3D Face Recognition Using Joint Differential Invariants

Marinella Cadoni[1], Manuele Bicego[1,2], and Enrico Grosso[1]

[1] Computer Vision Laboratory, DEIR, University of Sassari, Italy
[2] Computer Science Dept., University of Verona, Italy
{maricadoni,bicego,grosso}@uniss.it

Abstract. Stemming from a sound mathematical framework dating back to the beginning of the 20th century, this paper introduces a novel approach for 3D face recognition. The proposed technique is based on joint differential invariants, projecting a 3D shape in a 9-dimensional space where the effect of rotation and translation is removed. As a consequence, the matching between two different 3D samples can be directly performed in the invariant space. Thus the matching score can be effectively used to detect surfaces or parts of surfaces characterised by similar when not identical 3D structure. The paper details an efficient procedure for the generation of the invariant signature in the 9-dimensional space, carefully discussing a number of significant implications related to the application of the mathematical framework to the discrete, non-rigid case of interest. Experimental evaluation of the proposed approach is performed over the widely known 3D_RMA database, comparing results to the well established Iterative Closest Point (ICP)-based matching approach.

Keywords: Biometrics, 3D Face Recognition, Invariants.

1 Introduction

The typical approach in the automatic analysis of faces context consists in analysing the intensity image of a face – the so called 2D face. Nevertheless, in uncontrolled environments, illumination and pose variations may severely impair recognition systems based on sole 2D information. To overcome these problems, different alternatives have been proposed in recent years, based on 3D information. The approaches presented in the literature – see [1] for an comprehensive review – can be broadly divided into two classes: the first contains methods that perform recognition on the sole basis of shape, while the second comprises the so-called mutimodal algorithms which use both 3D and 2D data (namely the texture of the 3D shape) [2]. Here we concentrate on the former class: in this context, different are the features that may be used, ranging from the raw clouds of points to the curvatures or the face profiles, ending with the well known depth (range) images. In this paper we propose a novel approach for 3D face recognition, solely based on the analysis of the shape of a face represented as a raw cloud of points. The method presented here follows the typical scheme for point

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 279–288, 2009.
© Springer-Verlag Berlin Heidelberg 2009

cloud-based systems [3]: given two clouds of points representing two faces, the matching score is obtained by a registration process where the two clouds are aligned, and the matching score is given by the registration error. In this context, the widely employed registration technique is the general Iterative Closest Point (ICP) method [4], which aligns two shapes by iteratively minimizing the pairwise distances between corresponding points. Even if some interesting and smart extensions have been proposed (e.g. to deal with non rigid variations [5,6,7]), the registration obtained with ICP is prone to be erroneous and time consuming. In fact the registration is obtained with an optimization process, which, starting from an initial alignment of the two shapes, iteratively minimises a closeness measure. Such process is a local optimiser, which converges to the nearest local minimum. It is therefore evident that a coarse pre-alignment is crucial to ensure a proper registration. Moreover, it should be noted that the registration time increases with the number of iterations, which could make the method unsuitable for real time recognition purposes. The face recognition approach presented in this paper aims to overcome these drawbacks, by proposing to perform the registration using joint differential invariants [13]. This sound mathematical framework provides an optimal way to project all points belonging to the two faces in a invariant space where the effects of rotation and translation are removed. The matching is then performed in such space. The resulting registration process is not based on an optimization procedure but directly derives from the invariant space. Moreover, no pre-alignment is required, since the effect of rigid variations is by definition removed. Finally, it should be emphasised that even in case of poor overlapping of the two faces (30% or less), the method has the potential to find the correct registration. It is important to note that such invariants are defined on smooth manifolds, assuming continuous surfaces. Clearly, the acquisition process samples the surface, and a too heavy sub-sampling may reduce the accuracy of the registration. Although this registration process reaches perfection in the ideal rigid-object noise-free case, in this paper we will show that it can be fruitfully exploited in the face case, where non rigid variations and noise may be present. In order to reduce the computational requirements, the matching is performed on specific points, detected around maxima of curvature. Preliminary experiments made on the well known 3D_RMA dataset [9] show promising results, also with respect to the standard ICP.

2 Joint Invariants for Surface Classification

In this section the theory that leads to Joint Invariants is presented, starting with an outline of the Moving Frames Theory of Olver [10], following with the description of the invariant signature adopted here and ending by detailing its application to the point cloud case.

