Feature Extraction and Registration – An Overview

S. Seeger, X. Laboureux

Chair of Optics, University of Erlangen-Nuremberg, Staudstrasse 7/B2, 91058 Erlangen, Germany
Email: sns@undine.physik.uni-erlangen.de, xl@undine.physik.uni-erlangen.de

Abstract

The purpose of this paper is to present a survey of *rigid* registration (also called matching) methods applicable to surface descriptions. As features are often used for the registration task, standard feature extraction approaches are described in addition. In order to give the reader a framework for his present registration problem, this report divides the matching task into three major parts (feature extraction, similarity metrics and search strategies). In each of them the reader has to decide between several possibilities, whose relations are in particularly pointed out.

1 Introduction

Registration or equivalently *matching* is the process of bringing two data sets into best possible alignment. This is reached by determining the transformation that transforms *corresponding areas* or *points* into each other. The terms *best possible alignment* and *corresponding areas/points* of two data sets are intuitively quite easy to understand but need a precise mathematical definition for a computational approach (see 6). Since features are one possibility to define corresponding areas/points, we discuss in addition some feature extraction methods (see 4). Features can be defined as modified data formed from a collection of the original data set which might be combined in linear or non-linear ways [4].

In this section only registration methods applicable to surface descriptions are described. In order to illustrate the above definitions, before starting the discussion of registration methods, we present some typical examples of data sets, corresponding areas/points and transformations.

2 Examples of data sets to be registered

In the following examples we present three kinds of data structures:

- intensity pictures
  $I : (n_x, n_y) \mapsto I[n_x, n_y] \in \mathbb{R}$, e.g. from photo cameras,
- surface descriptions (range images)
  $F : (n_x, n_y) \mapsto (x[n_x, n_y], y[n_x, n_y], z[n_x, n_y])^T \in \mathbb{R}^3$, e.g. from a tactil or optical sensor,
- volume data
  $\rho : (n_x, n_y, n_z) \mapsto \rho[n_x, n_y, n_z] \in \mathbb{R}$, e.g. from medical 3-D scanners like CT (Computer Tomography), MR (Magnetic Resonance) or sonography (ultrasound) which show anatomical structure, or PET (Positron Emitting Tomography), SPECT (Single Photon Emission Computed Tomography) or MRS (Magnetic Resonance Spectroscopy) which show functional and metabolic activity [29].

Intensity images and volume data sets are introduced in addition, since the registration methods used in their areas may also be used for the registration of surfaces:

- Typical optical 3-D sensors supply in addition to a range image a pixel identical intensity image. In this way the registration of the two intensity pictures is already sufficient to match the underlying range images.
- Since iso-surfaces ($\rho(n_x, n_y, n_z) = \text{const}$) are often extracted to register volume data sets, the registration methods used can be directly applied to the surface matching task.

Examples

- The data sets to be registered could be two intensity images of an object taken from different viewpoints. Due to illumination variations between the images corresponding points do not have the same intensities which makes the matching process even more difficult. The searched transformation is a 3-D rotation and a translation also known as *rigid transformation*. To find this transformation at least seven corresponding points in the images have to be found. The related range image can be calculated from the given intensity images [52].
- The data sets are two or more range images of an object taken from *different* viewpoints. The searched transformation is a rigid one. With at least three corresponding points in the range images the transformation can be found. In this way a complete model of the object can be generated from several views [35].
- The data sets are two intensity images of *different* objects e.g. two different human faces. Corresponding areas may be manually defined by features like eyes, the mouth and the nose. The searched transformation depends on the application. For example, in a face recognition system it may be useful to find the best rigid transformation between the two given faces [33]. The magnitude of the difference decides whether the same face or different faces are presented in the images. On the other hand for morphing (the smooth transition from one data set into another) a *non* rigid transformation is searched which allows to locally match all parts of the object in the two images to each other [18].
- The data sets are two range images of an object taken from the same viewpoint but at *different* times so that the object may have changed its shape in the meantime. For example the object might be a human face before and after a dental operation [11]. Corresponding areas result from features that can be detected in both range images and have not changed in the meantime e.g. the eyes, the nose and/or the forehead. The searched transformation depends again on the application. For example, if the detection of the post operational swelling is to be visualized, it might be useful to find the rigid transformation that transforms the unchanged parts of the object into each other. With the help of this rigid transformation the difference volume of the two range images can be visualized. In other applications it might be useful to parameterize the time variations of the object by a non rigid transformation which allows to match all parts of the object in the two range images to each other [15, 40]. Such a non rigid transformation can be a (global or local) affine, projective or curved transformation [27], depending on the magnitude of difference between the data sets.
- The data sets are two volume data sets of the same object measured by the same device but at *different* times. With a rigid transformation variations of the object (e.g. a skull or a brain) can be determined (e.g. to study the evolution of a disease) [27].
- The data sets are two volume data sets of *different* objects measured by the same device (e.g. the heads of different patients). Corresponding areas seem obvious for a human being. The searched transformation is a non rigid one. The registration of different patients'

images could allow to contrast a healthy and a sick person [40].

- The data sets are two volume data sets of the same object measured by *different* devices, e.g. a CT and a PET scan of the human head. It may be difficult even for a non expert human being to define corresponding areas. The searched transformation is a rigid one (e.g. to improve the diagnosis by using multimodality data) [26].
- The data sets to be registered can be of different dimensions, e.g. it is possible to match an intensity image $I : (n_x, n_y) \mapsto I[n_x, n_y] \in \mathbb{R}$ with a range image $F : (n_x, n_y) \mapsto (x[n_x, n_y], y[n_x, n_y], z[n_x, n_y])^T \in \mathbb{R}^3$ of the same object [47]. The searched transformation is a rigid one.

The above examples can be classified according to two basic criteria:

1. Nature and domain of transformation
   - rigid (local, global)
   - non rigid (local, global)
2. Modalities involved
   - monomodal
   - multimodal

Now we come to the presentation of several registration methods.

3 Overview of registration methods applicable to geometry data

The task of determining the best spatial transformation for the registration of data sets can divided into four major components [5]:

- feature space
- search space
- similarity metric
- search strategy.

The choice of feature space determines what is matched. Since features can be independently found in each data set in a preprocessing step, the amount of data to be matched can thus be reduced. Some examples are:

- raw data (intensities in intensity images, 3-D points in range images, density values in volume data sets),
- attributes defined for all points: curvatures, principal frames, point signatures [15, 42, 10],
- special collections of points: edges, surfaces, crest lines [30, 7, 14, 40],
- salient point features: corners, line intersections, points of high curvature, extremal points [36, 42],
- statistical features: moment invariants, centroids, principal axes; (they refer to measures over a region that may be the outcome from a segmentation preprocessing step) [16],
- higher-level structural and syntactic descriptions [12, 51].

The search space is the class of transformations from which we want to find the optimal transformation to align the data sets (global/local, rigid/nonrigid). The similarity metric determines how matches are rated (e.g. sum of squared euclidian distances, normalized cross-correlation, mutual information). The search strategy describes how to find this optimal transformation and depends on the search space (e.g. ICP, Hough method (clustering), correlation, relaxation, prediction-verification, indexing schemes, tree + graph matching).

We begin the discussion of registration methods by presenting the extraction of more sophisticated features like principal curvatures, point signatures, principal frames, crest lines and

extremal points from surface descriptions. These features can be found in all kinds of free-form surfaces and therefore are more general than corners or edges. In the whole discussion of registration methods we restrict ourselves to the search space of global rigid transformations (3-D rotations + translations). Therefore we give an overview of representations of rotation with an emphasis on quaternions. Then we present some widely used similarity metrics. At the end we describe some search strategies, especially the ICP-algorithm and some of its extensions. The ICP algorithm needs no feature extraction and has become the standard for the precise registration of two point clouds.

4 Extracting features from surface descriptions

4.1 Principal Curvatures

Minimal and maximal curvatures (also called principal curvatures) are shift and rotation invariant local features of an object surface. Given a parametrical suface description $\boldsymbol{p}(u,v) = (x(u,v), y(u,v), z(u,v))^T$, the principal curvatures κ_i $(i = 1, 2)$ at $\boldsymbol{p}(u_0, v_0)$ can be calculated as the eigenvalues of the Weingarten map (also called shape operator) [13],

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} L & M \\ M & N \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \kappa_i \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \tag{1}$$

where E, F, G, L, M and N depend on first and second order partial derivatives to u and v at $\boldsymbol{p} = \boldsymbol{p}(u_0, v_0)$:

$$\begin{array}{lll} E = \boldsymbol{p}_u \cdot \boldsymbol{p}_u & , \quad F = \boldsymbol{p}_u \cdot \boldsymbol{p}_v & , \quad G = \boldsymbol{p}_v \cdot \boldsymbol{p}_v, \\ L = \boldsymbol{p}_{uu} \cdot \boldsymbol{n} & , \quad M = \boldsymbol{p}_{uv} \cdot \boldsymbol{n} & , \quad N = \boldsymbol{p}_{vv} \cdot \boldsymbol{n}. \end{array} \tag{2}$$

$\boldsymbol{n}$ is the normal at $\boldsymbol{p}(u_0, v_0)$,

$$\boldsymbol{n} = \frac{\boldsymbol{p}_u \times \boldsymbol{p}_v}{\|\boldsymbol{p}_u \times \boldsymbol{p}_v\|}. \tag{3}$$

It is straightforward to determine the curvatures κ_i from (1) as the roots of the characteristic polynomial [49],

$$\kappa^2 - \frac{NE - 2MF + LG}{EG - F^2}\kappa + \frac{LN - M^2}{EG - F^2} = 0. \tag{4}$$

4.2 Principal frames

With the help of the components of the eigenvectors α_i and β_i the directions of minimal and maximal curvatures can be determined as:

$$\boldsymbol{e}_i = \frac{\alpha_i \boldsymbol{p}_u + \beta_i \boldsymbol{p}_v}{\|\alpha_i \boldsymbol{p}_u + \beta_i \boldsymbol{p}_v\|}. \tag{5}$$

It can be shown that the unit vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are perpendicular ($\boldsymbol{e}_1 \cdot \boldsymbol{e}_2 = 0$). Since both $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ lie in the tangential plane (defined by $\boldsymbol{p}_u, \boldsymbol{p}_v$), they are perpendicular to the normal $\boldsymbol{n}$ at $\boldsymbol{p}(u_0, v_0)$. Therefore $\boldsymbol{e}_1, \boldsymbol{e}_2$ and $\boldsymbol{n}$ define a local orthogonal frame at $\boldsymbol{p}(u_0, v_0)$ also called the principal frame or trihedron. Note that the principal frame is not uniquely determined, since there is no way to choose between the frames $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{n})$ and $(-\boldsymbol{e}_1, -\boldsymbol{e}_2, \boldsymbol{n})$ [15].

