3D Face Recognition using Sparse Spherical Representations*

Roser Sala Llonch, Effrosyni Kokiopoulou, Ivana Tosic and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory - LTS4
CH - 1015 Lausanne, Switzerland
{roser.salallonch,effrosyni.kokiopoulou,ivana.tosic,pascal.frossard}@epfl.ch

Abstract

This paper addresses the problem of 3D face recognition using spherical sparse representations. We first propose a fully automated registration process that permits to align the 3D face point clouds. These point clouds are then represented as signals on the 2D sphere, in order to take benefit of the geometry information. Simultaneous sparse approximations implement a dimensionality reduction process by subspace projection. Each face is typically represented by a few spherical basis functions that are able to capture the salient facial characteristics. The dimensionality reduction step preserves the discriminant facial information and eventually permits an effective matching in the reduced space, where it can further be combined with LDA for improved recognition performance. We evaluate the 3D face recognition algorithm on the FRGC v.1.0 data set, where it outperforms classical state-of-the-art solutions based on PCA or LDA on depth face images.

1 Introduction

Automatic recognition of human faces is one of the most challenging topics in the field of image processing and computer vision. It has numerous applications such as automated screening, surveillance, authentication or human-computer interaction. There has been a considerable progress in the area of two-dimensional(2D) face recognition where intensity/color images of human faces are employed. However, these systems are sensitive to illumination, pose variations, facial expressions and make-up. Most of these limitations can be overcome in 3D face recognition methods, which have
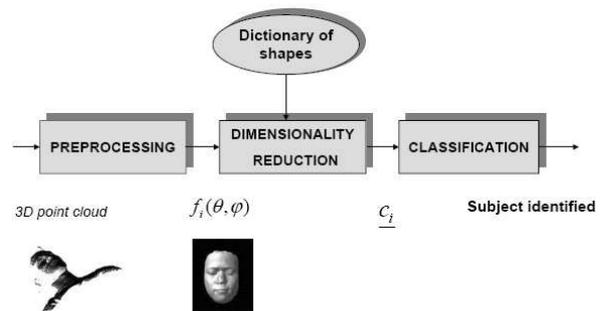

Figure 1. Block diagram of the 3D face recognition system.

shown potential for greater recognition accuracy (see e.g., [1, 2, 3]).

In this paper, we propose a 3D face recognition system based on the representation of faces on the 2D sphere, as illustrated in Figure 1. We propose first a fully automated facial region extraction and registration process. Contrarily to the vast majority of existing approaches where fiducial correspondences have been manually established beforehand, we propose a simple and effective methodology based on the automatic processing of the images that are represented as 3D point clouds. Then, we represent the pre-processed faces as spherical signals, which permits to preserve the geometry information and to provide important advantages with respect to depth images (i.e., 2D arrays of face to camera distance values). In order to cope with the limited size of the training set, and with possible registration errors, we enrich the training data by the addition of *virtual samples*. These samples represent artificial faces that are produced by slight pose variations of the

training faces, which are easily computed in the spherical representation.

We eventually perform dimensionality reduction on the sphere with simultaneous greedy Matching Pursuit algorithm [4, 5] in order to extract the most meaningful information of the spherical faces. In particular, the facial images are jointly represented using a few localized and oriented features in the sphere, which are able to capture discriminant geometrical characteristics. It results in a compact subspace representation that permits effective matching and recognition. We finally provide experimental results on the FRGC v.1.0 database that indicate the effective performance of the proposed system, that can be further combined with LDA for improved recognition. The novel algorithm is shown to outperform solutions that run state-of-the-art algorithms like PCA and LDA on the depth images.

2 Automatic facial region extraction and registration

Each 3D face scan typically forms a 3D point cloud representation, accompanied by a binary matrix A of valid points. We first use two lateral thresholds on the vertical projection curve (obtained by the column sum of the matrix A) in order to remove the subjects' shoulders. This is followed by thresholding on the depth histogram in order to separate the head region from the chest. Disconnected regions of outliers are further removed by morphological image processing on the corresponding binary matrix A (e.g., by keeping the region of largest area).

