2014 22nd International Conference on Pattern Recognition

Iconic Methods for Multimodal Face Recognition: a Comparative Study

Marinella Cadoni, Andrea Lagorio, Enrico Grosso
PolComIng Department
University of Sassari
Sassari, Italy

Email: maricadoni@uniss.it lagorio@uniss.it grosso@uniss.it

***Abstract*—When dealing with face recognition, multimodal algorithms, with their potential to capture complementary characteristics from the 2D and 3D data channels, can reach high level of efficiency and robustness. In this paper, we explore different combinations of iconic descriptors coupled with a shape descriptor and propose a fully automatic, multimodal, face recognition paradigm. Two iconic features extractors, the Scale Invariant Feature Transform (SIFT) and the Speeded-Up Robust Features (SURF), are used, in turn, to extract salient points from the images of the faces. The corresponding points on the scans are validated with Joint Differential Invariants, a 3D characterisation method based on local and global shape information. SIFT and SURF are then combined at feature level and the 3D Joint Differential Invariants used to validate them on the shape channel. The proposed method has been tested on the FRGCv2 database. Experimental results highlight the complementarity of the feature points extracted by SIFT and SURF and the effectiveness of their 3D validation.**

I. Introduction

Face recognition has seen major improvements in accuracy and computational feasibility over the last decade. The introduction of 3D techniques alone, or in conjunction with 2D methods can be identified as the main factor of improvement, and these are the techniques we intend to focus on. A comprehensive treatment of the state of the art in 3D and 2D+3D face recognition can be found in [5]. Recent and successful 3D and 2D+3D face recognition methods ([12], [15]) rely on the assumption that some predetermined facial points, which usually comprise the nose tip and the eye corners, can be robustly identified. To identify such points, however, heavy limitations on pose variation must be imposed. Moreover, facial expressions or acquisition noise can impair their search or prevent their robust localisation. In the cited works, the extracted points are subsequently used to initialise a registration of two face scans, so failure in their localisation will result in the impossibility to move on to the registration step. If we aim at improving recognition rates in less restrictive scenarios, where pose variations may occur, methods that automatically select and match interest points of a pair of image faces or scans are better suited. After two sets of matching feature points of a pair of faces have been collected, they are used to segment different areas of the faces, usually the most stable to expression variations, and these areas are coarsely registered using the correspondences of the points. The coarse registration is refined either by means of the ICP algorithm [3] or some of its variants, as in [12], [1], [11], or by random based searches for the optimal registration, based for instance on Simulated Annealing (SA) [15], or by both [9]. ICP is a computationally efficient algorithm but it does no guarantee that the transformed surface after the last iteration is the one that globally minimises the mean squared error (MSE). Indeed, depending on the coarse registration and the shape of the areas, the iteration process might well stop at a local minimum. On the other hand, the SA, being based on a random search, does not incur local minima, but it is more demanding from the computational point of view. The similarity measure between two faces is given by the MSE after the last iteration when the ICP is used, whereas in [15] a measure of the intersection of the two surfaces (the Surface Interpenetration Measure) is used. Both measures can be challenged by the presence of expressions or differences in sample density.

With these premises, in [7] we started to explore alternatives to the extraction of predetermined points and to the registration of the scans. To extract feature points from a face image we experimented the Scale Invariant Feature Transform (SIFT). Corresponding feature points were subsequently projected on the point clouds and used to generate joint differential invariants based on local and global shape information. The number of invariant matches under a fixed threshold was used as similarity score between scans. Experimental results showed that the 3D invariants improve recognition rates over SIFT alone, particularly when the number of SIFT matches is in the range $[3, 20]$. It also became evident that, in order to reach state of the art recognition rates, more iconic points needed to be extracted. In this paper, we adopt an alternative iconic points extraction method to SIFT, namely the Speeded-Up Robust Features (SURF) [2]. Our choice fell on SURF after the following consideration: SIFT provide accurate key-point localisation but the descriptor is based on weak information (local orientation histograms of image gradients); SURF, on the other hand, are based on integral images, which lead to a less stable key-point localisation, but adopt a descriptor which outperform SIFT in almost all cases. Once that corresponding SURF points are extracted, they are projected on the respective point clouds and can be used to generate the 3D joint differential invariants by following the same procedure as in [7]. A comparative analysis of the two procedures (SIFT augmented by 3D invariants and SURF augmented by 3D invariants), demonstrates the feasibility of a feature level fusion: SIFT and SURF points from two images are extracted and matched

