Thesis proposal

Simultaneous Point Matching and Recovery of Rigid and Nonrigid Shapes

Eduard Serradell Domingo

Advisor

Francesc Moreno Noguer

Co-Advisor

Alberto Sanfeliu Cortés

Institut de Robòtica i Informàtica Industrial
(CSIC-UPC)
Doctorat en Automàtica, Robòtica i Visió

July 29, 2011

Contents

1 Introduction

The registration problem consists in finding correspondences between two point sets and recovering the underlying transformation that maps one point set to the other. Registration of point sets is an important issue for many computer vision applications such as robot navigation, image guided surgery, motion tracking, camera pose estimation and face recognition. In fact, it is the key component in tasks such as object or image alignment, stereo matching, image segmentation and shape/pattern matching.

Shape matching is typically formulated as a point matching problem since points may be used as a global representation and are easy to extract. The 'points' are features, most often the locations of interest points extracted from an image (2D) or a volume (3D). Registration techniques may be rigid or nonrigid depending on the underlying transformation model. The key characteristic of a rigid mapping is that all distances are preserved. While some models are well defined by a homography or an affine transformation, the need for more general nonrigid registration occurs in many tasks, where complex nonlinear transformation models are required.

There are various factors that make the point matching problem difficult. One such factor is the existence of *outliers*: many point features that exist in one point set may not have a corresponding point in the other. Another factor is the presence of *occlusions*: part of the initial structure is not present in the target set. In addition, the presence of *repetitive patterns* in the point cloud may cause most of the existing algorithms to get stuck in local minima. And of course, the point positions are generally affected by *noise*. While things are a bit easier when the models are rigid, nonrigid registration remains a challenge in computer vision, especially if one wants the aforementioned issues to be addressed.

As said above, there are two unknowns in a point matching problem: the correspondence and the mapping between point sets. Since solving for either of them without information regarding to the other is quite difficult, most applications try to simultaneously solve for correspondence and mapping. Moreover, as in many cases finding the solution involves dealing with a non-convex problem, most approaches to rigid or nonrigid point matching use an iterated estimation framework.

When possible, additional cues have been taken into account to make the problems affordable. For this purpose, geometric relationships between point clouds such as graphs or spectral techniques have been explored. Alternatively, the correspondence of features can be estimated using their description (appearance), thus reducing the potential set of matches to a scored list.

Concerning the dimensionality of the transformation, in this project of research we will deal with two different kinds of problems. In the first one, the two point sets are defined by an in-space mapping such as $\mathbb{R}^2 \Leftrightarrow \mathbb{R}^2$ or $\mathbb{R}^3 \Leftrightarrow \mathbb{R}^3$. In the second, points are related by a projective transformation $\mathbb{R}^2 \Leftrightarrow \mathbb{R}^3$. In both cases, the mapping between the point clouds may be rigid or nonrigid. Some examples are given in Fig. 1.

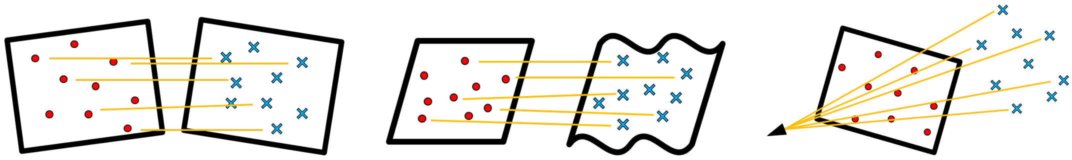

(a) (b) (c)

Figure 1: Different kinds of mappings: **a)** A homography is a rigid 2D-2D transformation. Any two images of the same planar surface in space are related by a homography. **b)** A nonrigid mapping between two point sets. **c)** In a projection, the 2D image points and 3D scene points are related by a nonlinear transformation.

Image/volume registration. One of the fundamental problems in computer vision is to register the structures appearing in images or volumes. Dealing with 2D-to-2D or 3D-to-3D transformations has also been object of research of many and extense studies. Iterative Closest Point (ICP) [6] is a very popular algorithm which iteratively assigns correspondences and finds the least squares transformation (usually rigid) relating these point sets. The algorithm then redetermines the closest point set and continues until it reaches the local minimum. Many variants of ICP have been proposed and we will come back to them later when detailing the state-of-art methods. Another family of algorithms which is able to work in either this framework or the camera pose recovery are the RANSAC-like algorithms [16]. When appearance is available, it can be used to speed up the search for consistent matches as proposed by PROSAC [11]. However, none of them solves entirely the problem as ICP needs a good initial estimate to reach a good solution and RANSAC methods, though avoiding local minima, are usually limited to a few number of point matches and do not scale-up very well to transformations with a large number of parameters.

Estimation of the camera pose. A particular problem which appears in many computer vision applications is estimating the pose of a camera from 3D-to-2D point correspondences between a 3D model and an image.

When the matches are given and the model is considered rigid, this is known as the Perspective-n-Point (PnP) problem and many effective methods have been proposed. However, when the correspondences are unknown one must simultaneously compute them along with the pose. An interesting algorithm is the SoftPOSIT [13], that looks for the solution as the minimum of a suitable objective function. Although its computational load is lower than other approaches, it is not able to guarantee the global minimum. An algorithm that performs a thorough search on the pose space is the Blind PnP algorithm [33]. It also solves the PnP problem without being given the correspondences and relies on the fact that prior information on the camera pose is often available. Such a prior is modeled as a Gaussian Mixture Model (GMM) and each component of the GMM is used to initialize a Kalman filter [25]. Then it explores the space of possible correspondences within a subset of potential matches and keeps the hypothesis that

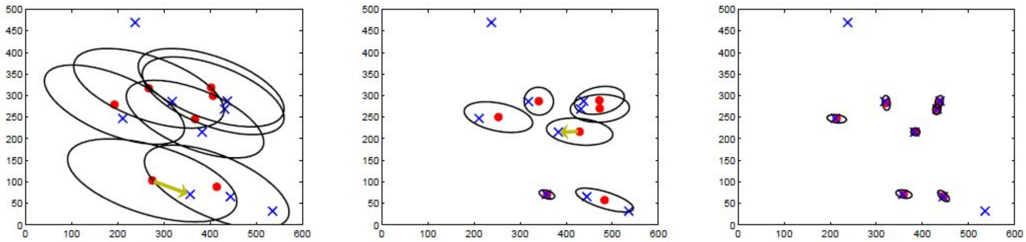
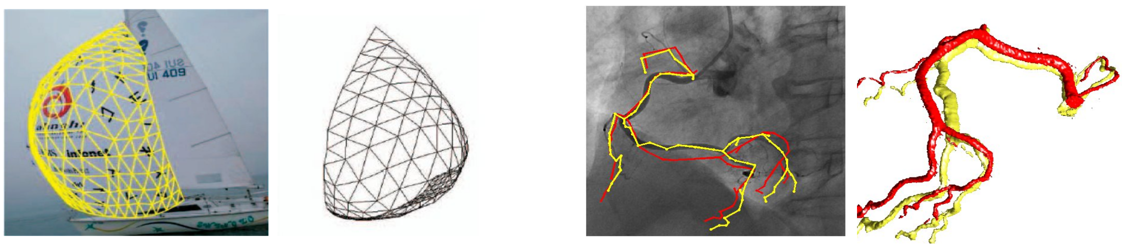


Figure 2: **Blind PnP.** The dots represent the projected 3D points and the ellipses the corresponding uncertainty regions given the pose prior. The crosses represent the 2D points. Even before hypothesizing any correspondence, the prior considerably reduces the set of potential matches. After some correspondence hypothesis, the uncertainty regions become so small that finding additional matches becomes easy.



