Affine invariant regions++

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of
Doctor of Technical Sciences

presented by
VITTORIO FERRARI
Dipl. Computer Science
born 8th September 1976
citizen of Lugano (TI)

accepted on the recommendation of
Prof. Dr. Luc Van Gool, examiner
Prof. Dr. David Lowe, co-examiner

2004
Abstract

Many problems in Computer Vision require the computation of correspondences between images. In order to cope with large differences in viewing conditions, several affine invariant region detectors have been developed in the last few years. These regions automatically adapt their shape so as to cover the same scene surface in any view. This dissertation builds upon existing detectors and develops various novel techniques which extend the power and functionalities of the regions.

The advances relate to different subfields of Vision, and can be summarized as four main contributions. First and foremost, the thesis presents a powerful Object Recognition system capable of working with large amounts of background clutter, severe occlusion, and strong viewpoint and scale changes. It can handle non-rigid deformations, and also finds the contours of the visible parts of the object.

The second innovation consists of a method to obtain region correspondences across several images taken from different viewpoints. These multi-view correspondences are important as they enable the automatic reconstruction of a 3D model given only a few still images. In contrast, traditionally this task requires a complete video as input.

Another branch of the thesis introduces a real-time algorithm which tracks the full affine shape of a region as it evolves through a video, and its application for markerless Augmented Reality. Most prior works instead rely on adding special markers to the scene.

Lastly, a technique to automatically find groups of regions correspondences lying on planar surfaces is presented. This allows to detect planar scene structures and their geometric transformation between views, which in turn can considerably simplify 3D reconstruction procedures, and is useful for robot navigation.
Zusammenfassung