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.789

separately, then joined in a set of iconic points from which the 3D invariants are calculated. Experiments run on the FRGCv2 database show that SURF are a suitable feature points extractor for our multimodal paradigm. Results show that the fusion of SIFT and SURF remarkably improves recognition rates, which demonstrate the complementarity of SIFT and SURF in our context.

II. The Iconic Driven Face Recognition Model

SIFT have been successfully used for key point localisation and 2D face recognition [4]. They are based on a scale-space representation obtained by successive smoothing of the original image with Laplacian of Gaussian kernels of different size. Note that the LoG operator is circularly symmetric and particularly suitable to detect blob-like structures. Recently, Lowe proposed an efficient algorithm for object recognition based on the approximation of the LoG by difference-of-Gaussian (DoG) filters [10]. This approximation can significantly accelerate the computation process, reaching real-time performances. SIFT descriptors are based on local orientation histograms of image gradients. They are invariant to scale and rotation transformations but they are challenged by strong illumination variations and face expressions. Moreover, a common drawback of the DoG representation is that local maxima can also be detected in the neighbourhood of contours or straight edges, where the signal change is only in one direction. These maxima are less stable because their localisation is more sensitive to noise or small changes and usually require additional processing steps to remove unreliable points. SURF [2] in turn, are based on a very basic Hessian-matrix approximation, obtained by the use of integral images. This approach drastically reduces the computation time and usually increases the number of key-points detected, at the cost of a less accurate localisation. On the other hand, the SURF descriptors are built on the distribution of first order Haar wavelet responses in x and y direction rather than the gradient. This has proved to be fast and to simultaneously increase the robustness of the descriptor. In particular, Bay and colleagues have shown that SURF descriptors outperform SIFT in case of blur and scale changes.

The above mentioned characteristics of SIFT and SURF make them the ideal iconic points extractors for our recognition paradigm. Indeed, the 3D joint differential invariants are calculated from triplet of surface points and produce a signature that characterises the surface up to Euclidean motions [13]. In the case of a point cloud, such as a face scan, computational time issues impose that the invariants are calculated from a small subset of scan points (see [6], where about 12 points were selected). If the feature points between two faces are already matched, however, the invariant step becomes much less demanding from a computational point of view and more feature points can be allowed (up to 100 per face). Since SIFT and SURF both come with descriptors able to match the points, we can take this burden off the 3D invariants and speed up the 3D processing step by limiting the role of the invariants to validating the matchings and defining a measure of similarity between the two faces. In the following sections, the proposed face recognition model is unfolded. In figure 1, a block diagram of the proposed recognition paradigm applied to a pair of faces is shown: it takes as input 2D and 3D face data and returns the similarity score of the two faces.

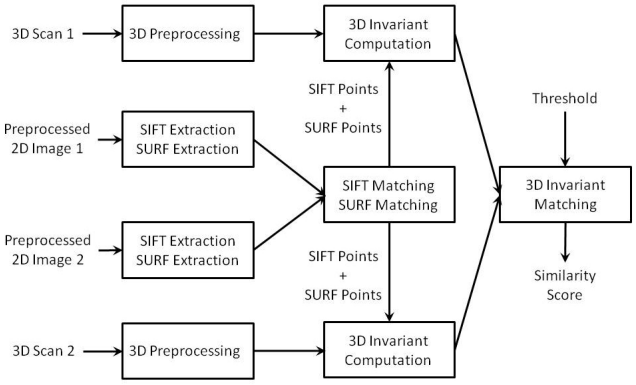

Fig. 1. Block diagram of the similarity evaluation of two faces

A. Face Preprocessing

Our proposed model is based on both the texture and shape information of a face. We require that these are registered: each pixel of the face image corresponds to a point of the scan. Before we extract SIFT and SURF points we need to preprocess the face images. A light preprocessing on the scans to remove spikes and noise is necessary.