(a) (b)

Figure 3: **(a) Deformable surfaces.** Given an image or a sequence of images, the goal is to recover the 3D nonrigid shape. The problem is ill posed as many potential representations can project correctly on the image. Thus, additional cues such as shape priors or illumination constraints shall be used. **(b) Articulated structures.** Given an input X-ray image (left), and a reference shape (shown in red at the right image) the goal is to retrieve the 3D configuration of the coronary tree that corresponds to the input image (yellow).

yields the smallest reprojection error. As illustrated by Fig. 2, the Kalman filter guides and speeds up the matching process, while preventing gross errors.

Yet, if the 3D model is assumed to be elastic becomes a challenging problem that still remains unsolved. The computer vision community has approached this problem using different strategies depending on the particularities of the context. The camera pose may be known or unknown, calibrated or not, and the model can be rigid or nonrigid. In the case of monocular pose and deformable 3D model recovery, two fields have been explored extensively but are still object of research: deformable surface and articulated body. Some examples are shown in Fig. 3.

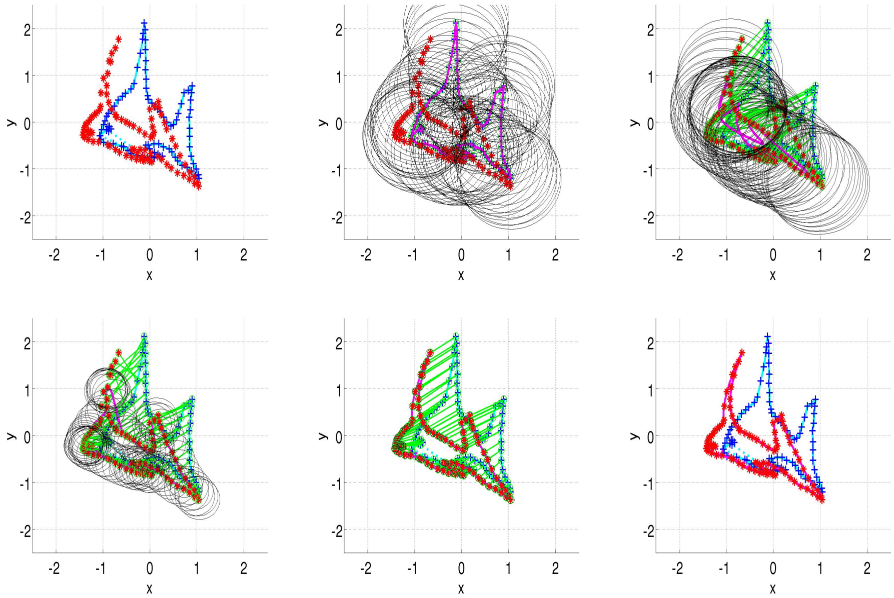


Figure 4: **Iterative Registration using Gaussian Processes.** The approach consists in simultaneously solve for point correspondences and to recover the nonrigid transformation by updating a Gaussian Process model. A GP is adjusted to the model structure (blue fish). Then, a simple matching is assigned using the Hungarian algorithm[34] on the Euclidean Distance Matrix and the process iteratively updates the hyperparameters until the uncertainty is less than a certain threshold. This particular experiment works as a nonlinear ICP because of the way the matches are assigned but it is easily extensible to a nonlinear robust point matching behavior.

Modeling the uncertainty. While the Kalman Filter approach is useful to introduce a probabilistic model for the point sets, to propagate the uncertainty and thus to reduce the number of potential matches, it can only be applied to rigid models with reduced number of parameters. We are exploiting new forms to propagate uncertainties in the context of our research. One that presents a strong mathematical background and an elegant formulation is the Gaussian Processes [40], which consists on a non-parametric Bayesian formulation of the probabilistic distribution of the model. Although the method is not new, only recently this approach has been applied to the field of computer vision. It is in the scope of the current research to find a way to represent the geometric constraints of the model point set by means of a Gaussian Processes and to use the uncertainty regions to reduce the search space while finding new correspondences. This kind of formulation would be of large interest for the computer vision community as it can be applied to both camera pose estimation and deformable shape matching. Preliminary results are shown in Fig. 4.

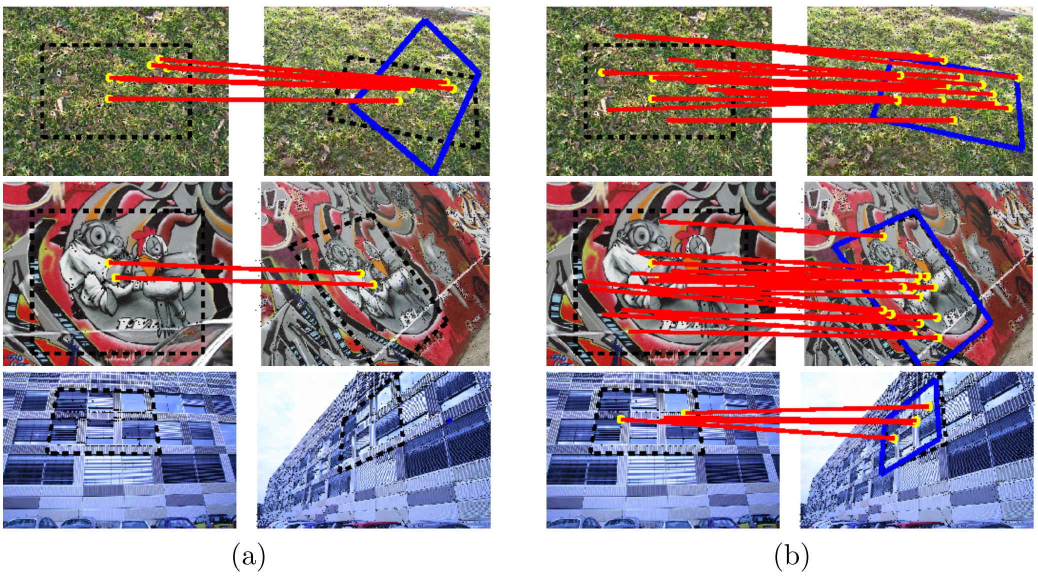


Figure 5: **Computing homographies.** The three pairs of images are related by a planar pattern. The homography recovery is difficult because of the highly textured images (Top), oblique angle (Middle) and repetitive patterns (Bottom). Using only appearance **(a)** is not enough to recover the correct transform and PROSAC fails due to the high number of outliers. Better results are shown in **(b)** using the Blind Homography algorithm [49], which is a result of this research. Here, geometric and appearance cues are combined to reassign correspondences.