2.1 Outline of the Moving Frame Theory

The classical differential invariant theory for surfaces embedded in the Euclidean space leads to invariants of second order derivative, the well known Gaussian

and Mean curvatures (see [14]). These invariants, together with their 1st order derivatives with respect to the Frenet frame, parameterise a signature manifold that completely determines the surface up to Euclidean transformations (see [10]). This implies that two surfaces are the same (up to an Euclidean motion) if and only if their signature manifolds are the same and thus gives us a way to characterise surfaces up to Euclidean motions. In practice, a given a surface and a copy of it obtained by simply rototranslating the original one, share the same invariants signature. However, if we apply this methodology to experimental cases, where the surface is approximated by a sampling of points, any noise in the points will be amplified by the second order derivatives, making comparison of signatures difficult.

The theory developed by Olver in [10] is based on the classical theory of Moving Frames first introduced by Cartan [12] and provides us with an algorithm for building functionally independent sets of invariants that characterise a surface up to Lie Group transformations (which include Euclidean motions). By prolonging the group action to various cartesian copies of the surface, we can build invariants that depend on smaller order derivatives, the so called joint differential invariants. In particular, if taking one copy of the surface leads to the classical Gaussian and Mean curvature at each point of the surface, by considering three copies of the surface the invariants will depend on three points, and will consist of the three distance between the points and 6 first order invariants. If we prolong the action to enough copies (7 in the case of surfaces), we will find 21 invariants that depend only on the inter-point distances of 7 points of the surface at a time. As in the case of the classical differential invariants, joint invariants parameterise a signature that characterises the original surface up to the considered transformations.

The theory of joint differential invariants therefore gives a very elegant and powerful way of constructing a minimal set of invariants that are necessary to define a signature of the surface. The signature will live in a space of dimension equal to the number of invariants that parameterise it whereas its dimension will depend on the number of points the invariants are defined on (i.e. on the prolongation). This means that, through prolongation, we can have a representation of the original surface that is invariant to transformations and dependent on low order derivatives, at the expense of high dimensionality and computational complexity. In practical terms, a zero order signature is parameterised by the 21 inter-point distances between all ordered subsets of 7 points of the surface. If the surface is a point cloud consisting of n points, then we would have n^7 subsets each of which generates 21 invariants: if n is large, both the generation of the signature and any further processing become computationally challenging.

2.2 Invariants and Signature Generation

To compromise between computational time and robustness we choose a 3-fold prolongation, so that our invariants will depend on three points at one time. As we will see, this choice leads to three invariants of order zero plus six of order

one. Let p_1, p_2 and p_3 be three points on the surface. If the surface is smooth we can define the normal vector n_i at each point p_i.

Furthermore, (see figure 2(a)) let r be the direction of the line between the first two points and n_t the normal to the plane through the 3 points:

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

The zero order invariants we find are the 3 interpoint distances $I_k(p_1, p_2, p_3)$ for $k = 1, 2, 3$:

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

The first order invariants are the following:

$$J_k(p_1, p_2, p_3) = \frac{(n_t \wedge r) \cdot n_k}{n_t \cdot n_k} \quad \text{for } \; k = 1, 2, 3$$

and

$$\tilde{J}_k(p_1, p_2, p_3) = \frac{r \cdot n_k}{n_t \cdot n_k} \quad \text{for } \; k = 1, 2, 3.$$

To each triplet (p_1, p_2, p_3) on the surface we can then associate a point of the signature given by $(I_1, I_2, I_3, J_1, J_2, J_3, \tilde{J}_1, \tilde{J}_2, \tilde{J}_3)$. As the invariants depend on three points, each of which has two degrees of freedom on the surface, the signature will be a 6-dimensional manifold embedded in 9-dimensional space.

2.3 Point Cloud Implementation

Our aim is to adapt the general framework for constructing an optimal set of invariants to our case of interest: the registration of (possibly partially overlapping) clouds of points obtained by sampling surfaces. Let $\mathcal{F}_1 = \{p_1, \ldots, p_n\}$ and $\mathcal{F}_2 = \{q_1, \ldots, q_m\}$ be two clouds of points sampled from two faces. For $i = 1, 2$ we consider all unordered triplets of points in the sets $\mathcal{F}_i$ and calculate the invariants described in 2.2. This will produce two discrete sets of points, $\mathcal{S}_1$ and $\mathcal{S}_2$, that would lie on the signature manifold theoretically generated by the continuous surfaces of the faces, i.e. two sub-samplings of the signature manifolds. The next step is the comparison of the two signatures.

