Expert Systems With Applications 45 (2016) 150–160

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Cooperative pose estimation of a fleet of robots based on interactive points alignment☆

Xavier Cortés, Francesc Serratosa*

Universitat Rovira i Virgili, Departament d'Enginyeria Informàtica i Matemàtiques, Spain

ARTICLE INFO

Keywords:
Robot cooperation
Human–Robot interaction
Social robots
Pose estimation

ABSTRACT

Given a fleet of autonomous robots performing a cooperative task, such as rescue of people, it is crucial for the robots to share their relative position. If the site has not been explored before, auto localise each robot through landmarks is not possible. Moreover, not always the GPS information is available or it has the desirable accuracy. Our framework is composed of a fleet of robots that have 2D and 3D cameras, a human coordinator and a Human–Machine Interface. 3D-images are used to automatically align them and deduct the relative position between robots. 2D-images are used to reduce the alignment error in an interactive manner. A human visualises both 2D-images and the current automatic alignment and imposes a new alignment through the Human–Machine Interface. Since the information is shared through the whole fleet, robots can deduct the position of other ones that do not visualise the same scene. Practical evaluation shows that in situations that there is a large difference between images, the cooperative and interactive processes are crucial to achieve an acceptable result.

1. Introduction

In recent years, interaction between robots and humans and also cooperation between robots has increased rapidly. Applications of this field are very diverse, ranging from developing automatic exploration sites to using robot formations to transport and evacuate people in emergency situations (Casper & Murphy, 2003), assembly lines (Unhelkar, Shah, 2015) or simply vehicle positioning (Ifthekhar, Saha, & Jang, 2015). Within the area of social and cooperative robots (Garcia et al., 2013; Kim, Taguchi, Hong, & Lee, 2014), interactions between a group of people and a set of accompanying robots have become a primary point of interest (Garrell & Sanfeliu, 2012). From the social robots applications, we highlight the previous paper (Garrell & Sanfeliu, 2012) and the hospital care application (Jeong et al., 2015).

One of the low level tasks that these systems have to face is the automatic pose estimation. If the information of GPS is not available or its accuracy is not enough, one of the usual methods is to localise the robots through detecting landmarks or identifying scenes previously classified. The problem of comparing or aligning two images is usually called *image registration* in the computer vision research field. Image registration tries to determine which parts of one image correspond to which parts of another image. This problem often arises at the early stages of many computer vision applications such as scene reconstruction, object recognition and tracking, pose recovery and image retrieval. Therefore, it has been of basic importance to develop effective methods that are both robust in the sense of being able to deal with noisy measurements and in the sense of having a wide field of application. An example of this research field is Hou, Sun, Jia, and Zhang (2012).

We present an interactive and cooperative method to deduct the relative pose of each robot with respect to the rest of the fleet. The method we present is part of a larger project in which social robots guide people through urban areas (Garrell & Sanfeliu, 2012) and they have tracking abilities (Montiel, Orozco-Rosas, & Sepulveda, 2015; Serratosa, Alquézar, & Amézquita, 2012). Fig. 1 represents three robots performing guiding tasks in an indoor environment. Robots fence the visitor group to force them to follow a specific tour. Robots need to work in a cooperative manner to keep a triangular shape in which people have to be inside. In these cooperative tasks, it is crucial to have a low-level computer vision task such that images extracted from the three robots are properly aligned to correctly deduct their relative pose. In this environment, there is a human that, through our Human–Machine Interface, gives orders to the robots and controls their tasks. Robots have embedded 2D and 3D cameras (http://www.optitrack.com/) and the human can visualise the 2D cameras.

The rest of the paper is organized as follows. In the next section, we summarise the image registration methods and the feature extraction methods that detect salient points on images. Moreover,

☆ This research is supported by Spanish projects DPI2013-42458-P & TIN2013-47245-C2-2-R.

* Corresponding author. Tel.: +34 636281015; fax: +34 977558710.
E-mail addresses: xavier.cortes@urv.cat (X. Cortés), francesc.serratosa@urv.cat, francesc.serratosa@gmail.com (F. Serratosa).

http://dx.doi.org/10.1016/j.eswa.2015.09.049

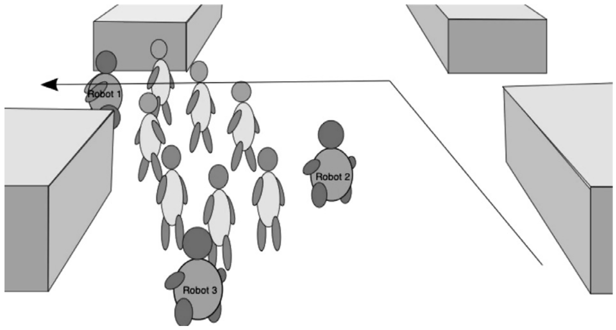

Fig. 1. Three robots performing guiding tasks. Robots are located to fence the group.

we show some results of the interactive image registration methods presented in Cortés and Serratosa (2015b) to demonstrate the validity of these interactive approaches. In Section 3, we present our model from a general point of view. In Sections 4 and 5, we concretise on the interactive relative pose estimation and cooperative methods. In Section 6, we experimentally validate our model and show that, on the one hand, with few human interactions, the accuracy of the estimated relative pose drastically increases. On the other hand, we demonstrate that it is crucial the robot cooperation to deduct the relative poses of all robots. We conclude the paper in Section 7.