We next introduce three of the cases of study of the proposed thesis and their respective contributions. They are particular problems of the general formulation we have just introduced and will help to get a detailed focus on the kind of issues we are dealing with.

1.1 Combining geometrical and appearance priors for robust homography estimation

Computing homographies from point correspondences has received much attention because it has many applications, such as stitching multiple images into panoramas [57] or detecting planar objects for Augmented Reality purposes [48, 62]. All existing methods assume that the correspondences are given *a priori* and usually rely on an estimation scheme that is robust both to noise and to outright mismatches. As a result, the best ones tolerate significant error rates among the correspondences but break down when the rate becomes too large. Therefore, in cases when the correspondences cannot be established reliably enough such as in the presence of repetitive patterns, they can easily fail.

The so-called *Blind PnP* approach [33] was designed to simultaneously establish 2D to 3D correspondences and estimate camera pose. To this end, it exploits the fact that, in general, some prior on the camera pose is often available. This prior is modeled as a Gaussian Mixture Model that is progressively refined by hypothesizing new correspondences. Incorporating each new one in a Kalman filter rapidly reduces the number of potential 2D matches for each 3D point and makes it possible to search the pose space sufficiently fast for the method to be practical.

Unfortunately, when going from exploring the 6-dimensional camera-pose space to the 8-dimensional space of homographies, the size of the search space increases to a point where a naive extension of the Blind PnP approach fails to converge. In general, any given model image point can be associated to several potentially matching target image points with progressively decreasing levels of confidence. To exploit this fact without having to depend on *a priori* correspondences, similarity of image appearance could be used to remove both low confidence potential correspondences and pose prior modes that do not result in promising match candidates.

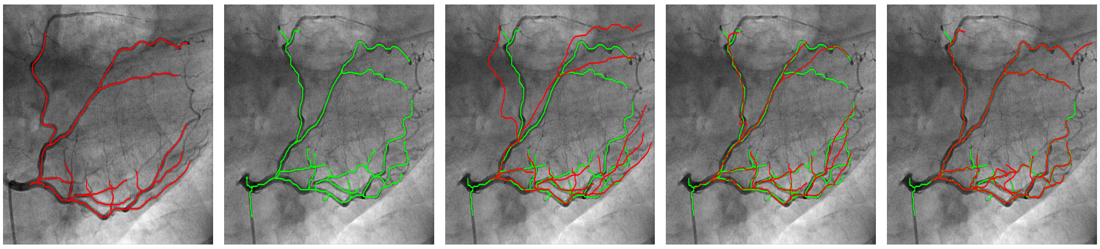
Expected contributions. Although homography estimation has been largely studied, difficult situations as in highly oblique views of planar scenes containing repetitive patterns are still complex problems to solve. In such scenes, interest point detectors exhibit very poor repeatability and, as a result, even such a reliable algorithm as PROSAC [11] fails because *a priori* correspondences are too undependable. We expect to address these kind of problems in our research. Actually, we have already been working on methods such the one that generates the results shown in Fig. 5, which outperforms PROSAC in situations where appearance is not fully reliable.

1.2 Nonrigid 3D model recovery from single images

Another kind of problem that remains unsolved is that of recovering a 3D nonrigid model from single images. This is a highly ambiguous problem since many different 3D configurations can virtually have the same projection. Thus, solving this problem requires prior knowledge about the type of deformations the structure can undergo.

Standard approaches within medical imaging assume a reference 3D scan of the model is known and that deformations in the input image are negligible. This reduces the shape recovering task to a rigid 3D-to-2D registration [22, 32, 42]. There exist a recent attempt of addressing the nonrigidity nature of the problem, although it has only been shown effective for relatively small deformations [20].

We may find other related areas in computer vision that essentially solve the same problem but in a different context, for instance, the techniques for 3D nonrigid surface reconstruction [15, 37, 46] and articulated human pose estimation from monocular images [1, 44, 50, 61]. In these approaches, though, it is often easy to obtain large amounts of training data and build detailed parametric models for specific deformations or 2D-to-3D mappings that directly link 2D observations with 3D configurations. Unfortunately, in some applications, producing these detailed models and mappings is simply not feasible and therefore, besides a weak 3D model of the reference shape, no further prior knowledge can be used.



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

Figure 6: **Elastic graph matching.** Two X-ray images **(a)** and **(b)** have been extracted at different times from a heart beat sequence. If we want to register the blood vessels, applying a simple rigid transformation is not enough as shown in **(c)**. Even state-of-art methods able to correct nonrigid transformations like Coherent Point Drift (CPD) [35] fail to recover the highly deformed coronary trees **(d)**. We are currently exploring new graph matching methods to relate the extracted vessel trees which achieve better registration results **(e)**.

Expected contributions. In this research, we intend to propose a novel approach that, given solely one single image and a reference 3D configuration, simultaneously recovers the 3D structure described in the input image and establishes correspondences with the reference shape. In addition we intend to deal with large amounts of noise and occlusions. The results shown on Fig. 3 e) are part of our current research work.

1.3 Elastic graph matching of 2D-to-2D or 3D-to-3D point sets

Graph-like structures are pervasive in biomedical 2D and 3D images. They may be blood vessels, pulmonary bronchi, nerve fibers, or dendritic arbors, which can be acquired at different times and scales, or using different modalities. This may result in vastly diverse image appearances.

Such drastic appearance changes make it impractical to use registration techniques that rely either on maximizing image similarity or finding image-based correspondences [67]. Since the graph structure may be the only thing that is shared across modalities, matching graphs, or subgraphs when the images have been acquired at different resolutions, therefore becomes the only effective registration means. Unfortunately, most existing techniques that attempt to do this rely on matching Euclidean or Geodesic distance between graph junction points [14, 18, 53]. This assumes that they are preserved and cannot handle true nonrigid motion or graphs whose topology might be different.

Expected contribution. In this work we intend to research methods to, using solely graph information, solve image registration of rigid and nonrigid transformations. In Fig. 6 are shown the benefits of these new methods.

The rest of the document is organized as follows. In Section 2, we perform a thorough study of the state-of-art which relates to our problem. In Section 3, we introduce the objectives and the scope of the research. Section 4 presents the expected contributions of the on-going research to the scientific community. In Section 5, we present a work plan for the remaining of the thesis. Section 6 is to account for the resources which permit the development of the research. And finally, in Section 7, we list the publications related to this research.

2 State of the art

Point set matching intends to establish a consistent correspondence between two point sets and recover the mapping function with the best alignment. Being a fundamental problem in the field of computer vision, a large amount of literature already exists. Next, the most relevant topics related to the robust point matching for both rigid and nonrigid transforms estimation in the context of computer image are reviewed, paying attention to the applications in which a contribution is expected to be done. Likewise, model recovery from single images and the reported approaches for its solution are discussed.

