Feature Fusion for Facial Landmark Detection

Panagiotis Perakis[a,c], Theoharis Theoharis[b,a], Ioannis A. Kakadiaris[c]

[a] *Computer Graphics Laboratory, Department of Informatics and Telecommunications, University of Athens, Ilisia 15784, Greece*
[b] *Department of Computer and Information Science, NTNU, 7491 Trondheim, Norway*
[c] *Computational Biomedicine Lab, Department of Computer Science, University of Houston, Texas 77204, USA*

Abstract

Facial landmark detection is a crucial first step in facial analysis for biometrics and numerous other applications. However, it has proved to be a very challenging task due to the numerous sources of variation in 2D and 3D facial data. Although landmark detection based on descriptors of the 2D and 3D appearance of the face has been extensively studied, the fusion of such feature descriptors is a relatively under-studied issue. In this paper, a novel generalized framework for combining facial feature descriptors is presented, and several feature fusion schemes are proposed and evaluated. The proposed framework maps each feature into a similarity score, combines the individual similarity scores into a resultant score, used to select the optimal solution for a queried landmark. The evaluation of the proposed fusion schemes for facial landmark detection clearly indicates that a quadratic distance to similarity mapping in conjunction with a root mean square rule for similarity

Email addresses: p.perakis@di.uoa.gr (Panagiotis Perakis), theotheo@di.uoa.gr (Theoharis Theoharis), ikakadia@central.uh.edu (Ioannis A. Kakadiaris)

fusion achieves the best performance in accuracy, efficiency, robustness and monotonicity.
Keywords: Facial Landmarks, Feature Extraction, Feature Fusion, Landmark Detection.

1. Introduction

Facial landmark detection is a crucial first step in facial analysis for biometrics and numerous other applications. However, it has proved to be a very challenging task due to the numerous sources of variation in 2D and 3D facial data. These variations can be environment-based (illumination conditions and occlusions), subject-based (pose and expression variations) and acquisition-based (image scale, distortion, noise, spikes and holes). Both 2D and 3D facial landmark detection suffers from occlusion and expression variations. In addition, 2D facial landmark detection suffers from pose and illumination variations.

2D and 3D facial landmark detectors have to possess the properties of robustness to data variations, repeatability and distinctiveness. To fulfill these properties and constrain the detection process, landmark detectors use trained landmark classifiers or 2D/3D appearance landmark models/templates and 2D/3D geometry models for global topological consistency. 2D landmark detectors use view-based 2D geometry and appearance models or 3D geometry models. 3D landmark detectors use solely 3D geometry and 3D appearance models. Fused 2D/3D landmark detection methods use 3D geometry and 2D+3D appearance models. 2D and 3D landmark detection is based mostly on variations of the seminal work on Active Appearance Models

of Cootes *et al.* [1, 2, 3, 4]. Fused 2D/3D landmark detection is presented in Boehnen & Russ [5], Jahanbin *et al.* [6], Lu & Jain [7], Passalis *et al.* [8] and Perakis *et al.* [9, 10].

Although many 2D/3D descriptors of facial features are used in the literature, a crucial issue has not been answered yet. How can these facial features be fused together in order to exploit their individual strengths and create a robust and accurate landmark detector?

Different feature descriptors can have complementary strengths and weaknesses, so combining them can increase system *accuracy*, *efficiency* and *robustness*, featuring *monotonicity*. Accuracy can be increased by exploiting data content from multiple sources (3D/2D) or the strengths of different data descriptors. In addition, using multiple descriptors can improve efficiency by limiting the landmarks' likelihood area. Finally, fusion can increase system robustness by limiting deficiencies inherent in using a single descriptor. For example a corner/edge detector is very sensitive in illumination variations, but the shape index is not. Thus, using multiple descriptors is a form of uncertainty reduction, since one descriptor may pick up what the other misses.

A landmark detector has four important levels (Fig. 1). At the *acquisition level* a sensor acquires the facial data. At the *feature extraction level* the data are transformed into features that represent the landmark classes. At the *matching score level* the extracted features are compared with feature templates that represent each landmark class in order to detect candidate landmarks with an associated matching score. Finally, at the *decision level* the matching scores (or ranks) are used to select a candidate landmark as the optimal solution for the queried landmark class.

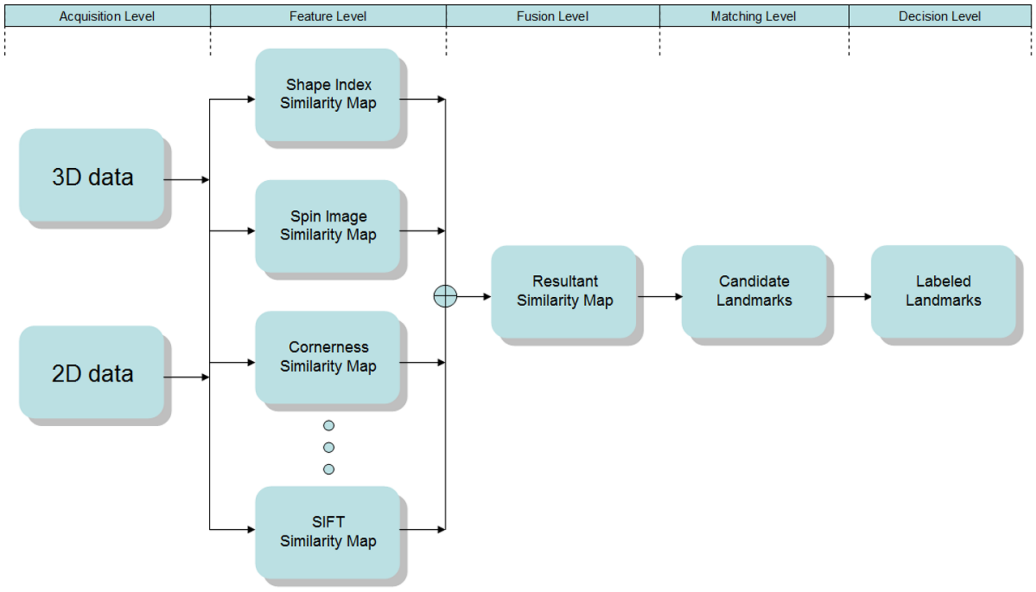


Figure 1: Pipeline of feature fusion procedure for landmark detection.

Fusion can be applied at the acquisition or feature extraction level (pre-classification fusion) and at the matching score or decision level (post-classification fusion) [11, 12]. Fusion at the matching score level can be viewed in two distinct ways. In the first, fusion is approached as a *classification* problem, while in the second, it is approached as a *combination* problem [11, 13]. In the classification approach, a composite feature vector (by weighted concatenation) is constructed using the values of the fused features, which is further classified by a composite classifier (e.g., Neural Network, K-NN, Decision Trees, SVM). In the combination approach, the matching scores of the fused features are combined to generate a single resultant feature score which is used for the final decision. The common characteristic of all combination techniques is that the individual feature classifiers are separately trained and the combination relies on simple fixed rules [13]. These rules are the *sum*

rule, *product rule*, *max rule*, *min rule*, *median rule* and *majority voting* [14]. The various schemes for combining classifiers can be grouped into three main categories according to their architecture: (i) *parallel*, (ii) *cascading* (serial), and (iii) *hierarchical* (tree-like) [15].

For landmark detection, although the construction of a composite feature classifier might be a potential solution, the combination method can be more easily applied to features whose values can be mapped to images, is more transparent (having also the strength of visualization), and possesses all the other fundamental properties required by a fusion scheme [16].

