

Visual Tracking Methods for Augmented Reality

Georg Klein

Scientist

Microsoft Corporation

Real world



A scenic view of a snow-covered mountain range under a clear blue sky. In the foreground, a valley is filled with snow and scattered evergreen trees. A prominent, jagged mountain peak rises in the background. A bright yellow starburst graphic is overlaid on the image, containing text and a yellow arrow pointing towards the mountain.

Apfelstrudel
50 CHF

Tracking works out where we are looking, so that graphics can be drawn in the right place



Good Tracking is:

- Accurate
- Precise
- Low-latency
- Agile
- Robust



Visual Tracking

- Establishes **correspondences** between the video feed and 3D positions in space

$$(u, v) \leftrightarrow (x, y, z)$$

- 6-DOF Position can be calculated from these correspondences

Lots of tools: Three-point-pose, RANSAC, N-point-pose, iterative nonlinear optimisation, robust M-estimation

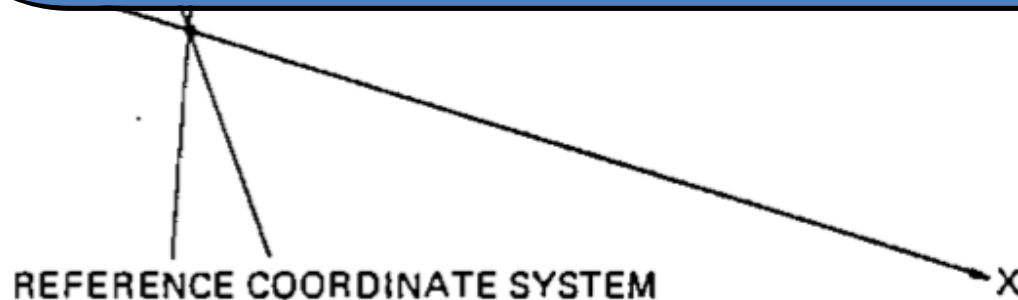


FIGURE 2 GEOMETRY OF THE LOCATION DETERMINATION PROBLEM

From Fischler & Bolles, 1980 (RANSAC / 3PP)

Approaches to Visual Tracking

- Use a **marker**
 - Corners of square give easy correspondences
- Use a **known textured object**
 - Coordinates of texture features are known
- Learn an **unknown environment on-line**
 - Coordinates of scene are computed on-the-fly

Marker-based Tracking

- Distinctive shapes which can be found using elementary image processing operations



ARToolkit (Kato & Billinghurst 1999)



Sony (Rekimoto et al)

Marker-based Tracking

- Best suited for tangible manipulation of virtual elements and untrained users
- Unsuitable for uncontrolled environments

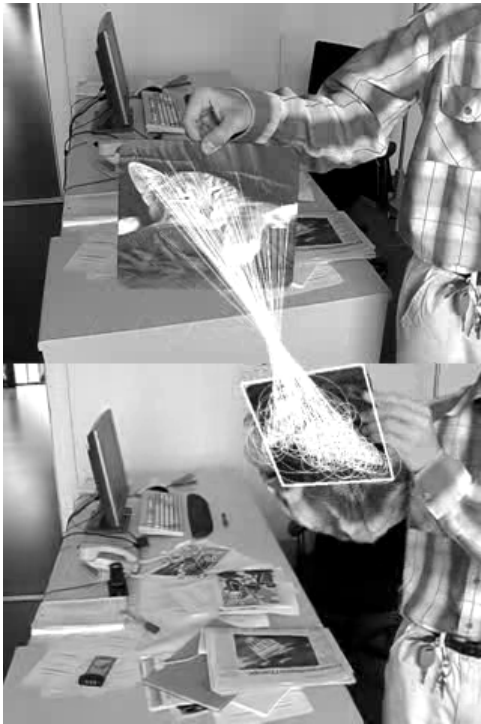


Known-Template Tracking

- Exploits advances in image processing: **Rapid feature extraction** and **invariant descriptor matching**
- Distinctive points of a textured object are matched to the image
- Video: FERNS (Özuysal, Fua & Lepetit 2007)
- Must be known in advance!