4.3 Crest lines

One of the principal curvatures is maximal in *absolute* value: it is called in [42] the *largest curvature* $\kappa_{\max}$, in order not to be mistaken with the maximal curvature. The associated principal direction is $\boldsymbol{e}_{\max}$. Crest lines are the loci of the surface where the largest curvature $\kappa_{\max}$ is locally maximal (in absolute value) in the associated direction $\boldsymbol{e}_{\max}$. In [42] crest lines are extracted from volume data. Then a crest line is the intersection of an iso-surface $\rho(n_x, n_y, n_z) = \text{const}$ with the implicit surface $\nabla\kappa_{max} \cdot \boldsymbol{e}_{\max} = 0$ (i.e. change of $\kappa_{\max}$ in direction $\boldsymbol{e}_{\max}$ is zero, which implies that $\kappa_{\max}$ is extremal in direction $\boldsymbol{e}_{\max}$)[1].

4.4 Extremal points

For the definition of crest lines only the largest curvature with its associated principal direction is used. In the same way there are also extremal lines associated with the curvature with minimal absolute value. Extremal points are defined as the points of intersection of such extremal lines with crest lines. In [42] extremal points are extracted from volume data. Then an extremal point is the intersection of an iso-surface $\rho(n_x, n_y, n_z) = \text{const}$ with the implicit surfaces $\nabla\kappa_1 \cdot \boldsymbol{e}_1 = 0$ and $\nabla\kappa_2 \cdot \boldsymbol{e}_2 = 0$. Note that the extremal points are generally not the points of the extremal lines whose curvature is locally maximal. It is only stated here that there are 16 different types of extremal points that can be distinguished. In addition there are several geometric invariants associated with extremal points: the geometric invariants of the surface (principal curvatures), the geometric invariants of the extremal lines (curvature, torsion) and the geometric invariants corresponding to the relative position of the extremal lines with respect to the underlying surface [42].

In order to extract all these features, partial derivatives have to be calculated on the given data. If no parametrical surface description but only point clouds are given, in order to calculate partial derivatives usually a polynomial surface is locally approximated at each data point. In this way the partial derivatives are always proportional to one of the polynomial coefficients [24, 49].

4.5 Point signatures

Similar to principal curvatures this kind of rotation and translation invariant feature can be defined for each surface point but with the great advantage that no derivatives have to be calculated [10]: For a given surface point $\boldsymbol{p}$ a sphere of radius r, centered at $\boldsymbol{p}$, is placed. The intersection of the sphere with the object surface is a 3-D space curve $\mathcal{C}$, whose orientation can be defined by an orthonormal frame formed by a "normal" vector $\boldsymbol{n}_1$, a "reference" vector $\boldsymbol{n}_2$, and the cross-product of $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$. $\boldsymbol{n}_1$ is defined as the unit normal vector of a plane $\mathcal{P}$ fitted through the space curve $\mathcal{C}$. In the limit r tends to zero, $\boldsymbol{n}_1$ approximates the surface normal at the point $\boldsymbol{p}$. A new plane $\mathcal{P}'$ can be defined by translating the fitted plane $\mathcal{P}$ to the point $\boldsymbol{p}$ in a direction parallel to $\boldsymbol{n}_1$. As well, if r tends to zero, $\mathcal{P}'$ approximates the tangential plane. The perpendicular projection of $\mathcal{C}$ to $\mathcal{P}$ forms a new planar curve $\mathcal{C}'$. The distances of the points of $\mathcal{C}$ to the corresponding projected points of $\mathcal{C}'$ form a signed distance profile that is called the signature of the point $\boldsymbol{p}$ in [10]. The reference direction $\boldsymbol{n}_2$ is defined as the unit vector from $\boldsymbol{p}$ to the projected point on $\mathcal{C}'$ which gives the largest positive distance. Note that $\boldsymbol{n}_2$ is orthogonal to $\boldsymbol{n}_1$ since it lies on $\mathcal{P}'$.

[1] Of course crest lines can also be directly calculated on surface descriptions.

4.6 Extended Gaussian Image (EGI)

The EGI is another way to represent the data of a surface data set. In this approach [19] the normal vector at each point of the data set is computed and mapped into a unit sphere where its tail is at the center of the sphere and its head lies on the surface. In addition each point on the surface of the sphere (i.e. head of a normal vector) is weighted by the Gaussian curvature of the corresponding point of the surface data set. In this way the EGI can be considered as the weighted orientation histogram of the data set.

This representation of the data sets has two interesting properties for the pose estimation problem:

- it is translation invariant,
- the EGI rotates in the same way as the corresponding data set.

Therefore the problem of finding the transformation in the 6-dimensional parameter space of rotations and translations can be divided into two reduced problems:

1. Finding the right rotation with the help of the EGI.
2. Finding the right translation with the help of an additional method.

However such an EGI approach assumes that the mapping between a point in the data set and a point on the sphere is uniquely defined. It can be shown that this is only the case if the object is convex.

One approach in which the distances from each data point to a given origin are additionally saved in each correspondence point on the sphere is the Complex EGI [23]. Note that also an extension to the EGI which can deal with all classes of surfaces, called UNSDLA, has already been formulated [25].

5 Representations of rotation

There are many ways to represent rotation. Some examples are: Gibbs vector, Euler angles [43], Pauli spin matrices [43], axis and angle [43], Cayley-Klein parameters [17], orthonormal matrices [43] , quaternions [20] and dual number quaternions [54]. Since quaternions are widely used in the computer vision community we want to give a short introduction. Some more details can be found in [20]. In the context of registration they are often used for a closed form solution to the problem of minimizing the least-squares sum of corresponding points.

A quaternion $\dot{q}$ can be represented in the complex number notation

$$\dot{q} = q_0 + iq_x + jq_y + kq_z \tag{6}$$

with real part q_0 and three imaginary parts q_x, q_y, q_z. For the imaginary units i, j, k the following equations hold:

$$\begin{array}{lllllllll} i^2 & = & -1, & j^2 & = & -1 & k^2 & = & -1, \\ ij & = & k, & jk & = & i, & ki & = & j, \\ ji & = & -k, & kj & = & -i, & ik & = & -j. \end{array} \tag{7}$$

With (7) the multiplication of quaternions $\dot{r}$ and $\dot{q}$ can be defined in terms of the products of their components,

$$\begin{aligned} \dot{r}\dot{q} \;=\; & (r_0q_0 - r_xq_x - r_yq_y - r_zq_z) \\ & + i\,(r_0q_x + r_xq_0 + r_yq_z - r_zq_y) \\ & + j\,(r_0q_y - r_xq_z + r_yq_0 + r_zq_x) \\ & + k\,(r_0q_z + r_xq_y - r_yq_x + r_zq_0)\,. \end{aligned}$$

In general $\dot{r}\dot{q} \neq \dot{q}\dot{r}$.

The dot product of two quaternions is the sum of products of corresponding components:

$$\dot{p} \cdot \dot{q} = p_0 q_0 + p_x q_x + p_y q_y + p_z q_z. \tag{8}$$

The square of the magnitude of a quaternion is the dot product of the quaternion with itself:

$$\|\dot{q}\|^2 = \dot{q} \cdot \dot{q}. \tag{9}$$

A unit quaternion is a quaternion whose magnitude equals 1.

The conjugate of a quaternion negates its imaginary parts:

$$\dot{q}^* = q_0 - iq_x - jq_y - kq_z. \tag{10}$$

Vectors can be represented by purely imaginary quaternions. If $\boldsymbol{r} = (x, y, z)^T$, we can use the quaternion

$$\dot{r} = 0 + ix + jy + kz. \tag{11}$$

Scalars can be similary represented by using real quaternions.

Using the fact that only rotations preserve dot products and cross products, we can represent a rotation by a quaternion if we can find a way of mapping purely imaginary quaternions (that represent vectors) into purely imaginary quaternions in such a way that dot and cross products are preserved. It can be shown that the composite product

$$\dot{r}' = \dot{q}\dot{r}\dot{q}^*, \tag{12}$$

where $\dot{q}$ is a unit quaternion, transforms the imaginary quaternion $\dot{r}$ into an imaginary quaternion $\dot{r}'$ and preserves the dot and cross products between $\dot{r}$ and a second imaginary quaternion $\dot{r}_2$. Since

$$(-\dot{q})\,\dot{r}\,(-\dot{q}^*) = \dot{q}\dot{r}\dot{q}^* \tag{13}$$

$-\dot{q}$ represents the same rotation as $\dot{q}$.

It is straightforward to verify that the composition of rotations corresponds to multiplication of quaternions:

$$\begin{aligned} \dot{r}'' &= \dot{p}\dot{r}'\dot{p}^* \\ &= \dot{p}\,(\dot{q}\dot{r}\dot{q}^*)\,\dot{p}^* \\ &\overset{\cdots}{=} (\dot{p}\dot{q})\,\dot{r}\,(\dot{p}\dot{q})^*. \end{aligned}$$

The overall rotation is represented by the unit quaternion $\dot{p}\dot{q}$.

It may be of interest to note that it takes fewer arithmetic operations to multiply two quaternions than it does to multiply two 3×3 matrices. Also, since calculations are not carried out with infinite precision on a computer the product of many orthonormal matrices may no longer be orthonormal, just as the product of many unit quaternions may no longer be a unit quaternion. However it is trivial to find the nearest unit quaternion, whereas it is quite difficult to find the nearest orthonormal matrix.

