

A Study on Representations for Face Recognition from Thermal Images

Yenisel Plasencia^{1,2}, Edel García-Reyes¹, Robert P.W. Duin²,
Heydi Mendez-Vazquez¹, César San-Martin³, and Claudio Soto³

¹ Advanced Technologies Application Center. 7a # 21812 e/ 218 y 222, Rpto. Siboney,
Playa, C.P. 12200, La Habana, Cuba

{yplasencia,hmendez,egarcia}@cenatav.co.cu

² Faculty of Electrical Engineering, Mathematics and Computer Sciences,
Delft University of Technology, The Netherlands

r.duin@ieee.org

³ Information Proc. Lab, DIE, Universidad de La Frontera. Casilla 54-D,
Temuco, Chile

csmarti@ufro.cl, c.soto04@ufromail.cl

Abstract. Two challenges of face recognition at a distance are the uncontrolled illumination and the low resolution of the images. One approach to tackle the first limitation is to use longwave infrared face images since they are invariant to illumination changes. In this paper we study classification performances on 3 different representations: pixel-based, histogram, and dissimilarity representation based on histogram distances for face recognition from low resolution longwave infrared images. The experiments show that the optimal representation depends on the resolution of images and histogram bins. It was also observed that low resolution thermal images joined to a proper representation are sufficient to discriminate between subjects and we suggest that they can be promising for applications such as face tracking.

Keywords: face recognition, dissimilarity representations, thermal infrared.

1 Introduction

The use of longwave infrared (LWIR) imagery ($8 - 12\mu\text{m}$) for developing face recognition systems has started to receive attention in the last years because of its robustness to illumination changes [1]. LWIR (also called thermal infrared) sensors collect the heat energy emitted by a body instead of the reflected light, this allows them to operate even in complete darkness. The use of the face thermal patterns has been validated as a biometric signature on short time scale. We are interested in the study of such imagery modality for face recognition at a distance, what implies to handle low resolution images.

The opaqueness to glass is one of the limitations of thermal infrared face recognition. Studies in the literature show that there is a preference for fusing

thermal and visible imagery to tackle this problem [2,3,4]. 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. However, for an application such as short term face tracking, the use of this imagery modality can be beneficial because temperature changes are negligible.

A number of algorithms has been proposed for the classification of faces from LWIR images, but in the representation level, only feature representations have been studied. Three major groups of methods can be distinguished for representation in the pattern recognition literature. The first group treats objects as vectors in Euclidean vector spaces. The second represents the object structure by graphs, grammars, etc. The third and newest group represents the objects by their nearness relation with respect to other objects. The vector space representation is the most developed one due to the existence of several statistical techniques that have shown a good performance on the different pattern recognition problems. Pekalska et al. [5] proposed the dissimilarity representation. It arises from the idea that the classes are constituted by similar objects, so (dis)similarities are more important than features for the class constitution. This dissimilarity approach has the potential of unifying the statistical and the structural approaches [6] because for example, statistical techniques can be used on a dissimilarity representation derived from graphs.

In this paper we will study classification performances on a dissimilarity space based on histogram distances, on the feature space of the histograms, and on the pixel space for face recognition from low resolution longwave infrared images. We will study in which conditions one representation is better than the other. Section 2 presents related work in face recognition using dissimilarity representations. Section 3 introduces the dissimilarity space representation. Histograms and the Chi Square distance are briefly described in Section 3. Section 4 presents the experimental results, including data, experimental setup, and discussion. The conclusions are drawn in Section 5.

2 Related Work

There are some studies in the literature where dissimilarity representations are introduced for face recognition, but none of them make use of thermal infrared imagery. In [7] after reducing dimensionality by Principal Component Analysis (PCA), 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 nearest neighbor (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 [8], 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.

3 Dissimilarity Space

The dissimilarity space was proposed by Pekalska et al. [5]. It was postulated as a Euclidean vector space. For its construction a representation set $R = \{r_1, r_2, \dots, r_n\}$ is needed, where the objects belonging to this set (also called prototypes) are chosen adequately based on some criterion that can be dependent of the problem at hand. Let X be the training set, R and X can have the following relationships: $R \cap X = \emptyset$, 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.

3.1 Histograms

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. 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 nonuniformity 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 [9] was used. This distance has been proving to be effective for histogram comparison. Let S and M be two histograms and n the number of bins in the histogram. The Chi square distance is defined as follows:

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



Fig. 1. Examples of LWIR images from the Equinox database

4 Experiments

Our goal is to compare a pixel-based, a histogram representation and a dissimilarity representation based on the histograms by means of classification accuracies of 1-NN and LDA classifiers for thermal infrared face recognition in the presence of low resolution images.

