Robotics and Computer-Integrated Manufacturing 30 (2014) 605–621

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

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

Object recognition and pose estimation for industrial applications: A cascade system

Luís F. Rocha [a,*], Marcos Ferreira [a], V. Santos [b], A. Paulo Moreira [a]

[a] INESC TEC - INESC Technology and Science (formerly INESC Porto) and FEUP - Faculty of Engineering, University of Porto, Portugal
[b] Department of Mechanical Engineering, IEETA, University of Aveiro, Portugal

ARTICLE INFO

Article history:
Received 5 June 2013
Received in revised form
28 February 2014
Accepted 29 April 2014

Keywords:
Pattern recognition
Flexible manufacturing
Autonomous systems
Robotics
Spray coating

ABSTRACT

The research work presented in this paper focuses on the development of a 3D object localization and recognition system to be used in robotics conveyor coating lines. These requirements were specified together with enterprises with small production series seeking a full robotic automation of their production line that is characterized by a wide range of products in simultaneous manufacturing. Their production process (for example heat or coating/painting treatments) limits the use of conventional identification systems attached to the object in hand. Furthermore, the mechanical structure of the conveyor introduces geometric inaccuracy in the object positioning. With the correct classification and localization of the object, the robot will be able to autonomously select the right program to execute and to perform coordinate system corrections. A cascade system performed with Support Vector Machine and the Perfect Match (point cloud geometric template matching) algorithms was developed for this purpose achieving 99.5% of accuracy. The entire recognition and pose estimation procedure is performed in a maximum time range of 3 s with standard off the shelf hardware. It is expected that this work contributes to the integration of industrial robots in highly dynamic and specialized production lines.

1. Introduction

1.1. Motivation and proposed approach

Industrial manufacturing has always taken into consideration competitive factors such as time cost and quality. However, manufacturing is more and more characterized by customization which can be accomplished by reducing lot sizes increasing variety and specific products manufactured in short periods of time. Another specificity of next-generation manufacturing is the increasing complexity of products. Instead of rigid infrastructures or inflexible approaches, adaptive manufacturing is envisioned as a new paradigm aiming for continuous and permanent adaptation of all entities inherent to manufacturing systems namely the ones related to manufacturing resources, materials and information flows. The understanding of these multidimensional challenges leads to the use of techniques and tools to improve manufacturing processes and to decrease and eliminate non-value activities (such as transportation tasks, machinery set-up times and others).

Now a days and due to the constant development of technology, sophisticated machinery is increasingly available which allows manufacturing firms to achieve significant process and set-up time reductions. Therefore, and to maintain their competitiveness in the market industrial manipulators must follow this technology evolution or otherwise they will only be used for repetitive processes or mass production scenarios. One of their most limiting features accepted as such from a flexible manufacturing point of view is their programming procedure. Typically this programming is a fairly time consuming process and represents a high investment unfordable for SMEs. These limitations are imposed by the complexity of their teach pendant (human–machine interface) that needs to contain all the commands available to interact with the industrial manipulator, as well as their programming simulator which is only accessible to expert personal. However, the programming procedure is not the only obstacle of industrial manipulators that prevent its widespread use in diversified fields of industry. The lack of capacity that they demonstrate in detecting and locating three-dimensional objects, as well as the stiffness of previously defined motion paths makes it impossible to be applied in highly dynamic production environments. These characteristics are at odds with the current state of the industry.

Therefore, in an industrial environment where an unstructured material flow is present (type of objects arrive randomly and/or

* Corresponding author. Tel.: +351 222 094 000; fax: +351 222 094 050.
E-mail addresses: luis.f.rocha@inescporto.pt (L.F. Rocha), marcos.ferreira@fe.up.pt (M. Ferreira), vitor@ua.pt (V. Santos), amoreira@fe.up.pt (A. Paulo Moreira).

http://dx.doi.org/10.1016/j.rcim.2014.04.005
0736-5845/

are randomly located) vision systems are a must enabling objects visual inspection needed for robot–object interaction.

The scientific developments achieved as part of this work focus on the flexibility enhancement of industrial cells where the key elements are industrial manipulators. The idea is to create a recognition and pose estimation system (robot independent) that indicates to the industrial manipulator the object Id and correspondent trajectory adjustments (varying accordingly to the objects pose). This work makes it possible to increase the flexibility associated to industrial manipulators by introducing new perception skills.

1.2. Literature review

Considering the addressed research problem, this section presents a revision of 3D Model Reconstruction Techniques and Object Classification streams. The focus of this paper is object recognition and pose estimation. Therefore, any scientific contributions were made regarding 3D modeling sensors.