Feature fusion techniques have been proposed in the past (see Section 2), but in an entirely different context, that of multimodal biometrics or that of abstract feature fusion. The problem that is investigated in this paper is the behavior of fusion schemes under the strict context of landmark detection on facial datasets, which is an entirely different problem, since fusion techniques for landmark detection have to be also "locally consistent", which means that they have to boost results on a constrained area on facial surfaces. This problem has not yet been investigated.

This paper provides a novel generalized framework of fusion methods and their application to landmark detection and comes as an extension to our previous work for landmark detection [10]. The proposed framework fills a gap in existing research, which is dominated by methods that use single landmark descriptors of 3D or 2D appearance of the face, without combining them (see Section 2). The fusion scheme proposed acts after the "feature extraction level", transforms features to similarities and then combines them to generate a resultant feature similarity, which is considered as the

matching score, and is used at the "matching level" for the detection of the queried landmarks (Fig. 1). The proposed approach of feature fusion offers significant dimensionality reduction and is easily extendable by adding new feature-components in feature space and changing the resultant similarity appropriately. This approach works equally well for any feature extracted either from 3D or 2D facial data. The only prerequisite is the availability of a common (u,v) parameterization so that the 3D and 2D data can be combined at the "acquisition level".

The rest of this paper is organized as follows: Section 2 presents related work in the field, Section 3 details the theoretical background of the proposed method, Section 4 presents its application to the detection of facial landmarks, Section 5 presents our results, and Section 6 summarizes our method.

2. Related Work

A number of studies showing the advantages of information fusion in pattern recognition and especially in multimodal biometrics have appeared in the literature.

Xu *et al.* [12] (1992) grouped different fusion methods into categories and proposed methods for classifier fusion at different levels (measurement, rank and abstract) for recognizing handwritten numerals. They reported a significant improvement over the performance of individual classifiers.

Kittler *et al.* [14] (1998) have developed a theoretical framework for the combination approach to fusion at the matching score level of multimodal biometric applications. In their approach the matching scores of individual

classifiers are interpreted as posterior probabilities and the resultant scores are the outcome of simple fixed fusion rules (sum rule, product rule, max rule, min rule, median rule and majority voting). They have experimented with face and voice biometrics and found that the sum rule outperformed the others.

Jain *et al.* [15] (2000) conducted experiments concerning the characteristics of combining twelve different classifiers using five different combination rules and six different feature sets generated from handwritten numerals (0-9). Reported results show that each case favors its own combining rule and that combining does not necessarily lead to improved peformance.

Ross and Jain [17] (2003) addressed the problem of information fusion in biometric verification systems by combining face, fingerprint and hand geometry modalities using sum, decision-tree and LDA based methods. They reported that the sum rule outperforms the others.

Jain *et al.* [11] (2005) presented a thorough classification of information fusion approaches in biometric systems. They experimented with different normalization techniques (i.e., min-max, z-score, median, sigmoid, tanh and Parzen) and fusion rules (i.e., sum rule, max rule, min rule and weighted-sum rule) to combine score from different matchers in a multimodal recognition system. They concluded that the tanh normalization is the most robust and efficient for a recognition system, and that weighted summation of matching scores resulted in a significant improvement in recognition rates.

Ross and Govindarajan [18] (2005) have experimented with fusion at the feature level in 3 different scenarios: (i) fusion of PCA and LDA coefficients of face; (ii) fusion of LDA coefficients corresponding to the R,B,G channels

of a face image; and (iii) fusion of face and hand modalities. They concluded that it is difficult to predict the best fusion strategy for a given scenario.

Snelick *et al.* [19] (2005) examined the performance of multimodal biometric authentication systems using fusion techniques over fingerprint and face modalities on a population approaching 1,000 individuals. They also introduced adaptive normalization techniques and weighted fusion rules. They concluded that multimodal fingerprint and face biometric systems can achieve better performance than unimodal systems.

Gökberk and Akarun [20] (2006) have presented fusion techniques for 3D face recognition. Their fusion schemes combine four face classifiers which are used for the comparison of gallery and probe faces. Reported results show that their serial fusion technique offers the best solutions.

Theoharis *et al.* [21] (2008) presented a multimodal biometric recognition system using the fusion of face and ear modalities. They reported that the fused multimodal system achieved better performance (99.7% rank-one recognition rate) than the unimodal systems. The high reported accuracy was attributed to the low correlation of the two modalities.

In landmark detection literature on the other hand the combination of landmark descriptors is an under-studied issue.

Lu and Jain [7] (2005) used the combination of shape index response derived from the range map (3D) and the cornerness response from the intensity map (2D) to determine the positions of the corners of the eyes and the mouth. They used a fusion scheme of a pixel-wise summation of the normalized shape index and cornerness response values, for the "resultant" feature values of mouth and eye corners.

Boehnen and Russ [5] (2005) used color images (2D) and range data (3D). A skin detection algorithm is applied using the YCbCr transformation of the initial RGB image. The face region that results from skin detection is refined by using z-erosion exploiting the range data. Thus, at first a face segmentation is applied; next, eye and mouth likelihood maps are calculated (using Cb and Cr values), to locate the corresponding landmarks. Thus this method is not a fusion method but merely a 2D/3D filtering method.

Jahanbin *et al.* [6] (2011) used Gabor jets to represent intensity (2D) and range (3D) data. Next, the jets of each pixel were compared (using the appropriate similarity measure) to a target bunch (describing the queried landmark) in order to create similarity maps for each modality and landmark class. Finally, intensity and and range similarity maps were combined into a "hybrid" similarity map ("resultant"). For the calculation of the "resultant" similarity map different approaches of fusion were examined such as taking the pixel-wise sum, product or maximum of the similarity scores. They concluded that summation is the most appropriate.

In our previous work [9, 8, 10], we presented a 3D facial landmark detection system using the fusion of shape index and spin image feature descriptors. Our previous fusion system operated in a cascade (sequential) fashion so that the candidate landmarks extracted from the shape index transformation were classified and filtered out according to their similarity with precalculated spin image templates. We also used a product rule fusion of landmarks' geometric distance to a landmark model and spin image similarities at the decision level. We reported high landmark detection accuracy under large facial yaw rotations.

3. Feature Fusion for Landmark Detection

The features used for facial landmark detection have very different characteristics, but in general can be distinguished in scalar features (such as the Shape Index and Cornerness/Edge Response), and vector features (1D/2D histogram features, such as the SIFT descriptor and Spin Images). For each scalar feature we can statistically compute a corresponding target value, while for each vector feature we can compute a corresponding vector target (template), which represent a landmark in feature space.

Thus, the main idea of the proposed method is that instead of fusing features by weighted concatenation, the features are first transformed to similarities with a target value or template, and then each similarity can act as a component in a normalized feature similarity space (Fig. 2), fused together to form a resultant similarity, using simple combination rules (such as sum, product, max, min, etc.). In this manner a dramatic dimensionality reduction is achieved since, instead of using multiple components for a vector feature, only the similarity with its template is used.

Each feature for a landmark class has a target value or template (t_f) that describes the landmark in its feature space. Furthermore, we can consider a cut-off value (c_f) for each feature to incorporate the notion of an outlier. Feature values out of the range $[t_f - c_f, t_f + c_f]$ can be filtered out, so that threshold masking is implemented. The cut-off value can also be considered as a scaling factor for the normalization of each feature's range (Fig. 2).

