3D-2D REGISTRATION OF CURVED OBJECTS

Ákos CZOPF *, Christian BRACK **, Michael ROTH **
and Achim SCHWEIKARD **

*Department of Electromagnetic Theory
Technical University of Budapest
H–1111 Budapest, Egri J. u. 18. Hungary

**Department of Computer Science IX
Technical University of Munich
D–81667 Munich, Orleansstr. 34, Germany

E-mail: czopf@in.tum.de
Tel: (49)-89-48095-115, Fax:(49)-89-48095-203

Abstract

We introduce here a new scheme for 3D-2D registration of curved objects from images. Having a 3D model of an object, the goal of the registration is to determine the object's 3D position on the base of 2D images of the object. The optimization methods generally used in this field need a quite precise initial position to avoid the effect of local optima. This paper focuses on the computation of the initial, approximate position, that is typically obtained manually, but it is provided automatically by our new registration method. We represent object models by 2D aspects. As the model aspects and the images are characterized by the same repertoire of local geometric features, a classification of images can be performed by comparing the detected set of features. After identifying the aspect of the object in the images, the parameters of the initial position can be easily determined. The exact position is then the result of an optimization procedure. We are motivated by clinical surgical registration problems involving the registration of human bones. The experiments seem to verify that the presented method provides an adequate initial position information for the optimization procedure in a reasonable intra-operative execution time on a conventional workstation.

Keywords: registration, 3D-2D, curved objects, geometric features, scale-space.

1. Introduction

Having a 3D model of an object, the goal of 3D-2D registration is to determine the object's 3D position on the base of one or more 2D images of the object. The 3D position here includes 6 parameters of location (3) and orientation (3) with respect to a 3D reference coordinate system. In order to achieve this goal, the registration procedure has to find the transformation which maps the object model onto a given image of that object. This transformation is a composition of a 3D transformation (translation and rotation) on the model and a 3D-2D projective transformation of the transformed model onto the image. If this transformation is found, the 3D transformation component provides the result of the registration, i.e., the object's 3D position.

Most of the existing related systems can be classified as follows. Some of them perform an automatic recognition of objects in images, neglecting an accurate determination of the object's position [1]. Others focus on an accurate registration, but they usually require manual interactions of the user [2][3]. Finally, some novel systems can solve the registration problem automatically, but they do not meet our requirements, because of the rough projective transformation model [4] or because of the assumption that the images contain nothing else than the object itself [5].

Our registration system can perform an automatic registration of curved objects from images. Object models are represented by 2D aspects. As both the model aspects and the images are characterized by the same set of local geometric features, a classification of images can be performed by comparing the detected set of features. After identifying the aspect of the object in the images, the parameters of the initial position can be easily determined. The exact position is then the result of an optimization procedure.

The automatic calculation of the initial, approximate object position is the main object of this paper.

1.1. Motivation

Our research is motivated by clinical surgical applications, e.g., femur osteotomy, total knee replacement (see *Fig. 1*). The operation is preceded by the construction of a CT-based 3D model of the bone(s) of interest and by the planning of the surgical intervention [6]. Intra-operatively an initial registration procedure is performed to determine the relation of the different coordinate systems (fixed to the bone, to the detector, to the robot hand and to the calibration plate). This registration updated permanently by a 3D-localizer is the basis for the operation partially executed by a robot. The robot is equipped with a sawing device developed specifically for such interventions [7]. The supervised use of a robot can be reasonable in interventions that demand a high precision.

The initial registration is supported by an enhanced calibration procedure [8]. The goal of the calibration is to determine the position of the imaging device with respect to the 3D reference coordinate system and to compute the parameters of the camera model that describes the projective transformation (distortion inclusive) provided by the imaging device. After the calibration, the distortion can be removed from the images.

1.2. Requirements

We postulated the following requirements for the registration scheme to construct:
Employ pictorial information extractable from X-ray images

- The modality of image acquisition limits the types of features that can be extracted from the image. By applying X-ray imaging we abandon the features derived from texture, colour etc.
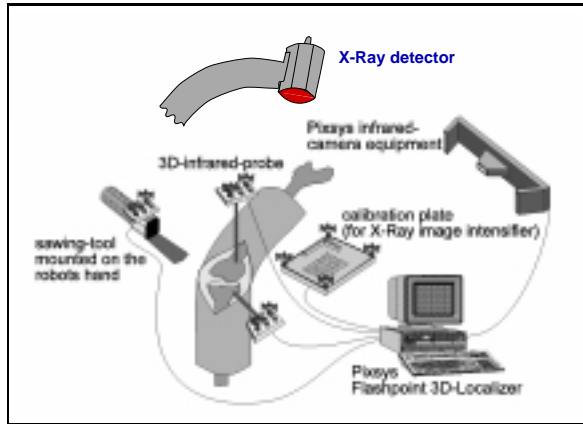
Fig. 1. Operation Scene. The calibration and the initial registration are supported by an X-ray imaging device (an image intensifier and a detector fixed to a C-arm); for simplicity only the detector is shown. After identifying the initial position a 3D-localizer tracks the bones (femur, tibia), the sawing-tool and the calibration plate, each equipped with an infrared probe.

Perspective projection between 3D and 2D

- There are some examples in the literature for using simplified projection models instead of perspective projection, e.g. [4], though the exact model would be the perspective one in fact. We want to handle the perspective projection explicitly.

Registration from parts

- To take the whole image of the bone is often impracticable with an X-ray detector, therefore the shots focus on the area of interest. Consequently, the system must be able to registrate the object if the whole object is not visible in the image, just a significant part of it.

Presence of additional objects

- The registration should work also in the presence of additional objects in the images (e.g. calibration landmarks or connected bones)

Minimally invasiveness

- Minimally invasiveness means for the registration to reduce the required radiation dose and the intra-operative execution time in order to protect the patient. Furthermore, the registration system must be based on natural landmarks, i.e., it must be able to handle the smooth form of a bone in order to avoid an invasive fixation of easily detectable artificial landmarks directly to the bone.

