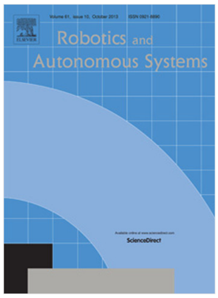
Accepted Manuscript



Coordinating robotic gaze using maximal octant difference election (MODE) visual features

Ruben Gonzalez

PII: S0921-8890(17)30039-8
DOI: http://dx.doi.org/10.1016/j.robot.2017.07.005
Reference: ROBOT 2876

To appear in: *Robotics and Autonomous Systems*

Received date : 17 January 2017
Revised date : 6 June 2017
Accepted date : 13 July 2017

Please cite this article as: R. Gonzalez, Coordinating robotic gaze using maximal octant difference election (MODE) visual features, *Robotics and Autonomous Systems* (2017), http://dx.doi.org/10.1016/j.robot.2017.07.005

Coordinating Robotic Gaze using Maximal Octant Difference Election (MODE) Visual Features

Ruben Gonzalez
Griffith University
Gold Coast Australia
r.gonzalez@griffith.edu.au

***Abstract*— In multi agent mobile robotic systems, coordinating the gaze of each autonomous robot without the use of pre determined gaze attractors requires the sharing of either a gaze vector together with knowledge of the location of each robotic or sharing of a visual description of what the gaze should be directed at. In the latter case, this should consist of the least amount of data possible to reduce transmission delay and processing requirements. Sharing a single unique visual feature is sufficient to direct gaze. Unfortunately, typical feature detection methods are tuned to detect many hundreds of features to maximize successful matches in corresponding images. This paper presents an alternate feature detection method that is tuned to select only a few, high quality features. Not only does this method achieve the accuracy of state of the art methods over the target range, it does so quicker and results in a significantly smaller amount of information that needs to be shared between agents for directing gaze to the same point of interest.**

Keywords—feature detection; gaze control; mobile robot coordination

I. INTRODUCTION

Controlling gaze is a known problem in vision endowed mobile robots. In the case of a single mobile robotic agent, the problem comes down to optimally controlling gaze for determining preferred navigation trajectories. This can be posed as a problem of reducing uncertainty [1, 2] such as via entropy optimisation [3]. Another approach might be to endow the robot with a suitably tuned gaze attractor or attention operator [4] based on the robot's objectives. Gaze control in the case of autonomous multi robot systems tracking an object of interest is altogether different. While each robot could be endowed with identical attention operators permitting them to independently determine where to direct their gaze, convergence by implicit means is not guaranteed. Even if full or partial convergence can be attained it would not necessarily be in a timely manner. While divergence may be the desired behavior in some cases, where timely convergence is required or where only a few "scout" robots specialised for reconnaissance have this capability the robot in the network that first senses the position and velocity of the target must broadcast this to all other nodes in the network [5, 6]. For this, each robot must be able to accurately determine its absolute position and orientation in 3D space. While this is a common capability, this approach cannot be used when positioning and orientation information is not available.

Given a gaze target, an alternate vision inspired approach, is for the lead node to broadcast a description or abstraction of what the gaze should be directed at to the other robots. Each node then searches for the object of interest by adjusting their gaze such that it most closely matches the shared description. For this to be successful, the description must be able to unambiguously identify the gaze target. It must also be invariant to the distance to the target, i.e., the size, location and relative orientation of the target. It is assumed that the robots are sufficiently distant to the target to minimize the effects of perspective distortion. This problem is closely related to that of object recognition in mobile robots [7], and vision based robot localization [8], which are hampered by the high computational complexity of common image recognition methods. As the location and orientation of robots relative to reference points are commonly determined by detecting and matching visual features, a similar approach can be used to determine the location of a given target relative to the robot.

A completely general solution, where the target orientation relative to the nodes is unconstrained, requires a full 3D description of the target, but this is can only be guaranteed if the target is predetermined and its description from all viewpoints is known *apriori*. When the target is not known beforehand, the only information initially available to the first observing node is a 2D image of the target from the perspective of the node. Hence the description of the gaze target shared with other nodes is, at least initially, constrained to being 2D.

In addition there may also be constraints on the amount of data that can be used for communicating the visual description with the other nodes. First, there is likely to be a preference to minimize reaction latency in the robots by reducing data transmission and processing delays. The perennial desire to conserve battery power is also dependent on reducing data transmission time. This sets the criteria for a compact, yet transformation invariant and discriminating, feature based description of the gaze target be used.

Detecting, extracting and matching image features is a well-explored problem in computer vision for applications such as image registration, camera calibration, depth