From the general theory it follows that if two signatures partially overlap (i.e. they intersect in a subset whose dimension equals their dimension), then also the surfaces that generated them will have the same property (they will overlap) after an Euclidean motion. In order to establish the intersection of $\mathcal{S}_1$ and $\mathcal{S}_2$, we need to define a metric in the embedding space R^9.

After normalising the values of the 9 invariants, we found the Euclidean metric was sufficient for reliably establishing matching points. Using a kd-tree algorithm [15], the search of the closest point $q_i \in \mathcal{S}_2$ for each point $p_i \in \mathcal{S}_1$ can be performed efficiently. Let $d_\star$ be a fixed threshold and M the set of pairs (p_i, q_i) that satisfy the inequality $\|p_i - q_i\| \leq d_\star$.

If we denote by $|M|$ the cardinality of M, the signatures intersection is defined to be $S_I = |M|/\min\{|S_1|, |S_2|\}$.

3 A Joint-Invariants Based Face Recognition Algorithm

In view of the theory outlined in 2, we could readily implement an algorithm for face recognition. Following the standard approach to point cloud-based 3D face recognition [3], given two faces to be compared, the idea is to register them and use the registration error as a matching score.

In details, suppose we have a training set G containing face scans of various subjects. Each face scan can be represented by its 6-dimensional signature embedded in E^9, which, as we have seen, characterises it up to Euclidean motion. When a test scan comes along and we want to compare it with the training scans we can build its signature and compare it to all signatures in the training set by evaluating the intersections S_I's. The unknown testing scan is then assigned to the subject with the most similar signature (highest value of the S_I's). The algorithm may be easily extended to the authentication scenario, where a testing face is authenticated if the template's signature is similar enough (given a threshold).

Using all points of the scans and invariants of maximum differential order equal to one, the matching would be robust and simple. Indeed, even considering facial expressions, it is reasonable to assume that there are enough stable points to distinguish an individual from another (since the method works also in case of partial overlapping) and no special metric is necessary to compare the signatures (see 2.3).

All this makes the framework very appealing. Unfortunately, computational complexity prevents us to readily apply it in this "full" form: the average cardinality of a face scan $\mathcal{F}$ in $3D$ databases can be beyond computationally capability. In the database we experimented on it is 4×10^3. Considering all these points would result in $4^3 \times 10^3$ signature points. This is computationally unfeasible, not to mention the processing time of such a bulk of data. To overcome this problem, we need a way to sub-sample the face scans. Since the invariants depend on distances and normals, we must ensure that, in common patches of different scans of the same subject, we sub-sample the same points.

3.1 Sub-sampling

Given a scan $\mathcal{F}$, we calculate an approximation of the mean curvature at each point p in this way. We fix a neighborhood radius r and consider all points within the sphere of radius r and centre p. We perform PCA analysis on the neighborhood points to find the principal directions and so the curvature value and the normal vector at the central point. Using the same neighborhood radius, we select the points of local maximum, and amongst them we take the 14 points of highest curvature (in norm). We found that this number of points was optimal to extract stable features like the nose tip and the eyes. In fact, to guarantee the presence of stable features, for each face we extract the maxima of curvature at different scales, by repeating the previous process 4 times using increasing neighborhood radii. As a result we have 4 sets of characteristic points on which to base the registration of the faces.