Precision

- The exact position provided by the registration is the basic spatial reference information for the operation partially executed by a robot. Robotic surgery has a sense when the precision of the operation can be increased by means of a robot. The precision of the robotic surgery is mainly influenced by the precision of the registration itself.

Robustness

- We need robustness to ensure that the registration works reliably for all patients under all possible environmental conditions. If it cannot do so, it should be also reported.

Automatic procedure

- The registration should be performed without user interactions.

It is apparent that some of the requirements are specific to our research environment, e.g., some of them meet demands corresponding to robotic surgery and X-ray imaging.

2. Previous Work

There are several methods developed during the last five years for the representation and recognition of curved objects. A general 2D shape representation method was suggested by SAUND [9]. He offers a scale-space blackboard architecture for maintaining geometric features as shape tokens and for manipulating them symbolically. See section 2.1 for details.

A viewer-centred shape representation method for 3D objects is that of the aspect graphs. After the aspect graphs were applied successfully for the shape class of polyhedra, some extensions were developed to build the aspect graph representation of curved objects analytically. While these extensions address still a restricted class of objects, like algebraic surfaces [10] or solids of revolution [11], our approach addresses a more general object class, namely the class of curved objects.

DICKINSON and PENTLAND combine object- and viewer-centred representations. They recover 3D volumetric primitives from 2D images and then assemble those primitives into an object-centred description [12]. However, in our application the X-ray images do not provide the proper information for region segmentation, and the 3D models built of such primitives could be only a rough estimate of the curved shape of a bone.

MAHMOOD's system [13] is an example for using additional properties such as colour and texture besides the usual geometric shape description. The interpretation of these additional properties is clarified for imaging modalities based on light reflection, but in the case of radiation transmission through the body as in X-ray imaging they are hardly interpretable. Consequently, the contour seems to

be the only reliable information in X-ray images which a segmentation and feature extraction can be based on.

Model-based object recognition methods need to find a correspondence between the object model and the image of the object. Some of these methods are based on point correspondences (e.g.[14][15]). Because the automatic extraction of point features from images of curved objects is somewhat difficult and unstable, these methods cannot be easily employed in the registration of curved objects like bones. Consequently we apply feature correspondences. Though after feature extraction one could define a mapping of the detected more complex features to point features, it is not favourable because of the loss of information stored in additional attributes (e.g. orientation, scale).

An enhanced appearance-based representation and recognition system was published by POPE and LOWE [1]. They do not handle 3D objects directly, but in terms of their appearance in 2D images, and so they can perform a 2D-2D matching for recognition. An appearance-based representation in our case would require one or more X-ray shots of each representative aspect. This is impracticable in our context and misses the aim to be minimally invasive. Furthermore they focus on object recognition while we focus on exact object registration.

The several matching methods differ according to the transformation model they use. We adopt a 2D similarity transformation model suggested by AYACHE and FAUGERAS [16]for 2D-2D transformations, because it preserves angles and results in a quickly solvable linear least squares problem, while for 3D-2D transformation we use a perspective transformation model.

Some methods assume that the images contain exclusively the object to registrate [5], while others allow the presence of other objects and occlusion in the images [1][4].

In most cases a precise registration is the result of an optimization procedure [2][4][3] that needs a proper initial position to avoid the effect of local optima. This initial position is typically obtained manually (e.g. [2][3], but it is provided automatically by our system.

2.1. General Purpose Shape Representation by Saund

The problem of visual shape representation is to determine what information about objects' shapes should be made explicit in order to support later visual processing tasks. Our visual knowledge can be built into a shape representation in the form of hierarchical descriptive vocabularies making explicit the important spatial events and geometrical relationships. In order to construct such a representation Saund suggests a *scale-space blackboard* architecture.

On the blackboard geometric features can be handled symbolically as *shape tokens* of several types. Shape tokens mark the occurrence of a shape fragment or some configuration of such fragments. Each token has location, orientation and size, and may possess additional attributes according to its type in order to depict the additional properties of the marked shape fragments. Attributes are represented in a scale-normalized (magnification-independent) way, e.g., geometric distances
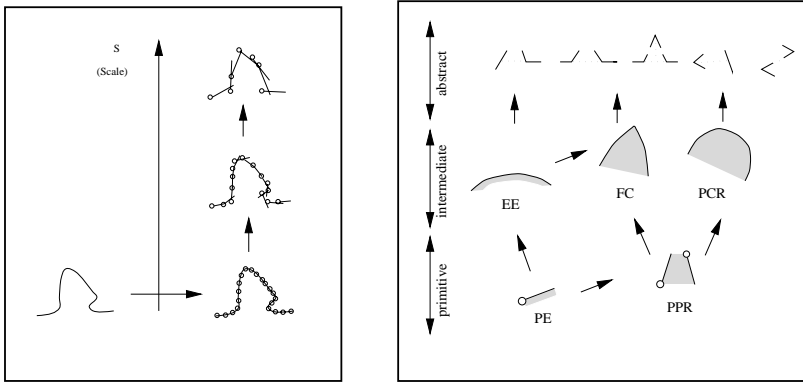
Fig. 2. Modalities of grouping a) fine-to-coarse: shape representation by PRIMITIVE-EDGEs of successive scales, b) primitive-to-abstract: previously extracted shape fragments support the extraction of more complex features; arrows imply grouping [9].

are given relative to the tokens' size. A strength attribute can be assigned to each token that indicates how correctly the model described by the token captures the given shape fragment. Scale-space means here that a (logarithmically graduated) scale dimension is added to the primal sketch. Consequently, the shape events can be indexed not just by spatial location, but also by size.

There are five types of shape tokens suggested by Saund, namely

PRIMITIVE-EDGE (PE): simple figure/ground boundary

PRIMITIVE-PARTIAL-REGION (PPR): region partially enclosed by a pair of PRIMITIVE-EDGEs

EXTENDED-EDGE (EE): collection of PRIMITIVE-EDGEs falling along a circular arc

PARTIAL-CIRCULAR-REGION (PCR): roughly circular region partially enclosed by the bounding contour; a collection of PRIMITIVE-PARTIAL-REGIONs

FULL-CORNER (FC): two contours roughly forming a wedge; asserted by two EXTENDED-EDGEs or by a collection of PRIMITIVE-PARTIAL-REGIONs

For a detailed description of the token types, see [9][17].

