Automatic coarse registration of three-dimensional surfaces by information theoretic selection of salient points

Nikolaus Schön and Gerd Häusler

We describe a new method to register surface data measured by optical three-dimensional (3-D) sensors from various views of an object. With our method, complete 3-D models of objects can be generated without user interaction. Circumferential acquisition of 3-D objects is done by taking several views from different directions. To generate a complete 3-D-model, the views must be aligned with each other. This process is called registration and is commonly done interactively by searching for so-called corresponding points in the different views and by use of these points to calculate the appropriate rotation and translation. Our approach is based on automatically finding points that are eye catching or salient compared with other surface points. We derive a quantitative measure of point salience and a feature definition for free-form surfaces by introducing a concept to measure pragmatic information. Experiments confirm that our salient points can be robustly located on general free-form surfaces, even if there are no corners or edges. Furthermore, the neighborhoods of the salient points are highly distinguishable from each other. This results in a large reduction of the complexity of the subsequent geometric matching. The computing time is only a few seconds. We present results from various fields of application.

OCIS codes: 100.0100, 100.5010, 100.5760.

1. Introduction

Our aim is to contribute to the automatic generation of complete three-dimensional (3-D) object models. We use new information in a theoretical approach for the registration of 3-D sensor views taken from different directions. Optical 3-D measurement technology is used in a wide variety of applications. In many cases, complete surface models of objects are required. In art conservation[1] the restoration of cultural heritage should be simulated on a computer to reduce the risk of damaging the original; originals are to be copied by computer numerically controlled, or stereolithography, machines. Dental technicians use 3-D models of teeth to create inlays. In industrial inspection[2] as well as in facial surgery,[3] one is interested in deviations between nominal and actual shapes. Other applications are model based image analysis[4] and virtual reality, e.g., for computer games.[5]

The generation of 3-D surface models is a complex process that consists of several sophisticated methods and algorithms.[6–8] The essential steps are as follows:

(1) A number of partially overlapping range images are acquired from several views of an object by use of an optical 3-D sensor.[9–11]
(2) The surface views are brought into the best possible alignment. This process is called surface registration, which differs from volume registration that is used especially for medical applications; see, e.g., Ref. 12.
(3) The registered views are integrated to one single, topologically closed surface description.[13]
(4) The amount of data and surface noise is reduced by specific surface modeling algorithms.[13]
(5) The color texture of the surface is reconstructed from color images of the object.[14]

Here, we concentrate on the second item, which is a necessary precondition for all the following steps. Registration is generally split into two steps: coarse registration and fine registration. Fine registration presupposes that two given data sets are already approximately aligned, and it minimizes iteratively the

N. Schön (nikolaus.schoen@optik.physik.uni-erlangen.de) and G. Häusler (gerd.haeusler@optik.physik.uni-erlangen.de) are with the Max-Planck Research Group, Institute of Optics, Information, and Photonics, Friedrich-Alexander-University of Erlangen-Nuremberg, Staudtstrasse 7/B2, D-91058 Erlangen, Germany.

Received 17 January 2006; revised 1 March 2006; accepted 23 March 2006; posted 28 March 2006 (Doc. ID 67272).
0003-6935/06/256539-12$15.00/0

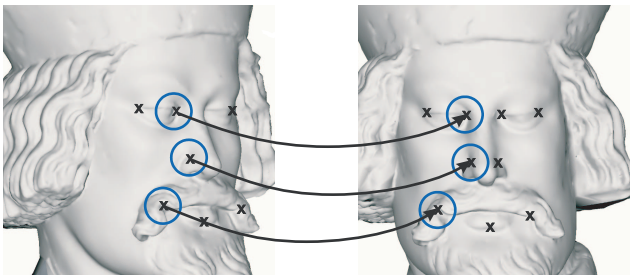

Fig. 1. (Color online) Two surface views with salient points for the human visual system (marked by an x). Three corresponding point pairs (circles) are assigned.

residual sum of distances between closest points on two surfaces.[15] We regard this problem as being solved; robust and efficient methods are known[16] and a theory exists to predict the achievable accuracy of fine registration.[17]

Much more difficult with regard to methodology is the initial coarse registration step. In contrast with fine registration, the correspondence of points on two surfaces is not given by their spatial closeness, but it must be established by their identification. The human visual system would solve this task by first looking for locations that are somehow eye catching or salient. In practice, two images of the surfaces are displayed on a computer screen. The user interactively searches both views for salient points that he can easily assign to their corresponding points by looking at the local surface structure around them (Fig. 1) and clicking on the corresponding points in the two images.

Our aim is to avoid user interaction and to achieve automatic registration. As in the interactive procedure, the most important step of the coarse registration method proposed in the following is to locate robust salient points on 3-D surfaces. Interesting questions arise: what is a salient point within the context of the registration problem? Can we give a quantitative measure of salience? Our answer is based on a general concept of information.

The basic idea is to look for locations with a high content of relevant information. Which information is relevant always depends on the context of the specific problem. In our case the context is the localization of salient points on free-form surfaces. We formulate this as a problem of geometric distinction. As a measure for distinguishability we propose an information measure. In this way, we demonstrate how the measurement of problem-relevant pragmatic information[18] can lead back to a Shannon-like entropy measure.[19] A coarse registration algorithm basically consists of the following four steps:

(1) Selection of salient points. Salient points are those selected from a small neighborhood that has a high and locally maximal content of relevant information.

(2) Reduction to similar point features. Invariant geometric properties are extracted from neighborhoods of the selected points and encoded as point features. The number of possible point correspondences is reduced to pairs of points with similar features.

(3) Geometric matching. The remaining ambiguities of point correspondences are reduced by geometric matching, i.e., finding geometrically consistent groups of points in different surface data sets.

(4) Rotation and translation. From the group of point correspondences the relative translation and rotation parameters between two surface data sets are computed and applied to align the surfaces.

The actual assignment of point correspondences is done by geometric matching. This step is computationally feasible only because the complexity of the problem is strongly reduced by the preceding steps. It is therefore important that the selection of salient points in two views leads to a small number of largely the same points and that these points form geometrically characteristic constellations. Additionally, to reduce the number of possible point correspondences by feature similarity, the point features must be highly distinguishable.

In Section 3 we explain the information theoretical considerations to select salient points on surfaces. Section 4 describes in detail the construction and comparison of local surface features. For estimation of the parameters relevant for surface analysis, see Section 5. Section 6 gives a discussion of the geometric matching step. Finally, we report on some applications of our method in Section 7.

2. Related Research

Coarse registration appears as an important step in many surface reconstruction approaches, and therefore has received more and more attention in recent years. There are several approaches that identify corresponding points on overlapping surfaces and compute the alignment transformation from the correspondences. These registration methods are called landmark based, following the terminology used in the survey of medical image registration given by Maintz and Viergever,[20] who also provide additional references on this topic. An alternative approach was described by Winkelbach *et al.*,[21] which is based on a random sample consensus (RANSAC) scheme. RANSAC relies mainly on exploiting geometric constraints that are determined by the shape of the data sets. Local surface features (curvatures) are used only supplementally.