Unit quaternions are closely related to the geometrically intuitive axis and angle notation. A rotation by an angle θ about the axis defined by the unit vector $\boldsymbol{e} = (e_x, e_y, e_z)^T$ can be represented by the unit quaternion

$$\dot{q} = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}(ie_x + je_y + ke_z)\,. \tag{14}$$

The relation of a unit quaternion $\dot{q}$ to the familiar orthnormal rotation matrix $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \begin{pmatrix} \left(q_0^2 + q_x^2 - q_y^2 - q_z^2\right) & 2\left(q_x q_y - q_0 q_z\right) & 2\left(q_x q_z + q_0 q_y\right) \\ 2\left(q_y q_x + q_0 q_z\right) & \left(q_0^2 - q_x^2 + q_y^2 - q_z^2\right) & 2\left(q_y q_z - q_0 q_x\right) \\ 2\left(q_z q_x - q_0 q_y\right) & 2\left(q_z q_y + q_0 q_x\right) & \left(q_0^2 - q_x^2 - q_y^2 + q_z^2\right) \end{pmatrix}. \tag{15}$$

Finally we want to note that a 3-D rigid motion (rotation + translation) can also be represented by a special dual quaternion. A dual quaternion or dual number quaternion $\hat{q}$ consist of two quaternions $\dot{q}$ and $\dot{s}$ so that

$$\hat{q} = \dot{q} + \epsilon \dot{s}, \tag{16}$$

where a special multiplication rule for ϵ is defined by $\epsilon^2 = 0$. In order to represent a 3-D rigid motion the following two constraints have to be satisfied:

$$\dot{q} \cdot \dot{q} = 1 \qquad \text{and} \qquad \dot{q} \cdot \dot{s} = 0. \tag{17}$$

6 Typical similarity metrics

The problem of aligning two data sets could be generally defined in the following way: Given two data sets $u(\boldsymbol{x}')$ and $v(\boldsymbol{x})$ describing parts of the same object at $\boldsymbol{x}'$ in the "u-frame" respectively $\boldsymbol{x}$ in the "v-frame", i.e.

$$v(\boldsymbol{x}) = F\left(u(\boldsymbol{x}')\right) \tag{18}$$

where F is the transfer function from u to v, we want to find the pose transformation T from $\boldsymbol{x}$ to $\boldsymbol{x}'$, i.e.

$$\boldsymbol{x}' = T(\boldsymbol{x}). \tag{19}$$

Combining (18) and (19) this is equivalent to resolve

$$v(\boldsymbol{x}) = F\left(u\left(T(\boldsymbol{x})\right)\right) \qquad \forall \boldsymbol{x}. \tag{20}$$

In general it is difficult to determine the transfer function F.

6.1 Correlation

However, if the effects of F can be neglected, e.g. u, v are intensity images from different views by negligible illumination variations between u and v, we get

$$v(\boldsymbol{x}) = u\left(T(\boldsymbol{x})\right) \qquad \forall \boldsymbol{x}. \tag{21}$$

Due to noise, different occlusions and partial overlapping a solution T valid for all $\boldsymbol{x}$ cannot be found. Therefore it is a common way to search for the transformation T that minimizes

$$E(T) = \sum_{\boldsymbol{x}} \left[v(\boldsymbol{x}) - u\left(T(\boldsymbol{x})\right)\right]^2 \tag{22}$$

$$= \sum_{\boldsymbol{x}} \left[v(\boldsymbol{x})\right]^2 - \sum_{\boldsymbol{x}} 2v(\boldsymbol{x})u\left(T(\boldsymbol{x})\right) + \sum_{\boldsymbol{x}} \left[u\left(T(\boldsymbol{x})\right)\right]^2. \tag{23}$$

Such a function that determines the ideal model parameters as the arguments that maximizes or minimizes the function is also often called a cost function or objective function. Since the first term in (23) is independent of T, minimizing $E(T)$ is equivalent to maximizing

$$C(T) = \frac{\sum_{\boldsymbol{x}} v(\boldsymbol{x})u\left(T(\boldsymbol{x})\right)}{\sum_{\boldsymbol{x}} \left[u\left(T(\boldsymbol{x})\right)\right]^2}, \tag{24}$$

called the normalized cross-correlation function [5]. A related measure, which is advantageous when an absolute measure is needed, is the correlation coefficient

$$\tilde{C}(T) = \frac{\sum_{\boldsymbol{x}} \left[v(\boldsymbol{x}) - \mu_v\right] \left[u\left(T(\boldsymbol{x})\right) - \mu_u\right]}{\sqrt{\sum_{\boldsymbol{x}} \left[v(\boldsymbol{x}) - \mu_v\right]^2 \sum_{\boldsymbol{x}} \left[u\left(T(\boldsymbol{x})\right) - \mu_u\right]^2}} \tag{25}$$

where μ_u and μ_v are the mean values of u and v. The denominator is given by the product of the standard deviations of u and v.

6.2 Mutual information

In the case where the effects of F must be taken into account, e.g. if u represents the normals at each point of a 3-D scan and v the corresponding intensity image, more sophisticated methods have to be applied. u and v can be interpreted as random variables with probability distributions P_u and P_v. Intuitively if u and v are well aligned the randomness of u given knowledge of v is maximally reduced. In statistics this intuition can be formalized as follows.

Firstly the randomness of a random variable X is measured by its entropy, defined by

$$H\left(X\right) \equiv -E_X\left\{\log\left(P(X)\right)\right\}. \tag{26}$$

Thereby $E_Z\left\{Z\right\}$ is the expected value of random variable Z and $P(X)$ the probability distribution of X. Secondly the randomness of random variable Y given knowledge of random variable X is measured by the conditional entropy

$$H(Y|X) \equiv -E_X\left\{E_Y\left\{\log P(Y|X)\right\}\right\}. \tag{27}$$

Then the intuition stated above can be formulated as maximizing

$$I\left(u(\boldsymbol{x}), v\left(T(\boldsymbol{x})\right)\right) = H\left(u(\boldsymbol{x})\right) - H\left(u(\boldsymbol{x})|v\left(T(\boldsymbol{x})\right)\right), \tag{28}$$

called the mutual information of u and v [46, 47].

6.3 Least-squares sum of corresponding points

In the special case where the data sets are two point clouds $\{\boldsymbol{p}_i\}$ and $\{\boldsymbol{p}'_i\}$ of an object measured by a 3-D sensor from two different viewpoints we describe the standard similarity metric in more details. For every pair of corresponding points $\boldsymbol{p}_i$ and $\boldsymbol{p}'_i$ we want to find the rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$ so that

$$\boldsymbol{p}_i = \boldsymbol{R}\boldsymbol{p}'_i + \boldsymbol{t} \tag{29}$$

with $\boldsymbol{p} = (x, y, z)^T$. For convenience we write (29) as

$$\boldsymbol{p}_i = T\left(\boldsymbol{p}'_i\right) \tag{30}$$

with the transformation T defined by

$$T\left(\boldsymbol{z}\right) = \boldsymbol{R}\boldsymbol{z} + \boldsymbol{t}. \tag{31}$$

The transformation T has six free parameters (e.g. three angles, three translation parameters). With at least three point correspondences of type (30) these six parameters are uniquely determined. However due to noise in the measurements the transformation calculated from three

arbitrary point correspondences is not the best one. To find the best transformation usually the least-squares solution to the overdetermined system of equations (30) is searched,

$$\sum_i \left\| \boldsymbol{p}_i - T\left(\boldsymbol{p}_i'\right) \right\|^2 \quad \longrightarrow \quad \text{minimum.} \tag{32}$$

We want to give a statistical explanation of (32) as the maximum likelihood model selection [46]. Our point sets $\{\boldsymbol{p}_i\}$ and $\{\boldsymbol{p}_i'\}$ can be interpreted as a sample $a = [\ldots\{\boldsymbol{p}_i, \boldsymbol{p}_i'\}\ldots]$ of the vector random variables $\boldsymbol{X}'$ and the functionally dependent vector random variable $\boldsymbol{X} = T\left(\boldsymbol{X}'\right) + \eta$, which is assumed to be perturbed by Gaussian measurement noise described by η. The likelihood of the sample is the conditional probability of the sample given the random variables $\boldsymbol{X}'$ and $\boldsymbol{X}$ and a model of their functional dependence $\hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)$, i.e.

$$\begin{aligned} p\left(a \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right) \quad &\equiv \quad p\Big(\quad \{\boldsymbol{X}, \boldsymbol{X}'\}_1 = \{\boldsymbol{p}_1, \boldsymbol{p}_1'\}, \qquad (33) \\ &\qquad\qquad \{\boldsymbol{X}, \boldsymbol{X}'\}_2 = \{\boldsymbol{p}_2, \boldsymbol{p}_2'\}, \ldots \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\Big) \\ &= \quad \prod_{(\boldsymbol{p}_i, \boldsymbol{p}_i') \in a} p_i\left(\boldsymbol{X} = \boldsymbol{p}_i \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{p}_i'\right)\right) \qquad (34) \end{aligned}$$

where we have assumed in (34) that the trials of the sample are independent. With the help of Bayes' law we can find the most likely model given the sample,

$$p\left(\hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right) \mid a\right) = p\left(a \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right) \frac{p\left(\hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right)}{p\left(a\right)}, \tag{35}$$

by maximizing $p\left(\hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right) \mid a\right)$ with respect to the parameters of T. The unconditional probability of the sample $p(a)$ could be arbitrary, since the sample is the same for all models. It is the assumption made by a maximum likelihood model selection that the prior probability of the model $p\left(\hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right)$ is the same for all models that are evaluated, i.e. $p\left(\hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right)$ is constant. Therefore maximizing $p\left(a \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right)$ is equivalent to maximizing $p\left(\hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right) \mid a\right)$. To simplify the maximization of $p\left(a \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right)$ from (34), usually the logarithm is taken which does not influence the position of the maximum since the logarithm is a monotonic function. So instead of (34) we maximize

$$\log p\left(a \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right) = \sum_{(\boldsymbol{p}_i, \boldsymbol{p}_i') \in a} \log p_i\left(\boldsymbol{X} = \boldsymbol{p}_i \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{p}_i'\right)\right). \tag{36}$$

We assume that the differences between the predicted points $\hat{\boldsymbol{p}}_i = T\left(\boldsymbol{p}_i'\right)$ and the actual trials of $\boldsymbol{X}$ are Gaussian,

$$p_i\left(\boldsymbol{X} = \boldsymbol{p}_i \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{p}_i'\right)\right) = g_{\psi_i}\left(\boldsymbol{p}_i - T\left(\boldsymbol{p}_i'\right)\right) \tag{37}$$

with

$$g_\psi\left(\boldsymbol{z}\right) \equiv \frac{1}{(2\pi)^{\frac{3}{2}} \left|\psi\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \boldsymbol{z}^T \psi^{-1} \boldsymbol{z}\right), \tag{38}$$

where ψ is the covariance matrix of the random vector $\eta = \boldsymbol{X} - \hat{\boldsymbol{X}}$ and $|\psi|$ its determinant.