The feature extraction starts with the *primal sketch* by indicating the contour fragments with PRIMITIVE-EDGE tokens at the finest scale of the blackboard. The description of a shape is realized by *grouping* operations over the shape tokens; a fine-to-coarse grouping along the scale dimension and a primitive-to-abstract grouping according to the level of abstraction (see *Fig. 2*). The role of fine-to-coarse grouping is to gain successively coarser descriptions of the contour by asserting tokens of the same type at larger scales, while the role of primitive-to-abstract

grouping is to assert other types of shape tokens representing more complex shape fragments.

Shape vocabularies of particular shape domains constitute the most abstract level of this description. A shape vocabulary consists of a set of shape descriptors addressing specific shape fragments of the given domain. A shape vocabulary of this type may support later visual tasks such as distinguishing shapes on the basis of subtle differences in geometry, if their overall configuration is common.

As a conclusion, we summarize some relevant advantages of this representation:

- symbolic manipulation of geometric features
- fast indexing mechanism
- magnification-independency of spatial configurations
- pyramid-style image representation

3. The New Registration Scheme

A sketch of our 3D-2D registration scheme is shown in *Fig. 3*. First an object model is constructed based on 3D data. Feature extraction is then performed on a representative set of 2D views of the model and selected features are stored for each view in the model description. These steps are arranged off-line. The image description is constructed after feature extraction and selection from a 2D image of the object. The model views and the image are characterized by the same repertoire of local geometric features. The matching between a model view and an image involves the search for model-image feature correspondences, the generation of matching hypotheses and the evaluation of such hypotheses by a comparison of the two corresponding feature sets. The goal of image classification is to find the model view with the best qualified match. After identifying the aspect of the object in the image this way, the parameters of the approximate position can be easily determined. While the approximate object position is typically obtained manually, by our method it is calculated automatically. The exact position of the object is a result of an optimization procedure. This optimization procedure needs a good initial estimate of the position, where the previously calculated approximate object position can provide this initial value. The accuracy of the registration can be improved by using more than one image, capturing the object from different viewpoints. After the exact initial registration the object position can be permanently updated by tracking.

Our new registration scheme involves, consequently, feature extraction from images (section 3.1), representation of images (3.2) and objects (3.3), classification of images (3.4) and calculation of the approximate position (3.5). (The other components of our registration procedure, namely the previous calibration and the
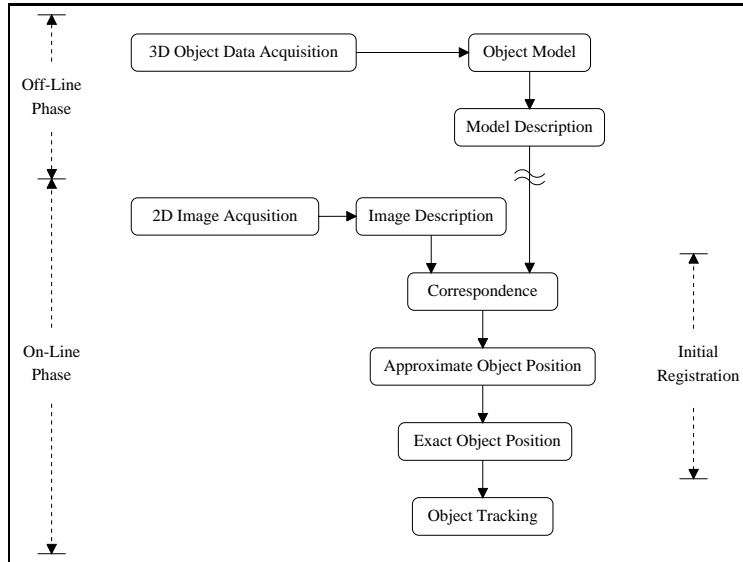
Fig. 3. Image processing steps of a model-based 3D-2D registration scheme.

following optimization have been already published in [8][2].)

3.1. Feature Extraction

With regard to feature extraction the base of our approach is an extension of the shape representation method suggested by Saund. Our extension manages also grey-scale images, not just binary ones and includes another representation of feature configurations.

3.1.1. Grey-Scale Images

The first challenge is to handle grey-scale images instead of binary ones to be able to extract features from X-ray images as well. The first step of the feature extraction is the generation of the primal sketch, where we transform a matrix of image intensities (in our case grey-scale values) into a set of symbols, namely a set of PRIMITIVE-EDGE tokens on the first level of the blackboard. These tokens represent a simple figure-ground boundary for binary images. However, an enhanced contour detection can provide similar information about the shape's contour for grey-scale X-ray images with the difference that due to the properties of X-ray imaging some internal contours of the shape may be also detected. Taking the results of the detection we can determine the parameters of the PRIMITIVE-EDGE tokens on the first level of the blackboard.

We apply at first a Canny–Deriche filter on the image to obtain smooth gradients [18]. Then the local maxima of the gradients are extracted in the direction of the gradient by the non-maximum suppression. The following step is the hysteresis thresholding [19] to gain connected components of the contour (see *Fig. 4*). This contour detection method provides a set of contour points with the parameters location, contour orientation and contour amplitude.

Alternatively we can use a semi-automatic segmentation with manual interactions that provides a closed silhouette of the shape and ignores internal contours. This can improve the results of the registration first of all, in the case of poor X-ray image quality. However, the later stages of the registration do not need such an exact segmentation definitely.
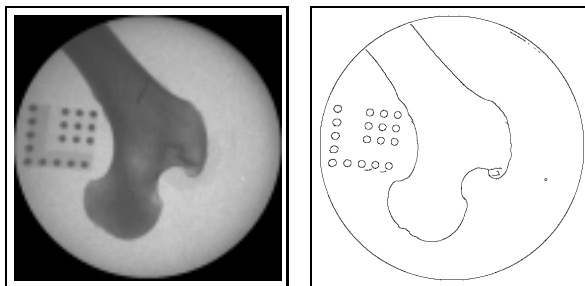
Fig. 4. A grey-scale X-ray image of the femur and the result of the contour extraction. Calibration plate lies next to the bone.