Registration of the pre-processed point clouds is then performed in a two-step approach. All faces are initially registered by means of the Iterative Closest Point (ICP) algorithm [6]. An Average Face Model (AFM) is built after resampling the data points on a 2D grid. The AFM is used as reference image in order to define an area-of-interest under the form of an ellipse that contains the salient facial region. Since all faces are registered, the facial region can be extracted from each 3D face by ellipse cropping, which discards at the same time the irrelevant points (e.g., hair, ears etc). Next, ICP is performed once more on the cropped 3D faces for even more accurate alignment. The resulting registered face data are mapped to signals defined in the space of square-integrable functions on the unit 2-sphere S^2, denoted as $L^2(S^2)$. The mapping is performed such that the value of the spherical signal $f(\theta, \varphi)$ represents the distance of a 3D point with spherical coordinates θ (azimuth angle) and φ (elevation angle) to the center of the sphere. The spherical facial signals are finally obtained from the registered faces using nearest neighbor interpolation on the spherical equiangular grid, which moreover removes any holes that may exist.

3 Spherical sparse representations

Let us introduce the concept of *sparse representations*. We assume the existence of a redundant dictionary (overcomplete basis) D that spans the space $L^2(S^2)$. The elements of the dictionary, which are indexed by $\gamma \in \Gamma$ i.e., $D = \{g_\gamma, \gamma \in \Gamma\}$, are usually called atoms and they have unit norm i.e., $\|g_\gamma\|_2 = 1$ where $\| \cdot \|_2$ denotes the L2 norm. A signal $s \in L^2(S^2)$ is called *k-sparse*, if it can be expressed as $s = \Phi_I \underline{c}$. Φ_I denotes a matrix composed of subset of atoms in $\mathcal{D}$ indexed by I, where $|I| \leq k$, and $\underline{c}$ denotes a vector of coefficients. Computing the sparsest approximation of a given signal is known in general to be NP-hard, but there exist suboptimal methods such as the Matching Pursuit (MP) algorithm that work sufficiently well in practice. In the case where a sparse representation is sought for a given set of spherical signals $s_i = s_i(\theta, \varphi)$, $i = 1, \ldots, n$, MP is extended to the Simultaneous Spherical MP (SSMP) algorithm. The goal of SSMP is to extract a set of atoms from the dictionary such that all signals are simultaneously represented as linear combinations of a few, say K, basis vectors, $s_i = \sum_{k=1}^K c_{ik} g_{\gamma_k}$, $i = 1, \ldots, n$.

SSMP is an iterative greedy algorithm, which extracts a subset G of the dictionary $\mathcal{D}$ for the approximation of all signals s_i. At the first step, the algorithm sets the residual signals as $R_i^0 = s_i$ and selects the atom g_{γ_1} from the dictionary that fits the best all the residual signals. Formally, the algorithm computes $\gamma_1 = \arg\max_{\gamma \in \Gamma} \sum_{i=1}^n |\langle R_i^0, g_\gamma \rangle|$. Recall that we are dealing here with spherical signals, and that $\langle \cdot, \cdot \rangle$ represents the inner product on the sphere. The spherical inner product between $f = f(\theta, \varphi)$ and $g = g(\theta, \varphi)$ is defined as:

$$\langle f, g \rangle = \int_\theta \int_\varphi f(\theta, \varphi) g(\theta, \varphi) \sin \theta d\theta d\varphi. \quad (1)$$

In the next step, all the residual signals are updated by subtracting the contribution of the selected atom, $R_i^1 = R_i^0 - c_{i1} g_{\gamma_1}, \forall i$, where R_i^1 is the residual of the i-th signal in the first iteration and c_{i1} is the corresponding contribution of the selected atom computed as,

$$c_{i1} = \langle R_i^0, g_{\gamma_1} \rangle. \quad (2)$$

2

Figure 2. Gaussian atoms

The same process of extracting the best atom and updating the residual is repeated iteratively. At step k, the best matching atom g_{γ_k} is chosen as follows,

$$\gamma_k = \arg \max_{\gamma \in \Gamma} \sum_{i=1}^{n} |\langle R_i^{k-1}, g_\gamma \rangle|. \qquad (3)$$

Typically, a few iterations are usually sufficient to capture most of the energy of the signals to be approximated.