In our definition of landmarks or salient points, we refer to a rather general concept of information. An overview of the history and the present status of the information concept is given by Capurro.[22] Gernert[18] proposes the concept of pragmatic information as a unifying concept that includes aspects of representation (syntax), meaning (semantics), and effect (pragmatics) of information. We follow this view in our considerations. However, in contrast to Gernert's graphic theoretical approach, we demonstrate how pragmatic information can be measured by entropy.

To encode the local surface data as feature vectors, we used a histogram method. This is not a new technique; similar approaches have already been used by Siggelkow and Burkhardt[23] for texture classification and by Johnson and Hebert[24] for surface registration. A comprehensive overview of other surface representations, especially for free-form object recognition, is given by Campbell and Flynn.[25] Surface analysis is typically affected by several parameters, e.g., the size of the analyzed region, the number of feature components, and the threshold for the feature similarity measure. The optimal configuration of these parameters depends strongly on surface type. Therefore we estimate the parameters automatically by optimization.

The final geometric matching of feature points is an adaptation and modification of the concept of matching attributed relational descriptions. The basic mathematical formalism, together with several algorithms, is given by Shapiro and Haralick.[26] Our geometric matching algorithm extends their concept to the matching of objects that show varying degrees of similarity rather than being just equal or not equal. Additionally, we impose an order on the possible point correspondences to use a best-first strategy for matching.

3. Selection of Salient Points

The most important step of the coarse registration algorithm is the reduction of the given point sets to a small number of salient points, because computation time and reliability of the subsequent steps highly depend on the robustness and effectiveness of this reduction. On the one hand we are searching for points that are salient on their own, i.e., their salience can be measured locally and in each 3-D view independently. On the other hand, salience is defined only in relation to the context of data. Therefore, data from a small region are considered for analysis. Since it is impossible to say in general what a salient point is, we turn the question around and ask "when is a point not salient"? We regard this question as equivalent to the problem of how to distinguish a point from its neighborhood. We begin our analysis with geometric considerations.

A. Geometric Considerations

Whether a point is salient depends on the geometry of its neighborhood. There are special types of surface on which points are not salient for symmetry reasons. In each point $\mathbf{x}$ on a free-form surface we can define a local coordinate system with axes $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$, such that $\mathbf{e}_z$ is identical to the normal $\mathbf{n}$ in $\mathbf{x}$. Then a description of the surface can be given relative to this coordinate system. This description will differ if we move to another surface point $\mathbf{x}'$ with normal $\mathbf{n}'$. The two points $\mathbf{x}$ and $\mathbf{x}'$ can therefore be distinguished by their associated surface descriptions.

There are, however, special cases of surfaces for which the surface description is identical in all local coordinate systems. These symmetrical cases are useless with respect to the localization of points. Figure 2 displays three such cases. On spherical, cylindrical, and planar surfaces the local coordinate system can be moved along two degrees of freedom without changing the surface description. In 3-D space, there are no additional types of surface that have this property. The constancy of the principal curvatures in all points of the surface is a necessary and sufficient condition that is shown to be valid exactly for the three mentioned cases (for proof, see, e.g., Putland[27]). A point lying on such a symmetric surface is not distinguishable from its neighborhood. Its pragmatic value and therefore its salience is zero. Now we formulate an information measure that reflects this fact.

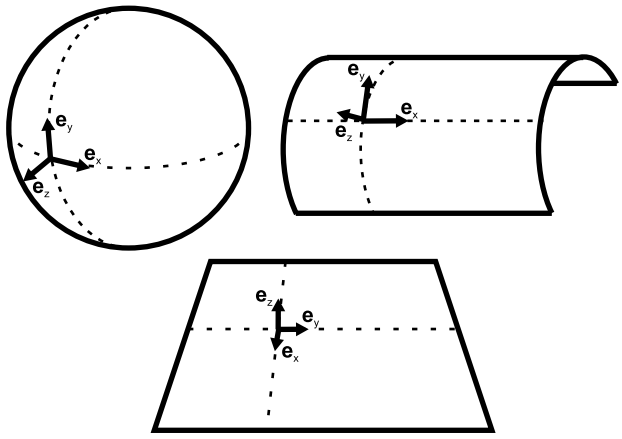

Fig. 2. Three surfaces on which points are not localized. The local frames ($\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$) can be moved along two independent directions (dotted curves) without changing the surface description with reference to the frame.

B. Information Theoretic Abstraction

From the considerations above we believe that a quantitative salience value has to be a measure of geometric distinguishability. This will now lead us to an information measure in the form of an entropy.

Classical information theory[19] provides a way to express the information content of data quantitatively by considering only statistical properties of information units (characters, data values, etc.). An information source X that can generate N different information units $\{x_1, x_2, \ldots, x_N\}$ would deliver an expected information content (entropy) of

$$H(X) = -\sum_{i=1}^{N} p(x_i)\log_2[p(x_i)] \qquad (1)$$

per information unit, where $p(x_i)$ is the probability of the occurrence of x_i. However, Shannon's information theory does not distinguish between relevant and meaningless data; it measures only syntactic information and separates it from semantic and pragmatic information.[22] It is presupposed that the data are given in a form that immediately represents relevant information. Generally, data are not given in such a form. The data have to be interpreted, i.e., transformed in such a way that Eq. (1) becomes a measure for the content of pragmatic information of the data. Boundary conditions for the formulation of

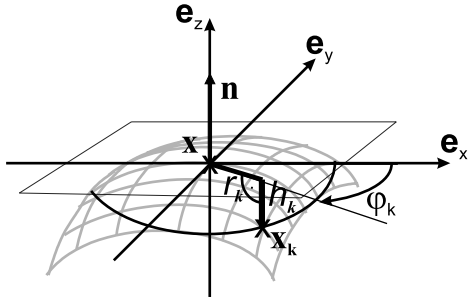

Fig. 3. Local coordinate system ($\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$) on a surface point $\mathbf{x}$ with normal $\mathbf{n}$. The neighborhood of $\mathbf{x}$, bounded by the intersecting line between the surface and the bounding cylinder, is adumbrated. Each point $\mathbf{x}_k$ in the neighborhood has local cylindrical coordinates of r_k, φ_k, h_k.

this transformation $X \to X'$ are that $H(X')$ has to be zero for the useless cases. Within the context of our problem, these are the symmetric surfaces. To quantize pragmatic information, our basic idea is to construct models that represent exactly the useless cases. When the given data are described by the parameters of such a model, the set of these parameters is totally redundant, i.e., their entropy equals zero. We note that $H(X')$ does not measure the actual information that we obtain from the data. Instead, it is an upper limit of potential information under the pragmatic aspect. For the difference between actual and potential information, refer to Weizsäcker.[28]

C. Method

We now propose such a transformation $X \to X'$ of the surface data. We will show that the entropy of the transformed data is actually a measure of salience. Local maxima of that salience on the surface define those salient points that are selected for coarse registration.