1) 2D Face Localisation and Equalisation: To segment the oval of the face we first use the Viola-Jones algorithm ([16]) to locate the ROI of the face. From the ROI of the face we extract the ellipse centred in the centre of the ROI and with axes equal to the edges of the ROI. This approximates the face oval. SIFT and SURF are sensitive to illumination variations. Since many databases (FRGCv2 included) were collected in uncontrolled illumination environments, we apply an histogram equalisation algorithm to the images.

2) 3D Filtering: Having extracted the face oval from the 2D image, we extract the 3D face from the scan by considering the subset of 3D points corresponding to the pixels of the face oval. 3D scans can arise from a variety of instruments: structured light systems, laser scanners etc. Different acquisition systems lead to different types of acquisition errors. For instance, a reflective patch of skin can cause spikes in the scan. The 3D validation step relies on local information to estimate the normals to the face at the characteristic points. To get a robust estimation we need to remove outliers (spikes and isolated points) and attenuate noise. To remove outliers from a scan F of n points, we estimate the average sampling density of the scan as $\delta_r(F) = 1/n \sum_{j=1}^{n} \delta_r(p_j)$, where $\delta_r(p_j) = |\{q | q \in U_r(p_j)\}|$ and $U_r(p_j) = \{q \in F \,|\, \|q - p_j\| < r\}$ and then remove the points $p_i \in F$ such that $\delta_r(p_i) < \delta_r(F)/4$. Noise is mostly due to acquisition error which occurs along the z-coordinate. After removing outliers we found a mean filter along the z-coordinate to be the most effective way to attenuate noise in view of the normals estimation.

B. Iconic Feature Points Extraction and Matching

From each image face we extract two sets of characteristic iconic points using SIFT and SURF.

1) SIFT Extraction: SIFT points are extracted using the original algorithm implemented by Lowe (see [10]). For faces

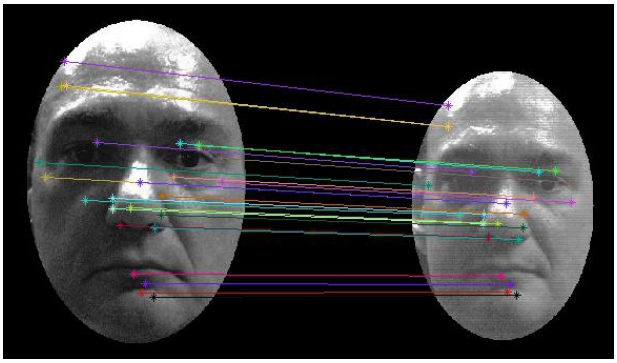

Fig. 2. Matched SIFT points of two images of the same subject.

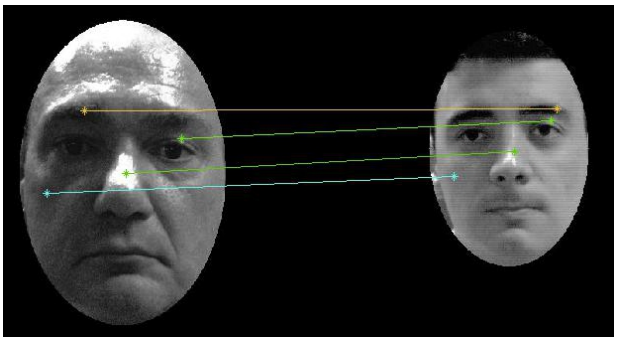

Fig. 3. Matched SIFT points of two images of different subjects.

