Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 26 (2005) 1620–1631
Pattern Recognition Letters
www.elsevier.com/locate/patrec

A fast and robust feature-based 3D algorithm using compressed image correlation

Sheng S. Tan *, Douglas P. Hart

Mechanical Engineering Department, Massachusetts Institute of Technology, 77 Mass Avenue, 3-243, Cambridge, MA 02139, USA

Received 19 March 2004; received in revised form 19 December 2004
Available online 19 April 2005
Communicated by E. Backer

Abstract

Two objectives of 3D computer vision are high processing speed and precise recovery of object boundaries. This paper addresses these issues by presenting an algorithm that combines feature-based 3D matching with Compressed Image Correlation. The algorithm uses an image compression scheme that retains pixel values in high intensity gradient areas while eliminating pixels with little correlation information in smooth surface regions. The remaining pixels are stored in sparse format along with their relative locations encoded into 32-bit words. The result is a highly reduced image data set containing distinct features at object boundaries. Consequently, far fewer memory calls and data entry comparisons are required to accurately determine edge movement. In addition, by utilizing an error correlation function, pixel comparisons are made through single integer calculations eliminating time consuming multiplication and floating point arithmetic. Thus, this algorithm typically results in much higher correlation speeds than spectral correlation and SSD algorithms. Unlike the traditional fixed window sorting scheme, adaptive correlation window positioning is implemented by dynamically placing object boundaries at the center of each correlation window. Processing speed is further improved by compressing and correlating the images in only the direction of disparity motion between frames. Test results on both simulated disparities and real motion image pair are presented.

Keywords: Compressed image correlation; Gradient-based compression; Feature matching; Adaptive window; 3D vision

1. Introduction

Features such as edges and corners play an important role in human vision. The visual cortex is especially responsive to strong features in a

* Corresponding author. Tel.: +1 617 253 0229; fax: +1 240 214 8665.
E-mail address: shtan@alum.mit.edu (S.S. Tan).

0167-8655/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2005.01.009

scene (Sharma et al., 2003). Together with related abilities such as correspondence matching and tracking, humans are able to react quickly to the environment and focus attention on objects of interest.

The significance of features is fully recognized in computer vision. For example, one traditional class of techniques applied to facial recognition is based on the computation of a set of geometrical features from a picture of the face such as the sizes and relative positions of eyes, mouth, nose and chin (Brunelli and Poggio, 1993; Gao and Leung, 2002). There is even belief that edge representations may contain all of the information required for the majority of higher level tasks (Elder, 1999).

Feature-based 2D tracking is extensively implemented in automated surveillance, robotic manipulation and navigation. Because real-time processing is a necessity in these applications, only perceptually significant information such as contours is retained from video feeds. If the target's 3D model is known, its detected contours are compared against its geometrical model to determine the object's current position and orientation (Drummond and Cipolla, 2002). If there is no a priori knowledge of the target, it is tracked by finding the contours' disparity between frames using cross correlation (Deriche and Faugeras, 1990) or level set (Mansouri, 2002).

Passive 3D imaging can be reduced to the problem of resolving disparities between image frames from one or several cameras. Some key issues involved are lack of texture, discontinuity and speed. Numerous algorithms have been proposed to address these issues that fall into three broad categories: feature-based, area-based and volume-based algorithms (Scharstein and Szeliski, 2002). Same as in 2D tracking, feature-based 3D imaging techniques are able to process an extensive amount of video data in real-time while providing enough latency for high level tasks such as object recognition (Hoff and Ahuja, 1989; Olsen, 1990). This group of methods generates sparse but accurate depth maps at feature points and excels at determining object boundary position where area-based techniques often fail. When a full-field depth map is desirable, the sparse 3D representation provides a solid foundation for additional area- or volume-based algorithms to fill in the voids when there is ample surface texture; otherwise, when texture is scarce or highly repetitive, object segmentation methods and interpolation are preferable (Izquierdo, 1999; Mahamud et al., 2003).

In the emerging field of image-based 3D editing, which has many applications in architectural design and entertainment, long and tedious human efforts are required to manually extract layers and assign depths for a 2D image (Oh et al., 2001). Automatic feature-based depth detection would greatly facilitate this process.

Broad adoption of 3D imaging technology is currently limited by speed and robustness. Applications such as robotic surgery (Hoznek et al., 2002) and autonomous navigation or tracking, demand real-time processing. As an example, a texture mapped 3D scene would greatly aid in a surgeon's tactile sense. 3D reconstructed views enable better object recognition without turning the camera and taking more images. 3D object tracking would be much more robust than its 2D counterpart if reliable depth information is available.

1.1. Relation to prior work

Current methods of finding feature correspondence can be categorized into global or local techniques. Global approaches to the sparse correspondence problem handle the entire set of sparse points by forming a global optimization function. Various constraints such as color constancy, continuity, uniqueness and epipolar constraints guide the search of a global solution. (Bartoli et al., 2003; Maciel and Costeira, 2003). Global techniques are usually robust but relatively slow due to the iteration and optimization process.

Local methods find each pixel's correspondence by computing a cost function in a small interrogation window around the pixel of interest. Popular cost functions include SSD (Sum of Squared Differences) and cross correlation. Sparse Array Image Correlation is commonly implemented in the field of Particle Image Velocimetry (PIV), where fluid fields are seeded with fluorescent tracer particles and illuminated with a laser sheet. Flow motion is measured by tracking particle displacement. PIV images are comprised of millions of bright spots