Finally, we propose to use the dictionary $\mathcal{D}$ introduced in [7], where the atoms are constructed by applying geometric transformations on a generation function, denoted hereby as $g(\theta, \varphi)$. We choose as a generating function the 2-D Gaussian function in $L^2(S^2)$, defined as, $g(\theta, \varphi) = \exp(-\tan^2 \frac{\theta}{2})$. It represents an isotropic Gaussian function, centered at the North Pole. The geometric transformation consists of translation, rotation and anisotropic scaling of g on the sphere. In this case, the atom index is $\gamma = (\tau, \nu, \psi, \alpha, \beta) \in \Gamma$, where (τ, ν) is the position of the atom in the sphere, ψ is its orientation and (α, β) its scaling parameters. Figure 2 shows a few sample Gaussian atoms that are used in the representation of spherical faces.

4 Experimental results

Test configuration	p	Number of subjects	Training set	Test set
T_1	1	200	200	673
T_2	2	166	332	474
T_3	3	121	363	308
T_4	4	86	344	187

Table 1. Test configurations.

In our methodology, the spherical facial images are projected in the reduced SSMP subspace by computing the corresponding low dimensional coefficients vectors. Those are computed by successive projections of the residual signals over the atoms, in a way that "mimics"

the MP process. For this reason, we call them the MP coefficients. Recognition is performed in the reduced space (among coefficient vectors) by NN classification i.e., assigning the test vector to the class corresponding to its nearest neighbor among the training data. The matching uses the L1 distance metric, which has been found to outperform other related distances in our experiments.

We evaluate the performance of the proposed algorithm on the University of Notre Dame (UND) Biometric database[1] (named also FRGC v1.0 database). The database consists of 277 subjects, and each subject has between one and eight scans. We defined several configurations for our experiments, driven by the number of samples p per class (subject) in the formation of the training set (see Table 1 for more details). For the sake of completeness, we include in our comparisons a hybrid method of SSMP followed by LDA, denoted as SSMP+LDA. Note finally that we use virtual samples in the configurations T_1, T_2 and T_3, where the training set is limited. In particular, each spherical training image is expanded by generating 8 slightly shifted versions of it (one pixel in each possible direction). The inclusion of virtual samples compensates for possible registration errors.

We compare our methods with PCA and LDA on the (preprocessed) depth images. Additionally, we provide recognition results with the Euclidean distance (EUC) on the depth images as well as the Mean Square Error (MSE) on the spherical functions. We report rank-1 recognition results across random experiments. For each experiment, we split randomly the samples between the training and the test sets. In Figure 3 we show the average classification error rate obtained for each configuration, over all 10 random splits. First, we observe that Euclidean distance on the depth images (EUC) is inferior to Mean Square Error (MSE) on the spherical signals. This is the main observation that encourages the use of spherical signals for representing 3D faces. Next, notice that SSMP outperforms PCA, which in turn mainly explains the fact that SSMP+LDA outperforms LDA (on the depth images), since the latter implements a PCA decomposition at first step. Finally, we can see that SSMP+LDA gives the best performances among all the tested schemes.

5 Conclusions

We have proposed a methodology for 3D face recognition based on spherical sparse representations. Spher-

