International Journal of Computer Vision 48(2), 75–90, 2002

Geotensity: Combining Motion and Lighting for 3D Surface Reconstruction

ATSUTO MAKI, MUTSUMI WATANABE* AND CHARLES WILES*
Toshiba Corporation, Corporate Research and Development Center, 1 Komukai-Toshiba-cho, Saiwai-ku, Kawasaki, 212-8582 Japan
atsuto.maki@toshiba.co.jp

Received December 29, 1998; Revised February 5, 2002; Accepted February 6, 2002

Abstract. This paper is about automatically reconstructing the full 3D surface of an object observed in motion by a single static camera. Based on the two paradigms, structure from motion and linear intensity subspaces, we introduce the *geotensity constraint* that governs the relationship between four or more images of a moving object. We show that it is possible in theory to solve for 3D Lambertian surface structure for the case of a single point light source and propose that a solution exists for an arbitrary number point light sources. The surface may or may not be textured. We then give an example of automatic surface reconstruction of a face under a point light source using arbitrary unknown object motion and a single fixed camera.

Keywords: geotensity, linear image subspaces, structure-from-motion, surface reconstruction

1. Introduction

Recovering 3D surface structure of an object has been a central issue in computer vision. In order to extract 3D information out of 2D images, it is useful to have a set of images showing the target object viewed from different directions. Stereo naturally provides two images each viewing the object from a different direction and it has been successfully applied to 3D reconstruction. The essential problem that must be solved is to match corresponding points in two or more images so that these point are the projections of an identical point on the surface of the object, and this typically exploits the constraint that the corresponding parts of the images have the same image intensities. Indeed, this constraint is often valid for stereo where two images are taking simultaneously as the lighting of the object is identical in each image. Various authors have used this constraint to achieve quite accurate surface reconstructions of human faces under stereo (Devernay and Faugeras, 1994; Fua, 1995).

A natural progression then is to replace the stereo camera with a single camera observing an object in motion. For example, a camera fixed on top of a workstation may observe a person shaking their head from side to side. Unfortunately, the constant image intensity constraint is nearly always invalid as non-uniform lighting causes the intensity at a specific location on the surface of an object to change as the object moves. Indeed Fig. 1 shows four views of a head selected from a sequence spanning only two seconds and show that the combined effect of motion and lighting can dramatically effect the intensity of a surface point.

However, if it were possible to derive a constraint on how the intensity of a surface point may change between images we could replace the constant intensity constraint used in stereo surface reconstruction algorithms with a more complicated, but just as appropriate constraint that we could use for surface reconstruction under object motion. Moreover, to be practically useful, we must be able to derive the parameters of this constraint automatically from arbitrary images of the

*Work performed while at Kansai Research Laboratories, Corporate Research and Development Center, Toshiba Co.