Using (37) in the expression for $\log p\left(a \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right)$ from (36) we get

$$\begin{aligned}\log p\left(a \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right) &= -\frac{1}{2}\sum_{(\boldsymbol{p}_i,\boldsymbol{p}_i')\in a}\log\left((2\pi)^3|\psi_i|\right) \\ &\quad -\frac{1}{2}\sum_{(\boldsymbol{p}_i,\boldsymbol{p}_i')\in a}\left(\boldsymbol{p}_i - T\left(\boldsymbol{p}_i'\right)\right)^T\psi_i^{-1}\left(\boldsymbol{p}_i - T\left(\boldsymbol{p}_i'\right)\right).\end{aligned} \tag{39}$$

Therefore maximizing $p\left(a \mid \hat{\boldsymbol{X}} = T\left(\boldsymbol{X}'\right)\right)$ is equivalent to minimizing

$$\sum_{(\boldsymbol{p}_i,\boldsymbol{p}_i')\in a} D_{\psi_i}\left(\boldsymbol{p}_i - T\left(\boldsymbol{p}_i'\right)\right), \tag{40}$$

where

$$D_\psi\left(\boldsymbol{z}\right) = \boldsymbol{z}^T\psi^{-1}\boldsymbol{z} \tag{41}$$

is the so called squared Mahalanobis distance. Assuming that all ψ_i are diagonal with equal variances on the diagonal ($\sigma_i^2 = \psi_{i_{11}} = \psi_{i_{22}} = \psi_{i_{33}}$), we get from (40) the standard weighted least-squares problem also known as the chi-square fitting problem [32],

$$\chi^2 = \sum_{(\boldsymbol{p}_i,\boldsymbol{p}_i')\in a}\frac{\left\|\boldsymbol{p}_i - T\left(\boldsymbol{p}_i'\right)\right\|^2}{\sigma_i^2}. \tag{42}$$

Assuming that all σ_i are equal, the simple least-squares problem (32) follows directly from (42).

Let us continue with the solution to the weighted least-squares problem (42). Using the explicit form of the transformation T from (31) we have to minimize,

$$\sum_i \frac{1}{\sigma_i^2}\left\|\boldsymbol{p}_i - \boldsymbol{R}\boldsymbol{p}_i' - \boldsymbol{t}\right\|^2. \tag{43}$$

If using Euler angles for the representation of $\boldsymbol{R}$, the rotation will be given by the following matrix product,

$$\boldsymbol{R} = \begin{pmatrix}\cos\alpha & -\sin\alpha & 0\\ \sin\alpha & \cos\alpha & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\cos\beta & 0 & \sin\beta\\ 0 & 1 & 0\\ -\sin\beta & 0 & \cos\beta\end{pmatrix}\begin{pmatrix}\cos\gamma & -\sin\gamma & 0\\ \sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{pmatrix}. \tag{44}$$

Therefore the parameters α, β, γ are not quadratic in (43) and the solution to the minimization problem cannot be reduced to a simple system of linear equations (by calculating the partial derivatives with respect to the parameters and setting them to zero i.e. solving the so called system of normal equations). Such least-squares problems are also called nonlinear least-squares problems. Of course the problem can be solved by standard optimization techniques like gradient descent, conjugate gradients, Newton's method or the Levenberg-Marquardt algorithm (that has become the standard technique for nonlinear least-squares problems [32]). Amazingly there are however closed form solutions to this problem that are significantly faster (2 to 5 times dependent on the number of point correspondences) than iterative approaches [1]. The known solutions are based on

- singular value decomposition[2] (SVD) [1, 22, 45],

[2]The SVD method used here should not be confused with the standard SVD method which is used in favor of solving the system of normal equations when we have a linear least-squares problem. Here a completely different matrix is decomposed.

- quaternions [20, 14],
- dual number quaternions [48, 54],
- orthonormal matrices/polar value decomposition [21, 22].

Let us present the SVD based method (refer to [32] for a motivation of the term SVD) in some details.

To simplify the problem in (43) it is convenient to compute the centroids of each point set and translate the point clouds so that the centroids coincide at the origin. In this way, in order to get the parameters of the rotation, we only have to minimize[3],

$$\sum_i \frac{1}{\sigma_i^2} \left\| \boldsymbol{p}_i - \boldsymbol{R}\boldsymbol{p}_i' \right\|^2 . \tag{45}$$

The translation can then be calculated as the difference between the centroids[20]. Expanding the square in (45) results in

$$\sum_i \frac{1}{\sigma_i^2} \left(\|\boldsymbol{p}_i\|^2 - 2\boldsymbol{p}_i^T \boldsymbol{R}\boldsymbol{p}_i' + \|\boldsymbol{R}\boldsymbol{p}_i'\|^2 \right) . \tag{46}$$

The first term in (46) does not depend on $\boldsymbol{R}$. Since a rotation is an operation that preserves lengths, i.e. $\|\boldsymbol{R}\boldsymbol{p}_i'\| = \|\boldsymbol{p}_i'\|$, only the second term depends on $\boldsymbol{R}$. Therefore minimizing (46) is equivalent to maximizing

$$\sum_i \frac{1}{\sigma_i^2} \boldsymbol{p}_i^T \boldsymbol{R}\boldsymbol{p}_i'. \tag{47}$$

Using the relation for two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ that $\boldsymbol{a}^T\boldsymbol{b} = \mathrm{trace}\left(\boldsymbol{a}\boldsymbol{b}^T\right)$ and the fact that the trace of a matrix product is cyclic, $\mathrm{trace}\,(\boldsymbol{A}\boldsymbol{B}) = \mathrm{trace}\,(\boldsymbol{B}\boldsymbol{A})$ for matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, (47) can be rewritten as

$$\mathrm{trace}\left(\boldsymbol{R}^T\boldsymbol{K}\right) \tag{48}$$

where the so called correlation matrix $\boldsymbol{K}$ is defined by

$$\boldsymbol{K} = \sum_i \frac{1}{\sigma_i^2} \boldsymbol{p}_i \boldsymbol{p}_i'^T. \tag{49}$$

Solutions to the maximization problem in (48) can be found based on singular value decomposition (SVD), polar value decomposition or quaternions. Based on singular value decomposition, we decompose the correlation matrix $\boldsymbol{K}$ into the form

$$\boldsymbol{K} = \boldsymbol{U}\boldsymbol{\Lambda}\,\boldsymbol{V}^T \tag{50}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are 3×3 orthonormal matrices, and $\boldsymbol{\Lambda}$ is a 3×3 diagonal matrix with nonnegative elements, the so called singular values. An algorithm for performing such a decomposition, which is a standard task in numerical mathematics, can be found in [32]. Now the matrix

$$\boldsymbol{R}_{\max} = \boldsymbol{U}\boldsymbol{V}^T \tag{51}$$

maximizes (48) due to the following lemma:

Lemma: For any positive definite and symmetric matrix $\boldsymbol{A}$ and any orthonormal matrix $\boldsymbol{B}$,

$$\mathrm{trace}\,(\boldsymbol{A}) \geq \mathrm{trace}\,(\boldsymbol{B}\boldsymbol{A})\,. \tag{52}$$

[3]From now on $\boldsymbol{p}_i$ and $\boldsymbol{p}_i'$ are relative to the centroids of the point sets.

A simple proof for this lemma based on the Schwarz inequality can be found in [1]. $\boldsymbol{A}$ is given in our case by $\boldsymbol{R}_{\max}^T \boldsymbol{K}$ (compare (48)) which fulfills the conditions of the lemma since

$$\begin{aligned} \boldsymbol{R}_{\max}^T \boldsymbol{K} &= \boldsymbol{V}\boldsymbol{U}^T \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T & (53) \\ &= \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T & (54) \end{aligned}$$

is symmetrical and positive definite. Thus, due to the lemma, for any 3×3 matrix $\boldsymbol{R}$,

$$\operatorname{trace}\left(\boldsymbol{R}_{\max}^T \boldsymbol{K}\right) \geq \operatorname{trace}\left(\boldsymbol{R}\boldsymbol{R}_{\max}^T \boldsymbol{K}\right), \tag{55}$$

so that $\boldsymbol{R}_{\max}^T$ really maximizes (48)[4].

7 Search strategies

7.1 Hough method

Before presenting the application of the Hough method to the registration of 3-D point sets, we describe its basic idea in a general framework.

Let $\{\boldsymbol{x}_i\}_{i=1\ldots N}$, $\boldsymbol{x}_i \in \mathbb{R}^l$ be a data set and $\{q_j\}_{j=1\ldots m}$, $q_j \in \mathbb{R}$ a parameter set related by a function $f : \mathbb{R}^m \times \mathbb{R}^l \rightarrow \mathbb{R}^w$

$$f\left(q_1, \ldots, q_m, \boldsymbol{x}_i\right) = 0 \quad \forall i. \tag{56}$$

In addition let suppose that all parameters q_i are uniquely determined by a subset of k samples of $\{\boldsymbol{x}_i\}_{i=1\ldots N}$

$$\{\boldsymbol{x}_i\}_{i=1\ldots k} \quad \Rightarrow \quad q_1, \ldots, q_m. \tag{57}$$

We search for the parameters $\{q_j\}_{j=1\ldots m}$ which realize (56) "as best as possible"[5] for all $\boldsymbol{x}_i$.

Now for the Hough method the following steps have to be performed: For each subset of k samples of $\{\boldsymbol{x}_i\}_{i=1\ldots N}$ (there are $\frac{N!}{(N-k)!k!}$ possibilities) the parameters $q_1, \ldots, q_m$ are calculated and at the corresponding position $(q_1, \ldots, q_m)$ in a m-dimensional accumulation table (Hough table) a counter is incremented by one. In this way every subset of k samples of $\{\boldsymbol{x}_i\}_{i=1\ldots N}$ resulting in the same parameter set $q_1, \ldots, q_m$ contributes to the same position counter in the table. Therefore the position whose counter has the highest score corresponds to the parameter set that is in best accordance with the given data set $\{\boldsymbol{x}_i\}_{i=1\ldots N}$.

To illustrate the Hough method we give a simple example: Let $\{\boldsymbol{x}_i\}_{i=1\ldots N}$ be a set of 2-D points $\boldsymbol{x} = (x, y)^T$ from an intensity image which are the outcome of a feature detection algorithm. Assume we know that some of the $\boldsymbol{x}_i$ describe circles in the original intensity image but we do not know which ones. We want to find the centers of the circles and their radii. Thus the searched parameters q_i , $i = 1, \ldots, m = 3$ are the center position coordinates x_c and y_c, and the radius R. For each point on a circle the following equation holds:

$$(x - x_c)^2 + (y - y_c)^2 - R^2 = 0. \tag{58}$$

Therefore (58) determines the function f from (56). Since three points uniquely define a circle, $k = 3$ in the general description above. If we now apply the Hough method we will find the circles present in the data from the positions of the accumulations in the 3-dimensional Hough table.

[4]Since orthonormal matrices build a group (the so called SO(3)) [43] $\boldsymbol{R}\boldsymbol{R}_{\max}^T$ represents an arbitrary orthonormal matrix if $\boldsymbol{R}$ also represents an arbitrary orthonormal matrix.