2.1 Rigid Point Matching

Correspondence-based approaches to computing homographies between images tend to rely on a RANSAC-style strategy [16] to reject mismatches that point matchers inevitably produce in complex situations. In practice, this means selecting and validating small sets of correspondences until an acceptable solution is found. The original RANSAC algorithm remains a valid solution, as long as the proportion of mismatches remains low enough. Early approaches [4, 19] to increasing the acceptable mismatch rate, introduced a number of heuristic criteria to stop the search, which were only satisfied in very specific and unrealistic situations. Other methods, before selecting candidate matches, consider all possible ones and organize them in data structures that can be efficiently accessed. Indexing methods, such as Hash tables [8, 26] and Kd-trees [5], or clusters in the pose space [36, 56] have been used for this purpose. Nevertheless, even within fast access data structures, these methods become computationally intractable when there are too many points.

Several more sophisticated versions of the RANSAC algorithm, such as Guided Sampling [59], PROSAC [11] and ARRSAC [39], have been proposed and they address the problem by using image-appearance to speed up the search for consistent matches. When the images contain repetitive structures resulting in unreliable keypoints and truly poor matches even they can fail. In these conditions, simple outlier rejection techniques [55] also fail.

In the context of the so-called *PnP* problem, which involves recovering camera pose from 3D to 2D correspondences, the SoftPOSIT algorithm [13] addresses this problem by iteratively solving for pose and correspondences, achieving an efficient solution for sets of about 100 feature points. Yet, this solution is prone to failure when different viewpoints may yield similar projections of the 3D points. This is addressed in the Blind PnP [33] by introducing weak pose priors, that constrain where the camera can look at, and guide the search for correspondences. Although achieving good results, both these solutions are limited to about a hundred feature points, and are therefore impractical in presence of the number of feature points that a standard keypoint detector would find in a high resolution textured image.

One objective of this research is to show that the response of local image descriptors, even when they are ambiguous and unreliable, may still be used in conjunction with geometric priors to simultaneously solve for homographies and correspondences. This lets us tackle very complex situations with many feature points and repetitive patterns, where current state-of-the-art algorithms fail.

2.2 Nonrigid Point Matching

Many registration techniques rely on finding a rigid or nonrigid transformation that maximizes image-similarity, often expressed in terms of correlation or mutual-information [67]. However, when dealing with hard datasets, the dissimilarity is such, that these methods do not do well.

One of the earliest algorithms, Iterative Closest Point (ICP) [6], iteratively assigns correspondences based on a closest distance criterion and optimizes a transformation (which can be nonrigid [21, 28]) relating the two point sets. The algorithm then redetermines the correspondences and continues until it reaches the local minimum. Another class are the probabilistic methods. In Robust Point Matching [17], a soft assignment of correspondences is introduced and a global optimum of the transformation is found by deterministic annealing. This was proved equivalent [9] to what could be achieved by an Expectation Maximization (EM) algorithm for GMM, where one of the point sets is treated as GMM centroids and the other as data points and the problem is formulated as a maximum likelihood (ML) estimation [23, 29, 30]. Recently, in [10] the authors proposed an optimization based approach, the TPS-RPM algorithm, in which two unknown variables (transformation and correspondence) are combined into an objective function. The soft assignment technique and deterministic annealing algorithm are used to search for an optimal solution. All those methods can cope, at different levels, with a small percentage of outliers, occlusions and noise and some of them even perform nonlinear registration. In Coherent Point Drift [35], the displacement field of the feature points is constrained, adquiring motion coherence and a better robustness to outliers. However, these approaches only perform well when the estimate for the initial position of the two point sets is good enough.

Another class of algorithms rely on that a graph structure can be obtained to link the features and guide the search for correct matches. Finding feature correspondence via Graph Matching is a challenging optimization problem which received considerable attention in the literature. A current assumption is that graphs can be related by a low-dimensional geometric transformation, such as a rigid one. In this case, using either RANSAC [16] or an improved version of it that takes appearance into account [11] is an attractive option. Unfortunately RANSAC-style approaches do not scale-up very well to transformations with higher number of parameters.

One therefore has to explicitly match the graph structures. Proposed techniques include: spectral methods [12, 27] and graph edit distance techniques [31, 41]. Among these papers, no practical solution is reported to obtain optimal solutions in reasonable computational time costs.

Applied to particular fields in medicine imaging, like retinal fundus, lung or liver vessels, techniques benefit from a combination of cues. A good example is the work presented in [53] where the euclidean distance matrix and the geodesic distance matrix are used to define a soft assignment between correspondences linked by a rigid tree and the final matching is done using spectral techniques. In [2] the features are structured using a proximity graph which aids in the correspondence search but the achieved complexity is similar to RANSAC. More adapted is [14], where the graph is particularly suited to express the vascular structure and used to recover the correspondences of the slightly nonlinear field of the retinal fundus. A final nonlinear ICP [43] is used to refine the transformation, yet achieving good results only to small nonrigid deformations.

The objective of the present research is to recover a high-dimensional elastic deformation to match graph structures, which not require an initial estimate of this deformation, and performs well even if the local appearances of the graph nodes are very different.

2.3 Monocular Nonrigid Reconstruction

Recovering the 3D structure from single images involves dealing with many different issues. Besides the inherent ambiguity of the monocular nonrigid reconstruction, the problem is further accentuated due to the presence of noise in the images and partial occlusions between different parts of the object. In the medical imaging literature this complexity has been traditionally alleviated by considering the system as a rigid structure [22, 32, 42] and using multiple views [60, 63]. To the best of our knowledge, [20] is the only approach that considers the nonrigid nature of the problem. They introduce 3D priors and inextensibility constraints into a steepest descent scheme to solve for the shape. Yet, their optimization procedure is only effective under relatively simple deformations.

When the model is considered nonrigid, various regularization methods have been proposed, such as Thin-Plate Spline [58], data embedding methods, and Finite Element Models (FEM) [24, 38, 64]. Thin-Plate Spline is a well-known interpolation method widely used in point set registration, which mainly penalizes the second order derivatives [58]. Besides, the data embedding techniques, such as Principal Component Analysis (PCA) [7, 45, 65], are also engaged as the regularization technique, although PCA requires a large number of training samples to obtain sufficient generalization capability. Finally, the FEM based regularization approach has also been extensively studied [24, 38]. However, for the techniques using the FEM models, the nonrigid surface must be explicitly represented by a triangulated mesh. Since those approaches have limitations, there is a need for developing new techniques to resolve these challenges. Gaussian processes [40] enjoy a solid foundation in statistics and machine learning, which provide a promising non-parametric Bayesian approach to regression problems and offer probability predictions. They have been recently applied to computer vision problems [44, 66] achieving a certain level of success and opening a path which is worth to explore.

Nonrigid surfaces. Some approaches are focused to reconstruct nonrigid 3D surfaces from monocular images. It has been shown that 3D shape can be retrieved by imposing local inextensibility and constraints introduced by a set of 3D-to-2D correspondences between the input image and a reference shape [15, 37, 46, 47]. If the 3D-to-2D correspondences are unknown the same kind of assumptions can be used, although additional constraints may be needed since in some contexts the matching has to be resolved simultaneously with the shape. In addition, many of these approaches impose strong shape priors based on previously acquired training data [46, 47] while in some contexts training data is hard to obtain and a feasible approach would have to rely on very weak shape priors.