The target and cut-off values can be estimated by examining the probability density function (pdf) of feature values or set to specific values based on a priori knowledge. A good choice for the target value could be the mean

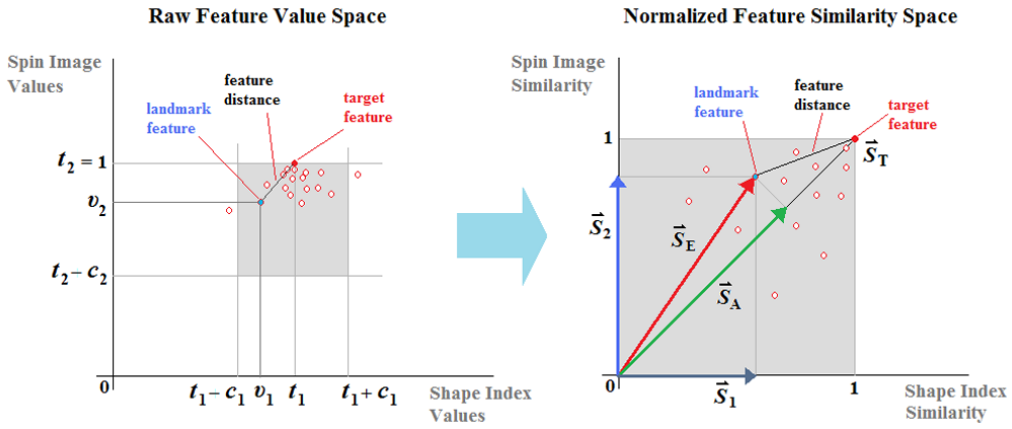


Figure 2: Example of the transformation from raw feature value space to normalized feature similarity space. Shape Index (v_1) and Spin Image (v_2) raw values are mapped onto Shape Index ($\vec{S_1}$) and Spin Image ($\vec{S_2}$) normalized similarity vectors. Note that the raw Spin Image values represent un-normalized similarity to the corresponding template.

of the pdf of feature values and for the cut-off value could be a multiple of standard deviation (std) (e.g., 3 $\times$ std as a first approximation), although the distribution of the values of every feature is not a Gaussian. Another choice for the target value could be the mode or the median of the pdf and the cut-off value could be determined so that a certain proportion of feature values (e.g., 99%) are within the range $[t_f - c_f, t_f + c_f]$.

For a good normalization scheme, the estimates of target (location), cut-off (scale) parameters and of the normalization function must be robust and efficient, and has to closely simulate the initial pdfs. In addition, a properly designed fusion method exploits information from each descriptor without degrading performance below that of the most accurate descriptor (monotonicity). This is the major challenge of adopting a fusion scheme.

3.1. Feature similarity mapping

Given a feature value v_f, a target value t_f and a cut-off value c_f for each feature descriptor f, we introduce a *normalized distance measure* to target D_f for each of the N feature descriptors of each landmark point:

$$D_f = \begin{cases} \frac{|v_f - t_f|}{c_f} & \text{if } |v_f - t_f| \leq c_f \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

Note that the above definition is a generalization of the z-score normalization and median normalization, presented in [11].

A *normalized similarity measure* to target S_f can be derived from D_f as:

a. Linear mapping:

$$S_f = 1 - D_f \ . \quad (2)$$

This is the classic linear distance to similarity transformation [13].

b. Quadratic mapping:

$$S_f = 1 - D_f^2 \ . \quad (3)$$

We introduce quadratic mapping, which favors close to target feature values. Note that D_f^2 behaves like the potential energy of elasticity.

c. Gaussian mapping:

$$S_f = exp(-\alpha D_f^2) \ , \quad (4)$$

where α is the drop-off parameter. We introduce Gaussian mapping, for smoothing out large distance measures. Note that the Gaussian tails can be cut at the cut-off values.

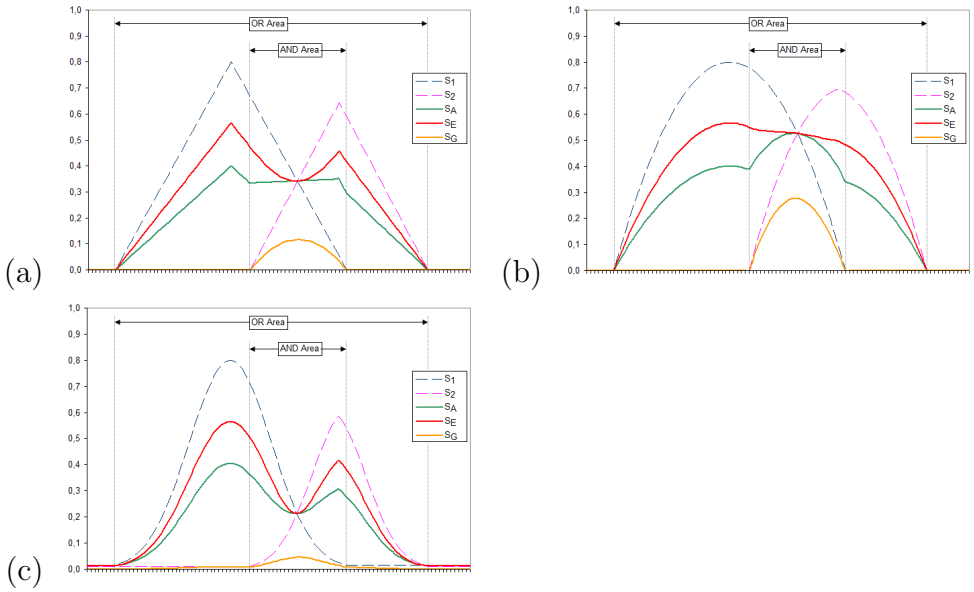


Figure 3: Depiction of fusion of similarities: (a) after linear mapping; (b) after quadratic mapping; and (c) after Gaussian mapping.

3.2. Feature similarity fusion

The resultant similarity measure to the target vector in the normalized similarity space describes the way by which the N feature descriptors can be fused together or combined into a resultant feature similarity for each queried landmark class:

a. Sum rule:

$$S_A = \frac{1}{N} \sum_{f=1}^{N} S_f \ , \tag{5}$$

which is the arithmetic mean or the Manhattan (L_1) metric (Fig. 2). Note that if the similarity measure is considered as the probability that the sample point is similar to the target, then this metric is equivalent to the *sum rule* for feature fusion [14, 13].

b. Root-mean-square rule:

$$S_E = \frac{1}{\sqrt{N}} \left(\sum_{f=1}^{N} S_f^2 \right)^{\frac{1}{2}} , \tag{6}$$

which is the root mean square (rms) of the similarities and actually a Euclidean (L_2) metric in the resultant similarity space. We introduce this novel *rms rule* so that feature similarities to targets can be considered as vectors and added according to vector addition (Fig. 2).

c. Product rule:

$$S_G = \left(\prod_{f=1}^{N} S_f \right)^{\frac{1}{N}} , \tag{7}$$

which is the geometric mean metric. Note that if the similarity measure is considered as the probability that the sample point is similar to the target, then this metric is equivalent to the *product rule* for feature fusion [14, 13].

d. Max rule:

$$S_{max} = \max_{f=1}^{N} \left(S_f \right) , \tag{8}$$

which is the L_∞ metric or *max rule* [14] and favors the feature with maximum similarity. Note that if the similarity measure is considered as a fuzzy variable, then this metric is equivalent to a fuzzy *OR rule* for feature fusion [13].

e. Min rule:

$$S_{min} = \min_{f=1}^{N} \left(S_f \right) , \tag{9}$$

which is the *min rule* [14] and favors the feature with minimum similarity. Note that if the similarity measure is considered as a fuzzy variable, then this metric is equivalent to a fuzzy *AND rule* for feature fusion [13].