2. Basic methods

In this section, we comment some Image registration models and also some models based on the Human Interactivity. Note that it is not the aim of this section to make a strict survey of these two research fields.

2.1. Image registration

The three typical steps involved in the solution of the image registration problem are the following (Xu & Petrou, 2011). First, some salient points are selected from both images. Second, a set of tentative matches between these sets of points is computed together with the image alignment. And third, a process of outlier rejection that eliminates the spurious correspondences can further refine these tentative matches and the initial alignment. Two interesting image registration surveys are Zitová and Flusser (2003) and Salvi, Matabosch, Fofi, and Forest (2007).

Salient points, which play the role of parts of the image to be matched, are image locations that can be robustly detected among different instances of the same scene with varying imaging conditions. These points can be corners (intersection of two edges) (Tomasi & Kanade, 1991), maximum curvature points (Han & Brady, 1995) or isolated points of maximum or minimum local intensity (Rosten & Drummond, 2006). There is an evaluation of the most competent approaches in (Mikolajczyk & Schmid, 2005). When salient points have been detected, several correspondence methods can be applied that obtain the alignment (or homography) that maps one image into the other (Zhang, 1994), discards outlier points (Fischler & Bolles, 1981) or characterises the image into an attributed graph (Sanromà, Alquézar, Serratosa, & Herrera, 2012a; Sanromà, Alquézar, & Serratosa, 2012b; Serratosa, 2014; Serratosa, 2015a; Serratosa, 2015b; Serratosa & Cortés, 2015; Solé, Serratosa, & Sanfeliu, 2012). Typically, these methods have been applied on 2D images but recently, 3D shape retrieval methods have appeared (Lia, 2015).

The main drawback of image registration methods is that their ability to obtain the correspondence parameters strongly depends on the reliability of the initial tentative correspondences. Moreover, it is needed to jointly estimate the image alignment parameters and correspondence parameters.

Considering the alignment parameters, there are two basic strategies. The first one is to consider a rigid deformation and the second one is to consider non-rigid deformation. In the first case, it is assumed the whole image (and so, the extracted salient points) suffers from the same deformation and so the image alignment parameters are applied equally to the whole salient points or image pixels. Some examples are Sanromà et al. (2012a), Gold and Rangarajan (1996), Luo and Hancock (2003), Rangarajan, Chui, and Bookstein (1997). In the second case, each salient points suffers a different projection and there are different alignment parameters applied to each salient point or image region. Some examples are Chui and Rangarajan (2003), Myronenko and Song (2010). Usually, the rigid strategy is applied to detect objects on outdoor images in which the deformation is mostly due to the change of the point of view. The non-rigid strategy is mostly applied to object detection or matching in medical or industrial images due to it is assumed objects suffer from deformations although the point of view is the same.

2.2. Human interactivity

Humans are very good at finding the correspondences between local parts of an image regardless of the intrinsic or extrinsic characteristics of the point of view. Human interactivity on image registration has been applied on medical images (Khader & Ben Hamza, 2012; Pfluger, et al., 2000; Pietrzyk, 1994) and two systems have been patented (Gering, 2010; Von & Neitzel, 2010). These papers and patents are really specific on some medical environments and for this reason cannot be applied on our problem. In Pfluger, et al. (2000), they show a comparison of 3-D images on MRI-SPECT format and they concretise on images from the brain. In Pietrzyk (1994), authors present a method to validate the 3D position given 3D medical images. Finally, in Khader and Ben Hamza (2012), the aim is to solve the registration problem given similar medical images extracted from different sensors or technologies. Patent (Von & Neitzel, 2010) defines a system for registration thorax X-Ray images such that it does not depend on bony structures. Finally, patent (Gering, 2010) defines a multi-scale registration for medical images where images are first aligned at a course resolution, and subsequently at progressively finer resolutions; user input is applied at the current scale. Another usual application of human interaction is semi-automatic video annotation (Bianco, Ciocca, Napoletano, & Schettini, 2015).

Current automatic methods to extract parts of images and their correspondences in non-controlled environments are far away of having the performance of a human. Fig. 2 shows two images extracted from RESID database (http://www.featurespace.org). In each image 50 salient points have been extracted by method (Harris & Stephens, 1988). Outlier detector (Fischler & Bolles, 1981) has considered 43 salient points where outliers and only 7 where inliers. The correspondence detector has missed 6 of the 7 points (red lines) and only has hit 1 point (green line). This is because of the large differences between both images and more precisely, due to the lack of ability of the initial correspondence detector to find a good initial correspondence.

For this reason, in this paper, we propose a semi-automatic method in which humans can interact into the system when it is considered the quality of the automatically found correspondences is not good enough and then they impose a partial and initial correspondence between some local parts of two scenes.

A Human–Machine Interface (HMI), also named Human–Computer Interface, provides more natural, powerful, and compelling interactive experiences. For decades, HMI has been an active research field closely related to new technology advances. Paper published in provides an interesting review of the current trends and key aspects of HMI, for instance in non-desktop computing. Some of