Let us assume a local frame in a surface point $\mathbf{x}$ that is defined up to an angle of rotation around the normal $\mathbf{n}$. Equivalently, one can state that we have a local cylindrical coordinate system with its origin in $\mathbf{x}$, where each point $\mathbf{x}_k$ in the neighborhood of $\mathbf{x}$ has the description $\mathbf{x}_k = (r_k, \varphi_k, h_k)$ and where the radius coordinate r_k and the height coordinate h_k are given, but the angular coordinate φ_k is unknown (see Fig. 3). Therefore, we use only r_k and h_k for surface analysis. We restrict the analysis of the information content to a small neighborhood around point $\mathbf{x}$, bound by a cylinder around surface normal $\mathbf{n}$ in $\mathbf{x}$ that is defined by

$$r_k \leq r_{\max}. \tag{2}$$

The size of the chosen neighborhood, defined by $r_{\max}$, is critical. If it is too small, the regions may contain not enough discriminative information; if it is too large, it is likely that the analyzed region is not complete in both of the two considered views. Additionally, the analysis of larger regions requires more computation time. Extraction of rotational symmetric patches makes the extracted neighborhood and therefore the calculated salience invariant against object rotation, which is a necessary condition.

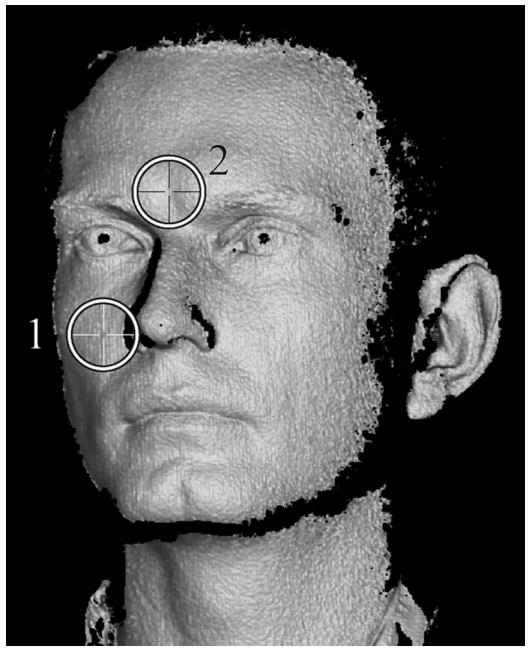

Fig. 4. Rendered range image of a face; see the Fig. 5 caption for a detailed explanation.

For each point $\mathbf{x}_k$ in the neighborhood of $\mathbf{x}$, we define a spherical surface that contains $\mathbf{x}_k$ and $\mathbf{x}$ and has the same normal $\mathbf{n}$ in $\mathbf{x}$. The value κ_k of the sphere's curvature κ can be computed from the local coordinates

$$r_k = \left\{\|x_k - x\|^2 - [\mathbf{n}(\mathbf{x}_k - \mathbf{x})]^2\right\}^{1/2},$$
$$h_k = \mathbf{n}(x_k - x) \tag{3}$$

by

$$\kappa_k = \frac{2h_k}{{r_k}^2 + {h_k}^2}. \tag{4}$$

The probability density $p(\kappa)$ is approximated by counting its relative frequencies in a histogram. The interval $\kappa \in [-\kappa_{\max}, \kappa_{\max}]$ is discretized in steps of $\Delta\hat{\kappa}$ to a number $\hat{N}_\kappa = 2\lfloor \kappa_{\max}/\Delta\hat{\kappa} \rfloor + 1$ of discrete values $\hat{\kappa}^{(i)}$, $i = 1, \ldots, \hat{N}_\kappa$. The (normalized) entries of the histogram are $p_\kappa(\hat{\kappa})$. Then the local entropy is

$$H_{\text{spherical}}(\hat{\kappa}) = -\sum_{i=1}^{\hat{N}} p\left[\hat{\kappa}^{(i)}\right] \log_2\left\{p\left[\hat{\kappa}^{(i)}\right]\right\}. \tag{5}$$

Computing this salience value for each measured point of a range image results in a salience image. Figure 4 displays a rendered range image of a face; Fig. 5 is the corresponding salience map. For any spherical surface, values κ_k are completely redundant, i.e., $H_{\text{spherical}}(\hat{\kappa}) = 0$. This includes the case of a planar surface ($\kappa = 0$).

In a salience image we look for local maxima, i.e., for points with a higher salience value than their

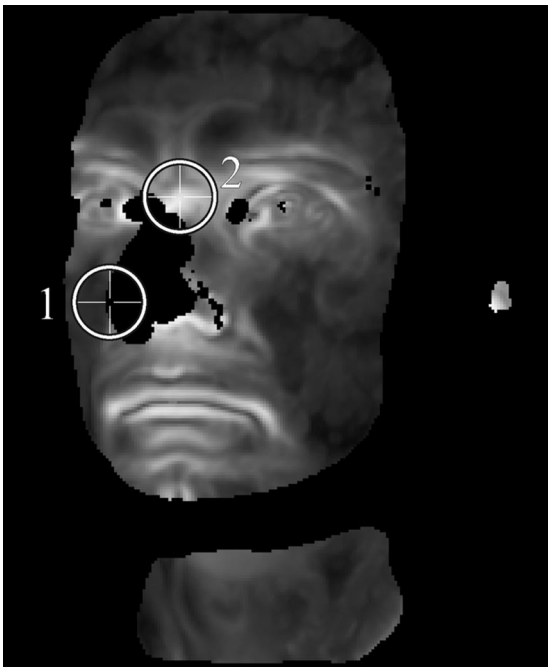

Fig. 5. Salience image of the 3-D view shown in Fig. 4. The salience values are displayed as intensity values. The size of the neighborhoods of two exemplary surface points is marked by circles. For point 1, for example, no salience value is computed. In neighborhood 1 in Fig. 4, points are missing on the right because of shadowing during the measurement. This effect explains why there is much less surface area covered by the salience image than by the range image. The nose, e.g., drops out almost completely. At point 2 there is a local maximum of the salience because the surface in its neighborhood is highly structured.

adjacent points in the image raster. From this set of local salience maxima, only the number N_{salient} of points with the highest salience values is selected. We can find these most salient points on any type of surface. The method works even with surfaces that contain no corners, edges, or any other sharp structures. $N_{\text{sal}} = 100$ appeared to be an appropriate compromise between the computation requirements of the matching process (Section 6) and the probability of getting points in the overlap region of the measurements.

The transformation in Eq. (4) already relates our information measure to planar and spherical reference surfaces. Cylindrical surfaces could basically be handled in the same manner but by a two-dimensional entropy. However, we chose a more simple method. On a cylindrical surface, the neighborhoods of adjacent points are identical and therefore result in the same $H_{\text{spherical}}$ values. To inhibit the selection of points with cylindrical neighborhoods, only those local salience maxima are selected for which the quotient of its salience values s_x and the salience values s_{x_l} of its N_{adj} adjacent points $\mathbf{x}_l$ exceed a certain threshold d_s:

$$\frac{s_x}{s_{x_l}} > d_s = 1.05, \quad \forall l \in \{1, \ldots, N_{\text{adj}}\}, \tag{6}$$

Fig. 6. Rendering of the complete surface model of a horse sculpture by Franz Marc. Photo courtesy of the Staatliche Galerie Moritzburg Museum, Halle (Saale), Germany.

i.e., s_x must be at least 5% higher than all adjacent s_{x_l}. The chosen value of d_s is motivated by experimental experience.