The PRIMITIVE-EDGE tokens on the first level have size, location, orientation and strength parameters. The size is set to one concerning to one pixel. The locations and orientations of the edges are provided by the locations and orientations of contour points found by our contour detection method. The strength parameter can be set to the value of the contour amplitude at the corresponding point that we get as the output of the contour detection method or an appropriate function of it. This way we have gained all the parameters that we need in the primal sketch.

3.1.2. Basic Features

We handle local geometric shape features in terms of *shape tokens* on a *scale-space blackboard*. Apart from some minor modifications the five basic shape tokens we use are the same as by Saund (see *Fig. 5* below the dotted line). We applied a fast curve fitting algorithm [20] in the grouping of EXTENDED-EDGES and limited the number of FULL-CORNER attributes.

The strength attribute holds two pieces of information: the contour strength (amplitude) of the contour fragment that the token is based on, and the goodness

of the representation, i.e. how correctly the model described by the token captures
the given shape fragment.
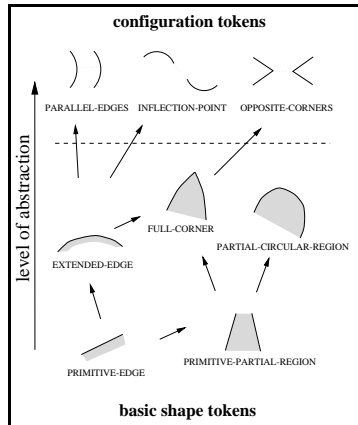
Fig. 5.   Types of shape tokens. All types of basic tokens and some representatives of
configuration tokens are shown. Arrows imply grouping.

3.1.3. Feature Configurations

According to Saund a feature configuration constituting a specific shape fragment
can be stored as a vocabulary descriptor. A shape vocabulary composed of such
descriptors may support a visual task such as distinguishing shapes on the basis of
subtle differences in geometry, if their overall configuration is common, but it is not
suitable for distinguishing shapes, if their overall configuration differs significantly.
For example, if our task were to distinguish dbifferent patients on the basis of a given
aspect of their thigh bone, we could apply a vocabulary of this type. But our task
is to distinguish different aspects of the thigh bone of a given patient. Due to this
difference we use another representation of feature configurations. We represent
feature configurations as tokens at more abstract levels. These *configuration tokens*
are not planned to address a specific, explicitly named shape fragment (like top
corner of a fish notch by Saund), but to address a typical shape moment of a particular
shape domain by defining a restricted class of spatial relations of nearly-located
basic shape tokens (see *Fig. 6*). This shape domain is that of the curved objects in
our case and the newly developed configuration tokens used in our experiments are
(see also *Fig. 5* above the dotted line)

PARALLEL-EDGE**s:** a pair of roughly parallel EXTENDED-EDGEs
INFLECTION-POINT: a pair of EXTENDED-EDGEs assigning an inflection-point
OPPOSITE-CORNER**s:** a pair of oppositely located FULL-CORNERs

The attributes and grouping conditions of these configuration tokens are detailed in the Appendix.

Image Moments Represented by Configuration Tokens

PARALLEL-EDGES   INFLECTION-POINT   OPPOSITE-CORNERS

EXTENDED-EDGE   FULL-CORNER   PARTIAL-CIRCULAR-REGION

PRIMITIVE-EDGE (scale 5)   PRIMITIVE-EDGE (scale 6)   PRIMITIVE-EDGE (scale 7)
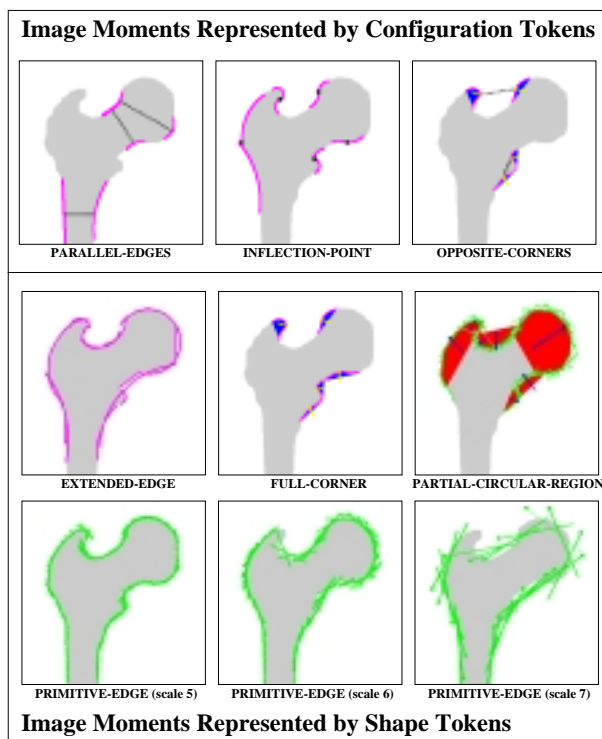
Image Moments Represented by Shape Tokens

Fig. 6. Result of token grouping. Tokens indicate several shape fragments of the femur's projection.

If the configuration tokens are planned suitably, *data reduction* can be achieved, too. Dealing with curved objects, it seems to be particularly useful in the representation of arcs. After grouping, the number of EXTENDED-EDGE tokens is typically between 100 and 500 in our experiments, while the number of PARTIAL-CIRCULAR-REGIONs and FULL-CORNERs is below 30. As defining the thresholds in pruning of arcs can be critical, we prefer providing the data reduction through the definition of configurations. The grouping procedure of configuration tokens is fast, and the time used for this grouping is recovered multiply in the matching phase. (Though essentially all token types, except the PRIMITIVE-EDGE, are derived from a configuration of other tokens, we use the term "configuration token" exclusively for tokens of non-basic types.)