1.2.1. Model reconstruction sensors

Over the last decade there has been a major development in 3D modeling. In [34] it is presented a reasonable set of three-dimensional image reconstruction techniques: (1) Structured light sensor, (2) Stereo Vision, Photometry, and (3) Time of flight and others.

(1) In the field of Structured Light Sensors, laser beam plus camera triangulation systems are usually used. Considering industrial applications, Pinto et al. in [28] sense and measure the position and dimensions of a cork piece in a conveyor belt. In [7] the authors present the same system this time in a coating application. In these approaches the precision values are directly related with the thickness of the laser line and the camera resolution. The major disadvantage is the need of object/robot motion so the 3D model can be created. Furthermore, a well-structured light environment is also required.

In more recent works, as in [16], a new approach is presented for 3D modeling that makes it possible to extract the 3D shape and color. This system is based on a three-dimensional Laser Range Finder (LRF) and a camera, which is relevant if the color feature is important to differentiate objects.

Using a very similar concept, projection of an infra-red pattern (structured light), the Microsoft Kinect[1] now receives most of the attention for 3D modeling especially because it is an affordable tool. Although the sensor has a high potential because it is capable of extracting 3D points model adding the color feature, its resolution falls shortly comparatively to other low cost solutions. The authors in [9] use the Kinect to perform the segmentation of objects present in a scene. In their application, the Kinect RGB camera is used to perform object color segmentation and the depth information used to differentiate objects that are not in the same plane of interest.

(2) In the field of stereo vision, [23] presents an object detection system that uses 3D information provided by stereo reconstruction. According to the authors the resulting system is a high-accuracy far distance obstacle detector, covering a wide range of real scenarios. In [19], Li et al. propose a 3D reconstruction approach based on stereo vision and texture mapping that it could be used in a vast application areas (such as visual navigation of robots or 3D games).

(3) Finally for the time of flight approaches, Cui et al. [3] describe a method for 3D object scanning by aligning depth scans that were taken from around an object. The authors refer that their approach overcomes the sensor's random noise and the presence of a non-trivial systematic bias by showing good quality 3D models with a sensor that presents such low quality data. Due to the simple technology that these sensors have comparatively the authors see potential for low cost production in large volumes.

References [15,1] present an interesting research work with 2D (LRF) to perform 3D scene reconstruction. However, this is usually done in mobile navigation and not with industrial systems. The disadvantages of LRF are their high price for high precision measurements and the measurement variation with the object reflective properties. Although laser based solutions are the most used in the industrial environment, laser beam sensors are the common choice.

For the research work presented in this paper, the solution for 3D modeling must consider the unique characteristics that may distinguish each of the objects to recognize (such as color or/and shape) and the industrial environment limitations where the recognition system will be assembled.

1.2.2. Feature extraction, object recognition and pose estimation

3D models contain a significant amount of information that can be analyzed, which makes it possible to extract the fundamental characteristics of the scene foreground. Object recognition is coarsely composed by two steps: (1) feature extraction and (2) object classification.

(1) To differentiate objects it is simply necessary to distinguish the value of the parameters/features that belong to each class [26]. In the image analysis field (and for the feature extraction purpose) one of the most used techniques for evaluating object shapes is determining the invariant moments as they do not depend on scaling, translation and rotation [8]. Although that is one of the most well known approaches, others such as Fourier descriptors, Dirichlet Laplacian eigenvalues [12] and wavelet descriptors [13] have been developed to describe the shape of different patterns. Mingqiang et al. [21] present a survey of shape feature extraction techniques. Unlike the traditional classification, in the referred paper the authors divided the shape-based feature extraction according to their processing methods.

(2) So, after capturing unique features using some of the techniques referred before object classification algorithms need to be applied. In this research field the most used strategies are from pattern recognition like: machine learning (such as the k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Neural Networks (NN), Hidden Markov Models and Bayesian approaches) and point cloud analysis. There is a large number of research papers available in these areas, combining feature extraction and machine learning techniques.

In [5], it is presented a fingerprint matching scheme based on transform features. In order to extract the unique features, the researchers use the Discrete Cosine, the Fast Fourier and the Discrete Wavelet Transform. Then, the Euclidean distance is used to classify the fingerprint minimum and thus compare two feature vectors. The authors refer that with the combination of the Discrete Cosine and the Fast Fourier it were achieved better results (recognition rate of 87.5%) when compared with the Discrete Wavelet Transform.

In [30] is used a neural network to learn the complex relationship between the robots pose displacements and the observed feature variations on the image. Their objective is to create a visual positioning system that addresses features extraction issues for a class of objects that have smooth or curved surfaces. After the NN training, the authors use visual feedback to guide the robot manipulator to the target.

For its turn, Koker et al. present an industrial machine vision for object recognition [17]. This system was developed based on a CCD

[1] http://www.microsoft.com/en-us/kinectforwindows/.