To illustrate the behavior of the proposed distance to similarity mappings and fusion schemes we depict the various combinations in Fig. 3. For simplicity the fusion of similarity mapping functions is presented in a single dimension. We also depict in Fig. 4 the behavior of the proposed distance to similarity mappings and fusion schemes in the neighborhood of the Eye Outer Corner (EOC).

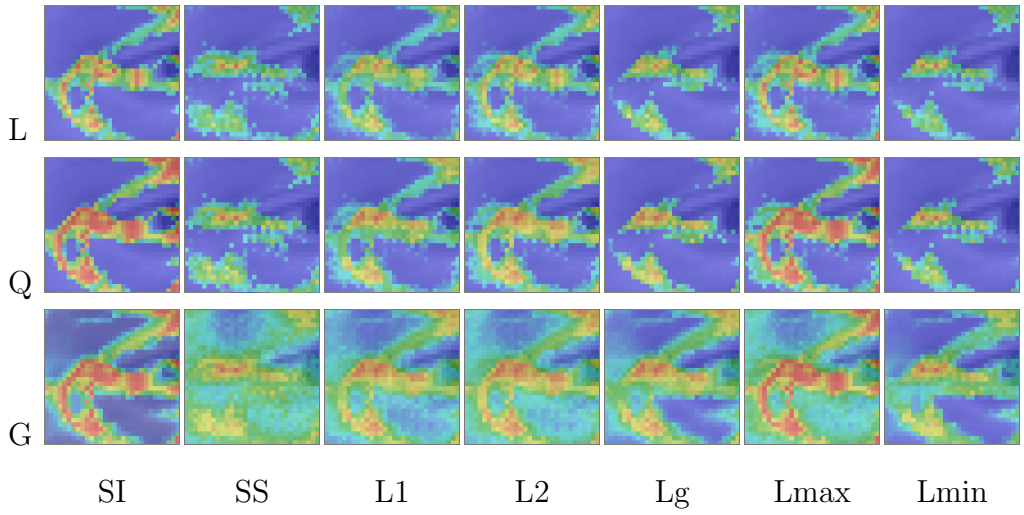


Figure 4: Depiction of the 2D similarity maps in the neighborhood of the Eye Outer Corner (EOC) for the various distance to similarity mappings and the various fusion methods: (blue) low similarity values (0.0); (green) medium similarity values (0.5); and (red) high similarity values (1.0). Rows depict: (top) L mapping; (middle) Q mapping; and (bottom) G mapping. Columns depict from left to right: SI and SS similarity; L1, L2, Lg, Lmax and Lmin fusion.

Remarks:

a. Linear mapping raises discontinuities in the superposed similarities, so it is expected to give unreliable results.

b. The "smoothest" results are given by the Gaussian and the Quadratic mapping, allowing a "locally smoother" combination of features.

c. S_G and S_{min} give results in the "AND Area" and S_A, S_E and S_{max} give results in the "OR Area". The "AND area" is more restricted and can be used to restrict the search space of candidate landmarks. The "OR area" is wider, which has the implication of a larger number of candidate landmarks to be detected, raising the "curse of dimensionality" at the decision level.

d. S_G and S_{min} give almost the same peak, approximately in the middle of the initial peaks of the fused features, having a similar behavior to an "AND operator". This peak is "smoother" for S_G and "sharper" for S_{min}.

e. S_{max} gives the same peaks as the initial peaks of the fused features, having a similar behavior to an "OR operator".

f. S_{max} gives as a result the similarity of the most "intensive" feature. Selecting the most "intensive" feature is unreliable, because it could be the one that makes the largest errors.

g. S_{min} gives as a result the similarity of the least "intensive" feature, which is not appropriate for landmark fusion, because it doesn't take into consideration the other features' similarities.

h. S_G and S_{min} may completely eliminate a feature's similarity peak which is not inside the "AND masking area", and thus are not appropriate for landmark fusion.

i. The S_A resultant similarity (L_1 metric) is equivalent to the normalized projection of the S_E similarity vector (L_2 metric) onto the target similarity vector S_T (see Fig. 2) (i.e. it is a normalized inner product metric, or the *cosine similarity measure* [13]). $\frac{\overrightarrow{S_E}}{\sqrt{N}} \cdot \frac{\overrightarrow{S_T}}{\sqrt{N}} = \frac{1}{N} \sum_{f=1}^{N} S_f \cdot 1 = S_A$.

4. Similarity mapping and fusion paradigms

To illustrate the characteristics of the proposed distance to similarity mappings and the fusion schemes we apply them for the detection of specific facial anatomical landmarks.

The landmark classes are: the Eye Outer Corner (EOC), the Eye Inner Corner (EIC), the Nose Tip (NT), the Mouth Corner (MC), and the Chin Tip (CT).

The descriptors that are used are: the Shape Index (SI), the Spin Image (SS), and the Edge Response (ER).

The distance to similarity mappings are: the Linear mapping (L), the Quadratic mapping (Q), and the Gaussian mapping (G).

The fusion schemes are: the sum rule using the arithmetic mean S_A (L1), the rms rule using the Euclidean mean S_E (L2), the product rule using the geometric mean S_G (Lg), the max rule using S_{max} (Lmax), and the min rule using S_{min} (Lmin).

4.1. Landmark Descriptors

To detect landmark points, we have used two 3D local shape descriptors that exploit the 3D geometry-based information of the facial datasets and one 2D local appearance descriptor that exploits the 2D intensity-based information: the *shape index*, the *spin images* and the *edge response*.

A facial scan belongs to a subclass of 3D objects which is a surface S expressed in parametric form with native (u, v) parameterization which also

(a) (b) (c) (d) (e)

Figure 5: Depiction of spin image templates: (a) EOC; (b) EIC; (c) NT; (d) MC; and (e) CT.

incorporates texture data. This parameterization allows to map 3D information onto 2D space and vice-versa, thus the 3D and 2D information can be cross-referenced [10, 8, 9].

The *Shape Index* [22, 23] is a continuous mapping of principal curvature values (k_{max}, k_{min}) of a 3D object point $\mathbf{p}$ onto the interval [0,1], and is computed as:

$$SI(\mathbf{p}) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{k_{max}(\mathbf{p}) + k_{min}(\mathbf{p})}{k_{max}(\mathbf{p}) - k_{min}(\mathbf{p})} \ . \tag{10}$$

The shape index captures the "local" shape of a surface. Five well-known shape types are: Cup = 0.0, Rut = 0.25, Saddle = 0.5, Ridge = 0.75, and Cap = 1.0.

A *Spin Image* [24] encodes the coordinates of points on the surface of a 3D object with respect to a so-called *oriented point* $(\mathbf{p}, \mathbf{n})$, where $\mathbf{n}$ is the normal vector at a point $\mathbf{p}$ of a 3D object surface. A spin image is a 2D grid accumulator of 3D points, as the grid is rotated around $\mathbf{n}$ by 360°. Thus, a spin image is a descriptor of the local shape of an object, invariant under rigid transformations.

The similarity measure between a spin image P and a spin image template

Q is expressed by the normalized linear correlation coefficient [24]:

$$SS(P,Q) = \frac{N\sum p_i q_i - \sum p_i \sum q_i}{\sqrt{[N\sum p_i^2 - (\sum p_i)^2]\,[N\sum q_i^2 - (\sum q_i)^2]}}\ , \tag{11}$$

where p_i, q_i denote each of the N elements of spin images P and Q, respectively.