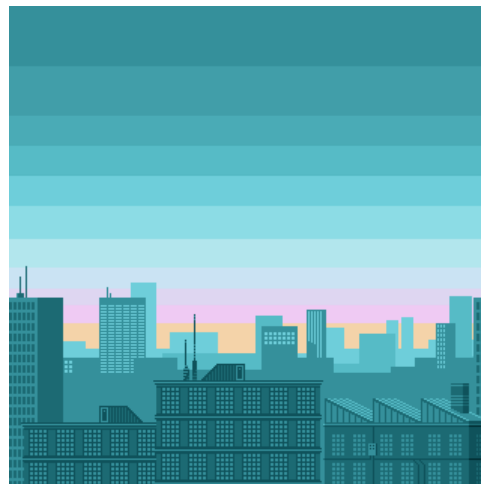
Video: Ferns

- Özuysal, Fua & Lepetit 2007



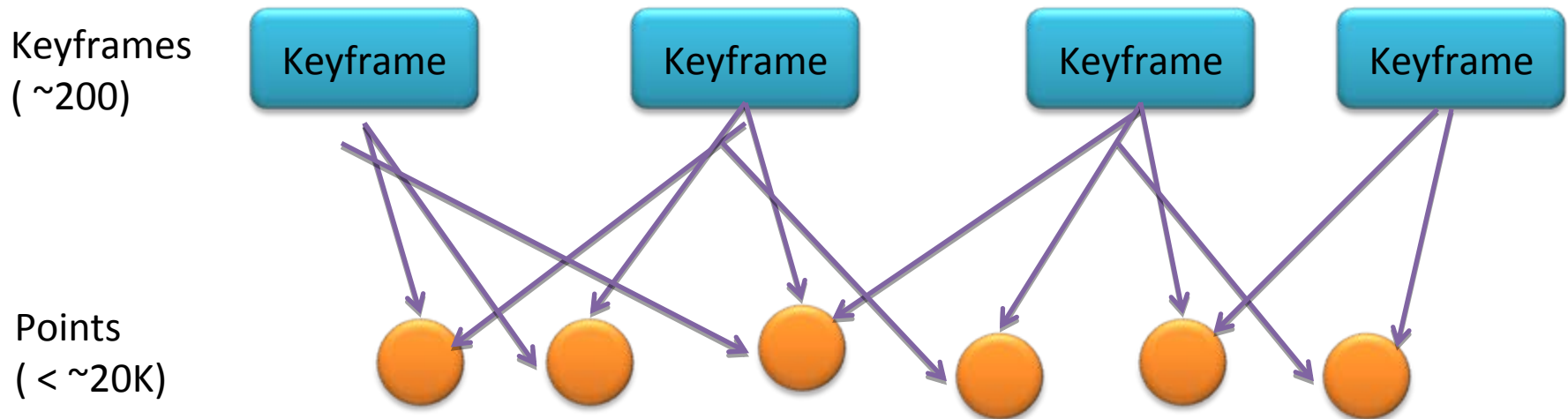
What about unknown environments?

- SLAM: Simultaneous Localization and Mapping
- Introduced to computer vision by Davison et al 2003
- Structure of world computed from image
- (x,y) observed directly – z computed from parallax



Parallel Tracking and Mapping (PTAM)

- Splits tracking and mapping into separate tasks. Map built in background from **keyframes**
- Draws from photogrammetry: **Bundle adjustment** and **epipolar geometry**