The effectiveness of the data reduction can be further improved by deliberate pruning of redundant configuration tokens. It is practical especially in the case of point features like the INFLECTION-POINT. The same inflection point may be

indicated by several INFLECTION-POINT tokens, arisen from pairs of EXTENDED-EDGEs at several scales. Nearby INFLECTION-POINT tokens are compared pairwise and redundant tokens are removed. The strength parameter can be extremely helpful in pruning.

3.2. Representation of Images

We represent images in terms of *feature sets*. Each entry of such a set holds the parameters of one token. In this representation we consider the configuration tokens and usually the tokens of the most abstract basic types (PARTIAL-CIRCULAR-REGION, FULL-CORNER). The parameters in one entry :

- index of token type
- position of token
  (location$_x$, location$_y$, orientation, size)
- additional attributes
  (e.g. angle, curvature, etc.)

If additional knowledge about the stability of features is available the ordering of entries according to stability can speed up the matching.

3.3. Representation of Objects

In our registration approach the identification of the position parameters is decomposed. The first step in determining the approximate position of the object is to identify which aspect of it can be seen in the current image. After identifying the aspect the further parameters can be calculated easily.

This first step could be realized also by a 3D-2D matching scheme, but without using additional knowledge about the possible set of aspects it can be too time-consuming to calculate the projections between 3D and 2D, to extract the features and to create the description all intra-operatively in our applications. Therefore we prefer using a *multiple-view representation* with a 2D-2D matching scheme. This approach enables us to calculate the projections and to arrange feature extraction and model description all pre-operatively. Multiple-view means that the object is represented by a set of its views. These representative views are projections of the 3D object model stored as images. Accordingly, the model description consists of a set of image descriptions for these representative views, and the classification of an arbitrary image of the object can be performed by 2D-2D matchings between the image and the representative views.

3.3.1. The Projective Transformation

The 3D-2D projective transformation model applied here can be a weak perspective (orthographic projection with scaling) or a perspective one. A *weak perspective projection* model can be used only for a restricted set of aspects, where the depth of the images is small, i.e, the object's extent perpendicular to the image plane is relatively small (or equivalently in our context the longitudinal axis of the bone is nearly parallel to the image plane). In this case the error caused by the coarse transformation model is minimal. To the weak perspective projection model applies that for a given viewing direction the projections differ only in scale while the image plane nears to the object. Consequently, the weak perspective projection model has the advantage that the estimation of the transformation arrangement can be avoided, i.e., a simple orthographic projection model extended with a scale parameter can be employed to create the projections for the representative views.

The *perspective projection* model represents the 3D-2D projective transformation perfectly in our case. The only drawback of the perspective projection model that we prefer is that the expected transformation arrangement (expected relative position of the imaging device and the object) must be estimated for the projections pre-operatively. This is practicable in contexts where the distance of the object from the imaging device is constrained. Alternatively one can use more than one representative views for each aspect with different transformation arrangements, see e.g. [1].

3.3.2. Aspect Model Database

The model description involving the descriptions of the representative views is stored in the aspect model database. The steps of creating the aspect model database:

1. `Generate the 3D model of the object`
   We reconstruct the model of the 3D object from computed tomography scans, but any other method for 3D model generation is acceptable that a projection can be derived from.

2. `Take an estimate of the expected transformation arrangement`
   Before calculating the object's perspective projection the expected relative position of the imaging device with respect to the object must be estimated. Representative views of the 3D object are gained by a Gaussian sphere representation as usual, i.e., the expected transformation arrangement is estimated by a sphere including the fixed object at the center. Viewpoints are specified on the sphere with a viewing direction pointing to the center. The radius of the Gaussian sphere is stored. Note that this representation implicates that the viewpoints are at the same distance from the center. In our context where the X-ray imaging device mounted on a C-arm can be moved around the object,

this representation of aspects is appropriate. However, for particular contexts one may have to choose a corresponding representation.

3. `Choose a reference point in the 3D model`
   The reference point of the 3D model is aligned with the origin of the viewing sphere. Commonly we choose the model's center of gravity as reference point. The three coordinates of the reference point with respect to the 3D model coordinate system are stored.

4. `Choose a rotation step between neighbouring viewpoints`
   The density of viewpoints for representative views is determined by a rotation step.

5. `Calculate the corresponding views of the 3D model`
   For each viewpoint the corresponding view is calculated and the two rotation parameters are stored. Though the Gaussian sphere representation implies that the object is fixed and the virtual detector moves, in fact, we fix the position of the detector and rotate the object accordingly. The resulted views are the same. The coordinate system of the detector is defined so that the X and Y axes are parallel to the image plane and the Z axis is perpendicular to it. Rotating the object around two axes, X and Y are sufficient now to generate the views, while rotation around the Z axis, that means a rotation in the image plane, is implicitly set to zero.

6. `Extract the features from each projection`
   Do feature extraction as described in section [3.1].

7. `Store the result in feature sets`
   `(one for each projection)`
   See section [3.2] for details.

The generation of the aspect model database is fully automated in our implementation. Steps 2-4 are regulated by input parameters. Steps 5-7 are executed in a cycle.

3.4. Classification of Images

By now we have the representative views and the current image described in terms of feature sets. The aim of the classification is to identify which aspect of the object can be seen in the current image. In order to classify the image, its feature set is compared to those of the representative views. The result of the classification are the first n representative views that are most similar to the current image, i.e., their feature sets are nearest in some sense to the feature set of the current image, where n is a free parameter of the algorithm. The main components of the classification procedure are as follows.

### 3.4.1. Feature Correspondences

In order to enable a matching between two feature sets, a correspondence between them must be established. This correspondence is realized in terms of model-image *feature pairs*. Such pairs consist of two features of the same type hypothesized to represent corresponding shape events. We apply a kind of *alignment* method first used by ULLMAN [21] to align the model with the image. At first an initial pair is sought that determines a transformation between the model and the image. Then model features are projected into the image according to this transformation and additional pairs are sought. Here a bounded error model is used, i.e., pairs are only accepted, if the distance of the image feature and the projected model feature is below a predefined value. In contrast to some known implementations (e.g. [22][1]) we apply full backtracking in the maintenance of feature pairings to try alternate pairings. Thus a replacement of the initial pair is also possible. This opportunity is definitely needed for images of curved objects, where the extractable features are not often robust enough to guarantee an adequate initial pair. Consistency checking [23] is applied to ensure that features of a new pair are still unused.