[5]The cost function is defined by the method itself.

We now apply the Hough method to the registration of two 3-D point sets $\{\boldsymbol{p}_i\}$ and $\{\boldsymbol{p}'_i\}$ of extracted point features. $\{\boldsymbol{x}_i\}_{i=1...N}$ is in this case the set of tupels $(\boldsymbol{p}, \boldsymbol{p}')$ of all combinations of points from the first set with points from the second set. The searched parameters are the $m = 6$ parameters of the rigid transformation $(\boldsymbol{R}, \boldsymbol{t})$ between both point sets, so that f from (56) is defined by

$$\boldsymbol{R}\boldsymbol{p}' + \boldsymbol{t} - \boldsymbol{p} = \boldsymbol{0}. \tag{59}$$

Since three non-collinear points uniquely define the 6 parameters of $\boldsymbol{R}$ and $\boldsymbol{t}$, $k = 3$ in the general description. Applying now the Hough method all transformations calculated from *correct* point correspondences result in the same transformation while all other transformations are distributed more or less randomly in the parameter space. As well as above the points of accumulation give us the right parameter set. Note that in the case where the Hough method is used to find the parameters of a transformation the method is also often called *clustering* [37].

7.2 Correlation

In the previous section we have already introduced the similarity metric correlation (24) that has to be maximized with respect to the transformation T

$$C(T) = \frac{\sum_{\boldsymbol{x}} v(\boldsymbol{x}) u\left(T(\boldsymbol{x})\right)}{\sum_{\boldsymbol{x}} \left[u\left(T(\boldsymbol{x})\right)\right]^2}.$$

One way to find the correlation maximum is to compute $C(T)$ for all possible transformations T. Since the number of possibilities may be very large, the complete search in the parameter space may not be feasible. For example, if T is a 3-D rigid transformation (rotation + translation) and each of the $d = 6$ dimensions of the parameter space is divided in M quantization steps, the expression in (24) has to be computed M^d times. Thus the complexity of such a calculation is $M^d N$, if N is the number of points in each data set. M determines the accuracy of the approach but can be reduced – when using the outcomes of a previous feature extraction process – by only considering the transformations that are in accordance with the extracted features. In addition, in order not to make the complexity prohibitively large, N and d have to be small. N can be reduced by only using data within a small window, usually called a template.

Let us present a typical example: Given two intensity images I_1 and I_2 describing parts of the same object ($d = 3$: 2 translations, 1 rotation), in which M_1 and M_2 features (e.g. corners) have been respectively extracted, we want to match them by finding corresponding features in I_1 and I_2. Using point features limits the search in the translation parameter space: only translations between extracted features are allowed. The neighborhood within a window centered at each extracted feature F_1 in I_1 is correlated with the neighborhood at every extracted feature F_2 in I_2, i.e. expression (24) has to be calculated $M_1 M_2 M$ times where M is the number of quantization steps in the rotation dimension. Using a window of size $n_x \times n_y = N_w$ the complexity of the correlation calculation is then $M_1 M_2 M N_w$.

7.3 Relaxation

Relaxation is a technique to resolve ambiguities between match candidates of two data sets. These candidates are the outcomes of a feature extraction process. There are ambiguities between the match candidates since in general a given feature attribute does not uniquely determine a candidate. Even after using a correlation technique for simple point features as

described above, a feature point in the first data set may be paired to several feature points in the second data set.

To overcome these ambiguities for a given feature the relaxation technique makes in addition use of relations (to features in the neighborhood) which are "more or less" preserved under the considered transformation. For point features these relations are typically the distances between two features in the same data set [52]. A further possibility could be the angles formed by three features. Since distances between features in intensity images (taken from 3-D objects) are distorted by perspective projections, these relations are only invariant for point features close to each other.

For the relaxation technique a similarity measure for each pair of match candidates is defined based on the following three criteria:

1. the goodness of the considered pair of match candidates (e.g. provided by correlation scores or the difference between feature attributes),
2. the goodness of all possible pairs of match candidates in the respective neighborhoods,
3. the agreements of the relations between the considered feature and the features in the neighborhood in the first data set with those in the second data set.

These similarity measures are iteratively changed by updating the relations defined in the third criterion until they converge. The relations are updated by successively canceling pairs of match candidates which have not reached a large similarity measure in the previous iteration.

Let us illustrate the relaxation approach in continuing the example presented in the previous paragraph: The correlation provides for each possible pair of match candidates (F_{1i}, F_{2j}), where F_{1i} and F_{2j} are extracted features in I_1 and I_2 respectively, a measurement of the goodness $c(F_{1i}, F_{2j})$ mentioned in the above criteria 1 and 2. The similarity measure S can be expressed as

$$S(F_{1i}, F_{2j}) = c\,(F_{1i}, F_{2j}) \sum_{F_{1k} \in \Omega_{F_{1i}}} \sum_{F_{2l} \in \Omega_{F_{2j}}} c\,(F_{1k}, F_{2l})\, \delta\,(F_{1i}, F_{2j}, F_{1k}, F_{2l}) \tag{60}$$

where $\Omega_{F_{1i}}$ and $\Omega_{F_{2j}}$ are the neighborhoods around F_{1i} and F_{2j} respectively, and $\delta(\ldots)$ is a function describing the agreement of the relations between (F_{1i}, F_{1k}) and (F_{2j}, F_{2l}). During the updating process some of the function values are set to zero and enable in this way a recomputation of a new similarity measure until all correspondences between match candidates are fixed.

7.4 Indexing schemes

In general indexing schemes precompute invariant feature values (e.g. principal curvatures, point signatures) in a data set and hash them into a look-up table (called hash table) with references to the corresponding feature positions [10]. In order to match two data sets $\{\boldsymbol{x}_i\}$ and $\{\boldsymbol{x}_i'\}$ the following steps are performed:

- Firstly the extracted invariant feature values in $\{\boldsymbol{x}_i\}$ are hashed in a table.
- Secondly for each point feature in $\{\boldsymbol{x}_i'\}$ with feature values $(v_1, ..., v_n)$ we find the possible corresponding positions in $\{\boldsymbol{x}_i\}$ (match candidates) by taking the points at the position $(v_1, ..., v_n)$ in the hash table.
- Thirdly to resolve ambiguities for the match candidates a relaxation, Hough or prediction-verification method is applied.

The advantage of the hash table is the access to possible match candidates in constant time.

A more sophisticated indexing scheme is *geometric indexing* [53]. In this case the indexing scheme is based on the geometrical relationships between extracted features. For each extracted point feature in the first data set $\{\boldsymbol{x}_i\}$ a basis for a coordinate frame is defined with the help

of further extracted point features (e.g. in 3-D space two other point features are necessary). Then the coordinates of all other point features are calculated with respect to this basis and are hashed into a look-up table which stores the k-tuple of features defining the basis. This is done for all combinations of point features defining a basis (e.g. for M extracted features in the first data set there are $M(M-1)(M-2)$ possibilities to define a basis in 3-D space and for each basis M feature coordinates have to be determined). In the second data set point features are similary extracted and an arbitrary basis $\mathcal{B}'_1$ formed by a k-tuple of features is chosen. Then all features $F_i \in \{\boldsymbol{x}_i\}$ are computed with respect to this basis, resulting in the coordinates $(x_1, \ldots, x_n)_i$. For each coordinate vector we find at the position $(x_1, \ldots, x_n)_i$ in the hash-table several basis from the first data set which can possibly correspond to the basis $\mathcal{B}'_1$. Each time a basis occurs it is voted for by incrementing a counter. After processing all coordinate vectors the basis with the highest vote is considered as a corresponding basis to $\mathcal{B}'_1$. This hypothesis can be confirmed with a prediction-verification scheme working on the whole data set. If the verification fails another basis $\mathcal{B}'_2$ will be chosen and tested in the same way.

7.5 Prediction-verification

The principle of a prediction-verification scheme is quite similar to the correlation approach: Just calculate a given similarity metric for a certain transformation. However there are some differences in practice:

- In the case of correlation all possible transformations are tested; in prediction-verification only the transformations resulting from a preprocessing step (e.g. from an indexing scheme or a feature based approach) are verified.
- In the case of correlation the transformation reaching an extremal value of the similarity metric (e.g. the minimal value of (22) or maximal of (24)) is accepted as the correct one; in prediction-verification the first transformation resulting in a value of the similarity metric better than a given threshold is considered as the right one.
- In the case of correlation in order to reduce the complexity the calculation is usually only applied to extracted features; in prediction-verification the transformation of the whole data set provides the most reliable verification.

Some examples for prediction-verification schemes can be found in [42, 10, 15].

7.6 Tree + graph matching

After a feature extraction preprocessing step a data set can be described by a tree (also called graph), where the nodes are defined by the features and the links by their geometrical relations. The matching of two data sets is then reduced to the mapping of two graphs. This search process is often called *subgraph isomorphism*. Some examples can be found in [51, 9].

7.7 Standard optimization techniques

The previously described search strategies are based on extracted features whenever they are used in practical applications. Otherwise their complexity would be too high. By contrast, standard optimization techniques try to find an extremum of a given similarity metric taking into account the whole data sets. An extremum can be either global or local. Although there is in practice till now no guarantee to find the global extremum, a few approaches deal with this problem: for example mean field theory [38], genetic algorithms [6] and simulated annealing [35, 32]. However, if a "good estimation" of the transformation between two data sets is known

there are several standard approaches to find the global extremum. A "good estimation" means in this context that the local extremum to which the method converges is in fact the global extremum. These techniques usually base on gradient information of the cost function. Typical examples are gradient descent (also called steepest descent), conjugate gradients or the Levenberg-Marquardt algorithm [32].