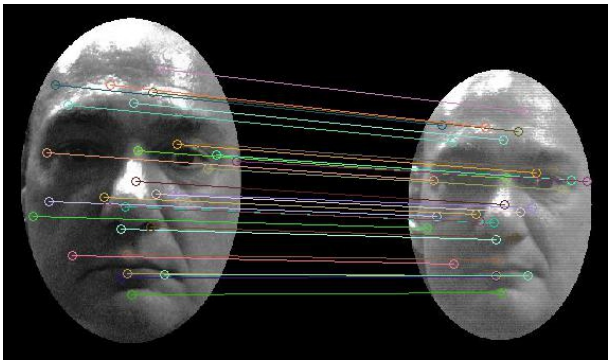

Fig. 4. Matched SURF points of two images of the same subject.

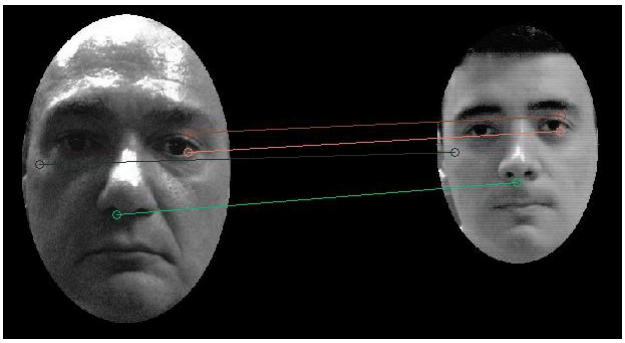

Fig. 5. Matched SURF points of two images of different subjects.

in the gallery database G, this step can be done off-line, so for each face $G_i \in G$ we will have a set of SIFT A_{G_i} which includes the location of the SIFT points and the description vector. When a probe face F comes in, SIFT A_F are extracted and matched against all gallery faces. After the SIFT matching we will have two sets $(A_F^m, A_{G_i}^m),\ G_i \in G$ of matching image points of equal cardinality (which in our experiment was in the range $[0, 64]$), for all faces in the gallery. In figure 2, the matching SIFT points of two images of the same subjects are displayed, whereas in figure 3 we see the matching SURF points of images relative to different subjects.

2) SURF Extraction: To compute the SURF the OpenSurf code is used [OpenSurf]. Analogously to the SIFT extraction, gallery faces can be processed off-line so for each face $G_i \in G$ we will have a set of SURF B_{G_i} which includes the location of the SURF points and the description vector. To calculate the distance between the descriptor vectors we implemented the same technique used for SIFT with a threshold $t = 0.6$. When a probe face F comes in, SURF B_F are extracted and matched against all gallery faces. After the SURF matching we will have two sets $(B_F^m, B_{G_i}^m),\ G_i \in G$ of matching image points of equal cardinality. In our experiments, the average cardinality was higher than the one of SIFT, ranging from 0 to 85. In figure 4, the matching SURF points of two images of the same subjects are displayed, whereas in figure 5 we see the matching SURF points of images relative to different subjects.

3) Iconic Points Extraction: Once SIFT and SURF points are extracted and matched from two faces F and G_i we can join them into two sets of undistinguished iconic points: $I_F = A_F^m \cup B_F^m$ and $I_{G_i} = A_{G_i}^m \cup B_{G_i}^m$. In figure 6, we can see on the left all SIFT and SURF extracted from a face, and on the right the surviving ones after the matching. SIFT are displayed with red stars, SURF with blue circles. The right image in figure shows there are 24 matching SIFT and 29 matching SURF, and they located the same (or very close, up to about 4 pixels) interest points 8 times, so there is an intersection between the two sets of points of about 30%. The remaining points are well spread in the face image which shows that the two iconic extraction methods are able to select complementary points. The union of the two types of features has the potential to increase their descriptiveness.

C. 3D Joint Invariants Validation

The proposed iconic extractors and descriptors are based on local information around a point, and an established correspondence derives from the similarity of the descriptors vectors. Neither the SIFT nor the SURF descriptors take into account the relative position of the points. Following the framework in [6] we validate the iconic correspondences by calculating 3D invariants on the corresponding points in the scans. The invariants we use arise from the Moving Frame Theory [13] and are based on the relative position of the points and local shape information. The procedure that leads to the generation of the invariants is discussed in full details in [6], here we just outline it. We generate invariants that depend on three points at a time. The invariant a triplet of points generates is a $9-$dimensional vector whose first three entries are the inter-