The remaining parameters of the selection process, $r_{\max}$, $\Delta\hat{\kappa}$, and $\kappa_{\max}$, are optimized by a criterium that quantifies their usefulness for the assignment of corresponding points. This is discussed in Section 5.

D. Evaluation

We verify experimentally that our method of point selection is applicable to free-form surfaces and that it selects largely the same points, independent of the viewing direction during acquisition. Figure 6 displays the reconstructed surface of a work of art by Franz Mark. The Museum Halle was interested in a comparison of the shapes of this bronze sculpture and a wax cast. In the example we focus on regions with smooth structures that barely show any corners or edges. The two salience images in Fig. 7 reveal that even surface areas that are only slightly structured lead to a robust localization of salient points. Additional experiments verified that the locations of the selected points correspond to high reliability. To estimate the reliability, we used pairs of previously registered surfaces and selected our salient points. Then for each salient point on one surface we computed the distance Δx_{csp} to the closest salient point on the other surface. If actual corresponding points are selected, the distance of closest salient points must be in the range of the mean sampling distance Δx of the measured data. We computed the histogram $p(\hat{\xi})$ with the discretized value $\hat{\xi}$ of the quotient.

$$\xi = \frac{\Delta x_{\text{csp}}}{\overline{\Delta x}}. \tag{7}$$

Figure 8 displays the histogram for a typical case. The diagram indicates that the selected points most

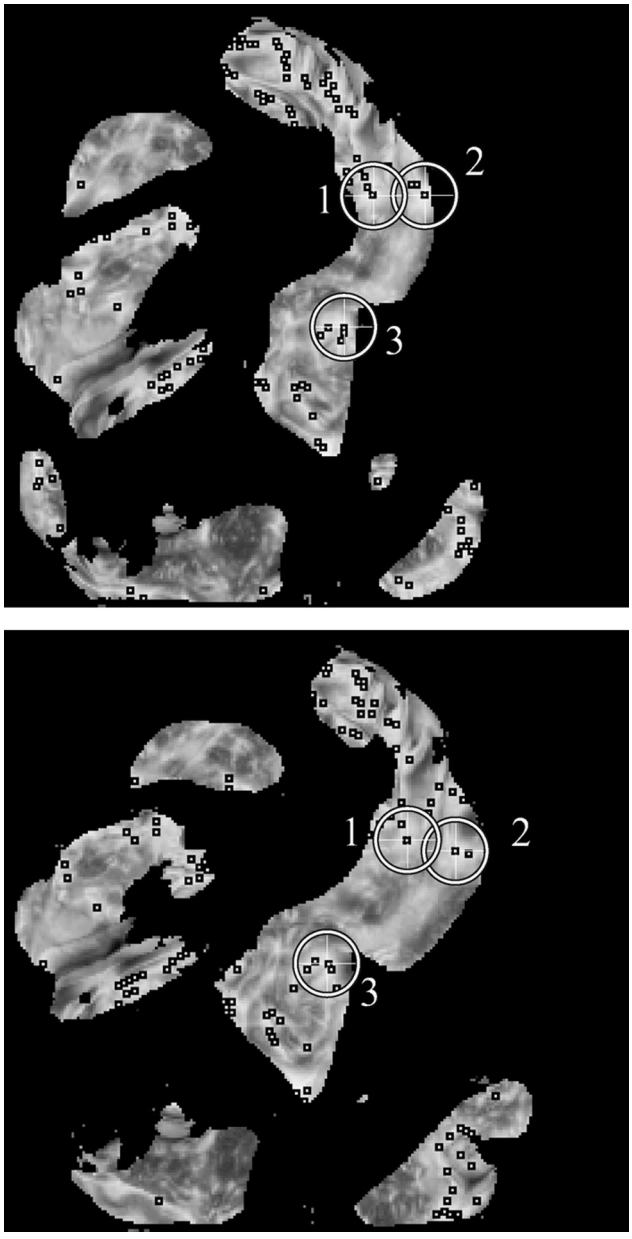

Fig. 7. Salience images of two different range images from regions of the object in Fig. 6 that are only slightly structured. The salience maxima are marked by black squares; some of the corresponding points by circles and numbers.

probably have distances between 0.5 and 1.0 times the mean sampling distance of the data sets, i.e., actual corresponding points are selected.

4. Feature Assignment

We have reduced the number of 250.000 points/view to 100 salient points, but we still do not know how they correspond to each other. Testing all possible correspondences by geometric matching is not feasible (see Section 6). Therefore we must reduce the number of possible correspondences per salient point in comparison with their characteristic feature vectors. These are extracted from the same local point

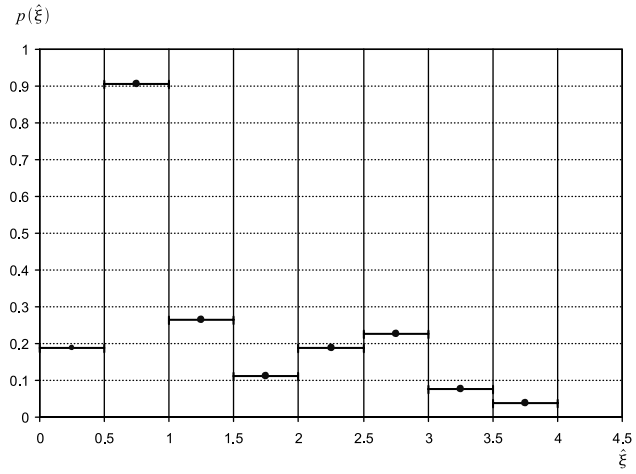

Fig. 8. Normalized frequencies $p[\hat{\xi}^{(i)}]$ of discretized normalized distances $\hat{\xi}^{(i)}$ of closest salient points in the overlapping regions of two data sets. The discretization intervals are $\Delta\hat{\xi} = 0.5$. The histogram is normalized to $\sum_{i=1}^{\infty} p[\hat{\xi}^{(i)}]\Delta\hat{\xi} = 1$. By far the most frequent distances are those that are smaller than the mean sampling distance, i.e., $\hat{\xi} < 1$.

neighborhoods that are already used for salience analysis.

A. Construction of Feature Vectors

As components of the feature vector of a surface point $\mathbf{x}$, we use the bin entries of a two-dimensional histogram $p_{\hat{u},\hat{v}}$ (histogram feature), with axes u and v. The histogram estimates the probability density of local invariants computed from the neighborhood of $\mathbf{x}$. For each point $\mathbf{x}_k$ in the neighborhood of $\mathbf{x}$, two invariants r_k and κ_k are computed, according to Eqs. (3) and (4). The values r_k and κ_k are mapped on the axes of the histogram according to

$$u_k = r_k^2, \tag{8}$$

$$v_k = \kappa_k. \tag{9}$$

Using two invariants per neighborhood point, instead of only one, increases the amount of information available for feature encoding. To compute the histogram, u is descretized as $\hat{u}$ in $\hat{N}_u$ steps of size $\Delta\hat{u}$ to discrete values $\hat{u}^{(i)}$, $i = 1, \ldots, \hat{N}_u$. Analogously, v is discretized as $\hat{v}$ in $\hat{N}_v$ steps of size $\Delta\hat{v}$ to values $\hat{v}^{(j)}$, $j = 1, \ldots, N_v$.

