Ridges, Valleys and Hausdorff Based Similarity Measures for Face Description and Matching.

Abstract

This paper proposes a system for recognizing face images independently of illumination and expression changes.The proposed system is based on the use of image ridges and valleys as illumination invariant feature descriptor, and Hausdorff based distances as flexible similarity measure. Image edges have been usually considered as an adequate face shape descriptor due to its robustness to illumination changes. This paper presents experimental evidence showing that, when face images are considered, image ridges and valleys behave in a more robust way to illumination changes and convey more identity information than edges.

1 Introduction

Automatic Face Recognition (AFR) has been a successful field of research mostly during the past two decades. This is evident when considering the continuous publication of reviews and surveys, from the earliest of Samal and Iyengar[11], to the latest of Grudin[8], passing through the works of Valentin et al[10], and Chellapa et al[9]. Two of the main problems when recognizing face images are those of face expression and illumination changes. The first source of variability produces a change in the intensities received by the image sensors, meanwhile the second one produces a rearrangement of the intensities across the sensing area. Currently, some of the most successful AFR systems rely on statistical methods to solve these two problems. Statistical methods try to derive (learn), using image samples, the smallest set of features that convey as much of the image identity information and at same time remains more constant (invariant) to usual face image changes (in our case both changes in illumination and facial expression). At difference of the statistical approach, the work presented in this paper uses as image feature extractor the MLSEC operator [2], obtaining image ridges and valleys which are, by definition, invariant to illumination changes. In addition, our method uses Hausdorff-based distances [5, 6] as similarity measure between the obtained feature descriptors.

Hausdorff based distances are flexible matching techniques that will allow us to measure distortion (displacement or position rearrangement of the features) rather than changes in image intensities. In this manner, the analysis of the

response of the system for isolated image changes (illumination or face expression) will allows us to analyze which of the two above parts (feature extractor or similarity measure) requires improvement.

The experimental results reported in this paper compare the performance (in terms of recognition ratios) of the proposed descriptors and similarity measures to those obtained using edges and gray level normalized images. We will show that, in the tested database, image valleys and ridges behaves in a more robust way, under illumination changes, than edges and gray level normalized representations, and that best recognition ratios are obtained when Hausdorff-based distances computed using ridges and valleys separately are combined in a joined distance.

The contents of the paper are distributed as follows. Section 2 reviews the image descriptors used in the comparative analysis. Section 3 introduces the Hausdorff measures we have considered. Section 4 presents the experimental results. Finally, Sect.5 summarizes the main conclusions of this paper.

2 Image Descriptors

Image analysis is commonly understood as a signal to symbol transformation which consists of computing low-level features that are then used by high-level tasks to automatically segment and classify the interesting objects appearing in the scene. Geometric descriptors are used as relevant low-level image features, among them, the well-known *edges* and the so-called *ridges* and *valleys* play an important role in many applications. In [3] we can find a comparison of edge detection methods and in [2] the same for ridges and valleys.

In the computer vision literature the Canny's edge detector [4] is considered as being one of the best, therefore, in the context of face recognition it is the detector we have chosen to represent the edge detection methods. Assuming additive zero-mean Gaussian noise, Canny looks for an optimal edge detector in the sense of three coupled criteria: detection (locate all real edges), localization (distance between the detected edges and the real ones) and thinness (only one response *per* real edge). In 2D, the approach of Canny basically consists of two steps: **i**) computing the magnitude of the image gradient, where the image has been smoothed by means of a Gaussian kernel; **ii**) delineating the maxima of this magnitude according to the so-called maximum suppression criterion and a threshold process with hysteresis.