7.8 ICP algorithm

In the subsection 6.3 we have mentioned that it is straightforward to find the best rigid transformation between two point clouds by minimizing (32)

$$\sum_i \|\boldsymbol{p}_i - T(\boldsymbol{p}'_i)\|^2 \tag{61}$$

where $\boldsymbol{p}_i$ and $\boldsymbol{p}'_i$ are corresponding points. However the point correspondences are not known in advance. In the case where the given data sets are already well aligned to each other (this should be possible with one of the feature based methods described above), the following heuristic assumption may be reasonable: *Corresponding points are the closest points between two given data sets.* In this way (61) can be directly derived from the correlation expression (22) by setting

$$v(\boldsymbol{x}) = \min_{\boldsymbol{p}'_k \in \{\boldsymbol{p}'_i\}} (\|\boldsymbol{p}'_k - \boldsymbol{x}\|), \tag{62}$$

$$u(T(\boldsymbol{x})) = \min_{\boldsymbol{p}_j \in \{\boldsymbol{p}_i\}} (\|\boldsymbol{p}_j - T(\boldsymbol{x})\|) \tag{63}$$

and by restricting the sum over all $\boldsymbol{x}$ in (22) to the data set $\{\boldsymbol{p}'_i\}$. Thus we get $v(\boldsymbol{x}) = 0$ and (22) becomes

$$\sum_{\boldsymbol{x} \in \{\boldsymbol{p}'_i\}} \left(\min_{\boldsymbol{p}_j \in \{\boldsymbol{p}_i\}} \|\boldsymbol{p}_j - T(\boldsymbol{x})\| \right)^2. \tag{64}$$

By defining the index i so that

$$\boldsymbol{p}_i = \arg \left[\min_{\boldsymbol{p}_j \in \{\boldsymbol{p}_i\}} \|\boldsymbol{p}_j - T(\boldsymbol{p}'_i)\| \right] \tag{65}$$

we get the expression (61). Since the transformation T is not known in advance but assumed to be small (the data sets are supposed to be well aligned), T is taken as the identity transformation. In this way corresponding points are defined as:

$$\boldsymbol{p}_i = \arg \left[\min_{\boldsymbol{p}_j \in \{\boldsymbol{p}_i\}} \|\boldsymbol{p}_j - \boldsymbol{p}'_i\| \right] \tag{66}$$

After calculating the corresponding points and the resulting transformation, it can be expected that – after applying the transformation – the data sets become closer to each other. Thus it seems reasonable to iterate this procedure (therefore the term ICP: Iterative Closest Points [3]) until convergence of the computed transformation:

1. find closest points according to (66),
2. calculate rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$ that minimize the least-squares sum of corresponding points (61),
3. apply the transformation to all points in the first data set.

However, if the estimation of corresponding points during the initialization step is "too bad" the algorithm will not converge to the right transformation[6] but will stuck in a local minimum. Nevertheless this algorithm has become the standard for the precise registration of two data sets, described in details in [3, 54] and used as the basis of more sophisticated algorithms in [44, 2, 15, 28, 42, 40]. Besl and McKay proposed an accelerated version of the ICP algorithm [3] in using linear and quadratic extrapolation of the registration parameters during the iterations. In this way they reduced the number of iterations until convergence of the transformation by a factor of 2 to 3.

In (66) corresponding points are defined as the points with minimal distances between points of two given data sets. Although this definition of corresponding points is quite often used [3, 54], another approach is possible: since a point set defines a surface (of course, not uniquely) the corresponding point to a point $\boldsymbol{p}'_0 \in \{\boldsymbol{p}'_i\}$ can also be defined as the point $\boldsymbol{p}_{s_0}$ (on the surface $\mathcal{S}$ defined by the points $\{\boldsymbol{p}_i\}$) having minimal distance to $\boldsymbol{p}'_0$, i.e.

$$\boldsymbol{p}_{s_0} = \arg\left[\min_{\boldsymbol{p}_s \in \mathcal{S}} \|\boldsymbol{p}_s - \boldsymbol{p}'_0\|\right]. \tag{67}$$

This definition of corresponding points is used for example in [8, 44, 28]. Since the surface resulting from a given point set is not uniquely defined, a more precise definition of corresponding points has to be given for actual calculations. For example Chen and Medioni [8] propose an iterative algorithm that takes 3 to 5 iterations to determine a closest point.

The worst case cost of finding the closest point of $\boldsymbol{p}'_0 \in \{\boldsymbol{p}'_i\}$ according to (66) is $O(n)$, where n is the number of points in $\{\boldsymbol{p}_i\}$. Therefore the total cost of finding the closest point for all $\boldsymbol{p}'_i \in \{\boldsymbol{p}'_i\}$ is $O(nn')$, where n' is the number of points in $\{\boldsymbol{p}'_i\}$. There are several methods which can considerably speed up the search process, e.g. bucketing techniques (in 3-D or in 2-D by projection), k-D trees (abbreviation for k-dimensional binary search tree; here $k = 3$) [54] or octree-splines [41]. The closest point search can be further accelerated by exploiting a coarse to fine strategy during the iterations of the ICP: during the first iterations closest points are only determined for some coarsely sampled points. Then a fine matching using more and more points follows [54], [44].

8 Robust registration

In order to improve the registration process with the ICP-algorithm it is recommendable to use the weight factors $1/\sigma_i^2$ in the least-squares sum (43). For example Turk and Levoy [44] use the dot product of calculated normals and a vector pointing to the light source as confidence values. However the derivation of the least-squares sum used in the ICP-algorithm is based on the assumption that deviations from the model (see (37)) are Gaussian distributed. This assumption is not valid in practice due to the following arguments:

- Usually there are many points in one data set that should not have a correspondence in the other data set due to different object occlusions in data sets from different views and/or since there is only a partial overlap between the data sets. However point correspondences (=closest points) are always found in the ICP-algorithm.
- Usually there are outliers in the data sets frequently due to some unknown reasons. Perhaps there was a percussion during the measurement process or the camera was overdriven due to light reflections.

[6] As described the algorithm always converges to the identity transformation. Here we mean that the combined transformation over all iterations does not converge to the right transformation.

Statisticians have developed various sorts of robust methods that can reduce the influence of outliers:

8.1 First simple approach

An intuitive method to eliminate outliers from the matching process of two data sets is to compute the mean value μ of the distances d_i between corresponding points, and to remove from the least-squares sum (32) each point pair whose distance is larger than a given threshold $\mathcal{D}$ (e.g. $\mathcal{D} = \mu + 3\sigma$ where σ is the standard deviation of the distances). Then one iteration of the ICP-algorithm is performed. Before the next iteration the mean value μ is calculated again on *all* points of the data sets to prevent that more and more points are eliminated [54]. However since the influence of corresponding points in the standard least-squares sum quadratically increases with their distances, outliers are much more weighted than "correct" data points and thus can strongly distort the estimation of the computed transformation parameters (one far-away outlier can make all other outliers have small distances d_i, so that "correct" data points are rejected instead of the outliers).

8.2 M-estimators

A generalization of this approach is the concept of M-estimators. In this case, instead of minimizing

$$\sum_i \|\boldsymbol{p}_i - T(\boldsymbol{p}'_i)\|^2 = \sum_i d_i^2, \tag{68}$$

we replace the squared distance (L_2-estimator) by a function ρ of d_i

$$\sum_i \rho(d_i) \tag{69}$$

where ρ is a symmetric, positive definite function with a unique minimum at zero and is less increasing than square. It can be shown that minimizing (69) is equivalent to minimizing the following iterated reweighted least-squares expression [52],

$$\sum_i \omega\left(d_i^{(k-1)}\right) d_i^2 \tag{70}$$

where $\omega(x) = \frac{1}{x}\frac{d\rho}{dx}$, the superscript $^{(k)}$ indicates the iteration number, and $\omega\left(d_i^{(k-1)}\right)$ has to be recalculated after each iteration in order to be used in the next iteration. In this way the first method described above can be considered as the special case of M-estimator minimization by taking $\omega(x) = \begin{cases} 1 & |x| \leq \mathcal{D} \\ 0 & \text{else} \end{cases}$. Note that M-estimators suffer from the same problems as described above. Examples of several weight functions ω can be found in [52, 34, 32].

8.3 Least Median of Squares

A really robust approach that overcomes the bad influence of outliers is the Least Median of Squares method (LMedS). This method is not affected by outliers up to a rate of 50% [28, 34]. In this approach the transformation parameters are estimated by minimizing

$$\text{median}_i\left(d_i^2\right) \tag{71}$$

in the following way:

- Since there is (probably) no straightforward formula for the median function, it is not possible to differentiate (71) and apply standard optimization techniques to find the minimizing transformation parameters. Thus the complete parameter space should be investigated.
- Since such an exhaustive search of the parameter space would not be feasible, only transformations based on corresponding points (closest points) can be considered: each combination of three point correspondences define a possible transformation (in 3-D space). In practice the number of such combinations is so large that it must be reduced by only randomly taking a few of them.
- For each random combination the corresponding transformation is applied and the median value of the distances $\|\boldsymbol{p}_i - T(\boldsymbol{p}'_i)\|$ between *all* corresponding points $\boldsymbol{p}_i$ and $\boldsymbol{p}'_i$ is computed. The transformation for which the median is minimal is retained.

The minimal number of randomly chosen combinations m depends on the rate of outliers ϵ in the data sets and on the desired reliability R that the LMedS solution is not corrupted by outliers. It can be easily seen that the reliability R is given by:

$$R = 1 - \left[1 - (1-\epsilon)^D\right]^m, \tag{72}$$

where D is the minimal number of points necessary to define a unique transformation (in 3-D space: $D = 3$). Therefore the minimal number of combinations is:

$$m = \frac{\log(1-R)}{\log\left[1-(1-\epsilon)^D\right]}. \tag{73}$$

(For example for $R = 0.99$ and $\epsilon = 0.4$ we get $m = 19$.)

8.4 Extended Kalman Filtering

Another robust technique coming from the signal processing theory is Extended Kalman Filtering (EKF). This method makes use of a priori knowledge and provides a recursive solution to the least-squares problem. Since it would be quite time consuming to explain this approach, we refer the interested reader for a basic introduction to [50] and for an application on 3-D registration to [31].

9 Registration of multiple point sets

Till now only the registration between two data sets was considered. For practical applications, such as virtual reality, CAD-processing or NC-manufacturing, it is however of interest to match several range views together to reconstruct a 3D-model of the original object. The techniques described below assume that:

- different views of the object to be modeled are pairwisely overlapping, i.e. each view has a common area with (at least) two other views.
- these views have already been transformed by a coarse registration, i.e. they nearly lie in the same coordinate system.

A first simple approach is to process the views sequentially [28]:

- Firstly one of the data sets is taken as reference.
- Secondly one of its (2 or more) connected views is registered, so that the integrated model forms the new reference.

- In this way all other views are added one by one building at each step a new integrated model.

The disadvantage of this method is that information of a subsequently registered view cannot improve the registration of previously matched views. Due to compounding registration errors a gap usually occurs in the area where the last view (of the closed object) is matched with the combined model of all other views registered so far.

One possibility to overcome this problem is to directly register all views to a *single* master view, for example a cylindrical scan of the whole object [44]. In this way the multiple view registration problem is reduced to several pairwise matching problems, which can be directly solved by one of the methods described above (typically ICP).

