments four different image sizes were considered: 320x240, 80x60, 32x24 and 16x12 pixels. An example of images with the highest and lowest resolution and their related histograms is shown in Fig. 2 and Fig. 3.

We constructed histograms of 256 bins. They were normalized with respect to the number of pixels of the image.

4.3 Results and Discussion

As a reference classifiers on feature representations we tested the 1-NN classifier using the Chi square distance between histograms of elements in test and training sets. This classifier can be seen as a dissimilarity based one since it interprets dissimilarities in pretopological spaces via the neighborhoods [15], but in our study we are only considering dissimilarity space vectors as dissimilarity representations. From now on the 1-NN classifier that interprets directly the Chi square distance matrix will be referred as the 1-NN on distance matrix or the feature based classifier.

As representation set for projecting the patterns in the dissimilarity space we used all the training objects. First, the Chi square distances of the histograms of the images in the training set with themselves were computed. With this procedure the training set was projected in the dissimilarity space. Then, the Chi distances between the histograms of the elements in the test set and the elements in the training set were computed and the test set was projected in the dissimilarity space. As a dissimilarity space classifier we used the 1-NN rule on the Euclidean distances between the Chi square dissimilarity vectors. This choice was made because in face recognition applications, the amount of subjects(classes) in the database grows regularly and an untrained classifier like 1-NN is needed to handle this. This classifier will be referred as the 1-NN on dissimilarity space or as the dissimilarity based classifier.

The correct classification rates with training sets in the columns and test sets in the rows are shown in Table 1. The average classification rates for the 4 selected sizes are presented in Table 2. Our classification results on the 320x240 images, using both the feature based and the dissimilarity based classifiers, are comparable to or better than previous results reported in the literature.

Some classification results obtained in [16] using the LDA method are showed in Table 3. This LDA method showed the best performance in a comparison with other methods in [16]. The authors of this approach did not consider some training/test combinations of the subsets since some images of one subset are included in the other. The average classification rates of the LDA and our dissimilarity based and feature based classifiers using the subsets taken into account in the LDA approach, are presented in Table 4.

With the high resolution images (320x240), the Chi square distance proved to be highly discriminative, and the 1-NN classifier took advantage of this, pro-

	VF	VA	VL
VF			98.8
VA	98.3		99.6
VL	97.4		
EF	94.6	97.2	95.6
EA	94.00	97.4	96.8
EL	93.7	97.4	97.4

Table 3: Correct classification rates with the LDA method on 320x240 images.

LDA method	1-NN on distance matrix	1-NN on dissimilarity space
96.78	99.91	99.34

Table 4: Average correct classification rates of the LDA method and our feature and dissimilarity based classifiers on 320x240 images.

viding excellent classification results. The 1-NN rule on the dissimilarity space performed worse. We think that this happens because not all the training set objects are good for representing our data. With the lowest image resolution (16x12) the 1-NN rule on the dissimilarity space outperformed the 1-NN rule on the dissimilarity matrix. We think that this is due to the fact that for the high resolution images the measure is sufficiently discriminative, but when the image resolution decreases the Chi square distance between the histograms becomes more noisy. As the representation set is large, dissimilarity vectors suffer less from noise than individual dissimilarities, and also the Euclidean distance in the dissimilarity space averages this noise.

5 Conclusions

In this paper we proposed the use of the Chi square measure computed on the image histogram for robust dissimilarity representation of thermal LWIR face images. The use of histograms in our approach arise as a good alternative for representing the face information since the thermal skin emission values are suitable encoded in the thermal LWIR image histograms. We think that this is due to the fact that there are only 89 persons in the database. For a very large database this may not hold because with pixel permutations we obtain the same histogram. The Chi square distance proved to be highly discriminative in spite of the fact that no face normalization was made. One advantage of this representation is that the thermal histograms are invariant to head rotations and horizontal shifts of the body in the scene. This is very important due to the lack of accurate techniques for finding landmark points for the alignment of LWIR face images. This representation also appeared to be insensitive to the non uniformity noise.