Contrarily to the case of edges, where people agree in their mathematical characterization and the effort is driven to find out efficient methods robust to noise, the case of ridges and valleys is more complex. In the literature we can find a plethora of mathematical characterizations that try to formalize the intuitive notion of ridge/valley (ridges and valleys are equivalent in the sense that the ridges of an image are the valleys of the inverted image, and the way around). In the context of face recognition, we have used the valleys obtained by thresholding the so-called *multi local level set extrinsic curvature* (MLSEC) [2, 1]. We have chosen the MLSEC due to their invariance both to rigid image

Figure 1: Response of the considered face image descriptors applied to two images(rightmost column) undergoing changes in expression and illumination.

motions and monotonic grey-level changes and, mainly, because its high continuity an meaningful dynamic range, in opposition to other measures with the same invariances.

Basically, the valleys based on the MLSEC are delineated by:

1. computing the normalized gradient vector field of the smoothed image (usually a Gaussian smoothing).

2. calculating the divergence of this vector field, which gives rise to a bounded and well-behaved measure of valleyness (negative values running from -2 to 0 in 2D) and ridgeness (positive values from 0 to 2).

3. thresholding the response of the above described operator, so that image pixels where the MLSEC response is smaller than -1 are considered valleys (black pixels in Fig. 1 fifth column), and those pixels larger than 1 (white pixels in the images at the leftmost column of 1) are considered ridges.

Besides of its desirable illumination invariant behavior, the relevance of valleys in the face shape description has been pointed out by some cognitive science works[14]. Among others, Pearson et al, hypothesize that this kind of filters are used as an early step by the human visual system (HVS). They found their assertions in the human behavior when recognizing faces, making these descriptors good candidates for the AFR systems.

3 Similarity Measures

Once we have the set of image points where relevant features have been detected, we need to measure how this set of feature positions is distorted with respect to the set of features detected in a model image. This measure will allow us to decide if the differences between images are due to changes of expression or identity. Hausdorff distance [5, 6] is a measure of similarity between two sets of points belonging to the same metric space. This measure has been used before both for face detection [13] and recognition [12] (using the response of an edge detector as face descriptor). The main advantage of this measure is that an explicit correspondence between points is not required. This requirement

is overcome through an implicit nearest-neighbor correspondence between the points of the sets. Formally, given two sets of points A, and B, the Hausdorff distance $H(A,B)$ between both sets is defined as:

$$\begin{array}{c} H(A,B) = \max(h(A,B), h(B,A)) \\ h(A,B) = \max_{a \in A}(\min_{b \in B}(d(a,b))) \end{array} \quad (1)$$

$h(A,B)$ being the direct Hausdorff distance from the set A to the set B, and $d(a,b)$ the distance, usually Euclidean, between the points a and b. It has to be noted that the direct Hausdorff distance, in general, is not symmetric. Several variations of the original Hausdorff distance (HD) can be found in the literature. All of them, can be conveyed in a general framework where Hausdorff based measures are defined as:

$$\begin{array}{c} H(A,B) = f^s(h(A,B), h(B,A)) \\ h(A,B) = f^i_{a \in A}(f^m(\min_{b \in B} d(a,b)) \end{array} \quad (2)$$

We will call "*symmetrizer*", "*integral*", and "*best-match*" function to f^s, $f^i_{a \in A}$, and f^m, respectively. Using this framework, HD would be obtained when *maximum* is considered both for the *symmetrizer* and *integral* functions, and *identity* is considered as the *best-match* function. Dubuisson and Jain [6], tested the robustness to synthetic noise of the Hausdorff distance and some of their possible modifications in an image edge matching experiment. The authors concluded that the most robust measure to noisy data was the Modified Hausdorff Distance (MHD). The MHD is obtained when *average* and *maximum* are considered , respectively, as *integral* and *symmetrizer* functions. In order to test the behavior of the HD modifications tested in [6] when noise is produced by face and acquisition image changes, a first set of experiments was conducted, obtaining slight enhancement in the results when *average* and *product* were considered as *integral* and *symmetrizer*. For this reason we have used it in the rest of our experiments. We will further refer to this measure as PAI-HM (standing for product (f^s), average (f^i) and identity (f^m) Hausdorff based measure).