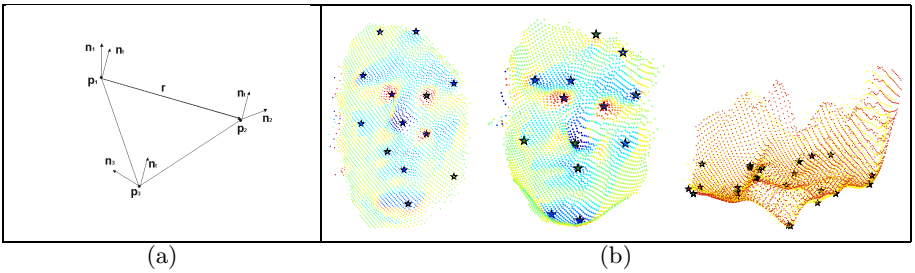
3.2 Signature Matching and Validation

For each set of characteristic points we consider all triplets, then for each triplet we determine the 9 invariants as in 2.2 and finally we collect them in a signature. At the end of the process we will have 4 signatures sets for each face. Given two faces $\mathcal{F}_1$ and $\mathcal{F}_2$, for each scale s_i, $i = 1, \ldots, 4$, we compare their signatures as in 2.3. For each s_i, we will have S_{Ii} matching pairs of signature points. As a final score we take the maximum of the S_{Ii}. In theory, this score should be enough for recognition purposes. In practice, however, the signatures are generated starting from a maximum of 14 points on the face scan, which leads to a heavy sub-sampling of the signature and so to the loss of certainty of properties that in the general theory follow from continuity and smoothness. Specifically, with no clues about proximity of matching pairs in the signature space, it might well happen that quite a few number of invariants are close enough to indicate a match, but this might just be the result of multiple transversal intersections of the two signatures. If our measure of signatures similarity solely consists on counting the number of matches over the total number of signature points we might be fooled by transversal matches. To prevent this, we need to validate the matches. Fortunately this almost comes for free: a match in the invariant space corresponds to a pair of matching triplets $\mathcal{T}_1 \subset \mathcal{F}_1$, $\mathcal{T}_2 \subset \mathcal{F}_2$. We find the roto-translation that takes $\mathcal{T}_2$ onto $\mathcal{T}_1$ and we apply it to all point of $\mathcal{F}_2$. To validate the match, we measure the "closeness" of the transformed scan to the other in this way. We start by looking for each point $q_i \in \mathcal{F}_2$ the closest point p_i in $\mathcal{F}_1$ and we save the Euclidean distance $d_i = \|q_i - p_i\|$ between the two. We get a set of distances $D = \{d_i\}_{i \in I}$ where $I = |F_2|$. However, since the acquisition viewpoint may change, even in the case of an accurate registration, some points might belong to only one of the scans and their distances from the closest points in the other scan could be relatively large. We therefore experimented two metrics. The first used a point to normal distance and is an approximation of volumetric distance: for each point $q_i \in \mathcal{F}_2$ we consider the normal line through it; if there is a point $p_i \in \mathcal{F}_1$ close enough to the line (where by close enough we mean comparable to the acquisition resolution), then we save their Euclidean distance d_i, otherwise q_i is ignored. In this way we remove outliers and we can evaluate the "closeness" of the transformed scan as the mean or median of distances (of remaining points). This measure proved to be reliable but time consuming for a practical experiment on a large number of scans. Also there are issues about the threshold on the point to normal distance since the face sampling might not be uniform (and it certainly is not if the cloud is the output of a single scan). A quicker but still reasonably robust alternative estimator of "closeness" of scans proved to be the median of distances computed over all the points of the clouds. This second metric has been used in extensive experiments.

4 Experimental Evaluation

We chose to test the proposed framework on the 3D_RMA database [9] which, despite being generally noisy, comprises scans of subjects taken from different

viewpoints and with varying facial expressions. The database contains a total of 106 subjects whose faces where scanned using a structured light acquisition system during two different acquisition campaigns, each of 3 scans. In total there are 617 scans each of which contains on average 4×10^3 points. We divided the dataset into a training set $G = \{S_1, \ldots, S_{106}\}$, where each S_i consists of the first 3 scans of subject i, and a test set $P = \{V_1, \ldots, V_{106}\}$, where each V_i contains the remaining scans of subject i acquired in the second campaign. For all scans in the training set, we extracted the points of maximum curvature without any pre-processing on the points and calculated the signatures at 4 different scales (see 2.2). Then, for $j = 1, \ldots, 106$, we considered all test and train subjects subsets pairs $G_j = \{S_1, \ldots, S_j\}$, $P_j = \{T_1, \ldots, T_j\}$ and proceeded as follows: one test scan at a time from P_j was compared to all scans in G_j. The training scan that achieves maximum score after signatures matching, registration and validation through the median distance (see 3.2) has been taken as a match. If the matching test and train scans belong to the same subject the match is validated.



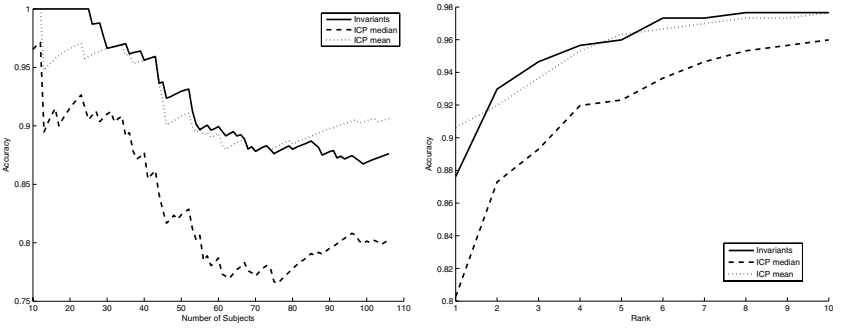
(a) (b)

Fig. 1. (a): Triplet with associated normals. (b): Characteristic points extraction and scans registration.