3.4.2. 2D Similarity Transformation

A 2D similarity transformation model is used to project model features into the image as in [16] , though it does not describe the transformation between two perspective views perfectly. This transformation can be decomposed into translation, rotation and scaling, and for a set of image-model feature pairs it results in a quickly solvable linear least squares problem. Thus the resulting transformation projects the set of model features into the corresponding set of image features providing minimal distance in least square sense. The position parameters are previously standardized. The 2D similarity transformation preserves a relatively large set of geometric properties, such invariants are ratio of lengths, ratio of areas, size of angles, parallelism and collinearity. The property of preserving the size of angles is crucial, because in our representation of images some attributes and grouping conditions are defined as angles.

3.4.3. Match Quality

In order to rank matches we need to evaluate the quality of a match. For the calculation of the match quality value we use a combined weighted least squares formula of the distance of corresponding model-image features after the similarity transformation, the difference in the feature attributes and the size of the set of feature pairings. Each representative view is matched to the image. The highest match quality value and the four parameters of the corresponding similarity transformation are stored for each. Then the representative views are ranked according to
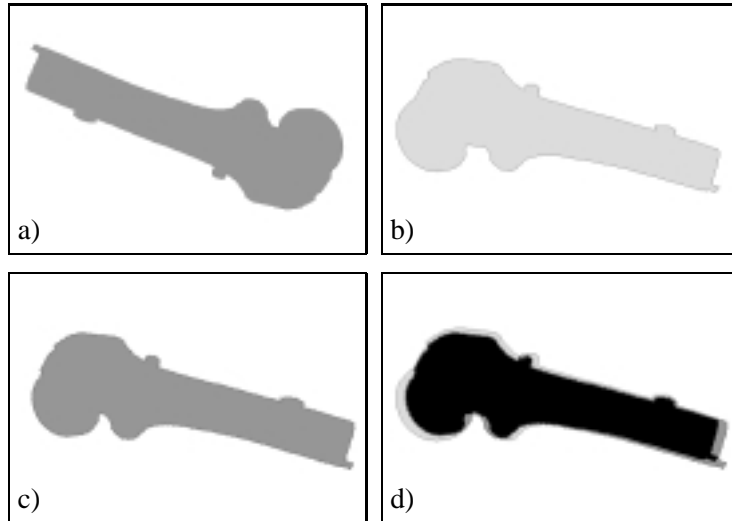
Fig. 7.   Image classification: a) The model view providing the best match for b) a synthetic
image of the femur. c) The model view after the similarity transformation. d) The
transformed model view aligned with the image; the joint region is indicated by
black. The model view and the image were characterized by a feature set of 23 and
25 features, respectively. The best match was provided by 14 model-image feature
pairs. The rotational difference between the best ranked model view and the image
is 17 degrees around the X axis and 7 degrees around the Y axis.

the match quality value and the first *n* of them are returned with the corresponding
transformation parameters as the result of the classification procedure. An example
for image classification is shown in *Fig. 7*.

3.4.4. Match Verification

We defined some heuristical thresholds for a match to be plausible, e.g., lower
bounds for the match quality value and for the proportion of matched model features.
If none of the representative views can provide a plausible match for an image, an
error is reported and other images of the object must be used for classification.

3.5. Approximate Object Position

The calculation of the approximate object position is simple after the classification
procedure. At first we reconstruct the 3D model's position with respect to the
coordinate system of the detector as stored in the aspect model database by the

representative view providing the best match quality. Let (x, y, z) and (α, β, γ) denote the model's location and orientation according to axes X, Y, Z, respectively.

The 3D object's reference point is translated so that its location in depth (z) is equal to the radius of the Gaussian sphere (r), while x and y are set to zero:

$$
\begin{aligned}
x &= 0, \\
y &= 0, \\
z &= r,
\end{aligned}
$$

reflecting that the viewing direction pointed to the center of the Gaussian sphere and the reference point was aligned with the center while the aspect models were recorded.

The model's two rotation parameters stored by the best representative view (rot_x, rot_y) can serve here as α and β, while γ is set to zero:

$$
\begin{aligned}
\alpha &= rot_x, \\
\beta &= rot_y, \\
\gamma &= 0.
\end{aligned}
$$

Now we have reconstructed the 3D model's position as recorded by the representative view providing the best match quality.

The 2D similarity transformation parameters of the classification hold the information how the model's representative view must be transformed to get the image of the object. Next we perform the corresponding 3D transformations on the 3D model to get the approximate position of the object. Let the notation for the similarity transformation be as follows: two translations (t_x, t_y), rotation θ and scale parameter s.

The rotation parameter of the similarity transformation is adopted for the object's orientation according to the Z axis:

$$
\gamma = \theta.
$$

The object's location in depth is in inverse proportion to the scaling parameter of the similarity transformation, e.g., if the scaling is equal to one the 3D model and the object have the same location in depth, while a scaling of two means that the object's location in depth is half of the location in depth of the model. As the model's location in depth was previously set to r we can write:

$$
z = \frac{r}{s}.
$$

According to the perspective transformation model 3D translations parallel to the image plane are proportional to the corresponding 2D translations in the image plane. This relation is determined by the object's location in depth and the focal

length of the camera model (f). In order to calculate the 3D location of the object we can write:

$$x = t_x \frac{z}{f},$$

$$y = t_y \frac{z}{f}.$$

This way we have gained the parameter set $(x, y, z, \alpha, \beta, \gamma)$ needed to reconstruct the object's 3D position in the coordinate system of the imaging device according to the image. Note that this object position may contain an error according to the last translation step that denotes that in contrast to the aspect model database the viewing direction does not point to the object's reference point in the image. This error is notable only if x or y is not significantly below z's order of magnitude.