The transformation of the r coordinate according to Eq. (8) ensures that discretization of $u \in [0, r_{\max}^2[$ in equally spaced intervals corresponds to partition of space in equal volumes (cylindrical hulls). Therefore it can be reasonably assumed that the expected content of relevant information in each interval $[\hat{u}^{(i)}, \hat{u}^{(i+1)}[$ is equal. The discretization of the v values is identical to the discretization of the κ values, as described in Subsection 3.C, therefore $v \in [-\kappa_{\max}, \kappa_{\max}[$, $v^{(j)} = \kappa^{(j)}$, and $\hat{N}_v = \hat{N}_\kappa$.

To avoid interference between the sampling of the 3-D surface data and the discretization intervals of the histogram computation, we use fuzzy histograms,

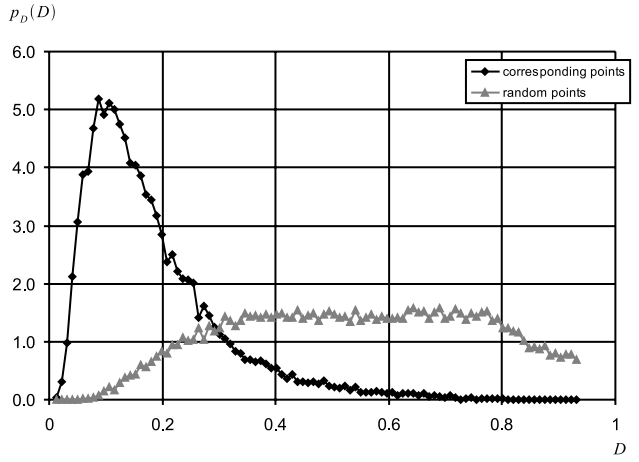

Fig. 9. Normalized frequencies p_D of discrete feature distances $D^{(i)}$ with $\sum_i p_D[D^{(i)}]\Delta D = 1$ computed for pairs of corresponding and randomly selected points.

as described by Siggelkow and Burkhardt[29] within the context of image retrieval, which results in fuzzy increments of the histogram cells that are real numbers instead of integers.

The local histograms are also independent of the varying local surface resolution in different views, which is achieved if we include the area of surface elements in the histogram computation. For each point, the increment in the local histogram $\mathbf{x}_k$ is not equal to 1, but to a point coverage value, which is calculated as the total area of surface elements that are connected to point $\mathbf{x}_k$. In this way, in regions with higher point density the bigger number of points is compensated for by smaller histogram increments.

B. Similarity Measure

To compare two histogram features $p_{\hat{u},\hat{v}}{}^{(1)}(\hat{u},\hat{v})$ and $p_{\hat{u},\hat{v}}{}^{(2)}(\hat{u},\hat{v})$, it is reasonable to use methods for comparing probability distributions. As feature distances $D[p_{\hat{u},\hat{v}}{}^{(1)}, p_{\hat{u},\hat{v}}{}^{(2)}]$ we examined the Kullback–Leibler distance (see Ref. 30) and the statistically well-founded χ^2 distance. We found that the robustness of feature assignment does not significantly depend on which of these feature distances is used. Furthermore, we obtained equally good results with a simple distance function, the (squared) Euclidean distance

$$D\big[p_{\hat{u},\hat{v}}{}^{(1)},\, p_{\hat{u},\hat{v}}{}^{(2)}\big] = \sum_{i=1}^{\hat{N}_u}\sum_{j=1}^{\hat{N}_v}\big\{p_{\hat{u},\hat{v}}{}^{(2)}\big[\hat{u}^{(i)},\, \hat{v}^{(j)}\big] - p_{\hat{u},\hat{v}}{}^{(1)}\big[\hat{u}^{(i)},\, \hat{v}^{(j)}\big]\big\}^2. \quad (10)$$

This feature distance has two advantages over the Kullback–Leibler distance and χ^2 distance: it is faster to compute and it is applicable to hierarchical search structures. The latter property is necessary for the extension of our method from a pairwise coarse registration to a registration among a large number of views. Here, similar histogram features must be found from the large set of histogram features of all the object views. The application of hierarchical search structures (for an overview see Gaede and Günther[31]) reduces the complexity and significantly increases the speed of similarity search.

A similarity measure is useful if it actually separates pairs of corresponding points from other, non-corresponding point pairs. The diagram in Fig. 9 demonstrates that, for pairs of corresponding points, small feature distance values are much more frequent than for randomly selected point pairs. For each N_{sal} salient point from the first data set, a number N_{sim} of matching candidates with a small feature distance is stored. In this way we obtain a number $N_P = N_{\text{sal}}N_{\text{sim}}$ of pairs that remain to be checked by geometric matching. It is essential for reliability of the point assignment that the threshold for the maximum allowable feature distance $D_{\max}$ is properly set to determine if a pair of features is similar or different.

5. Estimation of Analysis Parameters

In the previous sections, various parameters appeared that influence the results of the salience and feature analysis. We determine the values of these object-dependent parameters automatically by optimization. For this purpose, we give an adequate measure of performance that is maximized. The most direct way to optimize the parameters of a method is to maximize success. This is compliant with the objective to maximize pragmatic information. For the problem of correct assignment of points, success is quantified by recognition rate, i.e., the probability that a correct decision is made between the alternatives, points correspond and points do not correspond. This decision is to be made on the basis of similarity of the feature vectors of the points. If feature distance D is smaller than a certain threshold $D_{\max}$, $D \le D_{\max}$, the compared features are classified as similar, otherwise as nonsimilar or different.

For surface views that are already registered, the correct correspondence of points could be tested by their spatial closeness. Spatially close means that the spatial distance of two points is smaller than a certain threshold $\Delta x_{\max}$, which we set to the mean sampling distance of the surfaces, $\Delta x_{\max} := \overline{\Delta x}$. The problem is that we are dealing with unregistered views. To overcome this we assume that the histogram features of direct neighbor points on one surface are on average as different as those of two corresponding points on two surfaces. We randomly sample a number of points together with their direct neighbors on the surface (altogether approximately 1000 points) and build a set of randomly selected pairs of these points (10,000 pairs).

This set of sample point pairs can now be divided in to four subsets:

(1) A number N_{CS} of pairs of points that are spatially close and have similar feature vectors.

(2) A number $N_{\overline{C}S}$ of pairs of points are not spatially close but have similar feature vectors.

(3) A number $N_{\overline{C}\,\overline{S}}$ of pairs of points that are not spatially close and are classified as nonsimilar.