4.1 Data and Experimental Setup

For the experiments the Equinox face database was used, which is a benchmark database for thermal infrared face recognition. It was collected by Equinox Corporation under DARPA's HumanID program [10]. In total the database contains grayscale images of 89 subjects with 12 bits per pixel and 320x240 pixels of size.

The methodology described in [11] was followed for the experiments, but the subsets containing the pictures of the subjects wearing glasses were discarded. Different subsets were considered using (F)rontal and (L)ateral illuminations. VF and VL are (V)owel subsets including images of the subjects moving the mouth to pronounce the vowels. EF and EL are (E)xpression subsets including images of the subjects with different expressions. VA is composed of images taken from VF and VL. EA is composed of images taken from EF and EL. The experimental setup is shown in Table 1. Each subset used for training and test contains 267 images (3 images per 89 subjects).

In LWIR face images there is a lack of accurate techniques for detecting face landmark points. These points are needed for the geometric normalization of the face. We try to overcome this limitation using an histogram based representation that is robust to head rotations and horizontal shifts of the face in the scene. For the experiments 5 different image sizes were considered: 320x240, 80x60, 32x24, 16x12, and 6x8 pixels. An example of images with the 320x240 and 16x12 resolution and their related histograms is shown in Fig. 2 and Fig. 3.

As a reference we tested the pixel-based representation without a geometric normalization of the face. For the histogram representation the number of bins is 256. Histograms were normalized with respect to the number of pixels of the image. The dissimilarity representation was conformed using the Chi Square distance between the histograms. As representation set for projecting the patterns in the dissimilarity space we used the entire training set.

Table 1. Experimental setup

Setup	Training set	Test sets	Result
VL/VF	VL	VF1, VF2, VF3	$a_1 = \text{Average}(\text{VL}/\text{VF1}, \text{VL}/\text{VF2}, \text{VL}/\text{VF2})$
VF/VA	VF	VA1, VA2, VA3	$a_2 = \text{Average}(\text{VF}/\text{VA1}, \text{VF}/\text{VA2}, \text{VF}/\text{VA2})$
VL/VA	VL	VA1, VA2, VA3	$a_3 = \text{Average}(\text{VL}/\text{VA1}, \text{VL}/\text{VA2}, \text{VL}/\text{VA2})$
VF/VL	VF	VL1, VL2, VL3	$a_4 = \text{Average}(\text{VF}/\text{VL1}, \text{VF}/\text{VL2}, \text{VF}/\text{VL2})$
VF/EF	VF	EF1, EF2	$a_5 = \text{Average}(\text{VF}/\text{EF1}, \text{VF}/\text{EF2})$
VA/EF	VA	EF1, EF2	$a_6 = \text{Average}(\text{VA}/\text{EF1}, \text{VA}/\text{EF2})$
VL/EF	VL	EF1, EF2	$a_7 = \text{Average}(\text{VL}/\text{EF1}, \text{VL}/\text{EF2})$
VF/EA	VF	EA1, EA2	$a_8 = \text{Average}(\text{VF}/\text{EA1}, \text{VF}/\text{EA2})$
VA/EA	VA	EA1, EA2	$a_9 = \text{Average}(\text{VA}/\text{EA1}, \text{VA}/\text{EA2})$
VL/EA	VL	EA1, EA2	$a_{10} = \text{Average}(\text{VL}/\text{EA1}, \text{VL}/\text{EA2})$
VF/EL	VF	EL1, EL2	$a_{11} = \text{Average}(\text{VF}/\text{EL1}, \text{VF}/\text{EL2})$
VA/EL	VA	EL1, EL2	$a_{12} = \text{Average}(\text{VA}/\text{EL1}, \text{VA}/\text{EL2})$
VL/EL	VL	EL1, EL2	$a_{13} = \text{Average}(\text{VL}/\text{EL1}, \text{VL}/\text{EL2})$

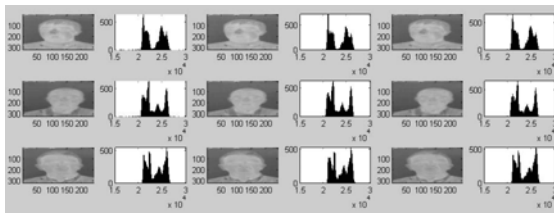


Fig. 2. Examples of images with resolution of 320x240 and their related histograms. Each row contains images of one subject.