In the absence of such a master view it seems reasonable to equally distribute the registration errors over the whole object: a similarity metric has to be defined, which takes into account the distances between corresponding points in all N views simultaneously. For each data set $\{\boldsymbol{p}_i\}^{(j)}, 1 \leq j \leq N$ which has correspondences with N_j other views the combined least-squares sum of corresponding points is given by (compare (45))

$$\sum_{k=1}^{N_j} \sum_i \frac{1}{\sigma_i^2} \left\| \boldsymbol{p}_i^{(j)} - T^{(kj)} \boldsymbol{p}_i^{(k)} \right\|^2 . \tag{74}$$

By defining a reference view to which *all* views are referred, the transformation from view $^{(k)}$ to $^{(j)}$ can always be written as

$$T^{(kj)} = T^{(j)^{-1}} T^{(k)} \tag{75}$$

where $T^{(j)}$ is the transformation from view $^{(j)}$ to the reference. In this way the transformation between any two data sets is then uniquely defined regardless of the path chosen to link the views [2]. To consider all views simultaneously the summation over all data sets has to be minimized [39]:

$$\tilde{E}\left(T^{(1)}, \ldots, T^{(N)}\right) = \sum_{j=1}^{N} \sum_{k=1}^{N_j} \sum_i \frac{1}{\sigma_i^2} \left\| \boldsymbol{p}_i^{(j)} - T^{(j)^{-1}} T^{(k)} \boldsymbol{p}_i^{(k)} \right\|^2 . \tag{76}$$

Note that $T^{(j_0)} = \mathrm{Id}$ if the j_0-th data set is the reference view. The minimization problem (76) can be solved by the standard optimization techniques presented in section 7.

References

[1] K.S. Arun, T.S. Huang, and S.D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698 – 700, Sept. 1987.

Given two point sets p_i and p'_i related by $p'_i = Rp_i + T$ (R rotation, T translation) and the correspondences between the points a closed-form solution to the least-squares problem of finding R and T is presented which is based on the singular value decomposition (SVD) of a 3×3 matrix. The performance of the algorithm is compared to the quaternion method and an iterative approach.

[2] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau. Towards a general multi-view registration technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(5):540 – 547, May 1996.

Given several range images from different viewpoints and good initial transformations

for these images the described algorithm considers the network of views as a whole and minimizes the registration errors of all views simultaneously via an iterated closest point (ICP) like algorithm. Experimental results are given.

[3] P.J. Besl and N.D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239 – 256, Feb. 1992.
Given two point sets and a good initial transformation between them an iterative closest point (ICP) algorithm in combination with a quaternion based method is described which minimizes the least-squares sum of corresponding points to find the exact displacement. In addition an accelerated ICP algorithm is presented (use linear and quadratic extrapolation of the registration parameters). The performance is compared to alternitive minimization approaches. A (non convincing) suggestion how to find a good initial transformation is made. Experimental results are given.

[4] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[5] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325 – 376, 1992.
A survey of image registration techniques published before 1992.

[6] K. Brunnström and A.J. Stoddart. Genetic algorithms for free-form surface matching. In *13th Int. Conference on Pattern Recognition*, pages D673 – 689, Vienna, Austria, 1996.
http://www.ee.surrey.ac.uk/pub/vision/papers/brunnstrom-icpr96.ps.Z .
An approach for finding the inital guess of the rotation and translation between the object and the model surface in the free-form surface matching problem is described. A fitness function that measures the match quality is designed and then used in a genetic algorithm. Experimental results are given.

[7] J. Canny. A computational approach to edge detection. *IEEE Trans. PAMI*, 8(6):679–698, 1986.

[8] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145 – 155, 1992.
Given several range images and good initial transformations between them an algorithm is presented that iteratively minimizes the least square sum of corresponding points of a range view to be registered and the merged data from all previous views registered so far. Only a subset of control points (50-200) is used for the registration process. The corresponding point of a point from one range image is defined here as the point on the surface (defined by the points of the second image) having minimal distance to the given point. An approximation to this definition of corresponding points is calculated as the point on the tangential plane (of the point in the second image being the intersection of the normal of the given point with the surface defined by the points of the second image) having minimal distance to the given point (hard to understand, I know ...). Experimental results are given.

[9] J. Cheng and T. Huang. Image registration by matching relational structures. *Pattern Recognition*, 17(1):149 –159, 1984.

[10] Chin Seng Chua and R. Jarvis. Point signatures: a new representation for 3d object recognition. *International Journal of Computer Vision*, 25(1):63 – 85, 1997.
Point signatures are rotation and translation invariant features that can be defined for each point of a (free-form) surface. Their calculation does not need the computation of partial derivatives. The matching of an object with an object database is done by an indexing scheme: The precomputed invariants for a given model are hashed into a lool-up table with references to the model from which they were extracted.

[11] L. Cloutot, X. Laboureux, and G. Häusler. Some medical applications of high speed PMT.

Chair for Optics, Annual report, 1998.

[12] S.V. Dickinson, A. P. Pentland, and A. Rosenfeld. From volumes to views: An approach to 3-d object recognition. *CVGIP: Image Understanding*, 55(2):130–154, 1992.

[13] M. P. do Carmo. *Differentialgeometrie von Kurven und Flächen*. Vieweg, 1993.

[14] O.D. Faugeras and M. Hebert. The representation, recognition, and locating of 3-d objects. *The International Journal of Robotics Research*, 5(3):27 – 51, 1986.
Given two surfaces to be registered a rough estimate of the displacement is found with the help of detected curves on the object where the surface undergoes a C^1 discontinuity and with the help of detected surface primitives (planes, quadrics). To match the primitives a Hough transform technique is suggested but rejected for their purpose in favour of a tree search technique. To find an exact transformation a minimization problem is formulated for corresponding points, line segments and planes. For points and planes with given correspondences a closed form solution is described with the help of quaternions; for line segments an iterative solution is proposed (no ICP).

[15] J. Feldmar and N. Ayache. Rigid, affine and locally affine registration of free-form surfaces. *International Journal of Computer Vision*, 18(2):99 – 119, 1996.
Given two surfaces to be registered a rough estimate of the displacement is found with the help of curvatures and principal frames. Then the best rigid displacement is found via an iterated closest point (ICP) algorithm. More sophisticated methods like searching the best affine transformation and looking for local affine transformations are explained. Experimental results are given.

[16] J.M. Galvez and M. Canton. Normalization and shape recognition of three-dimensional objects by 3d moments. *Pattern Recognition*, 26(5):667–681, 1993.

[17] H. Goldstein. *Classical mechanics*. Addison-Wesley, 1980.

[18] Jonas Gomes. Course notes #3. In *Warping and Morphing of Graphical Objects*, SIGGRAPH '95, 1995.

[19] Berthold K.P. Horn. Extended gaussian images. *Proc. IEEE*, 72(12):1671 – 1686, 1982.

[20] Berthold K.P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, A-4:629 – 642, 1987.
Given two point sets p_i and p'_i related by $p'_i = Rp_i + T$ (R rotation, T translation) and the correspondences between the points a closed-form solution to the (weighted) least-squares problem of finding R and T is presented which is based on quaternions.

[21] Berthold K.P. Horn. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America*, 5(7):1127 – 1135, 1988.
Given two point sets p_i and p'_i related by $p'_i = Rp_i + T$ (R rotation, T translation) and the correspondences between the points a closed-form solution to the (weighted) least-squares problem of finding R and T is presented which is based polar value decomposition.

[22] K. Kanatani. Analysis of 3-d rotation fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):543 – 549, May 1994.
Given two point sets p_i and p'_i related by $p'_i = Rp_i + T$ (R rotation, T translation) and the correspondences between the points a closed-form solution to the (weighted) least-squares problem of finding R and T is presented. The three existing methods – the method of singular value decomposition, the method of polar value decomposition, and the method of quaternion representation – are extended.

[23] S.B. Kang and K. Ikeuchi. The complex egi: A new representation for 3d pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(7):707 – 721, July 1993.

[24] X. Laboureux, S. Seeger, and G. Häusler. Computation of curvatures from 2.5d raster

data. *Chair for Optics, Annual report*, 1999.

[25] P. Liang and J.S. Todhunter. Representation and recognition of surface shapes in range images: a differential geometry approach. *Computer Vision, Graphics, and Image Processing*, 52:78 – 109, 1990.
Given an object model and a range image (sampled over a regularly spaced 2D rectangular grid) to be registered a rough estimate of the displacement is found with the help of a surface shape segmentation into regions with the same sign of Gaussian and mean curvatures and with the help of a relational graph representing the connectivity of the segmented surface regions. Each segmented surface region is represented as a 2D edge-node graph and each node represents a connected surface patch whose unit normals fall within a tessellated cell on the unit sphere (generalization of the EGI called UNSDLA). A node stores the average of one of the unit principal vectors and the average principal curvatures of the surface patch. Matching is formulated as an optimization problem by minimizing a matching error computed between principal vectors and principal curvatures of the surfaces. Experimental results are given.

[26] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE-Transactions-on-Medical-Imaging*, 16(2):187–198, April 1997.
A new approach to the problem of multimodality medical image registration is proposed, using a basic concept from information theory, mutual information (MI), or relative entropy, as a new matching criterion. The method presented in this paper applies MI to measure the statistical dependence or information redundancy between the image intensities of corresponding voxels in both images, which is assumed to be maximal if the images are geometrically aligned. Maximization of MI is a very general and powerful criterion, because no assumptions are made regarding the nature of this dependence and no limiting constraints are imposed on the image content of the modalities involved. The accuracy of the MI criterion is validated for rigid body registration of computed tomography (CT), magnetic resonance (MR), and photon emission tomography (PET) images by comparison with the stereotactic registration solution, while robustness is evaluated with respect to implementation issues, such as interpolation and optimization, and image content, including partial overlap and image degradation. Our results demonstrate that subvoxel accuracy with respect to the stereotactic reference solution can be achieved completely automatically and without any prior segmentation, feature extraction, or other preprocessing steps which makes this method very well suited for clinical applications.

[27] J.B. Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1 – 36, 1998.
A survey of recent (published in 1993 or later) publications concerning medical image registration techniques (based on more than 300 reviewed papers).

[28] T. Masuda, K. Sakaue, and N. Yokoya. Registration and integration of multiple range images for 3-d model construction. In *IEEE Proceedings of ICPR '96*, pages 879 – 883, 1996.
Given several point sets and good initial transformations between them an iterative closest point (ICP) algorithm is described which minimizes the median of squares of corresponding points (least median of squares (LMS) estimator) to find the exact displacements between the range images. With the help of the LMS estimator a robust classification of range data into inliers and outliers is performed. (No global optimization for the several point sets is done.) Experimental results are given.