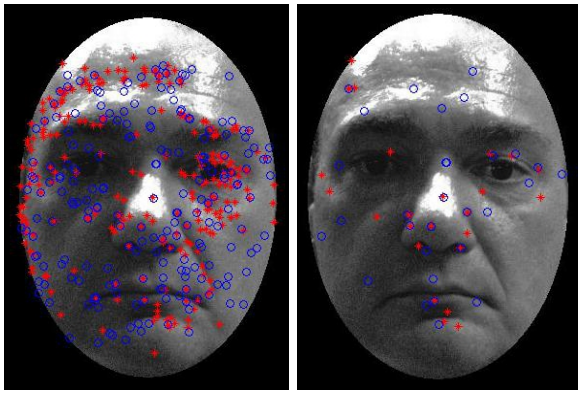

Fig. 6. Left: Extracted SIFT and SURF points on a subject image. Right:Matched SIFT and SURF points on a subject image.

point distances and the last six are functions of scalar and wedge products of various vectors, as defined by the formulae that follow.

1) Invariants Definition: Let $p_1, p_2, p_3 \in F$ and ν_i be the normal vector at p_i. The directional vector v of the line between p_1 and p_2 and the normal vector ν_t to the plane through p_1, p_2, p_3, are defined as:

$$v = \frac{p_2 - p_1}{\|p_2 - p_1\|} \quad \text{and} \quad \nu_t = \frac{(p_2 - p_1) \wedge (p_3 - p_1)}{\|(p_2 - p_1) \wedge (p_3 - p_1)\|}.$$

The zero order invariants are the 3 inter-point distances $I_k(p_1, p_2, p_3)$ for $k = 1, 2, 3$:

$$I_1 = \|p_2 - p_1\|, \quad I_2 = \|p_3 - p_2\| \quad \text{and} \quad I_3 = \|p_3 - p_1\|$$

whereas the first order invariants are

$$J_k(p_1, p_2, p_3) = \frac{(\nu_t \wedge v) \cdot \nu_k}{\nu_t \cdot \nu_k} \quad \text{for } k = 1, 2, 3$$

and

$$\tilde{J}_k(p_1, p_2, p_3) = \frac{v \cdot \nu_k}{\nu_t \cdot \nu_k} \quad \text{for } k = 1, 2, 3.$$

Each triplet (p_1, p_2, p_3) of points on the surface is now represented by a 9-dimensional vector whose coordinates are given by $(I_1, I_2, I_3, J_1, J_2, J_3, \tilde{J}_1, \tilde{J}_2, \tilde{J}_3)$

2) Invariants Generation and Matching: Given a set of iconic point features (either SIFT, SURF or their union) of a face F, and one of a face G_l we first generate the 3D joint differential invariants for both faces. If there are n matching points, there will be $\binom{n}{3}$ invariant vectors. The next step is to define a metric to compare the invariant vectors. Since the invariants $I_1, I_2, I_3, J_1, J_2, J_3, \tilde{J}_1, \tilde{J}_2, \tilde{J}_3$ are of different nature and can assume values in a variety of ranges ([7]) we proceeded by testing three different distances: Euclidean, Mahalanobis, Manhattan. The Euclidean distance was calculated both on the invariant vectors as defined in section II-C1, and on the same vectors normalised using standard deviation. We found that, amongst all, the Euclidean distance on the original vectors was the best suited one as it weighted slightly more the inter-point distances (the first three coordinates of the invariant vectors) which are more stable than the first order invariants (corresponding to the last six coordinates of the vectors). The invariant vectors of corresponding triplets are therefore compared using the Euclidean distance in $9-$dimensional space. The similarity measure between the probe face F and the gallery face G_l is given by the number of invariant vectors of F whose distance from the corresponding invariant vector in G_l is less than a fixed threshold σ: $\mathcal{S}_{G_l} = \left|\{\tilde{d}_i \,|\, \tilde{d}_i < \sigma\}\right|$. In an identification scenario, the matching gallery face is chosen to be the face G_j such that $\mathcal{S}_{G_j} = \max_{G_k \in G}\{\mathcal{S}_{G_k}\}$, i.e. the face with the greatest similarity measure.