---

[1]http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html

(a) Test Configuration T_1

(b) Test Configuration T_2

(c) Test Configuration T_3
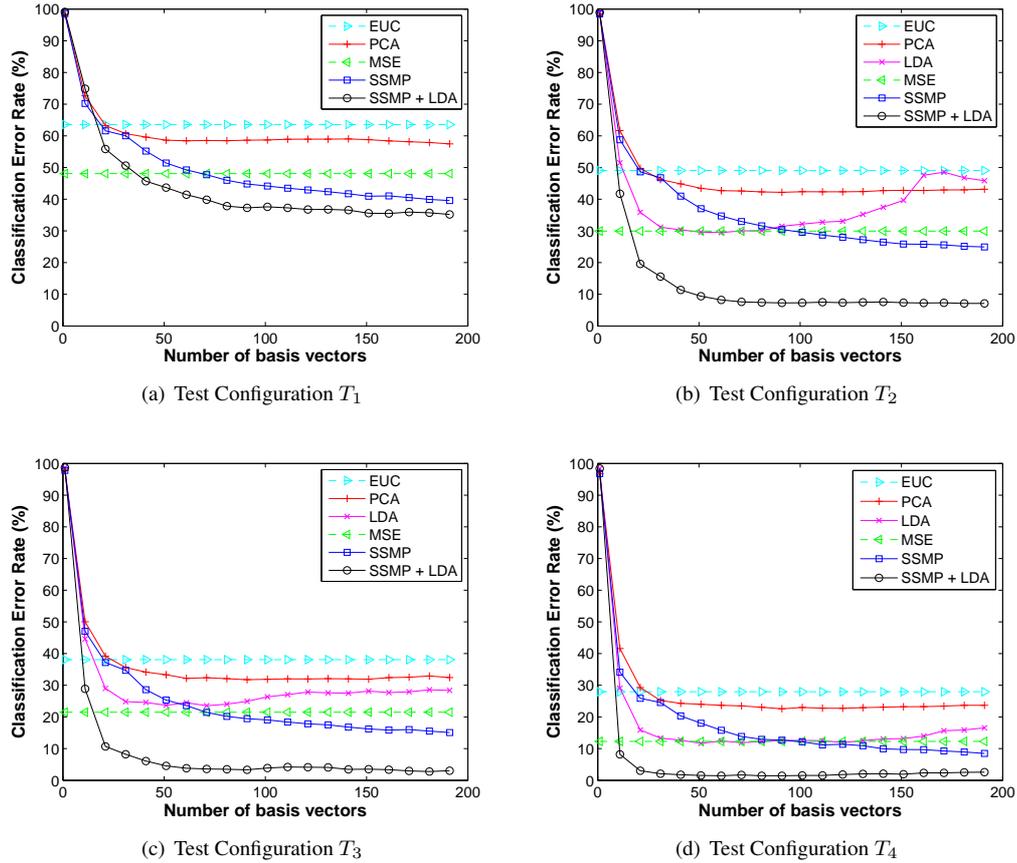
(d) Test Configuration T_4

Figure 3. Classification error rates on the FRGC v.1.0 database.

ical representation of faces allows for effective dimensionality reduction by preserving geometry information, which in turn leads to high performance matching in the reduced space. Experimental evidence indicates the advantages of the proposed approach over well known existing methods.

References

[1] K.W. Bowyer, K. Chang, and P. Flynn. "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition". *Computer Vision and Image Understanding*, vol. 101(1), pp. 1-15. Jan. 2006.

[2] C. Samir, A. Srivastava and M. Daoudi. "Three-dimensional face recognition using shapes of facial curves" *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, Nov. 2006.

[3] B. Gökberk, H. Dutağaci, A. Ulas, L. Akarun, B. Sankur. "Representation plurality and fusion for 3D face recognition". *IEEE Trans. on Systems, Man, and Cybernetics, part B*. vol. 38, no. 1, pp. 155-173. Feb. 2008.

[4] J. Tropp, A. Gilbert, and M. Strauss. "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit", *Signal Processing, special issue "Sparse approximations in signal and image processing"*, vol. 46, pp. 572-588, April 2006.

[5] E. Kokiopoulou and P. Frossard. "Semantic coding by supervised dimensionality reduction", *IEEE Transactions on Multimedia*, to appear.

[6] P. Besl, N. McKay, "A method for registration of 3D shapes", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, pp.239-256. Feb. 1992.

[7] I. Tosic, P. Frossard, and P. Vandergheynst. "Progressive coding of 3D objects based on overcomplete decompositions" *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 11, pp. 1338-1349, Nov. 2006.