The *Edge Response* is based on the well known Harris corner and edge detector [25], and encodes the intensity gradient of a point $\mathbf{p}$ on an image:

$$ER(\mathbf{p}) = I_x^2(\mathbf{p}) + I_y^2(\mathbf{p})\ , \tag{12}$$

where $I_x = \frac{\partial I}{\partial x}$ and $I_y = \frac{\partial I}{\partial y}$ denote the partial derivatives of the intensity image I in x and y respectively. $ER(\mathbf{p})$ is high in edge regions and close to zero in flat regions.

4.2. Training of the descriptors

To train the landmark descriptors we used 300 frontal facial datasets of different subjects, manually annotated at the specific landmark positions. These datasets come from FRGC v2 database [26, 27] and contain subjects with varying expressions and illumination conditions. The available 3D scans were used to train the shape index and spin image descriptors and the corresponding 2D texture images to train the edge response descriptor.

The pdf of the shape index values (SI) and edge response values (ER) for each landmark class were computed and used for the estimation of the shape index and edge response target and cut-off values. We computed spin image templates for each landmark class. Spin image templates represent the mean spin image associated with the five classes of landmarks (Fig. 5). The pdfs of the similarity values (SS) between the pre-computed spin image

Table 1: Target (t) and cut-off (c) values of the landmark descriptors for each landmark class

	EOC		EIC		NT		MC		CT	
	t	c	t	c	t	c	t	c	t	c
SI	0.32	0.53	0.12	0.60	1.00	0.40	0.09	0.68	0.96	0.70
SS	1.00	0.48	1.00	0.80	1.00	0.75	1.00	0.72	1.00	0.56
ER	0.20	0.72	0.16	0.62	0.10	0.40	0.22	0.70	0.02	0.17

templates and the spin images of each landmark class, were computed for the estimation of the cut-off values. The spin image target values are set to the maximum similarity (1.00).

The estimated target and cut-off values for each descriptor (SI, SS, ER) and for each landmark class (EOC, EIC, NT, MC, CT) are presented in Table 1, and the correlation coefficients between the landmark descriptors for each landmark class are presented in Table 2. Note that the introduction of distance to similarity mappings improves the correlation coefficients in comparison to the raw values.

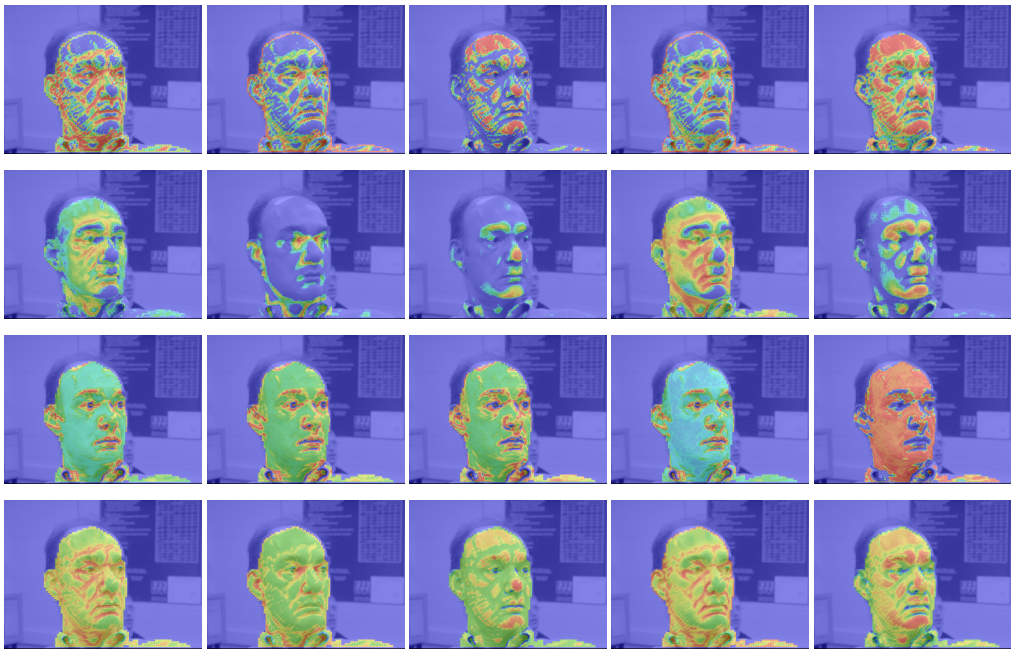
4.3. Landmark Labeling

General-purpose feature detection methods are not capable of identifying and labeling the detected candidate landmarks; some topological properties of faces must be taken into consideration. To address the problem of labeling the detected landmarks, we use our method presented in [10, 8, 9]. At the training phase, a Facial Landmark Model (FLM) is created by first aligning the training landmark sets and calculating a mean landmark shape using Procrustes Analysis, and then applying Principal Component Analysis (PCA)

Table 2: Correlation coefficients between landmark descriptors for each landmark class

	EOC	EIC	NT	MC	CT
Raw values					
SI / SS	0.0358	−0.1242	0.3202	−0.1823	0.1925
SI / ER	0.1458	0.0024	−0.0895	0.0000	0.0001
SS / ER	−0.0377	−0.1358	−0.1794	−0.2481	−0.0075
Linear mapping similarity values (L)					
SI / SS	0.1781	0.1806	0.3202	0.2669	0.2290
SI / ER	0.1665	0.0360	0.0638	0.1354	−0.0265
SS / ER	0.1080	0.0813	0.1002	0.1991	−0.0013
Quadratic mapping similarity values (Q)					
SI / SS	0.2095	0.1965	0.3098	0.2366	0.5241
SI / ER	0.1968	−0.0101	0.0572	0.0543	−0.0222
SS / ER	0.1184	0.0907	0.0370	0.1849	−0.0093
Gaussian mapping similarity values (G)					
SI / SS	0.2084	0.1921	0.3170	0.2508	0.3459
SI / ER	0.2023	0.0003	0.0524	0.0882	−0.0241
SS / ER	0.1205	0.0989	0.0614	0.2052	−0.0018

to capture the shape variations. The FLM serves as a 3D geometric model of the landmark points. At the detection phase, the algorithm first detects candidate landmarks on the queried facial datasets according to the previously described feature fusion schemes. The extracted candidate landmarks are then filtered out and labeled by matching them with the FLM.



(a) (b) (c) (d) (e)

Figure 6: Depiction of feature similarity maps with Q−L2 fusion: (blue) low similarity values (0.0); (green) medium similarity values (0.5); and (red) high similarity values (1.0). (**1st row**) SI similarity; (**2nd row**) SS similarity; (**3rd row**) ER similarity; and (**4th row**) Q−L2 resultant similarity. (**a**) eye outer corner; (**b**) eye inner corner; (**c**) nose tip; (**d**) mouth corner; and (**e**) chin tip.

5. Experimental Results

5.1. Test Databases

For the purposes of this evaluation, we used two databases:
(i) a database with 975 frontal facial datasets obtained from 149 different subjects, selected from the FRGC v2 database [26, 27], including subjects with varying degrees of expressions (45.44% "neutral", 36.41% "mild" and 18.15% "extreme"), acquired under varying illumination conditions. This

database will henceforth be referred as **DB00F**.

(ii) a composite database with the datasets of 39 common subjects found in the FRGC v2 database and in the UND Ear database [28]. This database consists of 117 (3x39) facial scans having three poses, frontal (39 scans) and 45 degrees left (39 scans) and right (39 scans), and will henceforth be referred as **DB00F45RL**.