III. Experimental Evaluation

The proposed method was tested on the FRGCv2 database. Our goals were the following:

- Verify the feasibility of using SURF to extract iconic features for our multimodal paradigm, i.e. to see if 3D validated SURF features outperform SURF alone.
- Compare its performance to the baseline model of 3D validated SIFT proposed in [7].
- Establish if SIFT and SURF features can be successfully joined and, if so, to measure the improvement over SIFT+SURF alone and over the use of a single iconic feature selector at a time.

A. FRGCv2 Description

The FRGCv2 database [14] collects scans and images of 466 subjects. There is a variable number of acquisitions per subject, the maximum being 22, for a total of 4007 acquisitions. Each acquisition consists of a scan and a registered image whose resolution is 640 x 480. Faces were not captured always from the same distance which results in different resolutions of the images and different sampling density of the scans. There is little variation in the pose of the subjects, whereas variations of illumination and expressions are relevant. Expressions are classified as "blank-stare" (i.e. neutral), "happiness", "sadness", "disgust", "surprise", "others". The FRGC provides a training set, Spring2003, on which to train algorithms or set thresholds. We used it to define a distance to compare invariant vectors and to set the threshold σ. Different values of σ in the range $[1, 6]$ were considered. For $\sigma = 4$ the best recognition rates were obtained so we set that value for subsequent experiments on the validation test.

B. Experimental Set Up

To test the proposed models and compare them to the baseline in [7] , we maintain the same experimental set up, which was also adopted in [8], [12], [15], [9] and it is usually named "First versus All" . The gallery set "First" consists of the pairs of scans and images relative to the first acquisition of each subject. There are therefore 466 pairs. Almost all of them are neutral. The probe set "All" consists of all the remaining acquisitions: 3541 pairs, of which 1984 are labelled as neutral and 1557 as non-neutral. We assume an identification scenario, where a probe subject has to be correctly matched to his instance in the gallery database.

Given this experimental set up we implemented different recognition paradigms:

TABLE I. COMPARATIVE RESULTS OF DIFFERENT RECOGNITION METHODS ON THE FIRST-V-ALL EXPERIMENTAL SET UP

Method	Recognition Rates
SIFT	76.6%
SIFT+Inv	81%
SURF	71.6%
SURF+Inv	80.1%
SIFT+SURF	82.5%
(SIFT+SURF)+Inv	**86.7%**
(SIFT+Inv)+(SURF+Inv)	85.2%

- SURF on face images
- 3D validated SURF on face images and scans, which will be called SURF+Inv
- SIFT+SURF on face images
- 3D validated SIFT and SURF points, which will be called (SIFT+SURF)+Inv

The first and third are purely 2D identification tests that use, in turn, the number of SURF and SIFT+SURF matches as a similarity score. Each face in the probe set was compared to all faces in the gallery set. The gallery face with the highest number of SURF (SIFT+SURF) matches was matched to the probe face. The second and fourth are two versions of the iconic driven face recognition model: SURF followed by 3D invariant validation and SIFT+SURF followed by 3D invariant validation. For each face of the probe set, the similarity scores with all faces in the gallery set were calculated. The gallery face with the highest similarity score was chosen as the correct match. All algorithms were implemented in MatLab and the running time for comparing a probe face to all 466 gallery faces was on average 120s for SURF+Inv and 160s for (SURF+SIFT)+Inv on a Intel Core i7 workstation.

C. Experimental Results

Recognition rates for the proposed recognition paradigms are shown in table I. In the first two lines, results from the baseline experiment with SIFT carried out in [7] are reported. From the third line onward, recognition rates for the paradigms SURF, SURF+Inv, (SIFT+SURF)+Inv are shown. SURF prove to be less reliable than SIFT, but when validated by invariants they score only 1 percentage point less than the SIFT+Inv, so the invariants are very effective at validating SURF matches, mainly by discarding false matches when comparing faces of different subjects. SIFT+SURF however, outperforms both SURF and SIFT, so the two iconic features extraction methods prove to be complementary. This is still evident in the experiment (SIFT+SURF)+Inv, where recognition rates are almost seven percentage points higher than those of the SIFT+Inv baseline experiment. The last row in the table refers to recognition rates obtained by fusing the scores of SIFT+Inv and SURF+Inv. The fusion at score level in this case is less effective than that at feature level, making SIFT and SURF ideal partners for the extraction of iconic feature points.