The first two images in figure 1(b) show the curvature variation in two scans of the same subject. The extracted points of local maximum absolute curvature values are represented by the stars. The third image is the result of the registration of the two scans; the profile view emphasises the accuracy of the registration.

To compare our results we set up the same experiment framework using Besl and McKay [4] original version of the ICP algorithm for registration and matching. This version of the ICP algorithm minimises point to point Euclidean distances (under a fixed threshold) through successive iterations.

To prevent ICP from converging to a local minima, a pre-registration has been provided by manual selection of nose tips. The final matching score is taken to be the mean or the median distance between corresponding points after the last iteration. The results of all tests are illustrated in figure 2(a). The figure should be read in this way: for each horizontal coordinate $j = 10, 11, \ldots, 106$, the vertical one expresses the accuracy (normalized to 1) of the recognition test



(a) (b)

Fig. 2. (a) Matching scores of the different techniques as a function of database cardinality. (b) Cumulative match characteristic curves for Invariants and ICP methods.

carried out on the pair of subsets $G_j = \{S_1, \ldots, S_j\}$ and $P_j = \{T_1, \ldots, T_j\}$, namely the data subset obtained by considering the first j subjects.

The invariants method clearly outperforms the ICP (both with median and mean final error measurements) for $10 \leq j < 30$. At $j = 30$ the accuracy of both methods is 96.7%. For $j > 30$ the invariant method performance gracefully degrades to the advantage of ICP. At $j = 106$ the Invariant method score is 87.7% against the ICP score of 90.6%. The reason for this is due to the nature of the 3D_RMA database: in the first of the two acquisition campaigns and for the first 30 subjects only, the matching of corresponding points detected by the projected light pattern was manually aided, leading to low noise point coordinates. From subject 31 onwards, the correspondences were found automatically [9]. Since the extraction of curvatures suffers from noisy input data, the registration through invariants might lose in accuracy, leading to wrong matches. In fact, when we performed the recognition test on the test subset $P_{30} = \{T_1, \ldots, T_{30}\}$ paired with the whole train set $G = \{S_1, \ldots, S_{106}\}$, the matching score was exactly the same for both methods, indicating that the presence of impostors $\{S_{31}, \ldots S_{106}\}$ does not affect the result. This did not hold true for $j > 30$.

In figure 2(b), we can see the Cumulative Match Characteristic (CMC) curves for the Invariant and ICP methods applied to the whole database (on all 106 subjects). It represents the classification accuracy of the three methods at different ranks. Rank k means that a correct classification is assigned if the correct label is found within the first k best matches. From the figure we can observe that ICP performs better at rank 1, whereas from rank 2 onwards the Invariant method is almost always ahead.

The performance of ICP when the median is used to measure the final registration error instead of the mean (see figures 2 is also worthy of remark: we expected similar results with the use of the median, but surprisingly we had a substantial performance drop. Since the median and the mean are evaluated on the same set of matching closest points, the only way to explain the difference is that the distances were not normally distributed within the interval limited by zero and the ICP threshold value.

In the joint invariants test, to validate a registration, we did not set a threshold to discard closest points whose distances were quite big compared to acquisition resolution. This is because our purpose was not only to obtain a measure of closeness of the registered scans, but also to validate the registration, and so we would have had to set two thresholds, one on the distance and the other on the number of closest points under that distance. To avoid setting two thresholds we used the median on the set of distances $D = \{d_i\}_{i \in I}$ as explained in 3.2. In the light of the ICP results, however, we could improve the matching score by using a different metric to validate the registration. This is supported by some experiments carried out using the volumetric (point to normal) distance defined in 3.2, in which we have improvements on false subjects matching. Alas, due to time limits we could not get results on the whole database thus far.

The Joint invariant and ICP algorithms together with the tests were all implemented in MatLab and carried out on AMD Opteron of 2 GB of RAM and 2.5 GHz CPU speed. The joint invariant method took on average 2 seconds to extract the maximum curvature points, generate the signatures, and validate the matches of a pair of scans. In order to speed-up the procedure, the extraction of the maximum curvature points and the generation the signatures of the training set can been performed off line. This took about 45 minutes in total, and reduced the average pairwise matching time to 1.5 seconds. The ICP proved to be slower due to the iteration process, and took on average 4 seconds per pairwise matching.