Articulated structures. Since some models may be regarded as an articulated structure, one might think in applying the techniques of articulated pose estimation to the problem [1, 50, 61]. These approaches rely on large amounts of data for learning a mapping from 2D image observations to 3D poses, and have the advantage of not requiring to solve the 2D-to-3D correspondence problem. Yet, as said above, while obtaining sufficient training data is feasible for applications such as human pose estimation [3, 51], it becomes unfeasible in some frameworks, particularly in medical imaging. Moreover, local distance constraints are much less restrictive when dealing with points linked through a tree-like structure than when dealing with neighboring points on a surface. Recent works suggest introducing similar constraints as those used for nonrigid shape recovery into the formulation of articulated pose estimation problems [44, 52, 54]. This allows fitting more detailed parametric 3D models [52] and reducing the dependency of articulated pose estimation techniques on the training data [44]. However, reducing the dependency on training data has the drawback of increasing the sensitivity to artifacts into the input data.

Drawing particular inspiration on these approaches, our research is oriented to combine tools from the techniques for articulated pose estimation and shape recovery. However, in order to tackle problems with much larger amounts of image noise and occlusions, we propose using a generative 3D model that progressively increases its complexity and adaptability while establishing correspondences and detecting and rejecting outlier points.

3 Objectives and scope

3.1 Objectives

The main goal of the proposed thesis is to develop methods for the robust matching of point sets under the effect of rigid and nonrigid deformations and develop applications adapted to the particularities of different kinds of set-ups and transformations. To this end, we pretend to take into account various types of information such as geometric constrains or appearance information. In any case, the proposed research is intended to obtain methods to work under the following circunstances:

a. Rigid Transformations. Two images of the same planar surface of a scene are related by a homography, which represents a rigid transformation of the feature points from the model image. An objective of the proposed research work is to obtain robust methods to simultaneously estimate the point correspondences and homography.

b. Nonrigid Transformations. An aim of this thesis proposal is to develop methods that not only deal with translation, rotation and scale changes between two point sets but correct elastic deformations. The objective is to consider image or volume matching (2D-to-2D and 3D-to-3D) but also model reconstruction from single images (3D-to-2D).

c. Priors integration. To study the transformations proposed above, we pretend to use different kinds of knowledge. Both geometrical and appearance information can be combined to improve state-of-art methods. The underlying graphical structure of the models shall be taken into account to build realistic algorithms.

d. Nonlinear probabilistic approaches to robust matching. Kalman Filter methods or RANSAC-like algorithms can only be applied to rigid models with reduced number of parameters. Using Gaussian Processes a nonrigid probability distribution can be inferred from the data. This fact can be exploited to guide the matching process through propagating the uncertainty regions using the transformation estimations. This would represent a great improvement to the existing methods in robust point matching and the culmination of this thesis research.

3.2 Scope

The thesis scope includes point set transformations, starting by understanding the nature of simpler rigid transformations and progressively increasing the level of difficulty introducing deformable priors on the models and finally dealing with large deformations between both point sets. As in many applications, correspondences are generally unknown, we will focus on methods which simultaneously recover the matching between feature points and robustly estimate the underlying transformations. We pretend to work with real datasets, and therefore some extra effort will be performed on image processing to extract information from images and volumes such as feature detection, image segmentation or graph extraction. We will consider the following steps:

1. **Simultaneous matching and robust homography estimation combining geometric and appearance information.** Any two images of the same planar surface in space are related by a homography. Recovering this transformation is a well studied problem and has many applications in the field of computer vision. The objective of the proposed thesis is to derive a method that recovers simultaneously correspondences and homography estimation. We pretend to use Kalman Filtering, in a similar manner of [33], to propagate the uncertainty modelled by geometric and appearance priors in the original image and progressively reduce the search regions in the target image.

2. **Simultaneous matching and robust transform estimation combining geometric and appearance information.** We pretend to extend the work above to deal with different kinds of models including pose, homography and fundamental matrix estimation. We will use the same Kalman Filtering approach and exploit geometrical and appearance information.

3. **Simultaneous correspondence and nonrigid 3D model reconstruction from single images.** Recovering the 3D structure of a nonrigid articulated pose from single images is a highly ambiguous problem since many different 3D configurations can virtually have the same projection. The plan of the research is to model a deformable prior of the 3D object and progressively adapt the shape of the structure related by the actual image. To reduce the complexity and make the problem affordable, the camera pose will be considered known except by small translation changes. We will extract tree structures from the 3D data and use the orientation of the feature points to guide the matching.

4. **Robust elastic geometric graph matching.** An objective of the proposed thesis is to parameterize the nature of transformations undergone by a structured point cloud. If this structure happens to be graph-like we could implement methods that, working upon both 2D or 3D, use the information of the extracted graph to simultaneously estimate correspondences and nonrigid transformations. In medical imaging, it is usual to come across structured data such as neurons, coronary vessels, retinal fundus, etc. Many applications would benefit from an improved matching.

5. **Simultaneously solve for point correspondences and recover the transformation by updating a Gaussian Process model.** A Gaussian Process may be learnt from the model structure modeling a nonlinear probability distribution. Then it can be progressively updated while hypothesizing point correspondences and iteratively adjusting the model hyperparameters until a good solution is found. This method is intended to outperform existing nonrigid point set matching and shape recovering algorithms.

4 Expected contributions

The achievement of the objectives presented in Section 3 would allow to:

1. Extend the methods available in the literature for both the rigid and nonrigid point set transformations. The resulting procedures would be of large interest for the scientific community as many applications would benefit from a robust solution to this kind of problem.

2. Contribute to the development of the techniques which exploit geometric information such as graph structures or uncertainty regions. In particular, contribute to the implementation of a new probabilistic formulation to relate nonlinear deformations between point sets.

3. Define methods to deal with the complex problem of simultaneous matching and transformation recovery in both rigid and nonrigid transformations, combining informations of different kinds and sources.

4. Find a global solution to point registration. This new method based on a non-parametric Bayesian approach would apply to both 2D-to-2D / 3D-to-3D and 2D-to-3D mappings, taking into account rigid and nonrigid transformations.

5 Work planning

The following is a description of the foreseen tasks in the development of the proposed research:

5.1 Task 0: Background

The author of this thesis proposal studied Telecommunication Engineering in the Universitat Politècnica de Catalunya from 1998 to 2005. The final thesis 'Méthodes statistiques de traitement d'image pour la détection d'exoplanètes' was performed under an European Erasmus grant at Université de Nice under the supervision of professor Andrea Ferrari. Of interest for the development of the proposed research work, the theoretical background resulted from these studies include.

- Classical methods in computer vision.
- Probabilistic background.

After a break of some years working on private enterprises, in September 2008, the author enrolled in the master program Automàtica i Robòtica of the Universitat Politècnica de Catalunya. The required ECTS credits were divided into 6 academic courses and a Master thesis. He took the academic courses in the fields of computer vision and robotics,

- Computer Vision

- Advanced Computer Vision
- Mobile Robots and Navigation
- Artificial Intelligence
- Planning in Robotics
- Pattern Recognition

The Master thesis ‘Combining Geometric and Appearance Priors for Pose Recovery’, performed under the supervision of Professor Francesc Moreno-Noguer, was defended in September 2009.