The search for other iconic points extractors arose when the baseline experiment showed that low performances occurred in the cases where a face did not have enough SIFT matches with its counterpart in the gallery database, see table II. We show the same analysis based on matches ranges in the case of SURF in table III. It turns out that SURF matches are more numerous on average. If we analyse the different matches ranges for the fusion of SIFT and SURF (see table IV) we notice that there are less faces with few iconic feature points. Indeed, for the SIFT matches alone were ≤ 10 for 937 faces out of 3541, whereas SIFT+SURF are ≤ 10 for only 196 cases. The addition of SURF points to the SIFT ones has therefore proved to provide complementary iconic points and boost recognition results.

TABLE II. IDENTIFICATION SCORES FOR DIFFERENT SIFT MATCHES RANGES IN THE FIRST-V-ALL CASE

Range of SIFT matches	Nr. of matches over Nr. of scans in range	
	SIFT	SIFT+Inv
$0 \leq N \leq 2$	0/111	0/111
$3 \leq N \leq 5$	4/252 1.6%	23/252 9.13%
$6 \leq N \leq 10$	211/574 36.76%	319/574 55.57%
$11 \leq N \leq 20$	1213/1320 91.89%	1237/1320 93.71%
$21 \leq N$	1284/1284 100%	1284/1284 100%

TABLE III. IDENTIFICATION SCORES FOR DIFFERENT SURF MATCHES RANGES IN THE FIRST-V-ALL CASE

Range of SURF matches	Nr. of matches over Nr. of scans in range	
	SURF	SURF+Inv
$0 \leq N \leq 2$	0/35	0/35
$3 \leq N \leq 5$	0/116 0%	1/116 0.86%
$6 \leq N \leq 10$	30/418 7.18%	115/418 27.51%
$11 \leq N \leq 20$	855/1309 65.32%	1060/1309 89.98%
$21 \leq N$	1651/1663 99.28%	1657/1663 99.64%

The Cumulative Match Curve (CMC) for all experiments carried out is presented in figure 7. At rank 2 the recognition rate for the (SIFT+SURF)+Inv experiment is 89.8%, at rank 10, 93.5%.

In table V, it is shown how different expressions in the probe set "All" affect recognition in the experiment (SIFT+SURF)+Inv. Confirming the finding in our previous work [7], faces labelled by the FRGCv2 as Blank-Stare (neutral faces) prove to be easier to recognise by the algorithm.

IV. CONCLUSION

The work presented in this paper was motivated by the work in [7] in which SIFT were coupled to 3D joint differential invariants to provide a multimodal face recognition model that would not require neither the location of predetermined

TABLE IV. IDENTIFICATION SCORES FOR DIFFERENT SURF+SIFT MATCHES RANGES IN THE FIRST-V-ALL CASE

Range of SURF+SIFT matches	Nr. of matches over Nr. of scans in range	
	SURF+SIFT	(SURF+SIFT)+Inv
$0 \leq N \leq 2$	0/12	0/12
$3 \leq N \leq 9$	0/148 0%	8/148 5.41%
$10 \leq N \leq 19$	141/461 30.59%	218/461 47.29%
$20 \leq N \leq 29$	545/675 80.74%	600/675 88.89%
$30 \leq N \leq 39$	680/689 98.69%	681/689 98.84%
$40 \leq N$	1556/1556 100%	1556/1556 100%

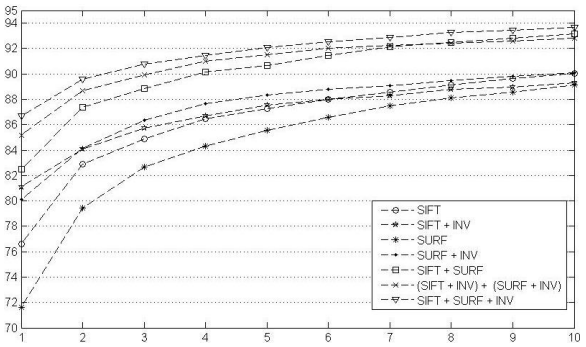

Fig. 7. Cumulative Match Curve for the experiments in table I.