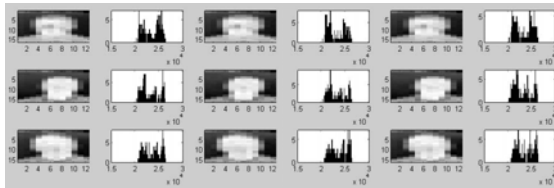


Fig. 3. Examples of images with resolution of 16x12 and their related histograms. Each row contains images of one subject.

Table 2 shows the results in terms of means and standard deviations of classification accuracies of 1-NN and LDA classifiers for the 3 different representations and the 5 different image resolutions. This classifiers are very different since LDA is linear and global and 1-NN is highly nonlinear and a local classifier. Each mean is calculated over the 13 results (a_1, a_2, \dots, a_{13}) in Table 1. In all the experiments the data was reduced to 48 dimensions with PCA in order to avoid regularization parameters for LDA and make the approaches comparable. The 1-NN classifier uses the Euclidean distance for all representations. The results where the accuracy is statistically significantly higher for each classifier and resolution for the different representations are in bold. This is evaluated using the formula $\gamma = |\mu_1 - \mu_2| - \frac{(\sigma_1 + \sigma_2)}{\sqrt{N}}$, where μ_1 and μ_2 are the two means of classifier

Table 2. Average classification accuracies and (standard deviations) for 1-NN and LDA classifiers on a dissimilarity, a histogram, and a pixel-based representation for different image resolutions

Resolution	Classifier	Dissimilarity Rep	Histogram Rep	Pixel-based Rep
320x240	LDA	91.88 (10.78)	86.01 (15.91)	78.20 (21.62)
	1-NN	99.34 (0.85)	99.73 (0.25)	95.17 (2.31)
80x60	LDA	95.24 (6.66)	91.52 (10.98)	62.61 (26.14)
	1-NN	99.64 (0.41)	99.67 (0.34)	95.24 (2.30)
32x24	LDA	99.53 (0.78)	97.14 (3.41)	66.81 (25.72)
	1-NN	99.57 (0.28)	99.27 (0.48)	95.40 (2.29)
16x12	LDA	98.07 (1.22)	98.38 (1.47)	68.95 (27.04)
	1-NN	97.34 (1.09)	94.26 (2.47)	95.83 (2.12)
8x6	LDA	75.53 (7.82)	76.21 (7.82)	71.79 (25.67)
	1-NN	66.73 (7.65)	53.64 (8.53)	96.32 (2.02)

accuracies, and σ_1 and σ_2 are the standard deviations and N is the number of experiments. Taking a threshold $\alpha = 0.5$ the difference on the means is statistically significant if $\gamma > \alpha$.

4.2 Results and Discussion

Our classification results using the 1-NN on the 320x240 images, using both the histogram and the dissimilarity representations, are comparable to or better than previous results reported in the literature. For example in [1] the authors tested the LDA method on the geometrically normalized 324x240 images and the average classification accuracy was 96.78. In our case, despite the fact that no geometric normalization was made, with the 1-NN on the dissimilarity space and the 1-NN on the histograms we achieved 99.34 and 99.73 of accuracy on the same data.

For the 320x240 resolution we can see that the two classifiers on the pixel-based representation suffer from the fact that the faces are not aligned, so classification results are the lowest. By using histograms, we achieve invariance to horizontal shifts and rotations of the face inside the image, leading to better classification results than with the pixel-based representation.

For 80x60, 32x24, and 16x12 resolutions, classification results on dissimilarity and histogram representations continue to be better than classification results on the pixel-based representations. Also for this low resolutions, classification accuracies are statistically significantly higher when using dissimilarity representations and in 2 of 3 cases the highest accuracies correspond to the 1-NN on the dissimilarity representation. By decreasing the resolution, histograms are more sparse so bins become less reliable features. Dissimilarities can handle this a little bit better. Also, the LDA classifier performs better because is a more global classifier and suffer less from this noise.

For the 32x24 resolution we can observe that classification results on both the histogram and the dissimilarity representations are similar to or better than classification results on the high resolution images. This may suggest that for recognizing faces from LWIR images we do not need a high resolution for the

Table 3. Average classification accuracies (and standard deviations) when decreasing the resolution of histograms to 50, 20, 10, and 5 bins for the 8x6 image resolution

Classifier	Dissimilarity Rep	Histogram Rep
LDA hist 50 bins	93.13 (3.29)	91.18 (4.40)
1-NN hist 50 bins	86.94 (4.36)	86.28 (4.39)
LDA hist 20 bins	87.50 (4.38)	92.97 (2.76)
1-NN hist 20 bins	86.01 (4.28)	84.23 (4.68)
LDA hist 10 bins	79.68 (5.13)	86.17 (3.92)
1-NN hist 10 bins	77.71 (6.16)	77.11 (6.12)
LDA hist 5 bins	30.73 (27.65)	64.56 (4.46)
1-NN hist 5 bins	55.22 (5.82)	58.30 (6.20)

images. This conclusion can only be applicable to databases with a controlled size like in our setups.