In September 2009, the author enrolled in the doctorate program Automàtica, Robòtica i Visió of the Universitat Politècnica de Catalunya, Barcelona, Spain. He joined to the Mobile Robotics research group of the Institut de Robòtica i Informàtica Industrial (IRI) to develop his doctoral research under the supervision of Professor Francesc Moreno-Noguer and being co-advised by Professor Alberto Sanfeliu Cortés.

5.2 Task 1: Simultaneous correspondence and robust estimation

The first step of the proposed thesis is to explore the possibilities of combining geometrical and appearance information when estimating different image transformations. Inspired by the work [33], the methods will propose to simultaneously recover matching and rigid transforms by the means of a Kalman filter approach. The Kalman filter is used to propagate the uncertainty from a geometric prior, formulated upon the model structure, to the target images where a search region for potential matches is defined. The appearance information is used to make the method computationally affordable. The proposed method consists of an extensive search of the point correspondences where both the uncertainty regions and the transformation estimation is updated iteratively. The efforts will be concentrated mainly on the following problems:

Task 1.1 Combine geometric and appearance priors to simultaneously recover correspondences and robust homography estimation.
This work has been already published in ECCV 2010 and corresponds to publication 1 in Section 7.

Task 1.2 Explore the possibilities of the proposed method to simultaneously solve point set matchings and robustly recover pose, homography and fundamental matrix.
This work is currently object of research and corresponds to work in progress 3 in Section 7.

Simultaneously to this study, the author will analyze the approaches reported in the literature including for instance the review of methods for robust point-set estimation and the analysis of different feature descriptors.

5.3 Task 2: Nonrigid model reconstruction from single images

Once explored the rigid transformation problem, the next natural step is to extend the research to nonrigid transformations. A deformable prior is introduced to model the original structures while a similar iterative method solves simultaneously for correspondence and global transformation. A 3D model reconstruction from single images is an ill posed problem. We will explore the use of tree-like structures to constrain the deformations. A complementary study will be performed on processing volumes and images to segment the datasets and extract the tree structures.

Task 2.1 Simultaneous correspondence and nonrigid 3D reconstruction of the coronary tree from single X-ray images. This work has been accepted for publication in ICCV 2011 and corresponds to publication 2 in Section 7.

Task 2.2 Simultaneous correspondence and nonrigid 3D-2D reconstruction of tree-like structures.
This work is currently object of research and corresponds to work in progress 2 in Section 7.

Task 2.3 Probabilistic segmentation of images.
A side task which consist on improving segmentation techniques in order to obtain better results in the recognition process.

5.4 Task 3: Elastic graph matching

The core research of this task is focused on the use of graph structures to guide point set matching embedded in both $\mathbb{R}^2$ or $\mathbb{R}^3$. To this end, we will explore the combination of global and local methods to solve first an affine transformation and then refine the estimation using a nonrigid approach. In a similar orientation we will study the possibilities of a probabilistic technique called Gaussian Processes to solve the same kind of problems.

Task 3.1 Robust nonrigid geometric graph matching.
This work is currently object of research and corresponds to submitted paper 1 in Section 7.

Task 3.2 Online search using Gaussian Processes for nonrigid graph matching.
This work is currently object of research and corresponds to work in progress 1 in Section 7.

The use of Gaussian Processes will provide a novel and elegant method to solve for nonrigid transformations in a similar manner than Kalman Filter for rigid and affine transformations.

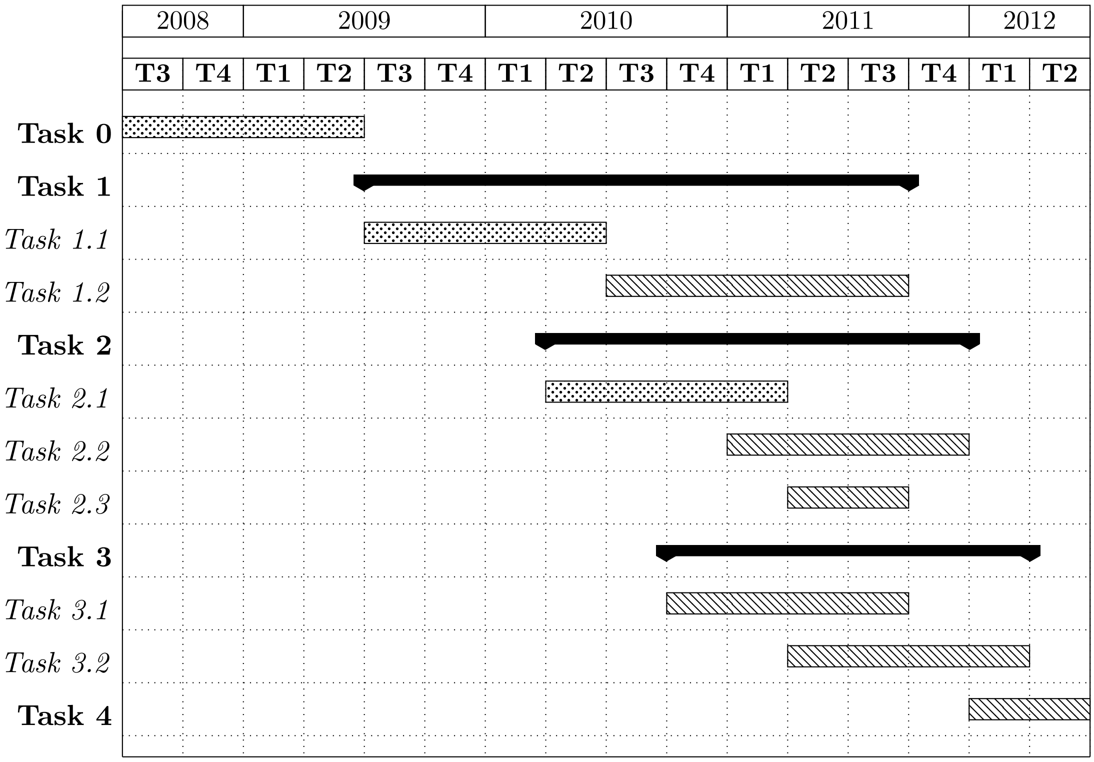


Figure 7: Work planning of the proposed thesis

5.5 Task 4: Compilation of results

The last task of this thesis proposal is assigned to the elaboration of the dissertation and the preparation of its public defense.

Figure 7 presents the schedule for the tasks described through the Sections 5.1 to 5.5 in a Gantt chart that spans over four years. In this, T stands for a three months period, so that *T1* represents the first three months of a year. The tasks already attained are shown with crosshatch points.

6 Resources

The proposed research work will be developed mainly at the Institut de Robòtica i Informàtica Industrial-IRI (UPC-CSIC) from Barcelona, working in the framework of the Mobile Robotics research group. Part of this research has also been done in collaboration with the Computer Vision Lab (CVLAB) at the 'Ecole Polytechnique Fédérale de Lausanne' (EPFL) in three alternative interships corresponding to the last period (T4) of 2009, 2010 and 2011. The author is being financed by the Spanish Ministry of Science and Innovation through a FPI grant under the UbRob project.