Note that valleys and ridges of an image define two disjoint sets (representing complementary image features). In this case, two face images, I_A and I_B, are described now by four sets of points A_V, A_R, B_V and B_R, (valleys and ridges obtained from the first image and valleys and ridges obtained from the second image respectively). These allows us to define four meaningful direct Hausdorff based measures $h(A_V, B_V)$, $h(A_R, B_R)$, $h(B_V, A_V)$, $h(B_R, A_R)$ that will be combined using the *symmetrizer* function. In this manner we use as similarity measure:

$$\text{PAI-HM}_{R\&V}(I_A, I_B) = h(A_V, B_V)h(A_R, B_R)h(B_V, A_V)h(B_R, A_R)$$

4 Experimental Results

We have tested the proposed joined ridge and valley similarity measure using a subset of 742 images from the AR-Face database [1]. These images correspond to

[1] Publicly available from http://www.cvc.uab.es/shared/arees/FaceDB.html

Figure 2: Examples of the acquisition conditions used to test the descriptor and similarity measures.

106 subjects (seven shoots per subject labeled from **(a)** to **(g)** in Fig.2). Each of the labels denotes an acquisition condition, that can be grouped into two sets of images: **i**) images conveying illumination changes where subjects show the same, neutral, expression(**(a)** diffuse illumination **(d)** left light **(e)** right light **(f)** frontal light); **ii**) images conveying facial expression changes maintaining the same, diffuse, illumination condition (**(a)** neutral expression, **(b)** smile, **(c)** angry, **(g)** image taken in a second session two weeks later). Eye location has been used in order to normalize the images in size, position, and orientation. Face images are then cropped and scaled to 78x68 pixels.

AFR systems are usually composed of two phases or steps. First, in a recruitment step, reference images of the subjects to be recognized (targets) are acquired. These images are processed, obtaining a set of feature descriptors, and stored in a database usually referred as the galley. Then, in a recognition step, image of subjects (query or probe) are presented to the system, and the system should decide if the acquired image belongs to one of the gallery subjects (recognition) and, in that case, to which subject does it belongs (identification).

When testing recognition systems where gallery and probe images can undergo different acquisition changes (illumination, expression, etc) it is worth to consider two types of test:

- Test with heterogeneous galleries (Test Type I), where images of the gallery have been acquired under different (illumination or expression) conditions.

- Test with homogeneous galleries (Test Type II), where all gallery images have been acquired under the same acquisition conditions, and probe images have been acquired under a different acquisition condition with respect to its gallery pairs.

Heterogeneous galleries are desirable when supervised methods are used, allowing to use the gallery samples to learn invariant face representations. The main drawback of these approach is that image gallery samples must be representatives of all the possible probe acquisition conditions. On the other hand, as shown in the experimental results reported later, homogeneous galleries (where differences between gallery images are only due to identity changes) are much more advantageous when the gallery has to be constructed using only one image per subject.

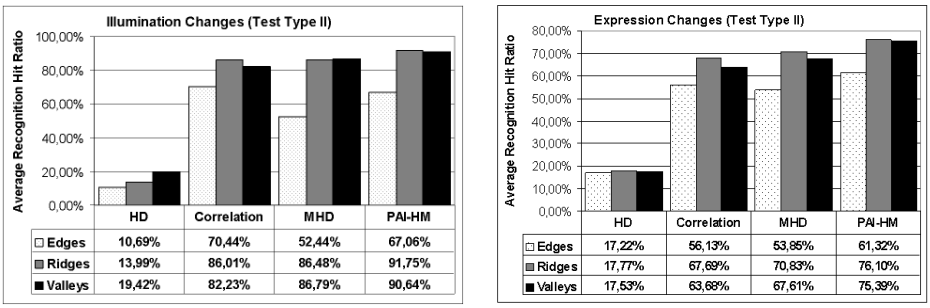

Figure 3: Average recognition hit ratios when changes between images is due to illumination (left) or facial expression (right).