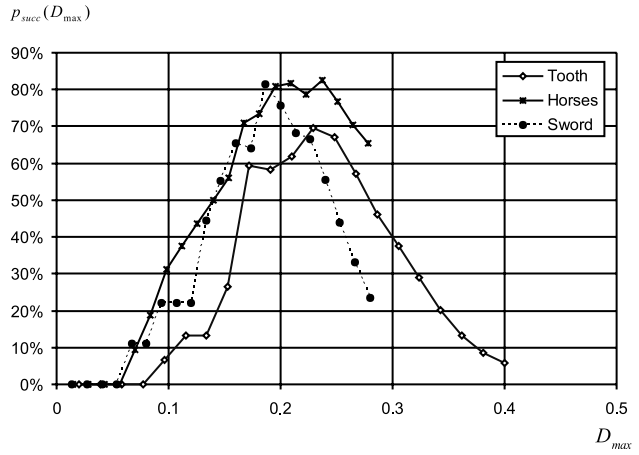

Fig. 10. Feature recognition rate versus maximum allowed feature distance for four objects.

(4) A number $N_{C\bar{S}}$ of pairs of points that are spatially close but are classified as nonsimilar.

Correct decisions are only cases (1) and (3). Our objective is to maximize probability p_{true} of correct decisions, which is calculated from the probability p_{CS} that a pair of spatially close points is classified as similar, and the probability

$p_{\bar{C}\bar{S}}$ that a pair of spatially far points is classified as nonsimilar:

$$p_{\text{succ}} = p_{CS}\, p_{\bar{C}\bar{S}}, \tag{11}$$

If we estimate p_{CS} and $p_{\bar{C}\bar{S}}$ by relative frequencies (N_C is the total number of spatially close pairs, $N_{\bar{C}}$ is the total number of spatially nonclose pairs), we obtain

$$p_{\text{succ}} = \frac{N_{CS}}{N_C}\,\frac{N_{\bar{C}\bar{S}}}{N_{\bar{C}}}, \tag{12}$$

where p_{succ} depends on configuration $\mathbf{P}$ of all the analysis parameters that are not computed analytically, $\mathbf{P} = (r_{\max}, \Delta\hat{\kappa}, \kappa_{\max}, D_{\max})$. This defines a four-dimensional parameter space in which we determine the optimal configuration $\mathbf{P}_{\text{opt}}$ by maximization of p_{succ}:

$$\mathbf{P}_{\text{opt}} = \underset{\mathbf{P}}{\operatorname{argmax}}\, p_{\text{succ}}(\mathbf{P}). \tag{13}$$

Fortunately, the optimization can be carried out by four one-dimensional parameter optimization steps, starting with a good initial configuration that we found by trial by using some of our test objects.

Figure 10 exemplifies that p_{succ} strongly depends on the correct choice of $D_{\max}$. If $D_{\max}$ is too small, it is likelier that no corresponding point pairs would be found. If $D_{\max}$ is too big, too many false point assignments will be made, which results in a significant increase of computation time for geometric matching. Figure 10 also suggests that the optimal value of $D_{\max}$ is not too sensitive to the type of object, a fact that enhances the robustness of the method.

6. Geometric Matching

Robust coarse registration requires the correct assignment (matching) of at least three point pairs from different views. The resulting rotation and translation parameters are estimated by a least-squares method and will therefore be more accurate when more assignments can be made. Point matching must take into account that the surface views overlap only partially, i.e., there are model points with no corresponding scene point as well as scene points with no corresponding model point.

We found the point assignment by searching for the largest geometrically consistent group of point correspondences (largest consistency group) among the N_P possible point correspondences remaining after feature assignment. This means that each pairwise distance between any two points in one view matches the distance between their corresponding points in another view.

The algorithm to construct these groups can be explained as follows: Consider a first pair of points $[\mathbf{x}_{\mathcal{A}}^{(1)}, \mathbf{x}_{\mathcal{B}}^{(1)}]$, with point $\mathbf{x}_{\mathcal{A}}^{(1)}$ from surface view $\mathcal{A}$ and point $\mathbf{x}_{\mathcal{B}}^{(1)}$ from view $\mathcal{B}$. The pair $[x_{\mathcal{A}}^{(1)}, x_{\mathcal{B}}^{(1)}]$ is added to the consistency group as the first element. Then consider a second pair of points, $[x_{\mathcal{A}}^{(2)}, x_{\mathcal{B}}^{(2)}]$. These two pairs are geometrically consistent, if the distance $d_{\mathcal{A}} = \|\mathbf{x}_{\mathcal{A}}^{(1)} - \mathbf{x}_{\mathcal{A}}^{(2)}\|$ between the points in data set $\mathcal{A}$ is (nearly) equal to the distance $d_{\mathcal{B}} = \|\mathbf{x}_{\mathcal{B}}^{(1)} - \mathbf{x}_{\mathcal{B}}^{(2)}\|$ in data set $\mathcal{B}$, i.e., if $|d_{\mathcal{A}} - d_{\mathcal{B}}| \le \overline{\Delta x}$ (with the mean sampling distance $\overline{\Delta x}$). If the two pairs are geometrically consistent, $[\mathbf{x}_{\mathcal{A}}^{(2)}, \mathbf{x}_{\mathcal{B}}^{(2)}]$ is added to the consistency group. If such a geometrically consistent pair is found in the set of N_P point pairs, a third pair of points is searched that is geometrically consistent with the two pairs already in the group, and so on. Group growth is continued until no more point pairs can be added.

This method can succeed only if the first pair in the group is actually a correct correspondence. Since this cannot be guaranteed, the group growth has to be repeated for each pair $(\mathbf{x}_A, \mathbf{x}_B)$ as the first. The matching process is formulated as a best-first search. We define as the best of the point correspondences those that (a) have a small feature distance and (b) consist of points with a high salience value. Therefore the set of N_P possible point pairs is sorted lexicographically, with the feature distance $D(\mathbf{x}_{\mathcal{A}}, \mathbf{x}_{\mathcal{B}})$ as the first criterium and with the salience value $s_x(\mathbf{x}_{\mathcal{A}})$ as the second. The largest group of point correspondences obtained in this way is chosen as the correct assignment of points. From their coordinates, the rotation and translation parameters can be computed by use of a least-squares method reported by Horn.[32]

The complexity of the matching process is reduced to $\mathcal{O}[(N_P)^2] = \mathcal{O}[(N_{\text{sal}}N_{\text{sim}})^2]$. Because of the robustness and selectivity of our method of point selection and because of the discriminacy of our histogram features, N_{sal} and N_{sim} are small numbers. With

Table 1. Properties, Parameters, Processing Times, and Registration Accuracy of Four Objects[a]

Property/Result	Face	Tooth	Horses	Sword
Original number of points per view (mean)	151400	61379	112004	83316
Number of points used per view	30000	22500	22500	22500
$\overline{\Delta x}$	1.10 mm	117 μm	1.07 mm	1.02 mm
r_{max}	15.03 mm	890 μm	6.00 mm	8.12 mm
k_{max}	0.0665 mm^{-1}	1.12×10^{-3} μm^{-1}	0.105 mm^{-1}	0.123 mm^{-1}
D_{max}	0.14	0.20	0.15	0.14
Histogram size ($\hat{N}_u \times \hat{N}_v$)	8×8	5×5	5×5	8×8
Preprocessing time per view	3.7 s	1.5 s	2.7 s	2.1 s
Time for registration (without preprocessing)	80 s	27 s	18 s	56 s
Deviation ($\overline{\Delta x_{cp}}$)	10 mm	50 μm	0.35 mm	0.32 mm

[a]($\overline{\Delta x_{cp}}$) is the mean distance of all pairs of closest surface points in the overlap of the surfaces. Preprocessing time means the time to smooth the surfaces and downsample them to the number of points used per view. Time for registration without preprocessing means the time used for salient point selection, feature assignment, and geometric matching.

