Optical flow from maxima in distance transform images

Charles Beumier, Xavier Neyt
Royal Military Academy of Belgium, Signal & Image Centre
beumier@elec.rma.ac.be

Introduction

The MoD *DAP-9* research project about "Persistent surveillance using a UAV for IED early warning" aims at tracking people and vehicle for suspicious activity detection from aerial surveillance in the context of Improvised Explosive Devices. Most trackers need camera motion compensation which is usually performed by matching descriptors of interest points (e.g. SIFT [1]). This short paper presents an alternative approach for the case of nearly static cameras.

Principle

In the case of nearly static cameras, acquired images share many areas with similar shape and texture corresponding to stationary objects. Most edge points in images are stable so that their image distance transforms ([2]) look alike. We found that T junctions in the skeleton of the distance map are valid key points for image registration.

Implementation

In each image, 5 % of the pixels with largest gradient are kept. The distance map is grown from these pixels by a fast 2-pass algorithm (Fig 1b). Local maxima of the distance map form the so-called image skeleton (Fig 1c) from which T-junctions are extracted. These precisely localised maxima appeared to be discriminant and rather stable in the case of image sequences acquired from a nearly static camera.

The coordinates of the T junctions are collected together with the following features: the local maximal distance, three orientations of the T-junction branches and the local mean image intensity.

T-junction points of two images are matched thanks to features in order to derive image local displacement vectors necessary to register or compare images. Matching is improved by looking for locally coherent translation vectors.

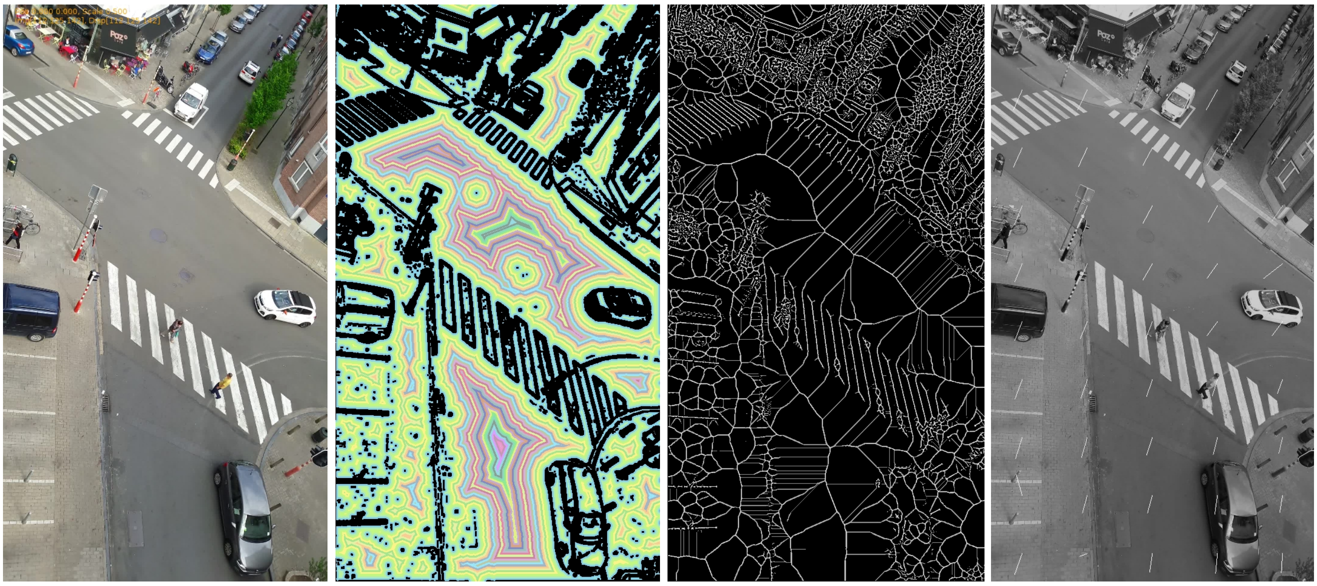

Fig 1. a) Source image; b) Distance transform; c) Skeleton of (b); d) Optical flow

Results

The automatically extracted displacement vectors, grouped locally to extract the most representative translation on a grid, as shown in Fig 1d, is the desired optical flow. This flow is globally coherent, also in the case of small camera rotation. It will help registering images to look for moving objects by image differencing. The image processing steps are fast but not real-time, especially due to a slow skeleton estimation. We are looking for a better implementation. Junction point matching concerns a few thousand points for most video images and can be handled at video rate.

The future work will first consider optimizing the code for near real-time processing, probably handling multi-resolution for higher speed and additional junction points from lower resolutions. The next phases for persistence surveillance will be to apply frame differencing to detect moving objects and track them. Thanks to their speed and trajectory, suspicious behaviours will be detectable.

References

[1] D. Lowe, "Object recognition from local scale-invariant features", *Int. Conf. on Computer Vision*, Corfu, Greece (Sep 1999), pp. 1150-1157.
[2] G. Borgefors, “Distance transformations in digital images”, Computer Vision, Graphics, and Image Processing, 34: 344-371, 1986.