Pattern Recognition 48 (2015) 2209–2226

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Fast and accurate surface alignment through an isometry-enforcing game

Andrea Albarelli [a,*], Emanuele Rodolà [b], Andrea Torsello [a]

[a] Department of Environmental Science, Informatics, and Statistics, Ca' Foscari University of Venice, Italy
[b] Department of Computer Science, Technische Universität München, Germany

ARTICLE INFO

Article history:
Received 17 July 2014
Received in revised form
9 January 2015
Accepted 23 January 2015
Available online 3 February 2015

Keywords:
Surface registration
Feature detection
3D scanning
Game theory

ABSTRACT

Surface registration is often performed as a two step process. A feature matching scheme is first adopted to find a coarse initial alignment between two meshes. Subsequently, a refinement step, which usually operates in the space of rigid motions, is applied to reach an optimal registration with respect to pointwise distances between overlapping areas. In this paper we propose a novel technique that allows to obtain an accurate surface registration in a single step, without the need for an initial motion estimation. The main idea of our approach is to cast the selection of correspondences between points on the surfaces in a game-theoretic framework, where a natural selection process allows matching points that satisfy a mutual rigidity constraint to thrive, eliminating all the other correspondences. This process yields a very robust inlier selection scheme that does not depend on any particular technique for selecting the initial strategies as it relies only on the global geometric compatibility between correspondences. The practical effectiveness of the approach is confirmed by an extensive set of experiments and comparisons with state-of-the-art techniques.

1. Introduction

Surface alignment is a pervasive problem throughout Computer Vision literature. In fact, it finds useful applications as a tool for in-line quality control [1], 3D object recognition [2], advanced human interfaces [3], and SLAM [4], just to name a few. For this reason, surface registration is one of the most studied topics in the field of 3D data acquisition and processing. With this paper we try to introduce a fresh view on the problem by proposing a game-theoretic approach that is robust to noise and allow to attain a very accurate registration without requiring an initial motion estimation.

1.1. Fine and coarse registration techniques

The distinction between fine and coarse surface registration methods is mainly related to the different strategies adopted to find pairs of corresponding points to be used for the estimation of the rigid transformation. Almost invariably, fine registration algorithms exploit an initial guess in order to constrain the search area for compatible mates and minimize the risk of selecting outliers. On the other hand, coarse techniques, which cannot rely on any motion estimation, must adopt a matching strategy based on the similarity between surface-point descriptors or resort to random selection schemes. The tension between the precision required for fine alignment versus the recall needed for an initial motion estimation stands as the main hurdle to the unification of such approaches.

The vast majority of current fine alignment methods are modifications to the original ICP proposed by Zhang [5] and Besl and McKay [6]. These variants generally differ in the strategies used to sample points from the surfaces, reject incompatible pairs, or measure error. In general, the precision and convergence speed of these techniques is highly data-dependent and sensitive to the fine-tuning of the model parameters. Several approaches that combine these variants have been proposed in the literature in order to overcome these limitations (see [7] for a comparative review). No matter which variant is used, ICP, being an iterative algorithm based on local, step-by-step decisions, is susceptible to the presence of local minima. Some recent variants mitigate this problem by avoiding hard culling assigning a probability to each candidate pair by means of evolutionary techniques [8] or Expectation Maximization [9]. Other, non-ICP-based, fine registration approaches include the well-known method by Chen [10] and signed distance fields matching [11].

Coarse registration techniques can be roughly organized into three main classes: global methods, feature-based methods and techniques based on RANSAC [12] or PROSAC [13] schemes. Global

* Corresponding author. Tel.: +39 041 2348465 (IT).
E-mail addresses: albarelli@unive.it (A. Albarelli), rodola@in.tum.de (E. Rodolà), torsello@dsi.unive.it (A. Torsello).

http://dx.doi.org/10.1016/j.patcog.2015.01.020
0031-3203/

methods such as PCA [14] or Algebraic Surface Model [15] exploit some global property of the surface and are thus very sensitive to occlusion. Feature-based approaches aim at the localization and matching of interesting points on the surfaces. They are more precise and can align surfaces that exhibit only partial overlap. Nevertheless, the unavoidable localization error of the feature points prevents them from obtaining accuracies on par with fine registration methods.

A completely different coarse registration approach is the one taken by RANSAC-based techniques. DARCES [16] is based on the random extraction of sets of mates from the surfaces and their validation based on the accuracy of the estimated transformation. The more recent Four Points Congruent Sets method [17] follows a similar route, but filters the data to reduce noise and performs early check in order to reduce the number of trials.

A recent and extensive review of many different methods can be found in [19]. Regardless of the criteria used to obtain pairs of mating points, the subsequent step in surface registration is to search for the rigid transformation that minimizes the squared distance between them. Many mature techniques are available to do this (see for instance [20]).