In the case of 8x6 images, classifiers on histogram and dissimilarity representations perform very poor compared to the 1-NN on the pixel-based representation. The pixel-based representation is performing better because by decreasing the resolution, the faces become aligned. On the other hand, the poor performance of histograms and dissimilarities can be attributed to the fixed number of histogram bins that leads to very sparse histograms when the resolution is as low as 8x6. To prove this, we conducted some experiments diminishing the number of bins of the histograms. The new bin values are the summations over neighboring bins of the larger histograms. The resolution of the histograms was decreased to 50, 20, 10, and 5. The results are shown in Table 3. We can observe that classification results can be improved if the histogram resolution is decreased while the image resolution is decreased (e.g. 50, 20, and 10 bins), but we need to take care of the selected histogram size because classification accuracy starts to decrease for very small sizes. When using histograms of 5 bins classification results are no longer better than those using 256 bins. The best result is obtained with LDA on the dissimilarity representation using histograms of 50 bins.

5 Conclusions

It is very common to find studies on complicated representations for face recognition from visual imagery. This is needed in order to achieve invariance to factors that affect visual images and degrade the recognition performance such as ambient illumination changes. By using the thermal imagery modality, invariance to illumination changes is achieved. Therefore, very simple representations can be suitable for this type of images.

In this paper we compared some representations such as a pixel-based, a histogram representation and a dissimilarity representation based on the histograms for face recognition from low resolution LWIR images. We find out that histograms characterize the subjects sufficiently, and dissimilarities on the histograms may improve this. For low resolutions images, histograms become sparse and results deteriorate. The pixel-based representation can now perform very well because the faces become more aligned. A good tradeoff between image

resolution and histogram resolution may be needed. It was also observed that low resolution thermal images joined to a proper representation are sufficient to discriminate between subjects and we suggest that they can be used in practical applications such as face tracking.

Acknowledgements. We acknowledge financial support from the FET programme within the EU FP7, under the SIMBAD project (contract 213250).

References

1. Socolinsky, D.A., Selinger, A.: A comparative analysis of face recognition performance with visible and thermal infrared imagery. In: ICPR 2002, Washington, DC, USA, vol. 4, p. 40217. IEEE Computer Society, Los Alamitos (2002)
2. Bebis, G., Gyaourova, A., Singh, S., Pavlidis, I.: Face recognition by fusing thermal infrared and visible imagery. *Image Vision Comput.* 24(7), 727–742 (2006)
3. Heo, J.: Fusion of visual and thermal face recognition techniques: A comparative study. University of Tennessee (2005)
4. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: Recent advances in visual and infrared face recognition - a review. *Computer Vision and Image Understanding* 97, 103–135 (2005)
5. Pekalska, E., Duin, R.P.W.: The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence). World Scientific Publishing Co., Inc., River Edge (2005)
6. Bunke, H., Riesen, K.: Graph classification based on dissimilarity space embedding. In: da Vitoria Lobo, N., Kasparis, T., Roli, F., Kwok, J.T., Georgiopoulos, M., Anagnostopoulos, G.C., Loog, M. (eds.) SSSPR 2008. LNCS, vol. 5342, pp. 996–1007. Springer, Heidelberg (2008)
7. Orozco-Alzate, M., Castellanos-Domínguez, C.G.: Nearest feature rules and dissimilarity representations for face recognition problems. In: Kurihara, K. (ed.) Face Recognition; International Journal of Advanced Robotic Systems, Vienna, Austria 337–356 (May 2007)
8. Kim, S.W.: On using a dissimilarity representation method to solve the small sample size problem for face recognition. In: ACIVS, pp. 1174–1185 (2006)
9. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: ECCV, vol. (1), pp. 469–481 (2004)
10. <http://www.equinoxsensors.com/products/HID.html>
11. Socolinsky, D.A., Wolff, L.B., Neuheisel, J.D., Eveland, C.K.: Illumination invariant face recognition using thermal infrared imagery. In: CVPR, vol. (1), pp. 527–534 (2001)