A dissimilarity representation was presented and compared with feature representations. In the case of very low resolution images (16x12 pixels), the dissimilarity space classification outperformed the feature based classification. In spite of the fact that the dissimilarity space classification proved to be a good alternative to classify the images, did not outperformed the feature based classifier on 320x240 image size. The remark on the suitability of dissimilarity representations for this low resolution images can be taken into account for building face recognition systems at a distance when only short periods of time have elapsed since the gallery picture was taken. One application can be the tracking of a person in a room containing more persons. In this application, a preprocessing should be done, in order to detect and segment all the faces that are present in the scene.

References

[1] Available at: http://www.equinoxsensors.com /products/HID.html.

[2] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. In *ECCV (1)*, pages 469–481, 2004.

[3] George Bebis, Aglika Gyaourova, Saurabh Singh, and Ioannis Pavlidis. Face recognition by fusing thermal infrared and visible imagery. *Image Vision Comput.*, 24(7):727–742, 2006.

[4] H. Bunke and K. Riesen. Graph classication based on dissimilarity space embedding. In *N. da Vitoria Lobo et al., editor, Structural, Syntactic, and Statistical Pattern Recognition, LNCS 5342*, pages 996–1008, 2008.

[5] RuFeng Chu, ShengCai Liao, and Lun Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):627–639, 2007. Senior Member-Stan Z. Li.

[6] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.

[7] Robert P. W. Duin and Elzbieta Pekalska. Open issues in pattern recognition. In *CORES*, pages 27–42, 2005.

[8] Aglika Gyaourova, George Bebis, and Ioannis Pavlidis. Fusion of infrared and visible images for face recognition. In *Lecture Notes in Computer Science 3024 2004*, pages 456–468, 2004.

[9] Jingu Heo. Fusion of visual and thermal face recognition techniques: A comparative study. In *University of Tennessee*, 2005.

[10] Sang-Woon Kim. On using a dissimilarity representation method to solve the small sample size problem for face recognition. In *ACIVS*, pages 1174–1185, 2006.

[11] Sang-Woon Kim and Jian Gao. On using dimensionality reduction schemes to optimize dissimilarity-based classifiers. In *CIARP*, pages 309–316, 2008.

[12] Seong G. Kong, Jingu Heo, Besma R. Abidi, Joonki Paik, and Mongi A. Abidi. Recent advances in visual and infrared face recognition - a review. *Computer Vision and Image Understanding*, 97:103–135, 2005.

[13] Mauricio Orozco-Alzate and César Germán Castellanos-Domínguez. Nearest feature rules and dissimilarity representations for face recognition problems. In Kenzo Kurihara, editor, *Face Recognition*, chapter 18, pages 337–356. International Journal of Advanced Robotic Systems, Vienna, Austria, May 2007.

[14] Elzbieta Pekalska and Robert P. W. Duin. Classifiers for dissimilarity-based pattern recognition. In *ICPR*, pages 2012–2016, 2000.

[15] Elzbieta Pekalska and Robert P. W. Duin. *The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence)*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2005.

[16] Diego A. Socolinsky and Andrea Selinger. A comparative analysis of face recognition performance with visible and thermal infrared imagery. In *ICPR '02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 4*, page 40217, Washington, DC, USA, 2002. IEEE Computer Society.

[17] Diego A. Socolinsky, Andrea Selinger, and Joshua D. Neuheisel. Face recognition with visible and thermal infrared imagery. *Comput. Vis. Image Underst.*, 91(1-2):72–114, 2003.

[18] Diego A. Socolinsky, Lawrence B. Wolff, Joshua D. Neuheisel, and Christopher K. Eveland. Illumination invariant face recognition using thermal infrared imagery. In *CVPR (1)*, pages 527–534, 2001.