Hauptseminar SS2007 (Englisch)

Title: Real-Time 3D Visual Tracking Techniques

Dr. Giorgio Panin

Abstract

From *Wikipedia*: definition of Visual Tracking

Visual tracking is the process of locating a moving object (or several ones) in time using a camera. An algorithm analyzes the video frames and outputs the location, optionally in real time. A visual tracking algorithm is based on a motion model which describes how the image of the target changes depending on a vector of motion parameters. When the target is a rigid 3D object, the motion model defines its aspect depending on its 3D position and orientation. The role of the tracking algorithm is to analyse the video frames in order to estimate the motion parameters. These parameters characterize the location of the target.

Applications of real-time 3D tracking cover many fields of interest, for which commercial products are increasingly becoming available.

Today there are several known approaches for this task, and we consider here some advanced techniques, by organizing them into

- Point-based tracking: Tracking of single *point* features, followed by least-squares object pose estimation.
- Contour-based tracking: Detection of the object boundary *line* (e.g. active contours, or Condensation algorithm) as it deforms with the roto-translation of the object in space
- Template-based tracking: Registration of the whole object *surface*, given as a triangular mesh, together with its texture (i.e. the surface appearance)

Lecture Slides from our WS06/07 Course "*Advanced 3D Tracking Methodologies*" provide also useful material for this Seminar. They are available at

http://atknoll1.informatik.tu-muenchen.de:8080/tum6/lectures/courses/ws0607/tracking

Please visit also the following Webpage for further interest (NOTE: the page is still "work in progress")

http://www.trackingsystems.org/

Proposed Themes for the Seminar

1) Point-based tracking: SIFT Algorithm + Pose estimation

A first modality to perform 3D tracking employs *local features* matching. This technique extracts relevant feature points from the current image, and tries to match them against a "database" of reference feature points, by using the features "descriptors".

SIFT (Scale Invariant Feature Transform, [Lowe04]) is a method for extracting distinctive invariant features from images, which can be used to perform reliable matching between different images of an object or scene. The features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, addition of noise, change in 3D viewpoint, and change in illumination. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images.

This can be a starting point for an approach that uses these features for object recognition. The recognition proceeds by matching individual features to a database of features from known objects using a fast nearest-neighbor algorithm followed by a Hough transform to identify clusters belonging to a single object, and finally performing verification through least-squares solution for consistent pose parameters (e.g. the POSIT algorithm [DeMenthon92]). This approach to recognition can robustly identify objects among clutter and occlusion while achieving near real-time performance.

Main References:

- David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* 60, 2 (2004), pp. 91-110.
http://www6.in.tum.de/~panin/seminar07/lowe04distinctive.pdf

- Daniel DeMenthon, Larry S. Davis, *Model-Based Object Pose in 25 Lines of Code*, ECCV '92: Proceedings of the Second European Conference on Computer Vision, 1992, (pp. 335-343), Springer-Verlag, London, UK
http://www6.in.tum.de/~panin/seminar07/Pose25Lines.pdf

2) Feature Points Tracking: The Shi-Tomasi Approach

No feature-based vision system can work unless good features can be identified and tracked from frame to frame. Although tracking itself is by and large a solved problem, selecting features that can be tracked well and correspond to physical points in the world is still hard. They propose a feature selection criterion that is optimal by construction because it is based on how the tracker works, and a feature monitoring method that can detect occlusions, un-occlusions, and features that do not correspond to points in the world. These methods are based on a new tracking algorithm that extends previous Newton-Raphson style search methods to work under affine image transformations. They test performance with several simulations and experiments.