Finally a coordinate transformation is performed from the coordinate system fixed to the detector into the reference coordinate system (in our case fixed to the calibration plate). If two or more images are used, then all the three location parameters can be calculated by determining the coordinates of the reference point in 3D space with a triangulation technique. A deliberate averaging of the provided rotation parameters can still improve the precision.

The resulting parameters provide the initial parameters of an optimization procedure [2]. The accurate location and orientation parameters returned by the optimization are the final results of the registration and can serve as reference information for the robotic intervention. If the optimization seems to converge to a local optimum, while the defined error is still over a threshold, the second parameter set is adopted and so on until the error gets below the desired threshold or all the n initial parameter sets are tried.

4. Experimental Results

We have developed a program to implement the described registration scheme. The pre-operative generation of an aspect model database of 648 aspect models (10-degree rotation steps around the X and Y axes) required approximately 2 hours. The intra-operative registration step of determining the approximate position of one image was always below 30 seconds. The average execution times of the several registration steps can be seen in *Table 1*.
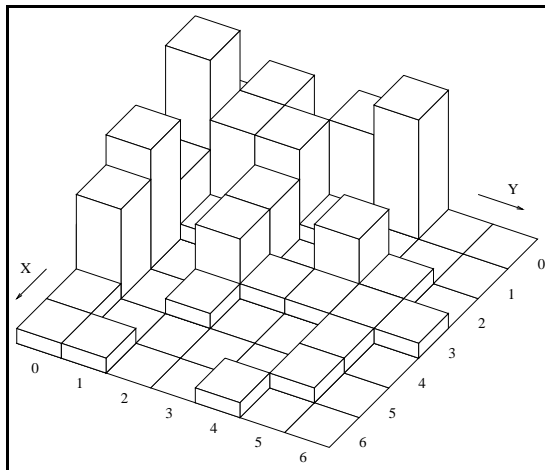
The approximation's accuracy for α and β, i.e., the rotation parameters around the X and Y axes that seems to be the crucial point of the registration was about 14 degrees for a test set of 60 images of the femur. A nearly correct approximation of α and β results in an even more accurate approximation of γ, and the approximation's error for x and y is less than one centimeter. When only one image is used for the registration the approximation of z is less precise, as it is estimated from the scaling parameter of the similarity transformation. This problem is easily solvable

Table 1. Average execution times of the several registration steps according to one image on a HP workstation (model 9000/735/B).

Projection	Primal Sketch	Feature Extraction	Configuration Extraction	Matching
1.52 sec	0.36 sec	9.87 sec	0.85 sec	17.43 sec
Pre-Operative Steps on Projections				
	Intra-Operative Steps on X-Ray Images			

by using two or more images and triangulation. *Fig. 8* shows the histogram of the approximation's accuracy for the rotation around the X and Y axes without verification. The verification produces a rejection rate of about 15%.

Fig. 8. Histogram of the approximation's accuracy for the rotation around the X and Y axes; differences between the best classified aspect and the real aspect of the bone for a test set of 60 images. X and Y axes are graduated in 10-degree rotation steps. Results with an accuracy less than 90 degrees (4 images of 60) are omitted.

5. Conclusion

We have introduced and implemented a new 3D-2D registration scheme for curved objects that is suitable among others for intra-operative registration of human bones from X-ray images. In order to evaluate our registration system we want now to reflect once more on the requirements (section 1.2) and examine how our implemented registration meets them. The system is based on image contours, so it employs pictorial information extractable also from X-ray images. Perspective projection between 3D and 2D is explicitly handled, but the reconstructed approximate position may contain significant error. A conflict emerges here between the requirements of accurate handling of perspective projection and minimally invasiveness that aims also to reduce intra-operative execution time. The chosen method is a compromise between accuracy and speed. Registration from parts and registration in the presence of additional objects is possible up to a certain limit adjusted by some heuristic thresholds. Minimally invasiveness is provided by the employment of natural landmarks and by the low radiation dose required for at most three X-ray images. The precision of the approximate registration is presented in the experiments. The robustness of the system must still be analyzed in the future. The registration is fully automated.

The first experiments show that the described method for determining the approximate position extended with the preceding calibration and the following optimization is suitable for intra-operative registration. In the future we want to focus on the analysis of the stability and robustness of our registration procedure.

Grouping Conditions and Attributes of Configuration Tokens
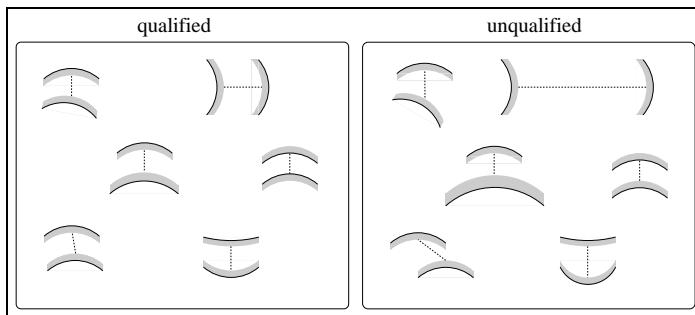
PARALLEL-EDGES

PARALLEL-EDGES are constructed from pairs of EXTENDED-EDGE tokens. Grouping conditions:

1. The EXTENDED-EDGEs must occur at the same scale.
2. The EXTENDED-EDGEs must have reverse orientations.
3. The EXTENDED-EDGEs' orientations must be perpendicular to the vector connecting their locations.
4. The EXTENDED-EDGEs must have the same scale-normalized curvature.
5. The EXTENDED-EDGEs must lie at a fixed, prespecified scale-normalized distance from one another.

The above requirements show the ideal grouping conditions. In fact, these configuration space requirements are defined as a range including the ideal case. Conditions 2–4 guarantee that the EXTENDED-EDGEs are roughly parallel. *Fig. 9* presents sample spatial configurations of two EXTENDED-EDGE tokens which do

and do not satisfy the configuration space requirements of the PARALLEL-EDGES token. Attributes:

- scale-normalized distance of the EXTENDED-EDGES
- average of scale-normalized curvature of the EXTENDED-EDGES

Fig. 9. Sample pairs of EXTENDED-EDGE tokens indicating the range of qualified spatial configurations for a PARALLEL-EDGES token.