5.2. Performance Evaluation of Fusion Schemes

Table 3: Qualitative evaluation of proposed fusion schemes

	Accuracy	Efficiency	Robustness	Monotonicity
L−L1	Fair	**High**	Fair	Fair
L−L2	Fair	Low	Fair	Fair
L−Lg	**High**	Fair	Fair	Fair
Q−L1	**High**	**High**	Fair	Fair
Q−L2	**High**	**High**	**High**	**High**
Q−Lg	**High**	Fair	Fair	Fair
G−L1	**High**	**High**	**High**	**High**
G−L2	**High**	**High**	Fair	Fair
G−Lg	**High**	Fair	Fair	Fair
L−Lmax	Low	Low	Low	Low
Q−Lmax	Low	Low	Low	Low
G−Lmax	Low	Low	Low	Low
L−Lmin	Unreliable	Fair	Fair	Low
Q−Lmin	Unreliable	Fair	Fair	Low
G−Lmin	Unreliable	Fair	Fair	Low

The evaluation of the performance of the proposed distance to similar-

ity mappings and fusion schemes for landmark detection is not a straightforward task, since there are many factors that characterize performance. As already stated, fusion techniques are expected to improve system's *accuracy*, *efficiency* and *robustness*. An equally important characteristic of a fusion scheme is that of *monotonicity*, i.e., the addition of a new feature descriptor should improve prior results.

Thus, we evaluate performance according to these four characteristics. *Accuracy* is evaluated according to the distance between the selected optimal landmark and the manually annotated landmark, which is considered as ground-truth. The selected optimal landmark is the 1^{st} rank candidate landmark for each landmark class (i.e., the candidate landmark which has the maximum resultant similarity score). *Efficiency* is evaluated according to the reduction of the likelihood area of a landmark class (see the high similarity areas in Figs. 4 and 6). The likelihood area of a landmark class is very important since its reduction means that fewer candidate landmarks have to be retained and fed to the "selection level". *Robustness* is evaluated by the use of testing datasets which contain subjects acquired under large yaw rotations, varying expressions and different illumination conditions, and also by the use of five different landmark classes. *Monotonicity* is evaluated according to the accuracy improvement between the use of individual descriptors, the fusion of the two richest descriptors, the shape index (SI) and the spin image (SS), and the fusion with the addition of a third poorer descriptor, the edge response (ER).

Table 4: Landmark localization distance-error (mm) results of Shape Index (SI), Spin Image (SS) and Edge Response (ER) fusion, in **DB00F** and **DB00F45RL**

	DB00F						**DB00F45RL**					
	EOC	EIC	NT	MC	CT	Mean	EOC	EIC	NT	MC	CT	Mean
SI	11.72	7.71	14.66	**5.98**	10.81	10.18	10.99	7.20	12.51	**4.68**	11.26	9.33
SS	**7.31**	**4.42**	**3.84**	8.47	**7.56**	**6.32**	**9.16**	**4.83**	**3.68**	7.03	**7.24**	**6.39**
ER	12.26	13.05	10.54	9.27	11.74	11.37	11.31	12.10	11.79	9.16	12.29	11.33
L−L1	6.40	4.60	4.12	**4.82**	**7.16**	5.42	**6.97**	**4.94**	4.40	**4.09**	**7.56**	**5.59**
L−L2	6.72	4.74	4.19	4.78	7.24	5.53	7.22	5.11	4.88	**4.09**	7.57	5.77
L−Lg	**6.31**	**4.52**	**4.08**	4.85	7.23	**5.40**	6.98	4.95	**4.20**	4.14	7.69	**5.59**
Q−L1	6.21	4.15	3.97	4.90	7.31	5.31	6.89	**4.59**	3.82	**3.83**	7.80	5.39
Q−L2	**6.19**	**4.14**	3.97	**4.87**	**7.28**	**5.29**	6.80	**4.59**	3.82	**3.83**	**7.73**	**5.35**
Q−Lg	6.20	4.15	**3.95**	4.92	7.29	5.30	**6.77**	**4.59**	**3.80**	**3.83**	7.79	5.36
G−L1	6.19	**4.14**	**3.97**	**4.86**	**7.28**	**5.29**	**6.80**	**4.59**	**3.82**	**3.83**	**7.73**	**5.35**
G−L2	**6.16**	4.15	3.98	4.89	**7.28**	**5.29**	6.85	4.64	3.84	**3.83**	**7.73**	5.38
G−Lg	6.21	4.15	**3.97**	4.90	7.31	5.31	6.89	**4.59**	**3.82**	**3.83**	7.80	5.39
L−Lmax	11.93	11.57	14.66	8.45	11.63	11.65	11.89	10.86	12.51	7.91	11.96	11.03
Q−Lmax	12.17	11.50	14.69	8.49	12.05	11.78	12.01	10.79	12.51	7.91	12.44	11.13
G−Lmax	12.17	11.50	14.69	8.49	12.05	11.78	12.01	10.79	12.51	7.91	12.44	11.13
L−Lmin	7.21	3.97	3.88	5.23	8.41	5.74	8.53	4.64	3.53	4.42	7.88	5.80
Q−Lmin	7.21	3.97	3.88	5.23	8.41	5.74	8.53	4.64	3.53	4.42	7.88	5.80
G−Lmin	7.21	3.97	3.88	5.23	8.41	5.47	8.53	4.64	3.53	4.42	7.88	5.80

A qualitative performance evaluation of the proposed fusion schemes according to the aforementioned characteristics is presented in Table 3. Detailed landmark localization errors are presented in Tables 4 and 5.

Table 5: Landmark localization distance-error (mm) results of Shape Index (SI) and Spin Image (SS) fusion, in **DB00F** and **DB00F45RL**

	DB00F						**DB00F45RL**					
	EOC	EIC	NT	MC	CT	Mean	EOC	EIC	NT	MC	CT	Mean
SI	11.72	7.71	14.66	**5.98**	10.81	10.18	10.99	7.20	12.51	**4.68**	11.26	9.33
SS	**7.31**	**4.42**	**3.84**	8.47	**7.56**	**6.32**	**9.16**	**4.83**	**3.68**	7.03	**7.24**	**6.39**
L−L1	7.58	4.81	**3.85**	5.85	7.30	5.88	8.82	5.11	**3.67**	5.04	7.38	6.00
L−L2	7.70	4.84	**3.85**	5.81	**7.16**	**5.87**	8.80	5.06	**3.67**	5.03	7.53	6.02
L−Lg	**7.54**	**4.80**	**3.85**	**5.80**	7.38	**5.87**	**8.53**	**5.05**	**3.67**	**4.99**	**7.35**	**5.92**
Q−L1	7.54	4.73	**3.84**	**5.84**	**7.28**	5.85	8.39	4.98	**3.62**	**4.72**	**7.53**	5.85
Q−L2	**7.52**	**4.72**	3.85	**5.84**	**7.28**	**5.84**	**8.33**	**4.97**	**3.62**	**4.72**	**7.53**	**5.83**
Q−Lg	7.53	4.73	3.85	5.87	7.29	5.85	8.39	**4.97**	**3.62**	**4.72**	7.54	5.85
G−L1	**7.52**	**4.72**	3.85	**5.84**	**7.28**	**5.84**	**8.33**	**4.97**	**3.62**	**4.72**	**7.53**	**5.83**
G−L2	7.53	**4.72**	**3.84**	**5.84**	**7.28**	**5.84**	8.34	**4.97**	3.67	**4.72**	**7.53**	5.85
G−Lg	7.54	4.73	**3.84**	**5.84**	**7.28**	5.85	8.39	4.98	**3.62**	**4.72**	**7.53**	5.85
L−Lmax	11.72	7.71	14.66	6.06	10.81	10.19	11.00	7.23	12.51	4.68	11.26	9.34
Q−Lmax	11.72	7.72	14.66	6.06	10.81	10.19	10.99	7.20	12.51	4.68	11.26	9.33
G−Lmax	11.72	7.72	14.66	6.06	10.81	11.78	10.99	7.20	12.51	4.68	11.26	9.33
L−Lmin	7.34	4.61	3.84	5.91	7.39	5.82	8.53	4.64	3.53	4.42	7.88	5.80
Q−Lmin	7.34	4.61	3.84	5.91	7.39	5.82	9.20	4.88	3.51	5.03	7.27	5.98
G−Lmin	7.34	4.61	3.84	5.91	7.39	5.82	9.20	4.88	3.51	5.03	7.27	5.98