7 Publications

The following is the list of the international publications resulting from the current state of research: [1]

Published

1. E. Serradell, M. Özuysal, V. Lepetit, P. Fua and F. Moreno-Noguer, *Combining Geometric and Appearance Priors for Robust Homography Estimation*, 11th European Conference on Computer Vision, 2010, Crete, in Computer Vision - ECCV 2010, Vol 6313 of Lecture Notes in Computer Science, pp. 58-72, 2010, Springer, Berlin. Acceptance rate: 24.5%.

2. E. Serradell, A. Romero, R. Leta, C. Gatta and F. Moreno-Noguer, *Simultaneous Correspondence and Non-Rigid 3D Reconstruction of the Coronary Tree from Single X-ray Images*, accepted for publication in International Conference on Computer Vision, ICCV 2011. Acceptance rate: ??.??%.

[1] Please note that the top three computer vision conferences (ICCV, ECCV, CVPR) are highly competitive with low acceptance rates $< 30\%$. ICCV and ECCV have CiteSeer impact factor rankings in the top 5% and 7%, respectively of all computer science journals and conferences. MICCAI has also an acceptance rate $< 30\%$.

Submitted

1. E. Serradell, J. Kybic, F. Moreno-Noguer and P. Fua, *Robust Elastic 2D/3D Geometric Graph Matching*, submitted to Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2011.

In progress

We are currently working in some of the tasks presented in 5. We expect to present our work as follows:

1. *Active Learning with Gaussian Processes for Nonrigid Graph Matching.* We intend to use Gaussian Processes to constrain the search for matches in an Active Learning approach to robustly estimate nonrigid deformations. We are preparing a conference paper to send to the next CVPR (Conference on Computer Vision and Pattern Recognition).

2. *Simultaneous Correspondence and Non-Rigid 3D Reconstruction of the Coronary Tree from Single X-ray Images.* We are extending the work already accepted for publication in ICCV 2011, by performing a better segmentation of the blood vessels in both 2D and 3D. As a result, improved tree-like structures will be extracted and the recovered models will contain less error. We are preparing a journal paper to send to TMI (IEEE Transactions on Medical Imaging).

3. *Combining geometric and appearance priors for robust shape recovery.* We are currently researching a method to extend the results presented in our ECCV 2010 paper to other rigid transformation models, i.e. fundamental matrix. We are preparing a journal paper to sent to PAMI (IEEE Transactions on Pattern Analysis and Machine Intelligence.

8 Bibliography

[1] A. Agarwal and B. Triggs. 3D Human Pose from Silhouettes by Relevance Vector Regression. In *Conference on Computer Vision and Pattern Recognition*, 2004.

[2] W. Aguilar, Y. Frauel, F. Escolano, M.E. Martinez-Perez, A. Espinosa-Romero, and M.A. Lozano. A Robust Graph Transformation Matching for Non-Rigid Registration. *Image and Vision Computing*, 27:897 – 910, 2009.

[3] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape Completion and Animation of People. *ACM SIGGRAPH*, 24:408–416, 2005.

[4] N. Ayache and O. Faugeras. Hyper: A New Approach for the Recognition and Positioning of Two-Dimensional Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):44–54, 1986.

[5] J. Beis and D.G. Lowe. Indexing without Invariants in 3D Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):1000–1015, 1999.

[6] P. Besl and N. McKay. A Method for Registration of 3D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, February 1992.

[7] V. Blanz and T. Vetter. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[8] J.B. Burns, R.S. Weiss, and E.M. Riseman. View Variation of Point-Set and Line-Segment Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(1):51–68, 1993.

[9] H. Chui and A. Rangarajan. A Feature Registration Framework Using Mixture Models. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, MMBIA '00, page 190, 2000.

[10] H. Chui and A. Rangarajan. A New Point Matching Algorithm for Non-Rigid Registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, 2003.

[11] O. Chum and J. Matas. Matching With PROSAC - Progressive Sample Consensus. In *Conference on Computer Vision and Pattern Recognition*, pages 220–226, June 2005.

[12] T. Cour, P. Srinivasan, and J. Shi. Balanced Graph Matching. In *Neural Information Processing Systems*, 2006.

[13] P. David, D. Dementhon, R. Duraiswami, and H. Samet. SoftPOSIT: Simultaneous Pose and Correspondence Determination. In *European Conference on Computer Vision*, May 2002.

[14] K. Deng, J. Tian, J. Zheng, X. Zhang, X. Dai, and M. Xu. Retinal Fundus Image Registration via Vascular Structure Graph Matching. *International Journal of Biomedical Imaging*, 2010.

[15] A. Ecker, A. D. Jepson, and K. N. Kutulakos. Semidefinite Programming Heuristics for Surface Reconstruction Ambiguities. In *European Conference on Computer Vision*, 2008.

[16] M.A Fischler and R.C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting With Applications to Image Analysis and Automated Cartography. *Communications ACM*, 24(6):381–395, 1981.

[17] S. Gold and A. Rangarajan. A Graduated Assignment Algorithm for Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:377–388, 1996.

[18] S. Gold, A. Rangarajan, C. Lu, and E. Mjolsness. New Algorithms for 2D and 3D Point Matching: Point Estimation and Correspondence. *Pattern Recognition*, (31): 1019–1031, 1998.

[19] E. Grimson. The Combinatorics of Heuristic Search Termination for Object Recognition in Cluttered Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):920–935, 1991.

[20] M. Groher, D. Zikic, and N. Navab. Deformable 2D-3D Registration of Vascular Structures in a One View Scenario. *IEEE Transactions on Medical Imaging*, 28: 847–860, 2009.

[21] D. Hahnel, S. Thrun, and W. Burgard. An Extension of the ICP Algorithm for Modeling Nonrigid Objects with Mobile Robots. In *International Joint Conference on Artificial Intelligence*, pages 915–920, 2003.

[22] J. Hipwell, G. Penney, R. McLaughlin, K. Rhode, P. Summers, T. Cox, J. Byrne, J. Noble, and D. Hawkes. Intensity Based 2D-3D Registration of Cerebral Angiograms. *IEEE Transactions on Medical Imaging.*

[23] W.M. Wells III. Statistical Approaches to Feature-Based Object Recognition. *International Journal of Computer Vision*, 21:63–98, 1997.

[24] S. Ilić and P. Fua. Implicit Meshes for Surface Reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):328–333, 2006.

[25] R.E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering*, pages 35–45, 1960.

[26] Y. Lamdan and H.J. Wolfson. Geometric Hashing: A General And Efficient Model-based Recognition Scheme. In *International Conference on Computer Vision*, pages 238–249, 1988.

[27] M. Leordeanu and M. Hebert. A Spectral Technique for Correspondence Problems Using Pairwise Constraints. In *International Conference on Computer Vision*, pages 1482–1489, October 2005.

[28] H. Li, R.W. Sumner, and M. Pauly. Global Correspondence Optimization for Non-Rigid Registration of Depth Scans. In *Proceedings of the Symposium on Geometry Processing*, SGP '08, pages 1421–1430, 2008.

[29] B. Luo and E.R. Hancock. Structural Graph Matching Using the EM Algorithm and Singular Value Decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1120–1136, 2001.

