AI Communications 25 (2012) 373–375
DOI 10.3233/AIC-2012-0520
IOS Press

Thesis

Computer vision methods for robot tasks: Motion detection, depth estimation and tracking

E. Martínez-Martín
Robotic Intelligence Lab, Jaume-I University, Avda. Vicente Sos Baynat s/n, E-12071, Castellón, Spain
E-mail: emartine@icc.uji.es

Abstract. On the way to autonomous robots, perception is a key point. Among all the perception senses, vision is undoubtedly the most important for the information it can provide. However, it is not easy to identify what is seen from the provided visual input. On this regard, inspired by humans, we have studied motion as a primary cue. Particularly, we present a computational solution for motion detection, object location and tracking from images captured by perspective and fisheye cameras. The proposed approach has been validated with an extensive set of experiments and applications using different testbeds of real environments with real and/or virtual targets.

Keywords: Robotics, motion detection, tracking, depth estimation

1. Introduction

Robots are aimed at helping human beings in their daily tasks. So, from industrial robots, which act in restricted scenarios, robot development has been conducted to emulate alive beings in their natural environment. On that way to autonomy, one of the key issues is how a robot perceives its environment and how to use that information to make useful decisions.

In this context, visual perception plays a main role since it can be used in almost all the tasks performed by a robot such as *navigation* (e.g., obstacle avoidance), *manipulation* (e.g., object identification and safety issues), *cooperative behaviour* and *human–robot interaction*, to name some. Nevertheless, although images can contain a huge amount of information, it is not easy to identify what we see from that visual input since we live in a 3D world, but images are 2D. So, from the issues to be solved, we have focused on perception by investigating motion as a primary cue. This is inspired by human vision, where motion is a dominant cue. In addition, motion cue allows to obtain other characteristics such as object's shape, speed or trajectory, which are meaningful for detection and recognition.

For that reason, motion detection is the core of multiple automated visual applications by providing a stimulus to detect objects of interest in the field of view of a sensor. However, the motion observable in a visual input could be due to different factors: movement of the imaged objects, observer movement, changes in the light sources or a combination of (some of) them. Therefore, image analysis for motion detection will be conditional upon the considered factors. In particular, we have focused on motion detection from images captured by perspective and fisheye still cameras. Note that, egomotion has not been considered in this Thesis regarding image processing, although all the other factors have been considered to design and implement the proposed approach, abling to occur at any time.

With that assumption, we propose a *complete sensor-independent* visual system that provides a robust target motion detection. So, firstly, the way sensors obtain images of the world, in terms of resolution distribution and pixel neighbourhood, is studied. In that way, a proper spatial analysis of motion can be performed. Then, a novel background maintenance approach for robust target motion detection is imple-

mented. On this matter, two different situations have been considered: (1) a fixed camera observing a constant background where interest objects move; and (2) a still camera observing interest objects moving within a dynamic background. The reason for this distinction lies on developing, from the first analysis, a surveillance mechanism which removes the previously existing constraints when a reliable initial background model is obtained, since their absence cannot be guaranteed when a robotic system works in an unknown environment. Furthermore, on the way to achieve an ideal background maintenance system, other canonical problems are addressed such that the proposed approach successfully deals with (gradual and global) changes in illumination, the distinction between foreground and background elements in terms of motion and motionless, non-uniform vacillating backgrounds, among others.

Once moving objects are robustly detected in an image, the ability to visually perceive where the targets are, is investigated. For that, more than one image is needed. So, two or more images of the same object from slightly different locations are combined to infer object's position in the 3D world. So, each detected target should be represented in a way allowing to properly establishing image correspondences in different time steps. Note that there are a million of appearances of the same object, and changes in image contrast, intensity or colour can lead to a mismatch. As a solution, we propose an invariant object representation. It identifies an object among a broad range of objects, even when they leave and re-enter the scene, being robust to partial (and total) occlusions and able to learn new targets from only one frame. Moreover, the designed representation includes a feature array that helps to discard false matches and to make correct decisions.

Then, depth estimation is studied from two different points of view by depending on the accuracy level required at any time. On the one hand, the active vision paradigm is used to build a relative scene representation (in terms of *disparity*). From a robotic point of view, this binocular depth estimation can be used for motion control, since it provides the knowledge required to properly interact with the surrounding environment. On the other hand, a *comparative* depth perception (i.e., with respect to other objects) can be more appropriate for certain tasks. As a result, a reasoning inference process for distance estimation from a visual input is presented. It provides context-dependent comparisons that helps the system make decisions about the objects to focus on. Furthermore, the analysed model is abstracted such that a general model to solve the reasoning process of all qualitative models based on intervals has been developed.

2. Contributions

Thus, this Thesis [1] addresses various problems in Computer Vision - specifically, image acquisition with different image devices, robust motion detection, depth estimation from both a quantitative and a qualitative point of view, and the design of a new object representation leading to a reliable tracking process. The proposed approach has been implemented and validated with different testbeds designed for that purpose. As results have shown, the proposed approach introduces important advantages with respect to the state-of-the-art methods by being its major contributions the following:

- Design and implementation of a robust motion detection algorithm for different image devices (i.e., perspective and fisheye cameras). To be successful in such task, the proposed approach includes a dynamic, adaptive method for automatically setting segmentation thresholds based on image resolution distribution.
- Foreground and background object distinction in terms of motion and motionless situations and adaptation to changes in illumination.
- Depth estimation of targets from a disparity map based on phase-difference.
- Qualitative depth estimation from a general reasoning inference process.
- Simultaneous tracking of several individuals, dealing with (partial) occlusions and changes in image characteristics and viewpoint.
- Test within several scenarios and applications such as animal behaviour analysis, multi-robots, traffic monitoring, sports analysis and robot navigation.

Therefore, the methods proposed in this Thesis provide important advances with respect to state-of-the-art computer vision approaches in terms of robot reliability. The motion detection algorithm allows a good environment adaptation of the system as it properly deals with most of the vision problems when dynamic, non-structured environments are considered. In addition, the proposed object representation satisfies the requirements of a recognition task, even when movements are accompanied by changes in shape and/or size. Moreover, the integration of depth estimation into the sys-

tem improves the matching process as well as helps to arrange targets in order of proximity. In that way, the system can focus on those objects closer to it. All these contributions are validated with an extensive set of experiments and applications using different testbeds of real environments with real and/or virtual targets.

Reference

[1] E. Martínez-Martín, Computer vision methods for robot tasks: motion detection, depth estimation and tracking, PhD thesis, Jaume-I University, 2011.