Our experimental findings are similar to those of [15], which are summarized in the following:

i) There is no single combination rule that scores best for all cases.

ii) Combining does not necessarily lead to improved performance.

iii) There are cases where none of the combining rules does better than the

best individual detector.

Despite these general findings a more detailed examination of the results shows that there are some fusion schemes that perform better in most cases and can be adopted, and others that perform quite poorly and should be avoided (see also the Remarks of Section 3.2).

Our results show that, in general, the Quadratic (Q) and Gaussian (G) mappings behave better than the Linear (L) mapping. For the Linear mapping the product rule (Lg) behaves better than other rules. For the Quadratic mapping the rms rule (L2) behaves better than other rules. For the Gaussian mapping the sum rule (L1) behaves better than other rules. Quadratic and Gaussian mappings have almost the same performance.

The introduction of the Edge Response (ER) descriptor improves the results for the EOC, EIC and MC landmarks, but degrades the results for NT and CT. Note that, although ER is a poor descriptor, the improvement in accuracy is more dramatic in MC and EOC where the ER descriptor is more correlated with the SI and SS descriptors. Also note that the decline in accuracy is more dramatic in NT and CT where the ER descriptor is uncorrelated with the SI and SS descriptors (Table 2).

Accuracy improvement is more dramatic when the information fused is correlated. In correlated features the performance of one descriptor predicts to some extent the performance of the other and strengthens the results. On the other hand highly uncorrelated features have similarity peaks that do not coincide and degrade the results. Efficiency improvement is achieved by excluding obvious non-matches, reducing the number of candidate landmarks, for each landmark class. Fusion, also, reduces system sensitivity to

sample-specific, poor-quality or erroneous descriptors.

We can thus deduce that the best performance in terms of accuracy is exhibited by the Q-L2 and G-L1 fusion schemes, with the Q-L2 exhibiting a slight better performance than the G-L1 in landmarks' likelihood area reduction. Q-L2 and G-L1 also exhibit high robustness in yaw, expression and illumination variations, and strong monotonicity.

5.3. Performance Evaluation of Landmark Detection

General-purpose feature detection methods are not capable of identifying and labeling the detected candidate landmarks; some topological properties of faces must be taken into consideration. To have a fully functional landmark detector the detected, through the feature fusion procedure, candidate landmarks have to be identified and labeled. For this purpose we utilized a Facial Landmark Model (FLM) [10].

The performance of the proposed landmark detection method against other state-of-the-art methods, landmark localization errors are presented in Tables 6 and 7. Note that each method uses a different facial database, making direct comparisons difficult. However, these results indicate that our method for landmark detection by feature fusion outperforms previous methods.

Landmark localization using the Q-L2 fusion scheme improved the accuracy and robustness of the landmark detector (with mean landmark localization error within a range of $3.5 - 5.5\ mm$), indicating the superiority of the fusion approach against our previously published approaches.

Table 6: Comparison of the proposed landmark detection method using Fusion Schemes against state-of-the-art on almost-frontal complete facial datasets

Mean Localization Error (mm)									
Method		REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Yu *et al.* [29]	(GA model)	4.74	5.59	-	-	2.18	-	-	-
Nair *et al.* [30]	(w/o PDM)	25.01	26.68	31.84	34.39	14.59	-	-	-
	(w PDM)	12.11	11.89	20.46	19.38	8.83	-	-	-
Lu *et al.* [31]	(3D)	8.30	8.20	9.50	10.30	8.30	-	6.00	6.20
Lu *et al.* [7]	(3D+2D)	6.00	5.70	7.10	7.90	5.00	-	3.60	3.60
Colbry [32]	(w/o CFDM)	5.50	6.30	-	-	4.10	11.00	6.90	6.70
	(w CFDM)	5.60	6.00	-	-	4.00	11.70	5.40	5.40
Perakis *et al.* [9]	(SISI–NP)	7.02	7.46	8.13	9.21	5.23	6.71	8.30	9.83
Passalis *et al.* [8]	(UR3D-S)	5.03	5.48	5.79	5.62	4.91	6.31	5.65	6.47
Perakis *et al.* [10]	(SISI–NPSS)	4.15	4.41	5.58	5.83	4.09	4.92	5.56	5.42
Fusion scheme	Q–L2(SI+SS)	4.00	4.13	4.78	5.39	3.65	4.13	4.21	4.48
Fusion scheme	Q–L2(SI+SS+ER)	3.53	4.26	4.48	5.53	3.78	4.09	3.91	4.51

Table 7: Comparison of the proposed landmark detection method using Fusion Schemes against state-of-the-art on mixed (frontal and profile) facial datasets

Mean Localization Error (mm)									
Method		REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Lu *et al.* [31]	(3D)	9.00	7.10	13.60	13.30	6.40	-	6.70	5.20
Passalis *et al.* [8]	(UR3D-S)	5.97	6.87	6.51	6.71	4.60	6.59	5.52	6.10
Perakis *et al.* [10]	(SISI–NPSS)	4.65	4.90	5.32	6.06	4.41	4.80	5.01	4.91
Fusion scheme	Q–L2(SI+SS)	4.35	4.86	4.74	5.39	4.14	4.86	4.02	3.73
Fusion scheme	Q–L2(SI+SS+ER)	4.10	5.23	4.60	5.48	4.41	5.29	3.54	4.21

Comparative results of landmark localization errors on almost-frontal facial datasets are presented in Table 6. Note that the FRGC v1 database used in Yu *et al.* [29], Lu *et al.* [31], Lu *et al.* [7], and Colbry [32] is con-

sidered less challenging than the FRGC v2 used in our experiments, since FRGC v1 contains subjects with neutral expressions, while FRGC v2 contains subjects with various facial expressions. Furthermore, the database used by Colbry [32] contains a small portion ($\approx 5\%$) of proprietary datasets with pose variations, occlusions and expressions. The BU-3DFE database [33] used in Nair *et al.* [30] contains frontal only 3D facial datasets, which were created by the fusion of facial data acquired at $\pm 45^{\circ}$ yaw, from 100 subjects that perform seven universal expressions.

Comparative results of landmark localization errors on mixed (frontal and profile) facial datasets are presented in Table 7. Note that the proprietary MSU database used in Lu *et al.* [31] contains 300 3D facial scans from 100 subjects, three scans for each subject captured at 0 and $\pm 45^{\circ}$ yaw angles. The DB00F45RL database used in our experiments, despite having fewer subjects, is considered more challenging, since yaw angles lie in the range $[-65^{\circ}, +67^{\circ}]$ (for more details see [10]).