5 Conclusions and Future Work

The main original contribution of this paper is the introduction of a novel approach for 3D face recognition, based on the sound mathematical framework of Moving Frames. In this context, a single signature for a cloud of 3D points is generated using joint differential invariants. Even if this concept of signature is defined for continuous surfaces, here it has been adapted to the considered discrete set of points, projecting a 3D shape in a 9-dimensional space where the effect of rotation and translation is removed.

In the paper we have presented an efficient procedure for the generation of an invariant signature in 9-dimensional space, suitable to be employed for registration-based matching. Experimental evaluation over the 3D_RMA database showed that the proposed method performance is in line with the well known ICP-based matching approach and outperforms it in the case of low noise input data. It should be noted that, contrary to the ICP, the proposed method does not require pre-registration and can work in case of limited overlap between surface scans.

Performance improvements, in terms of computational speed and robustness, are foreseeable with a more robust extraction of the points of maximum curvature. This is especially true in the case of noisy input data. Also, it is reasonable to think that pre-processing the data, e.g. cropping it to remove spikes due to hair or acquisition artefacts, would positively affect the results. Additional work will also be devoted to the implementation of a more sophisticated metric to

validate the registration; in particular, an implementation of the volumetric distance described in 3.1 will be evaluated against the whole database. Tests will be run on other databases, in order to further evaluate the performance of the method in the presence of noise, facial expressions etc.

Finally, the ability to automatically capture and segment rigid parts of a face is expected to be a main outcome of this research effort in the short time.

References

1. Bowyer, K., Chang, K., Flynn, P.: A survey of approaches and challenges in 3d and multi-modal 2D+3D face recognition. Computer Vision and Image Understanding 101(1), 1–15 (2006)
2. Chang, K., Bowyer, K., Flynn, P.: An evaluation of multimodal 2d+3d face biometrics. IEEE Trans. Pattern Anal. Mach. Intell. 27(4), 619–624 (2005)
3. Gokberk, B., Akarun, L.: Comparative analysis of decision-level fusion algorithms for 3d face recognition. In: Proc. of Int. Conf. on Pattern Recognition, pp. 1018–1021 (2006)
4. Besl, P., McKay, N.: A method for registration of 3D shapes. IEEE Trans. on Pattern Analysis and Machine Intelligence 14, 239–256 (1992)
5. Chang, K., Bowyer, K., Flynn, P.: Adaptive rigid miltiregion selection for handling expression variations in 3D face recognition. In: IEEE Workshop on FRGC (2005)
6. Irfanoglu, M., Gokberk, B., Akarun, L.: 3D shape-based face recognition using automatically registered facial surfaces. In: Proc. of Int. Conf. on Pattern Recognition, pp. 183–186 (2004)
7. Lu, X., Jain, A.: Deformation analysis for 3d face matching. In: Proc. Int. Workshop on Applications of Computer Vision, pp. 99–104 (2005)
8. Lu, X., Jain, A., Colbry, D.: Matching 2.5D Face Scans to 3D Models. IEEE Trans. on Pattern Analysis and Machine Intelligence 28, 31–43 (2006)
9. Beumier, C., Acheroy, M.: Automatic 3d face authentication. Image and Vision Computing 18(4), 315–321 (2000)
10. Olver, P.J.: Joint Invariants Signatures. Found. Comput. Math. 1, 3–67 (2001)
11. Olver, P.J.: A survey of moving frames. In: Li, H., J. Olver, P., Sommer, G. (eds.) IWMM-GIAE 2004. LNCS, vol. 3519, pp. 105–138. Springer, Heidelberg (2005)
12. Cartan, É.: La Méthode du Repère Mobile, la Théorie des Groups Continue, et les Espaces Géneralisés. Hermann, Paris, France (1935)
13. Cadoni, M.I., Chimienti, A., Nerino, R.: Automatic Coarse Registration by Invariant Features. In: The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2006)
14. Do Carmo, M.: Differential Geometry of Curves and Surfaces. Prentice-Hall, Englewood Cliffs (1976)
15. de Berg, M., van Kreveld, M., Overmars, M., Schwarzkopf, O.: Computational Geometry, 2nd revised edn., pp. 99–105. Springer, Heidelberg (2000)