Reprinted from

Robotics and Autonomous Systems

Robotics and Autonomous Systems 19 (1997) 337–346

Visually guided grasping in unstructured environments

Roberto Cipolla *, Nick Hollinghurst [1]
Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

ELSEVIER

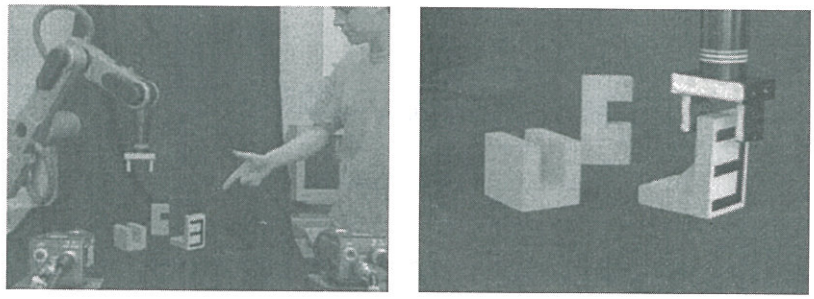

Fig. 1. The arrangement of the uncalibrated stereo cameras and the robot. The operator points at an object, and the robot picks it up under visual control.

to the specific application because there are no fixed constraints on the camera geometry.

Operation comprises three distinct stages. In the first stage, an operator indicates the object to be grasped by simply pointing at it. Next, the vision system segments the identified object from the background by grouping edges into planes and plans a suitable grasp strategy. Finally, the robotic arm reaches out towards the object and executes the grasp. Here we present the three vision algorithms used for these three separate stages.

Fig. 1 shows a typical set-up. The robot arm has 5 degrees of freedom (DOF) and a parallel-jawed gripper. The robot has its own controller for the low-level control and provides a Cartesian kinematic model. Two cameras are placed 1.5–3 m from the robot's workspace. The angle between the cameras is in the range of 15–30°.

2. Indicating the target by pointing

For the first part of the experimental system, we use machine vision in conjunction with a human operator who indicates the object to be grasped by pointing at it. A pair of cameras view the pointing hand in stereo. Unlike most existing gesture-based interfaces (e.g. [9,22]) our system requires no special gloves or markers. Active contours are used to track the hand in real time. A simple result from projective geometry allows us to estimate where the hand is pointing to on the robot's work table, using just four coplanar reference points. The Cartesian coordinates of the reference points are also unknown and camera calibration is not required.

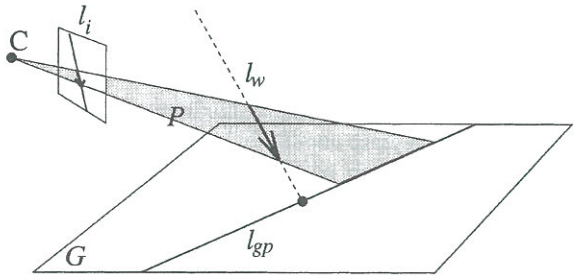

Fig. 2. Relation between lines in the world, image and ground planes.

2.1. Geometrical framework

A single view of a pointing hand is ambiguous: its distance from the camera cannot be determined, and the 'slant' of its orientation cannot be measured with any accuracy. This means that the 'piercing point', where the line defined by the hand intersects the work surface, is constrained to a line, which is the projection of the hand's line in the image (see Fig. 2). A second view is needed to fix its position in two dimensions [18].

Consider a pair of pinhole cameras viewing a planar surface (such as the robot's work surface). The viewing transformations can be modelled by plane projectivities, and there exists a projective transformation that maps one image to the other. This transformation can be computed by observing a minimum of four points on the plane. We exploit this to transform the constraint lines into a common 'canonical' view of the plane, and hence find their intersection (Fig. 3).