Main References:

- Jianbo Shi and Carlo Tomasi. Good Features to Track. IEEE Conference on Computer Vision and Pattern Recognition, pages 593-600, 1994
http://www6.in.tum.de/~panin/seminar07/shi04.pdf

3) Contour-based tracking with Particle Filters – The Condensation Algorithm

The problem of tracking curves in dense visual clutter is challenging. Kalman filtering is inadequate because it is based on Gaussian densities which, being unimodal, cannot represent simultaneous alternative hypotheses. The CONDENSATION algorithm uses “factored sampling", previously applied to the interpretation of static images, in which the probability distribution of possible interpretations is represented by a randomly generated set. CONDENSATION uses learned dynamical models, together with visual observations, to propagate the random set over time. The result is highly robust tracking of agile motion. Notwithstanding the use of stochastic methods, the algorithm runs in real-time.

The paper by Klein and Murray demonstrates a real-time, full-3D edge tracker based on a particle filter. In contrast to previous methods this system is capable of tracking complex self-occluding three-dimensional structures. The system exploits graphics hardware in a novel manner, allowing it not only to perform hidden line removal for each particle but also to evaluate pose likelihoods directly on the graphics card. This approach allows video-rate filtering with hundreds of particles on a standard workstation.

Main References:

- Michael Isard, Andrew Blake, *Condensation -- conditional density propagation for visual tracking*, International Journal of Computer Vision (IJCV), 1998, vol. 29 n. 1 (pp. 5-28)
http://www6.in.tum.de/~panin/seminar07/ijcv98.pdf

- Georg Klein, David Murray, Full-3D Edge Tracking with a Particle Filter In Proc. British Machine Vision Conference (BMVC'06, Edinburgh - Poster)
http://www6.in.tum.de/~panin/seminar07/klein_murray_bmvc2006.pdf

4) Contour-based tracking with Local image statistics – The CCD Algorithm

The task of fitting parametric curve models to boundaries of perceptually meaningful image regions is a key problem in computer vision with numerous applications, such as image segmentation, pose estimation, 3-D reconstruction, and object tracking. The Contracting Curve Density (CCD) algorithm and the CCD tracker are solutions to this problem. The CCD algorithm solves the curve-fitting problem for a single image whereas the CCD tracker solves it for a sequence of images.

The CCD algorithm extends the state-of-the-art in two important ways. First, it applies a novel likelihood function for the assessment of a fit between the curve model and the image data. This likelihood function can cope with highly inhomogeneous image regions because it is formulated in terms of local image statistics that are learned on the fly from the vicinity of the expected curve. Second, the CCD algorithm employs blurred curve models as efficient means for iteratively optimizing the posterior density over possible model parameters. Blurred curve models enable the algorithm to trade-off two conflicting objectives, namely a large area of convergence and a high accuracy.

The CCD tracker is a fast variant of the CCD algorithm. It achieves a low runtime, even for high-resolution images, by focusing on a small set of carefully selected pixels. In each iteration step, the tracker takes only such pixels into account that are likely to further reduce the uncertainty of the curve. Moreover, the CCD tracker exploits statistical dependencies between successive images, which also improves its robustness. This can be achieved without substantially increasing the runtime.

Real-Time 3D tracking with this algorithm has been recently demonstrated and presented in [Panin06].

Main References:

- Robert Hanek, Michael Beetz, *The Contracting Curve Density Algorithm: Fitting Parametric Curve Models to Images Using Local Self-adapting Separation Criteria*, International Journal of Computer Vision (IJCV), 2004, vol. 59 n. 3, (pp. 233-258)
http://www6.in.tum.de/~panin/seminar07/14_Hanek04.pdf

- Giorgio Panin, Alexander Ladikos, Alois Knoll, *An Efficient and Robust Real-Time Contour Tracking System*, IEEE International Conference on Vision Systems 2006, New York (USA), January 2006
http://www6.in.tum.de/~panin/seminar07/ICVS06.pdf

5) Contour-based Tracking based on the image Edge Map

3D Tracking can be performed by using an edge map, directly extracted from the image. An earlier work by Chris Harris describes how to obtain 3D pose tracking by matching the wire-frame CAD model of the object and the extracted edge map.