In this way, when considering images acquired under two different acquisition conditions, namely condition (c_1) and (c_2), we can conduct two experiments:

- Experiment Type I: a pool of 212 images is constructed using the images labeled as (c_1) and (c_2). Recognition hit ratio is then computed as the percentage of pool images for which its nearest image belongs to the same subject. In this case, the nearest image to an image labeled (c_1), can be one labeled either as (c_1) or (c_2).

- Experiment Type II: Considering the 106 images labeled as condition (c_1) and the 106 images labeled (c_2) as gallery and probe respectively. Recognition hit ratio for the experiment is then obtained as the percentage of probe images for which probe and its nearest gallery image belong to the same subject.

This allows us to compute 42 different experiments, when combinations of two different labels taken from **(a)** to **(g)** are considered, and 12 experiments when differences between images are restricted to those only due to illumination (labels **(a)**, **(d)**, **(e)**, and **(f)**) or expression changes (labels **(a)**, **(b)**, **(c)**, and **(g)**). The maximum and minimum obtained recognition hit ratio of these 42 or 12 experiments, as well as their average, are used to measure the performance of each descriptor-similarity measure configuration.

Figure 3 shows the average hit ratios when ridges, valleys or edges are considered as descriptor, and Hausdorff Distance (HD), normalized correlation, Modified Hausdorff Distance (MHD), and the proposed PAI-HM are used. These results have been obtained using homogeneous galleries (test type II), and illumination changes (left figure), or facial expression changes only (right figure). Left figure, assess the enhancement on robustness to illumination changes of ridges (grey bar) and valleys (black bar) with respect to the use of edges (white doted bar). It can be seen how ridges perform slightly better than valleys, and both perform significatively better than edges (24 and 22 points improvement when similarity is measured using PAI-HM and correlation respectively), ob-

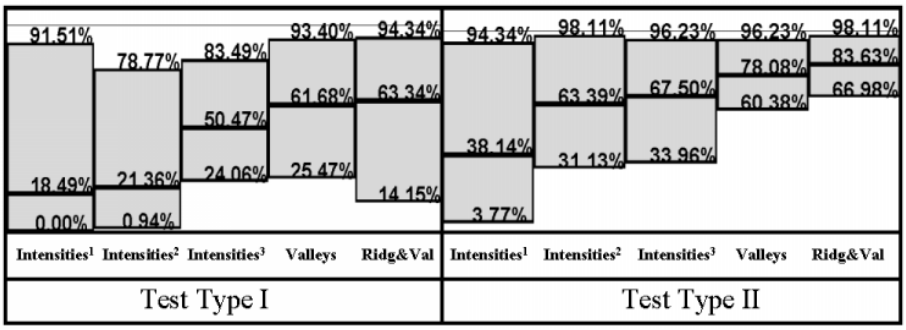

Figure 4: Comparison of recognition results obtained using original images, gray level normalized images, valleys and ridges descriptors, both for test type I and II. For each test, the maximum, minimum and average of the 42 experiment hit ratios are shown.

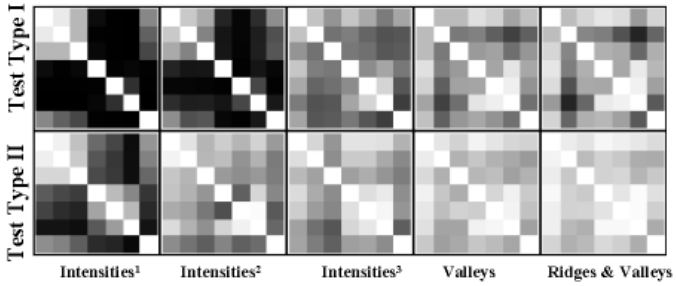

Figure 5: Comparison of recognition results obtained using original images, gray level normalized images, and valley and ridges descriptors, both for test type I and II. For each test, the hit ratios obtained in the 42 experiments are depicted.

taining an average recognition hit ratio of **91.75%** (PAI-HM and ridges). This ratio raises to **94.42%** when the proposed combination of ridges and valleys distance ($PAI\text{-}HM_{R\&V}$) is used.