5.4. Computational Cost

For the evaluation of the presented method's computational efficiency, a PC with the following specifications was used: Intel Core i5 2.5 GHz with 4 GB RAM. Using this PC, 13.76 s (on avg.) was required to locate and label the landmarks for each facial scan. The average time taken for each step of the method is: Data loading 0.04 s, shape index similarity maps computation 0.66 s, spin image similarity maps computation 3.34 s, edge response similarity maps computation 0.42 s, resultant similarity map computation and candidate landmarks detection 2.45 s, FLM5L-FLM5R matching and landmark labeling 3.85 s, and FLM8 matching and optimal landmark set

selection 3.02 s. Speedups through parallelization are possible and thus the computational efficiency of the presented landmark detector makes it applicable to real-world applications.

6. Conclusion

A novel generalized framework for feature fusion and its application to landmark detection has been presented. The proposed fusion scheme acts after the "feature extraction level", transforms features to similarities and combines them to generate a resultant feature similarity, which is considered as the matching score used at the "matching level" for the detection of queried landmarks. The proposed fusion scheme is easily extendable to new feature-components in feature space, offers significant dimensionality reduction and works equally well for features extracted from 3D or 2D facial data.

For the proposed fusion scheme different distance to similarity mappings (linear, quadratic and Gaussian) and different fusion rules (sum rule, rms rule, product rule, max rule and min rule) have been evaluated according to *accuracy*, *efficiency*, *robustness* and *monotonicity*. The results indicate that the quadratic distance to similarity mapping in conjunction with the rms rule for fusion (Q-L2) exhibits the best performance. Landmark localization using this fusion scheme achieved state-of-the-art accuracy (with $3.5-5.5\,mm$ mean landmark localization error), indicating the superiority of the fusion approach against other previous methods.

Acknowledgment

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

References

[1] T. Cootes, C. Taylor, Statistical models of appearance for computer vision, Tech. rep., University of Manchester (Oct. 2001).

[2] T. Cootes, C. Taylor, D. Cooper, J. Graham, Active shape models - their training and application, Computer Vision and Image Understanding 61 (1) (1995) 38–59.

[3] T. Cootes, K. Walker, C. Taylor, View-based active appearance models, in: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2002, pp. 227–232.

[4] D. Cristinacce, T. Cootes, Automatic feature localization with constrained local models, Pattern Recognition 41 (10) (2008) 3054–3067.

[5] C. Boehnen, T. Russ, A fast multi-modal approach to facial feature detection, in: Proc. 7^{th} IEEE Workshop on Applications in Computer Vision, Vol. 1, 2005, pp. 135–142.

[6] S. Jahanbin, H. Choi, A. Bovik, Passive multimodal 2-D+3-D face recognition using gabor features and landmark distances, IEEE Transactions on Information Forensics and Security 6 (4) (2011) 1287–1304.

[7] X. Lu, A. Jain, Multimodal facial feature extraction for automatic 3D face recognition, Tech. Rep. MSU-CSE-05-22, Michigan State University (Oct. 2005).

[8] G. Passalis, P. Perakis, T. Theoharis, I. Kakadiaris, Using facial symmetry to handle pose variations in real-world 3D face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (10) (2011) 1938–1951.

[9] P. Perakis, T. Theoharis, G. Passalis, I. Kakadiaris, Automatic 3D facial region retrieval from multi-pose facial datasets, in: Proc. Eurographics Workshop on 3D Object Retrieval, Munich, Germany, 2009, pp. 37–44.

[10] P. Perakis, G. Passalis, T. Theoharis, I. Kakadiaris, 3D facial landmark detection under large yaw and expression variations, IEEE Transactions on Pattern Analysis and Machine Intelligence 99 (Preprint) (2012) 1. doi:http://doi.ieeecomputersociety.org/10.1109/TPAMI.2012.247.

[11] A. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognition 38 (12) (2005) 2270–2285.

[12] L. Xu, A. Krzyzak, C. Suen, Methods for combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on System, Man, and Cybernetics 22 (3) (1992) 418–435.

[13] S. Theodoridis, K. Koutroumbas, Pattern Recognition, 3^{rd} Edition, Academic Press, 2006.

[14] J. Kittler, M. Hatef, R. Duin, J. Matas, On combining classifiers, IEEE Transactions on Pattern Analysis an Machine Inteligence 20 (3) (1998) 226–239.

[15] A. Jain, R. Duin, J. Mao, Statistical pattern recognition: A review, IEEE Transactions on Pattern Analysis and Machine Inteligence 22 (1) (2000) 4–37.

[16] E. Bossè, A. Guitouni, P. Valin, An essay to characterise information fusion systems, in: Proc. 9^{th} International Conference on Information Fusion, Florence, Italy, 2006, pp. 1–7.

[17] A. Ross, A. Jain, Information fusion in biometrics, Pattern Recognition Letters 24 (13) (2003) 2115–2125.

[18] A. Ross, R. Govindarajan, Feature level fusion using hand and face biometrics, in: Proc. SPIE Conference on Biometric Technology for Human Identification II, Orlando, USA, 2005, pp. 196–204.

[19] R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain, Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems, IEEE Transactions on Pattern Analysis an Machine Inteligence 27 (3) (2005) 450–455.

[20] B. Gökberk, L. Akarun, Comparative analysis of decision-level fusion algorithms for 3d face recognition, in: Proc. 18^{th} International Conference on Pattern Recognition, Vol. 3, Hong Kong, China, 2006, pp. 1018 – 1021.

[21] T. Theoharis, G. Passalis, G. Toderici, I. Kakadiaris, Unified 3D face and ear recognition using wavelets on geometry images, Pattern Recognition 41 (3) (2008) 796–804.

[22] C. Dorai, A. K. Jain, COSMOS - a representation scheme for 3D free-form objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (10) (1997) 1115–1130.

[23] J. Koenderink, A. van Doorn, Surface shape and curvature scales, Image and Vision Computing 10 (1992) 557–565.

[24] A. E. Johnson, Spin Images: A Representation for 3-D Surface Matching, Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (Aug. 1997).

[25] C. Harris, M. Stephens, A combined corner and edge detector, in: Proc. 4^{th} Alvey Vision Conference, 1988, pp. 147–151.

[26] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the Face Recognition Grand Challenge, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, 2005, pp. 947–954.

[27] P. Phillips, T. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, M. Sharpe, FRVT 2006 and ICE 2006 large-scale experimental results, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2010) 831–846. doi:10.1109/TPAMI.2009.59.

[28] UND, University of Notre Dame Biometrics Data Sets, http://www.nd.edu/∼cvrl/CVRL/Data_Sets.html (2012).

[29] T. Yu, Y. Moon, A novel genetic algorithm for 3D facial landmark localization, in: Proc. 2^{nd} IEEE International Conference on Biometrics: Theory, Applications and Systems, Arlington, VA, 2008.

[30] P. Nair, A. Cavallaro, 3-D face detection, landmark localization, and registration using a point distribution model, IEEE Transactions on Multimedia 11 (4) (2009) 611–623.

[31] X. Lu, A. Jain, Automatic feature extraction for multiview 3D face recognition, in: Proc. 7^{th} International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 2006, pp. 585–590.

[32] D. Colbry, Human face verification by robust 3D surface alignment, Ph.D. thesis, Michigan State University (2006).

[33] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, A 3D facial expression database for facial behavior research, in: Proc. 7^{th} International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 2006, pp. 211–216.