1.2. Feature detection on 3D objects

Feature detection and characterization is a key step in many tasks involving the recognition, registration or database search of 2D and 3D data. Specifically, when suitable interest points are available, all these problems can be tackled by working with the set of extracted features rather than dealing with the information carried by the whole data, which is less stable and noisier. For an interest point to be reliable it must exhibit two properties: repeatability and distinctiveness. A feature is highly *repeatable* if it can be detected with good positional accuracy over a wide range of noise levels and sampling conditions as well as different scales and transformations of the data itself. Further, description vectors calculated over interesting points are said to be *distinctive* if the descriptors related to different features lie far apart in feature space, while descriptors associated to multiple instances of the same point lie within a small distance from one another. These properties are somewhat difficult to attain since they are subject to antithetical goals: In fact, to achieve good repeatability despite of noise, larger patches of data must be considered, which unfortunately leads to a lower positional precision and a less sharp culling of uninteresting points. Moreover, for descriptor vectors to be distinctive among different features, they need to adopt a large enough basis, which, owing to the well known "curse of dimensionality," also affects their coherence over perturbed versions of the same feature. In the last two decades these quandaries have been addressed with great success in the domain of 2D images, where salient points can be localized with sub-pixel accuracy using detectors exploiting strong local variation in intensity, such as Harris Operator [21] and Difference of Gaussians [22], or using techniques that are able to locate affine invariant regions, such as Maximally Stable Extremal Regions (MSER) [23] and Hessian-Affine [24]. Among the most used descriptors are the Scale-invariant feature transform (SIFT) [25], the Speeded Up Robust Features (SURF) [26] and Gradient Location and Orientation Histogram (GLOH) [27]. While these approaches work well with 2D intensity images, they cannot be easily extended to handle 3D surfaces since no intensity information is directly available. On the other hand, there has been huge effort to use other local measures, such as curvature or normals. One of the first descriptors to capture the structural neighborhood of a surface point was described by Chua and Jarvis, who with their Point Signatures [28] suggest both a rotation and translation invariant descriptor and a matching technique. Later on, Johnson and Hebert introduced Spin Images [29], a rich characterization obtained by binning the radial and planar distances of the surface samples respectively from the feature point and from the tangent plane. Given their ability to perform well with both surface registration and object recognition, spin Images have become one of the most used 3D descriptors. More recently, Pottmann et al. proposed the use of Integral Invariants [30], stable multi-scale geometric measures related to the curvature of the surface and the properties of its intersection with spheres centered on the feature point. Zaharescu et al. [31] presented a comprehensive approach for interest point detection (MeshDOG) and description (MeshHOG), based on the value of any scalar function defined over the surface (i.e. curvature or texture, if available). MeshDOG localizes feature points by searching for scale-space extrema over progressive Gaussian convolutions of the scalar function and thus by applying proper thresholding and corner detection. MeshHOG calculates a histogram descriptor by binning gradient vectors with respect to a rotationally invariant local coordinate system. Finally, the recent SHOT descriptor [32], introduced by Tombari et al. exploits a novel 3D reference frame to offer enhanced descriptive power and robustness. The reader interested in a comparison between recent geometric descriptors and feature descriptions can find an in-depth coverage within [33].

In the following sections we introduce a novel pipeline that can be used to obtain an accurate surface registration without requiring an initial motion estimation. The contribution is twofold. First we propose very simple descriptors, named *Surface Hashes*, that span only 3–5 dimensions. As their name suggests, we expect Surface Hashes to be repeatable through the same feature point, yet to suffer from a high level of clashing due to their limited distinctiveness. In order to overcome this liability we also adopt a robust game-theoretic inlier selector which exploits rigidity constraints among surfaces to guarantee a global geometric consistency. The combination of these loosely distinctive features and our robust matcher leads to an effective and robust surface alignment approach.

2. Game theoretical matching

The use of tools coming from Game Theory for matching purposes is rather novel: it has been introduced as an effective heuristic in the domain of graph matching in [34] and has recently been applied successfully to other Computer Vision scenarios [38,35,37]. In particular, the concept of Game-Theoretic Matching (GTM) has a central role within the framework discussed in this paper and will be explored thoroughly in the next two sections.

2.1. Basic ideas

This kind of matching process is performed on the basis of two premises. The first is the ability to model a set of potentially matching features as strategies in a non-cooperative game. The second requirement is the availability of a suitable payoff function between strategies that measures how well two matches behave together. The main idea that underlies the proposed technique is quite simple: if both the strategy definition criterion and the payoff function are chosen properly with respect to a given problem, it is reasonable to think that a subset of strongly mutually compatible strategies (i.e. putative matches) can be found embedded in the initial set. Those matches must be found and isolated from the outliers with some selection technique.

In principle, those assumptions are not that different from those that subtend to the very popular RANSAC inlier selection method. In fact, any RANSAC or PROSAC flavor assumes that a subset of large consensus exists in correspondence with the correct solution. In addition, some compatibility function is always defined in order to evaluate the consensus that a proposed solution receives from all the remaining samples. However, the loose connection between GTM and