$N_{sal} = 100$ and $N_{sim} = 5$, the evaluation of group consistency has to be executed maximally 250,000 times. With our Pentium IV, 2.4 GHz test system the matching step for all the test objects was executed in less than 1 s. Most of the execution time for registration is spent on the determination of salient points,

Fig. 11. Top: salience images of preoperatively (left) and intraoperatively (right) acquired face data. The salient points are indicated by black squares; the corresponding points are emphasized by circles. Bottom: 3-D renderings of preoperative (left), intraoperative (middle), and registered (right) range images displayed in different shades of gray.

Fig. 12. Salience images and 3-D renderings of two views of a tooth; see the Fig. 11 caption for details.

which is necessary. Otherwise the time for geometric matching could not be performed in $\mathcal{O}[(N_P)^2]$ time but would increase exponentially with the number of points in the data sets.

One could think about an interesting alternative to feature based assignment followed by geometric matching. The application of a RANSAC-based scheme as in Winkelbach *et al.*[21] on the sets of salient points promises that the assignment procedure would become much more effective and could handle a larger number of points. This reduces the requirements on the robustness and uniqueness of the selected points and could allow for a faster point selection step.

7. Results and Applications

Our coarse registration method is usable in a large variety of applications. From cooperation with various projects concerned with 3-D surface reconstruction we achieved typical data sets that we used to test our method.[33] Table 1 shows some of the properties, analysis parameters, and registration times for four different objects. The execution times are comparable with those of Winkelbach *et al.*[21] who performed coarse registration with mean execution times between 4.85 and 64.94 s for noisy data.

The accuracy of our method is greater, the greater the resolution of the downsampled data sets. Primarily, accuracy depends on the type of surface and how many salient points it contains, since the rigid transformation is computed in a least-squares manner. This makes it difficult to compare it with the results of other authors.

The medical project of Benz *et al.*[3] aimed at correcting displacements of the eyeball in patients' faces. The special problem here is that preoperatively modeled nominal faces are to be registered to intraoperatively acquired faces that have been modified by a surgeon. In Ref. 34 the localization of the salient point in data sets of this kind was already reported. Now we have results from the complete registration process (see the face column in Table 1). This object took the longest time for registration because the second data set contributes a large environmental area of the actual object that must also be analyzed. The relatively high deviation of the registered data sets is caused mainly by deformations of the face, not by the registration method. Maier *et al.*[35] proposed a coarse registration algorithm specially adapted for these surgical data. At least for cases with minor modifications of the surface, however, our general method can provide an alternative (Fig. 11).

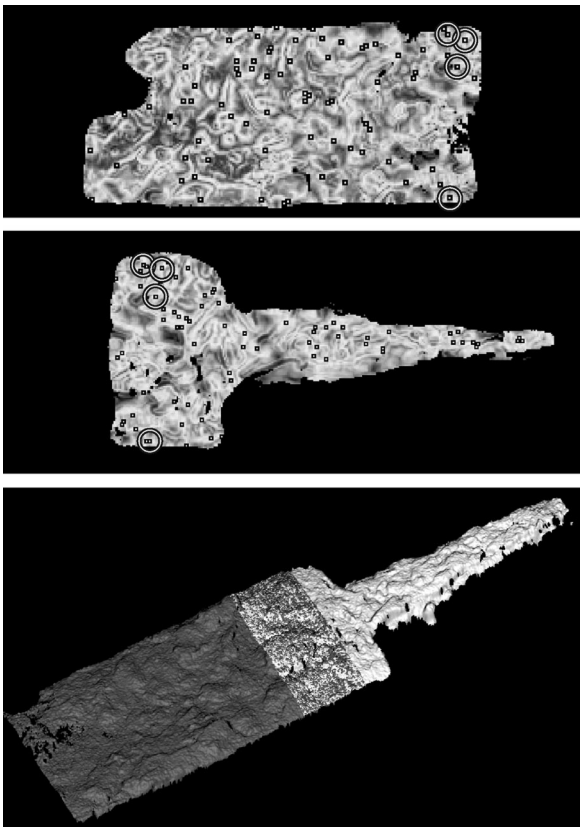

Fig. 13. Top and middle: salience images of two different views of a sword with the salient points indicated by black squares. The corresponding points are emphasized by white circles. Bottom: registered data sets displayed in different shades of gray.

Another application is for dental medicine. The surface of human teeth is reconstructed from optical measurements to model and manufacture inlays and crowns. We registered data provided to us by 3D-Shape GmbH[36] (see Fig. 12). The time required for the whole process is feasible for application in a dental laboratory (see the Tooth column in Table 1). The residual deviation is eventually reduced by subsequent fine registration. The accuracy reached by our method as an initial step is by far high enough to ensure the convergence of fine registration.

The method has also been tested with data from various projects with regard to art conservation. Figure 13 displays data sets of an ancient sword. It is an example that would be difficult to register for a human. The surface is jagged because of corrosion and therefore it is difficult to identify particular locations. For our information based selection, however, this is not a problem (see the Sword column in Table 1). Although the Sword object has too much structure for point identification by the human visual system, the surface of the Horses in Figs. 6 and 7 is too smooth. Both cases are reliably registered by our method.

8. Summary and Conclusions

We introduced a fast and reliable method for coarse registration of three-dimensional surfaces. Our method selects robust points of interest even from surfaces that have no corners or edges. We abstracted the concept of salience using the general concept of pragmatic information. We pointed out that spheres, cylinders, and planes are surfaces of low pragmatic information within the context of point localization on 3-D surfaces. Our quantification of salience is therefore based on the difference in these symmetric cases. We introduced a definition of highly discriminant local surface features by using histograms of transformed geometric invariants. We could strongly reduce the complexity of the final step of geometric matching by applying a best-first search, according to the salience of the points and the similarity of their features. We tested the method with data from applications in maxillofacial surgery, dental medicine, and art conservation. Our method can serve as an essential component of a general methodology of fully automatic and complete 3-D surface reconstruction.

This project was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) as subproject A1 of "Sonderforschungsbereich 603, Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten" (Special Research Sector 603, Model Based Analysis and Visualization of Complex Scenes and Sensor Data).[37] This paper is dedicated to Adolf Lohmann in celebration of his recent 80th birthday.