[29] H. Morneburg, editor. *Bildgebende Systeme fr die medizinische Diagnostik*. Siemens Ak-

tiengesellschaft, Publicis MCD Verlag, Erlangen, 1995.

[30] R. Nevatia and K. R. Babu. Linear feature extraction and description. *Computer Graphics and Image Processing*, 13:257–269, 1980.

[31] X. Pennec and J.P. Thirion. Validation of 3-d registration methods based on points and frames. Technical Report 2470, INRIA, Jan. 1995.

[32] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1996.

[33] P.J. Rauss, P.J. Phillips, M.K. Hamilton, and A.T. Depersia. The feret (face recognition technology) program. In *25th AIPR Workshop. Emerging Applications of Computer Vision*, volume 2962 of *Proceedings-of-the-SPIE – The-International-Society-for-Optical-Engineering*, pages 253–263, Washington, DC, USA, 16-18 Oct 1996.

The mission of the Department of Defense Counter-drug Technology Development Program Office's Face Recognition Technology (FERET) program is to develop automatic face-recognition systems for intelligence and law enforcement applications. To achieve this objective, the program supports research in face-recognition algorithms, the collection of a large database of facial images, independent testing and evaluation of face-recognition algorithms, construction of real-time demonstration systems, and the integration of algorithms into the demonstration systems. The FERET program has established baseline performance for face recognition. The Army Research Laboratory (ARL) has been the program's technical agent since 1993, managing development of the recognition algorithms, database collection, and conducting algorithm testing and evaluation. Currently, ARL is managing the development of several prototype face-recognition systems that will demonstrate complete real-time video face identification in an access control scenario. This paper gives an overview of the FERET program, presents performance results of the face-recognition algorithms evaluated, and addresses the future direction of the program and applications for DoD and law enforcement.

[34] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and outlier detection.* John Wiley, 1987.

The problem of the influence of outliers in regression analysis is discussed (regression outliers can't be discovered by looking at the least squares residuals!). Since the simple least squares estimator is extremley sensible to outliers (breakdown point of 0% which means that only one outlier can corrupt the whole regression) more robust estimators are described (e.g. LMS, LTS, S-estimators). The LMS (least median of squares) estimator (breakdown point of 50%) is the robust estimator with least computation times in experiments.

[35] H. Schönfeld, G. Häusler, and S. Karbacher. Reverse engineering using optical 3d sensors. In *Three-Dimensional Image Capture and Applications*, volume 3313 of *Proceedings of SPIE*. R. N. Ellson and J. H. Nurre, 1998.

[36] A. Singh and M. Shneier. Grey level corner detector: A generalization and a robust real time implementation. *Computer, Vision, Graphics, and Image Processing*, 51:54–69, 1990.

[37] G. Stockman, S. Kopstein, and S. Benett. Matching images to models for registration and object detection via clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(3):229 – 241, 1982.

Given two intensity images the 2-D rotation, translation and scaling between the images is found without any a priori knowledge. In a first step point features are detected in the images. Two point features define a vector. If n point features are given there are n^2 possible vectors in one image. One combination of vectors from both images define a transformation (rotation, translation, scaling). Each transformation is filled in a cluster space ($O(n^4)$). The maximum in the cluster space determines the right transformation.

This method uses the fundamental idea of the Hough method. Therefore clustering should be better named *Hough method*.

[38] A.J. Stoddart and K. Brunnström. Free-form surface matching using mean field theory. In *British Machine Vision Conference*, pages 33 – 42, Edinburgh, UK, 1996. http://www.ee.surrey.ac.uk/pub/vision/papers/stoddart-mft-bmvc96.ps.Z .
An approach for finding the inital guess of the rotation and translation between the object and the model surface in the free-form surface matching problem is described. An objective function that measures the match quality is designed and then used in an algorithm based on mean field theory (MFT). Experimental results are given.

[39] A.J. Stoddart and A. Hilton. Registration of multiple point sets. In *13th Int. Conference on Pattern Recognition*, pages B40 – 44, Vienna, Austria, 1996. http://www.ee.surrey.ac.uk/pub/vision/papers/stoddart-icpr96.ps.Z .
Given several range images from different viewpoints and good initial transformations for these images the described algorithm minimizes the registration errors of all views simultaneously with one cost function for all views. Experimental results are given.

[40] G. Subsol, J.P. Thirion, and N. Ayache. A scheme for automatically building three-dimensional morphometric anatomical atlases: application to a skull atlas. *Medical Image Analysis*, 2(1):37 – 60, 1998.
Given several surfaces (segmented by intensity thresholding from CT scans) of different subjects (different skulls) a feature registration algorithm is presented (for building a morphometric anatomical atlas). In a first step crest lines are extracted as features. To identify point and line correspondences between two surfaces an adapted ICP (iterated closest point) algorithm is performed on the points of the crest lines that takes into account the strong topological constraint of a line as an ordered list of points. The transformation used in the ICP is a rigid one during the first iterations, then an affine one and at the end a spline transformation to model more local and complex deformations. In addition a common feature identification is performed (common feature subsets form the structure of the atlas) and the average of the common feature positions and their variabilities are determined.

[41] R. Szeliski and S. Lavalle. Matching 3-d anatomical surfaces with non-rigid deformations using octree-splines. *International Journal of Computer Vision*, 18(2):171 – 186, 1996.

[42] J.P. Thirion. New feature points based on geometric invariants for 3d image registration. *International Journal of Computer Vision*, 18(2):121 – 137, 1996.
Given two 3D images (e.g. MRI) of the same subject, taken in two different positions and with the same acquisition device, a rough estimate of the displacement is found with the help of extremal points and geometric invariants characterizing these points (but not with the help of the principal frames in these points). In a prediction-verification scheme the rough transformation is found (3 points to 3 points means $O(n^6)$; complexity reduction with the help of 28 invariants). Then the best rigid displacement is found via an iterated closest point (ICP) algorithm. Experimental results are given.

[43] Wu-Ki Tung. *Group theory in physics*. World Scientific Publishing Co Pte Ltd., 1985.

[44] G. Turk and M. Levoy. Zippered polygon meshes from range images. In A. Glassner, editor, *Proceedings of SIGGRAPH '94*, Annual Conference Series, pages 311 – 318. ACM SIGGRAPH, July 1994.
Given several range images from different viewpoints and good initial transformations for these images a method for combining them into a single polygonal mesh is described. To find the exact rigid transformation between the images (first a triangle mesh is created for each range image) an iterative closest point (ICP) algorithm in combination with a quater-

nion based method and the acceleration proposed by Besl [3] (use linear and quadratic extrapolation of registration parameters) is used to minimize the weighted (confidence values given from dot product of normals and vector pointing to light source) least squares sum of corresponding points (corresponding points = closest points defined by minimal distance of point to C^0 continuous surface). In addition to Besl [3] a distance threshold (twice the spacing between range points) for outlier detection is introduced in the process of finding closest points and a restriction to points not lying on the boundary is added. To accelerate the algorithm a coarse to fine strategy in taking points during the iterations is used (mesh hierarchy in which each mesh uses one-forth the number of range points that are used in the next higher level). To overcome the problem of compounding registration errors when matching several range images, all meshes are registered to a single mesh created from a cylindrical range image. Experimental results are given.

[45] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:376 – 380, 1991.
Given two point sets p_i and p'_i related by $p'_i = Rp_i + T$ (R rotation, T translation) and the correspondences between the points a closed-form solution to the least-squares problem of finding R and T is presented. It is correction to the method of Arun et al. [1].

[46] P. Viola. *Alignment by maximization of mutual information.* PhD thesis, MIT Department Electrical Engineering and Computer Science, Cambridge, Mass., 1995.
Theory of mutual information is described in detail. Alignment experiments are presented.

[47] P. Viola and W.M. Wells III. Alignment by maximization of mutual information. In *Fifth International Conference on Computer Vision*, Cambridge, Massachusetts, USA, 1995. IEEE.
Given two data sets (of different modalities) a method for finding the pose of an object in an image is presented that is based on maximization of their mutual information. The technique does not require information about surface properties of the object, besides its shape, and is robust with respect to variations of illumination. Experimental results are given.

[48] M.W. Walker, L. Shao, and R.A. Volz. Estimating 3-d location parameters using dual number quaternions. *CVGIP: Image Understanding*, 54(3):358 – 367, 1991.
Given two point sets p_i and p'_i related by $p'_i = Rp_i + T$ (R rotation, T translation) and the correspondences between the points a closed-form solution to the least-squares problem of finding R and T is presented which is based on dual number quaternions.

[49] Wu Wang and S.S. Iyengar. Efficient data structures for model-based 3-d object recognition and localization from range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):1035 – 1045, Oct. 1992.
Given a model database and a range image, objects in the range image will be recognized and localized with the help of curvature information of surface segments. A detailed description how to get differential information from range images is given. The recognition process is done in a prediction-verification scheme using some heuristics: taking surface segments s of the range image one after the other it is searched for all segments in the model database which match to all taken and matched s so far(using that corresponding surface points have the same curvatures and the same distances to the center of the surface segment) until three surface segments of the range image match with surface segments of the same model in the database. To verify the recognition hypothesis the transformation between the model and the range image is calculated with the help of the central points of the surface segments.

[50] G. Welch and G. Bishop. An introduction to kalman filter. http://pc222.math.klte.hu:80/norbi/SpatStat/kalman.pdf.

[51] A.K.C. Wong, S.W. Lu, and M. Rioux. Recognition and shape synthesis of 3-d objects based on attributed hypergraphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(3):279 – 290, March 1989.
Given a range image of a (simple) object and a database of (simple) prototype objects an algorithm for object recognition is presented which is based on attributed hypergraph representations (AHR) of the objects. An attributed hypergraph is a collection of surfaces and primitive blocks represented by attributed graphs which are again collections of attributed vertices and attributed relations between them.

[52] Gang Xu and Zhengyou Zhang. *Epipolar Geometry in Stereo, Motion and Object Recognition*. Kluwer Academic Publishers, 1996.

[53] Y.Lamdan and H.J. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. *2nd Int'l Conf. Comput. Vision*, pages 238 – 249, 1988.

[54] Zhengyou Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(1):119 – 152, 1994.
Given two point sets and a good initial transformation between them an iterative closest point (ICP) algorithm in combination with the dual number quaternion method is described which minimizes the sum of squared residuals (least squares sum of corresponding points) to find the exact displacement. A statistical method based on distance distribution is used to recognize outliers (in each iteration of the ICP). The search process for closest points is accelerated with the help of k-D trees and a coarse-to-fine strategy during the iterations. Experimental results are given.