Right figure summarizes the results obtained when the same test is applied to the set of images conveying facial expression changes (labels **(a)**, **(b)**, **(c)**, and **(g)** in Fig.(2)). This test allows us to analyze the behavior of the similarity measure to pattern deformations. Examining the results both for ridges and valleys, it can be seen that the best well suited similarity measure is the proposed PAI-HM, obtaining an average recognition ratio of **76%**, that overcome in 6, 9 and 59 points to those obtained using MHD, correlation, and HD respectively. This result is improved again when the proposed ridges and valleys combined measure, $PAI\text{-}HM_{R\&V}$, is used, obtaining in this case an average recognition ratio of **81.37%**.

Further experiments have been conducted to see how does the proposed method compares against gray level or intensity based methods. Three kinds

of intensity level representations of face images have been tested (denoted respectively as $intensities^1$, $intensities^2$ and $intensities^3$ in Fig.5 and Fig.4): 1) direct gray level images; 2) images normalized to have zero mean and standard deviation equals to 1; 3) ratio between local intensity values and average intensity of their neighborhood [7]. Examples of these images can be seen respectively in the first, second and third column of Fig.1. Similarity between these images have been computed using Principal Component Analysis (PCA) and Euclidean distance. We have used PCA only to carry out the computations in an efficient way, for this reason, principal components have been determined using the full set of images (742 images), and the number of selected bases or eigenfaces [7] have been chosen to obtain almost perfect image reconstruction (98% of the total samples set variance).

These recognition results have been obtained using test with heterogeneous galleries (Fig.5 left), and homogeneous galleries (Fig.5 right), using in both cases the full set of image acquisition changes (labeled from **(a)** to **(g)**). For each test, the average, maximum, and minimum hit ratio obtained in the 42 test experiments are depicted. The values of the recognition hit ratios obtained for each of these 42 experiments as well as for each of the descriptors and test type, are also shown graphically in Fig.5. For each test, a grid of 7x7 elements is depicted. The gray level of a square at grid position (i, j), denotes the recognition hit ratio (black for 0% and white for 100%) obtained in the experiment done using the 106 images of the i-th and 106 images of the j-th acquisition condition.

It has to be noted the significative differences between the recognition hit ratios obtained using heterogeneous (Test Type I) and homogeneous galleries (Test Type II). Note too, that these results can not be directly compared against those shown in Fig.3 and Fig.6 because experimental results shown in Fig.5 have been constructed using all 42 experiments (i.e, gallery and probe images can differ at same time in both illumination and expression). From these figures it can be seen that $PAI\text{-}HM_{R\&V}$ average recognition ratios overcome those obtained with the tested gray level descriptors, both for test type I and II (an increment superior to 13 points in both cases). When test Type I is considered, the worst experiment result for $PAI\text{-}HM_{R\&V}$ is lower than the minimum recognition ratio obtained using PAI-HM between valleys, or normalized images. This recognition result is obtained when neutral frontal illuminated images and smiling faces acquired with diffuse illumination are considered (see black squares at leftmost upper grid of Fig.5).