References

1. G. Guidi, J.-A. Beraldin, and C. Atzeni, "High-accuracy 3D modeling of cultural heritage: the digitizing of Donatello's Maddalena," IEEE Trans. Image Process. **13,** 370–380 (2004).
2. C. Brenner, J. Böhm, and J. Gühring, "Experimental measurement system for industrial inspection of 3D parts," in *Machine Vision Systems for Inspection and Metrology VII,* B. G. Batchelor, J. W. Miller, S. S. Solomon, eds., Proc. SPIE **3521,** 237–247 (1998).
3. M. Benz, X. Laboureux, T. Maier, E. Nkenke, S. Seeger, F. Neukam, and G. Häusler, "The symmetry of faces," in *Proceedings of Vision, Modeling, and Visualization,* G. Girod, H. Niemann, T. Ertl, B. Girod, and H.-P. Seidel, eds. (Akademische Verlagsgesellschaft, 2002), pp. 43–50.
4. V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," IEEE Trans. Pattern Anal. Mach. Intell. **25,** 1063–1074 (2003).
5. T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized reality: constructing virtual worlds from real scenes," IEEE Multimedia **4,** 34–47 (1997).
6. I. Söderkvist, "Introductory overview of surface reconstruction methods," Research Report 10 (Department of Mathematics, Lulea University, S-97187 Lulea, Sweden, 1999), www.sm.luth.se/~inge/publications/surfrec.ps.
7. S. Karbacher, N. Schön, H. Schönfeld, and G. Häusler, "Digitizing 3D objects for reverse engineering and virtual reality," in *Principles of 3D Image Analysis and Synthesis,* B. Girod, G. Greiner, and H. Niemann, eds. (Kluwer Academic, 2000), Chap. 8.1, pp. 336–347.
8. S. Karbacher, X. Laboureux, N. Schön, and G. Häusler, "Processing range data for reverse engineering and virtual reality," in *Proceedings of Third International Conference on 3-D Digital Imaging and Modeling* (Institute of Electrical and Electronics Engineers, 2001), pp. 314–321.
9. M. Gruber and G. Häusler, "Simple, robust and accurate phase-measuring triangulation," Optik **89,** 118–122 (1992).

10. K. Veit and G. Häusler, "Metrical calibration of a phase measuring triangulation sensor," in *Vision, Modeling, and Visualization 2000,* B. Girod, G. Greiner, H. Niemann, and H.-P. Seidel, eds. (IEEE Signal Processing Society, 2000), pp. 33–38.
11. C. Wagner and G. Häusler, "Information theoretical optimization for optical range sensors," Appl. Opt. **42,** 5418–5426 (2003).
12. P. Hastreiter, C. Rezk-Salama, G. Soza, M. Bauer, G. Greiner, R. Fahlbusch, O. Ganslandt, and C. Nimsky, "Strategies for brain shift evaluation," Med. Image Anal. **8,** 447–464 (2004).
13. S. Karbacher and G. Häusler, "New approach for the modeling and smoothing of scattered 3D data," in *Three Dimensional Image Capture and Applications,* R. N. Ellson and J. H. Nurre, eds., Proc. SPIE **3313,** 168–177 (1998).
14. N. Schön, P. Gall, and G. Häusler, "Three dimensional acquisition of colored objects," in Workshop Color Image Processing (German Society of Color Science and Application, 2002), pp. 63–70.
15. P. J. Besl and H. D. McKay, "A method for registration of 3-D shapes," IEEE Trans. Pattern Anal. Mach. Intell. **14,** 239–256 (1992).
16. N. Diehl and H. Burkhardt, "Motion estimation in image sequences," in *International Workshop on High Precision Navigation,* K. Linkwitz and U. Hangleiter, eds. (Springer, 1989).
17. X. Laboureux and G. Häusler, "Localization and registration of three-dimensional objects in space—where are the limits?" Appl. Opt. **40,** 5206–5216 (2001).
18. D. Gernert, "Pragmatic information as a unifying concept," in *Information. New Questions to a Multidisciplinary Concept,* K. Kornwachs and K. Jacoby, eds. (Akademie-Verlag, 1996), pp. 147–162.
19. C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27,** 379–423, 623–656 (1948).
20. J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," Med. Image Anal. **2,** 1–36 (1998).
21. S. Winkelbach, M. Rilk, C. Schönfelder, and F. M. Wahl, "Fast random sample matching of 3d fragments," in *Pattern Recognition (DAGM Symposium),* C. E. Rasmussen, H. H. Bülthoff, B. Schölkopf, and M. A. Giese, eds., Vol. 3175 of Lecture Notes in Computer Sciences, (Springer, 2004), pp. 129–136.
22. R. Capurro and B. Hjørland, "The concept of information," Annu. Rev. Inf. Sci. Technol. **37,** 343–411 (2003), http://www.capurro.de/infoconcept.html.
23. S. Siggelkow and H. Burkhardt, "Invariant feature histograms for texture classification," in *Joint Conference on Information Sciences (JCIS),* Vol. 4, pp. 230–233 (Research Triangle Park, North Carolina, 1998).
24. A. E. Johnson and M. Hebert, "Surface registration by matching oriented points," in *International Conference on Recent Advances in 3-D Digital Imaging and Modeling* (Institute of Electrical and Electronics Engineers, 1997), pp. 121–128.
25. R. J. Campbell and P. J. Flynn, "A survey of free-form object representation and recognition techniques," Comput. Vision Image Understand. **81,** 166–210 (2001).
26. L. G. Shapiro and R. M. Haralick, "Relational matching," Appl. Opt. **26,** 1845–1851 (1987).
27. G. R. Putland, "Modeling of horns and enclosures for loudspeakers," Ph.D. dissertation (Department of Electrical and Computer Engineering, University of Queensland, 1995), http://www.users.bigpond.com/putland/phd/thes.ps.gz.
28. C. F. v. Weizsäcker, *Aufbau der Physik* (Hanser, München, 1985), p. 171.
29. S. Siggelkow and H. Burkhardt, "Image retrieval based on local invariant features," in IASTED *International Conference on Signal and Image Processing* (International Association of Science and Technology for Development, 1998), pp. 369–373.
30. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, 1991), p. 18.
31. V. Gaede and O. Günther, "Multidimensional access methods," ACM (Assoc. Comput. Mach.) Comput. Surveys **30,** 170–231 (1998).
32. B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," J. Opt. Soc. Am. A **4,** 629–642 (1987).
33. N. Schön and G. Häusler, "Automatic coarse registration of 3D Surfaces," in *Vision, Modeling, and Visualization 2005,* G. Greiner, J. Hornegger, H. Niemann, and M. Stamminger, eds. (Akademische Verlagsgesellschaft, 2005), pp. 479–486.
34. N. Schön, B. M. Maier, E. Nkenke, F. W. Neukam, and G. Häusler, "Information-soptimierte Merkmale zur Grobregistrierung von Freiform-Flächen," in *Bildverarbeitung für die Medizin 2004,* T. Tolxdorff, J. Braun, H. Handels, A. Horsch, and H.-P. Meinzer, eds., Informatik Aktuell (Springer, 2004).
35. T. Maier, M. Benz, N. Schön, E. Nkenke, F. W. Neukam, F. Vogt, and G. Häusler, "Automatic coarse registration of 3D surface data in oral and maxillofacial surgery," in *Perspective in Image-Guided Surgery,* T. M. Buzug and T. C. Lueth, eds. (RheinAhrCampus Remagen, 2004), pp. 51–58.
36. 3D-Shape GmbH, Henkestrasse 91, D-91052 Erlangen, Germany; http://www.3d-shape.com.
37. "Sonderforschungsbereich 603, Modellbasierte Analyse und Visualisierung komplexer Szenen und Sensordaten," http://sfb-603.uni-erlangen.de.