Subsequently, the paper by Drummond and Cipolla presents a novel framework for three-dimensional model-based tracking. Graphical rendering technology is combined with constrained active contour tracking to create a robust wire-frame tracking system. It operates in real time at video frame rate (25 Hz) on standard hardware. It is based on an internal CAD model of the object to be tracked which is rendered using a binary space partition tree to perform hidden line removal. The visible edge features are thus identified online at each frame and correspondences are found in the video feed. A Lie group formalism is used to cast the motion computation problem into simple geometric terms so that tracking becomes a simple optimization problem solved by means of iterative reweighted least squares. A visual servoing system constructed using this framework is presented together with results showing the accuracy of the tracker.

Main References:

- Chris Harris, “Tracking with rigid models”, Active vision, MIT Press, Cambridge, MA, 1993

- Tom Drummond , Roberto Cipolla, “Real-Time Visual Tracking of Complex Structures”, IEEE Transactions on Pattern Analysis and Machine Intelligence, v.24 n.7, p.932-946, July 2002

http://www6.in.tum.de/~panin/seminar07/Drummond02.pdf

6) Template-based tracking: Active Appearance Models

Active Appearance Models (AAMs) and the closely related concepts of Morphable Models and Active Blobs are generative models of a certain visual phenomenon. Although linear in both shape and appearance, overall, AAMs are nonlinear parametric models in terms of the pixel intensities. Fitting an AAM to an image consists of minimising the error between the input image and the closest model instance; i.e. solving a non-linear optimization problem. They propose an efficient fitting algorithm for AAMs based on the inverse compositional image alignment algorithm. They show that the effects of appearance variation during fitting can be precomputed using this algorithm and how it can be extended to include a global shape normalising warp, typically a 2D similarity transformation. They evaluate our algorithm to determine which of its novel aspects improve AAM fitting performance.

AAM are also used for 3D tracking, by providing an estimate of roto-translation parameters following the 2D template deformation, thus combining 2D+3D model, as in the paper below mentioned.

Main References:

- T.F. Cootes, G.J. Edwards, C.J. Taylor, *Active Appearance Models*, Lecture Notes in Computer Science, 1998.
http://www6.in.tum.de/~panin/seminar07/16_Cootes98.pdf

- I. Matthews and S. Baker *Active Appearance Models Revisited*, International Journal of Computer Vision, Vol. 60, No. 2, November, 2004, pp. 135 – 164
http://www6.in.tum.de/~panin/seminar07/17_Matthews04.pdf

- J. Xiao, S. Baker, I. Matthews, and T. Kanade *Real-Time Combined 2D+3D Active Appearance Models*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June, 2004.
http://www6.in.tum.de/~panin/seminar07/19_Xiao04.pdf

- http://www.ri.cmu.edu/projects/project_448.html

7) Template-based tracking : the EigenTracking technique

The paper by Black and Jepson describes an approach for tracking rigid and articulated objects using a view-based representation. The approach builds on and extends work on eigenspace representations, robust estimation techniques, and parameterized optical flow estimation. First, they note that the least-squares image reconstruction of standard eigenspace techniques has a number of problems and they reformulate the reconstruction problem as one of robust estimation. Second they define a “subspace constancy assumption” that allows them to exploit techniques for parameterized optical flow estimation to simultaneously solve for the view of an object and the affine transformation between the eigenspace and the image. To account for large affine transformations between the eigenspace and the image they define a multi-scale eigenspace representation and a coarse-to-fine matching strategy. Finally, they use these techniques to track objects over long image sequences in which the objects simultaneously undergo both affine image motions and changes of view. In particular they use this “EigenTracking” technique to track and recognize the gestures of a moving hand.

Main References:

- Michael J. Black , Allan D. Jepson, “EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation”, International Journal of Computer Vision, v.26 n.1, p.63-84, Jan. 1998
http://www6.in.tum.de/~panin/seminar07/black98eigentracking.pdf