INFLECTION-POINT

INFLECTION-POINT is constructed from pairs of EXTENDED-EDGE tokens. Grouping conditions:

1. The EXTENDED-EDGEs must occur at the same scale.
2. The EXTENDED-EDGEs must have the same orientations.
3. One of the EXTENDED-EDGEs must be convex, the other must be concave.
4. The EXTENDED-EDGEs must lie at a fixed, prespecified scale-normalized distance from one another.

Attributes:

- average of scale-normalized curvature of the EXTENDED-EDGES

OPPOSITE-CORNERS

OPPOSITE-CORNERS are constructed from pairs of FULL-CORNER tokens. Grouping conditions:

1. The FULL-CORNERs must occur at the same scale .
2. The FULL-CORNERs must have reverse orientations.

3. The FULL-CORNERs' orientations must be perpendicular to the vector connecting their locations.
4. The FULL-CORNERs must lie at a fixed, prespecified scale-normalized distance from one another.

Attributes:

- scale-normalized distance of the FULL-CORNERs
- average angle of the FULL-CORNERs.

References

[1] POPE, A – LOWE, D.: Learning Appearance Models for Object Recognition, In *Lecture Notes in Computer Science*, **1144** of *ECCV'96 Workshop on Object Representation in Computer Vision*, pp. 201–219, Springer, Berlin.

[2] LAVALLEE, S. – SZELISKI, R.: Recovering the Position and Orientation of Free Form Objects from Image Contours Using 3D Distance Maps, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **17**, 4, pp. 378–390, 1995.

[3] WEESE, J. – PENNEY, G. P. – BUZUG, T. M. – FASSNACHT, C. – LORENZ, C.: 2D/3D Registration of Pre-Operative CT Images and Intra-Operative X-ray Projections for Image Guided Surgery, *Computer Assisted Radiology and Surgery*, pp. 833–838 June, 1997, Elsevier Science, Amsterdam.

[4] CHEN, J. L. – STOCKMAN, G. C.: Determining Pose of 3D Objects with Curved Surfaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **18**, 1, pp. 52–57, 1996.

[5] FELDMAR, J. – AYACHE, N. – BETTING, F.: 3D-2D Projective Registration of Free-form Curves and Surfaces, Rapport de recherche no. 2434, Inria, France, 1994.

[6] MOCTEZUMA, J. – BERNASCH, J. – LOHMANN, G. – SCHWEIKARD, A. – GOSSÉ, F.: Robotic Surgery and Planning for Corrective Femur Osteotomy, *Proc. IEEE Workshop Intelligent Robots and Systems*, pp. 870–877, 1994.

[7] ROTH, M. – BRACK, CH. – SCHWEIKARD, A. – GOETTE, A. – MOCTEZUMA, J. – GOSSÉ, F.: New Less Invasive Approach to Knee Surgery Using a Vision-Guided Manipulator, *ISRAM'96, Proc. of the Sixth International Symposium on Robotics and Manufacturing*, pp. 731–738, May, 1996.

[8] BRACK, CH. – ROTH, M. – SCHWEIKARD, A. – GOETTE, H. – GOSSÉ, F.: Towards Accurate X-Ray-Camera Calibration in Computer-Assisted Robotic Surgery, *Proceedings of the International Symposium on Computer and Communication Systems for Image Guided Diagnosis and Therapy, Computer Assisted Radiology (CAR'96)*, pp. 721–728, June, 1996, Paris, France.

[9] SAUND, E.: Putting Knowledge into a Visual Shape Representation, *Artificial Intelligence*, **54**, pp. 71–119, 1992.

[10] PETITJEAN, S. – PONCE, J. – KRIEGMAN, D.J.: Computing Exact Aspect Graphs of Curved Objects: Algebraic Surfaces, *International Journal of Computer Vision*, **9** pp. 231–255, 1992.

[11] EGGERT, D. – BOWYER, K.: Computing the perspective projection aspect graph of solids of revolution, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **15**, 2, pp. 109–128, 1993.

[12] DICKINSON, S.J. – PENTLAND, A. P. – ROSENFELD, A.: 3-D Shape Recovery Using Distributed Aspect Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **14** 2, pp. 174–198, 1992.

[13] MAHMOOD, S. T. F.: Attentional Selection in Object Recognition, Cambridge: MIT, 1993.

[14] SEGEN, J.: Model Learning and Recognition of Nonrigid Objects, *CVPR'89, Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pp. 597–602, 1989.

[15] WERMAN, M. – SHASHUA, A.: The Study of 3D-from-2D using Elimination, ICCV'95, *Int. Conf. Computer Vision*, pp. 473–479, June, 1995.

[16] AYACHE, N. – FAUGERAS, O.D.: A New Approach for the Recognition and Positioning of Two-Dimensional Objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **8**, 1, pp. 44–54, 1986.

[17] SAUND, E.: Adding Scale to the Primal Sketch, *CVPR'89, Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pp. 70–78, 1989, San Diego.

[18] DERICHE, R.: Fast Algorithms for Low-Level Vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **12**, 1, pp. 78–87, 1990.

[19] CANNY, J.: Finding Edges and Rows in Images, MIT Artificial Intelligence Lab. Report, 1983, AI-TR-720, Cambridge.

[20] GANDER, W. – GOLUB, G.H. – STREBEL, R.: Fitting of Circles and Ellipses. Least squares solution, ETH Zuerich, Department Informatik, Institut fuer Wissenschaftliches Rechnen, *Tech. Re.* No. 217, 1994.

[21] ULLMAN, S.: Aligning Pictorial Descriptions: An approach to object recognition, *Cognition*, **32** pp. 193–254, 1989.

[22] LOWE, D. G.: Three-Dimensional Object Recognition from Single Two-dimensional Images, *Artificial Intelligence*, **31**, pp. 355–395, 1987.

[23] POPE, A.R.: Learning to Recognize Objects in Images: Acquiring and Using Probabilistic Models of Appearance, British Columbia, 1995.