Some real AFR applications allow to acquire the subject reference images (gallery) in a controlled environment. In this cases it is worth considering which illumination condition and face expression are more adequate for constructing it. Figure 6 show the average recognition hit ratio obtained when each of the acquisition conditions are considered as gallery and the rest of illumination conditions (left) and face expression images are considered as probe. These results have been obtained using the proposed $PAI\text{-}HM_{R\&V}$. From these results, it can be concluded that there is a slight advantage in using frontally illuminated images (assessing the robustness of the descriptor to illumination changes), meanwhile

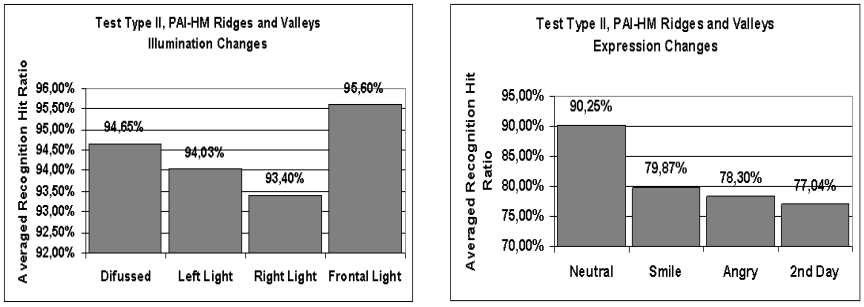

Figure 6: Average recognition hit ratio for each of the gallery acquisition condition.

using neutral face expression images is much more advantageous, as gallery images, than any other of the considered face expressions.

5 Conclusions

We have shown in this paper how ridges and valleys are more robust (in terms of recognition ratios) to illumination changes than both edges and usual gray level normalized image representations. In order deal with facial expression changes this paper proposes to use a Hausdorff based measure, that measures distortion of the image valleys and ridges locations rather than intensity changes. The recognition results obtained when images convey facial expression changes outperform those obtained with other measures, although it requires further improvement.

References

[1] Lopez, A.M., Lloret, D., Serrat, J., Villanueva, J.J.: Multilocal Creaseness based on the Level Set Extrinsic Curvature. In: Computer Vision and Image Understanding. Vol. 77, (2000) 111–144

[2] Lopez, A.M., Serrat, J., Lumbreras, F., Villanueva, J.J.: Evaluation of Methods for Ridge and Valley Detection. In: IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. 21 (1999) 327–335

[3] Heath, M., Sarkar, S., Sanocki, T., Bowyer, K.: A Robust Visual Method for Assesing the Relative Performance of Edge-Detection Algorithms. In: IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. 19, (1997) 1338–1359

[4] Canny, J: A Computational Approach to Edge Detection. In: Fischler, M., Firschein, O. (eds.) Readings in Computer Vision: Issues, Problems, Principles, and Paradigms. Morgan Kaufmann (1987) 184–203

[5] Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing Images Using the Hausdorff Distance. In: IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. 15, (1993) 850–863

[6] Dubuisson, M.P., Jain, A.K.: A Modified Hausdorff Distance for Object Matching. Proceedings of International Conference on Patter Recognition. (1994) 566–568

[7] Brunelli, R., Poggio, T.: Face Recognition: Features versus Templates. In: IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. 15, (1993) 1042-1052

[8] Grudin, M.A.: On internal representations in face recognition systems. In: Pattern Recognition. Vol. 33, (2000) 1161–1177

[9] Chellappa, R., Wilson, C.L., Sirohey, S.: Human and Machine Recognition of Faces: A Survey. In: Proceedings of the IEEE. Vol. 83, (1995) 705–740

[10] Valentin, D., Abdi, H., O'Toole, A.J., Cottrell, G.W.: Connectionist Models of Face Processing: A Survey. In: Pattern Recognition. Vol. 27, (1994) 1209–1230

[11] Samal, A., Iyengar, P.A.: Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. In: Pattern Recognition. Vol. 25, (1992) 65–77

[12] Takacs, B.: Comparing face images using the modified Hausdorff distance. In: Pattern Recognition. Vol. 31, (1998) 1873–1881

[13] Frischholz, R.W., Dieckmann, U.: BioID: A Multimodal Biometric Identification System. In: Computer. Vol. 21, (2000) 64–68

[14] Pearson, D. E., Hanna, E., Martinez K.: Computer-generated cartoons. In: Images and Understanding. Cambridge University Press, (1990) 46–60