[30] G. Mcneill and S. Vijayakumar. A Probabilistic Approach to Robust Shape Matching. In *International Conference on Image Processing*, pages 937 –940, 2006.

[31] B.T. Messmer and H.Bunke. A New Algorithm for Error-Tolerant Subgraph Isomorphism Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:493–504, 1998.

[32] C.T. Metz, M. Schaap, S. Klein, L. Neefjes, E. Capuano, C. Schultz, R.J. van Geuns, P.W. Serruys, T. van Walsum, and W.J. Niessen. Patient Specific 4D Coronary Models from ECG-Gated CTA Data for Intra-Operative Dynamic Alignment of CTA with X-Ray Images. In *Conference on Medical Image Computing and Computer Assisted Intervention*, volume 1, pages 369–376, 2009.

[33] F. Moreno-Noguer, V. Lepetit, and P. Fua. Pose Priors for Simultaneously Solving Alignment and Correspondence. In *European Conference on Computer Vision*, pages 405–418, October 2008.

[34] J. Munkres. Algorithms for the Assignment and Transportation Problems. *Journal of the Society of Industrial and Applied Mathematics*, 5(1):32–38, March 1957.

[35] A. Myronenko and X.B. Song. Point-Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (12):2262–2275, 2010.

[36] C. Olson. Efficient Pose Clustering Using a Randomized Algorithm. *International Journal of Computer Vision*, 23(2):131–147, June 1997.

[37] M. Perriollat, R. Hartley, and A. Bartoli. Monocular Template-based Reconstruction of Inextensible Surfaces. In *British Machine Vision Conference*, 2008.

[38] J. Pilet, V. Lepetit, and P. Fua. Fast Non-Rigid Surface Detection, Registration and Realistic Augmentation. *International Journal of Computer Vision*, 76(2), February 2008.

[39] R. Raguram, J.M. Frahm, and M. Pollefeys. A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus. In *European Conference on Computer Vision*, pages 500–513, 2008.

[40] C. E. Rasmussen and C. K. Williams. *Gaussian Process for Machine Learning.* MIT Press, 2006.

[41] A. Robles-Kelly and E.R. Hancock. Graph Edit Distance from Spectral Seriation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:365–378, 2005.

[42] D. Ruijters, B.M. ter H. Romeny, and P. Suetens. Vesselness-Based 2D-3D Registration of the Coronary Arteries. *International Journal of Computer Assisted Radiology and Surgery*, 4:391–397, 2009.

[43] S. Rusinkiewicz and M. Levoy. Efficient Variants of the ICP Algorithm. In *International Conference on 3-D Digital Imaging and Modeling*, 2001.

[44] M. Salzmann and R. Urtasun. Combining Discriminative and Generative Methods for 3D Deformable Surface and Articulated Pose Reconstruction. In *Conference on Computer Vision and Pattern Recognition*, June 2010.

[45] M. Salzmann, J. Pilet, S. Ilić, and P. Fua. Surface Deformation Models for Non-Rigid 3D Shape Recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1481–1487, February 2007.

[46] M. Salzmann, F. Moreno-Noguer, V. Lepetit, and P. Fua. Closed-Form Solution to Non-rigid 3D Surface Registration. In *European Conference on Computer Vision*, volume 4, pages 581–594, 2008.

[47] J. Sanchez, J. Ostlund, P. Fua, and F. Moreno-Noguer. Simultaneous Pose, Correspondence and Non-Rigid Shape. In *Conference on Computer Vision and Pattern Recognition*, pages 1189–1196, 2010.

[48] C. Scherrer, J. Pilet, V. Lepetit, and P. Fua. Souvenirs Du Monde Des Montagnes. *Leonardo, special issue on ACM SIGGRAPH*, 42(4):350–355, August 2009.

[49] E. Serradell, M. Özuysal, V. Lepetit, P. Fua, and F. Moreno-Noguer. Combining Geometric and Appearance Priors for Robust Homography Estimation. In *European Conference on Computer Vision*, pages 58–72, 2010.

[50] G. Shakhnarovich, P. Viola, and T. Darrell. Fast Pose Estimation with Parameter-Sensitive Hashing. In *International Conference on Computer Vision*, 2003.

[51] L. Sigal and M. J. Black. Humaneva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion. In *Technical Report*, 2006.

[52] L. Sigal, A. O. Balan, and M. J. Black. Combined Discriminative and Generative Articulated Pose and Non-Rigid Shape Estimation. In *Neural Information Processing Systems*, 2007.

[53] D. Smeets, P. Bruyninckx, J. Keustermans, D. Vandermeulen, and P. Suetens. Robust Matching of 3D Lung Vessel Trees. In *Conference on Medical Image Computing and Computer Assisted Intervention*, 2010.

[54] C. Sminchisescu, A. Kanaujia, and D. Metaxas. Learning Joint Top-down and Bottom-up Processes for 3D Visual Inference. In *Conference on Computer Vision and Pattern Recognition*, 2006.

[55] C. V. Stewart. Robust Parameter Estimation in Computer Vision. *SIAM Reviews*, 41:513–537, 1999.

[56] G. Stockman. Object Recognition and Localization via Pose Clustering. *Computer Vision, Graphics, and Image Processing*, 40(3):361–387, 1987.

[57] R. Szeliski. Image Alignment and Stitching: A Tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104, 2006.

[58] R. Szeliski and J. Coughlan. Spline-based Image Registration. *International Journal of Computer Vision*, 22(3):199–218, 1997.

[59] B. Tordoff and D.W. Murray. Guided Sampling and Consensus for Motion Estimation. In *European Conference on Computer Vision*, pages 82–98, 2002.

[60] G-A. Turgeon, G. Lehmann, G. Guiraudon, M. Drangova, D. Holdsworth, and T. Peters. 2D-3D Registration of Coronary Angiograms for Cardiac Procedure Planning and Guidance. *Medical Physics*, 32:3737–3749, 2005.

[61] R. Urtasun and T. Darrell. Local Probabilistic Regression for Activity-Independent Human Pose Inference. In *Conference on Computer Vision and Pattern Recognition*, 2008.

[62] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Pose Tracking from Natural Features on Mobile Phones. In *International Symposium on Mixed and Augmented Reality*, September 2008.

[63] J. Yao and R. Taylor. Assessing Accuracy Factors in Deformable 2D/3D Medical Image Registration Using a Statistical Pelvis Model. In *International Conference on Computer Vision*, volume 2, pages 1329–1334, 2003.

[64] J. Zhu and M.R. Lyu. Progressive Finite Newton Approach to Real-Time Nonrigid Surface Detection. In *International Conference on Computer Vision*, October 2007.

[65] J. Zhu, S. Hoi, and M.R. Lyu. Real-Time Non-Rigid Shape Recovery Via Active Appearance Models for Augmented Reality. In *European Conference on Computer Vision*, pages 186–197, 2006.

[66] J. Zhu, S. Hoi, and M.R. Lyu. Nonrigid Shape Recovery by Gaussian Process Regression. In *Conference on Computer Vision and Pattern Recognition*, pages 1319–1326, 2009.

[67] B. Zitova and J. Flusser. Image Registration Methods: A Survey. *Image and Vision Computing*, 11(11):977–1000, 2003.