TABLE V. (SIFT+SURF)+INV RECOGNITION RATES ON EXPRESSION SUBSETS OF THE PROBE SET ALL

Expression	Nr. of matches/Nr. of scans	Recognition Rates
Blank Stare	1821/1984	91.78%
Happiness	246/304	80.92%
Other	425/537	79.14%
Surprise	280/339	82.60%
Disgust	143/201	71.14%
Sadness	148/176	84.09%

iconic (or 3D) points nor the registration of the scans. The results were promising: the validation of SIFT through 3D invariants improved recognition rates. However, it was also evident that SIFT alone was not a robust enough method to extract enough salient points, which spurred us to search for more iconic points to be added to the SIFT ones. Here we have explored the use of SURF, which consist of a key-point detector based on integral images and a descriptor that is more robust to blur and scale changes than the SIFT one. SURF proved to be less descriptive when used alone, but reached almost the same recognition rates when coupled to 3D invariants. Moreover, the fusion a feature level of SIFT and SURF iconic points, followed by validation through invariants induced a remarkable improvement in recognition rates. Most recent 2D-3D techniques ([12], [15], [8], [9]) assessed their recognition method using the same protocol First vs All and report performances ranging from 92% [8] to 98% [15]. To try to reach similar scores, we can proceed in different ways: persevere in the search of other descriptive iconic points, for example by using edge and corner detectors, to completely exploit the 2D information of the face, or integrate the iconic points with 3D points. 3D feature points extraction is more demanding than its 2D counterpart but it would not be necessary to perform it for all faces; in fact it is clear from the results that the search of additional 3D key-points could be restricted to the small set of faces from which SIFT and SURF cannot extract enough iconic features.

REFERENCES

[1] F. Al-Osaimi, M. Bennamoun, and A. Mian. An expression deformation approach to non-rigid 3d face recognition. *Int. J. Comput. Vision*, 81(3):302–316, Mar. 2009.

[2] T. H. B. Bay, Herbert Surf: Speeded up robust features. *Computer Vision - ECCV 2006*, pages 404–417, 2006.

[3] P. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992.

[4] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of sift features for face authentication. *Computer Vision and Pattern Recognition Workshop, 2006.*, pages 35–35, 2006.

[5] K. W. Bowyer, K. I. Chang, and P. J. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.

[6] M. Cadoni, M. Bicego, and E. Grosso. 3d face recognition using joint differential invariants. *Advances in Biometrics, ICB 2009*, pages 279–288, 2009.

[7] M. Cadoni, A. Lagorio, and E. Grosso. Augmenting sift with 3d joint differential invarian for multimodal, hybrid face recognition. *The IEEE SIxt International Conference on Biometrics: Theory, Applications and Systems*, 2013.

[8] J. Cook, V. Chandran, S. Sridharan, and C. Fookes. Face recognition from 3d data using iterative closest point algorithm and gaussian mixture models. *3D Data Processing, Visualization and Transmission, 2004.*, pages 502–509, 2004.

[9] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):640–649, 2007.

[10] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[11] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni. Performance of geometrix activeidtm 3d face recognition engine on the frgc data. pages 154–154, 2005.

[12] A. S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007:1584–1601, 2007.

[13] P. J. Olver. Moving frames and joint differential invariants. *Regular and Chaothic Mechanics*, 4:3–18, 1999.

[14] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *Conference on Computer Vision and Pattern Recognition Vol. 1*, pages 947–954, 2005.

[15] C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. P. Segundo. 3d face recognition using simulated annealing and the surface interpenetration measure. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(2):206–219, 2010.

[16] P. A. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 2004.