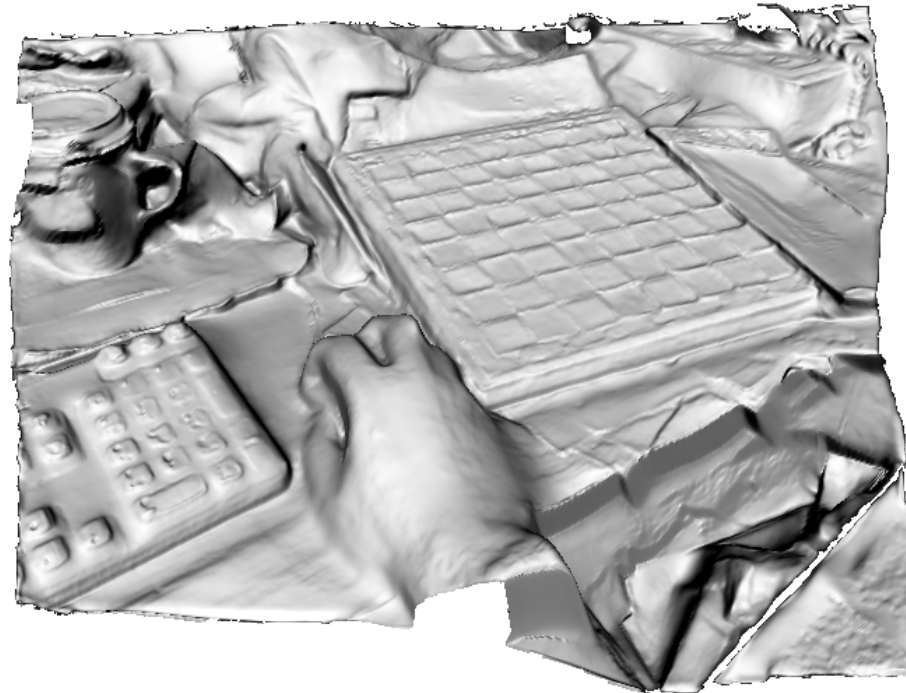
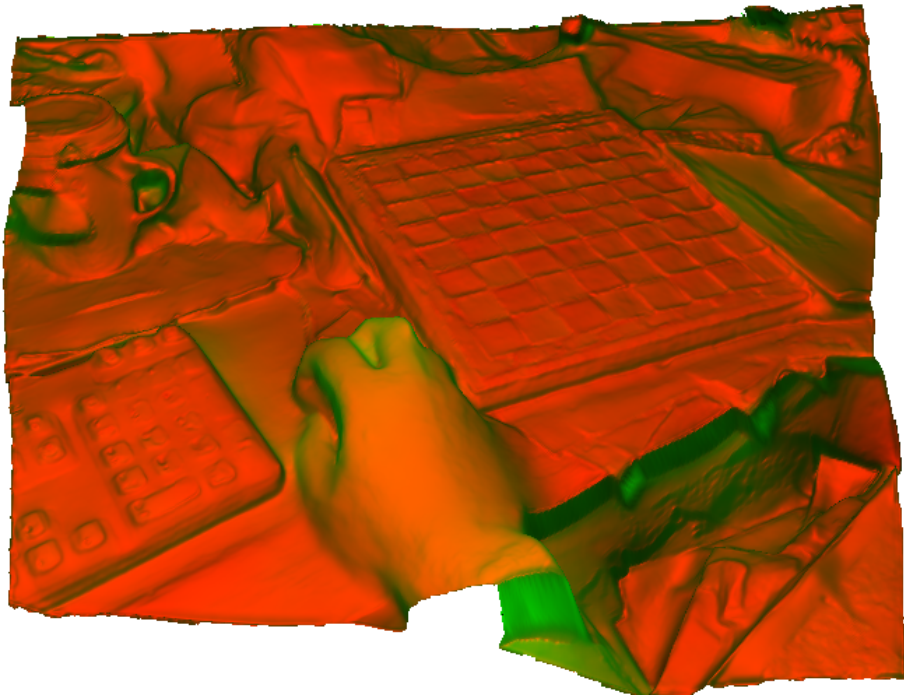
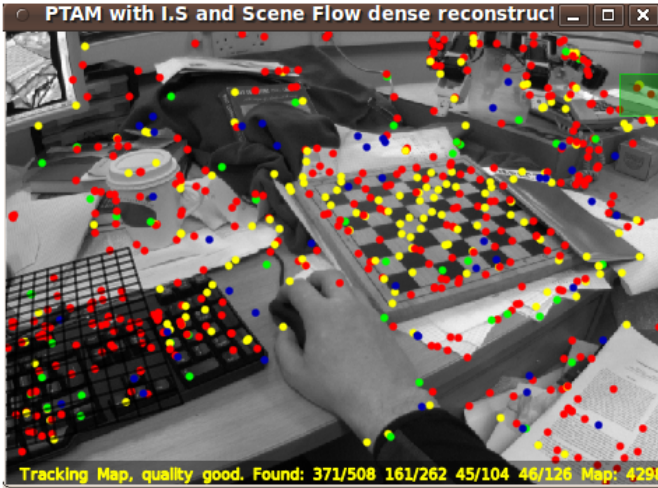
Demo: PTAM (Klein&Murray 2007)

New frontiers in visual tracking

- Live dense reconstruction
- Dynamic environments
- Interaction e.g. through hand tracking

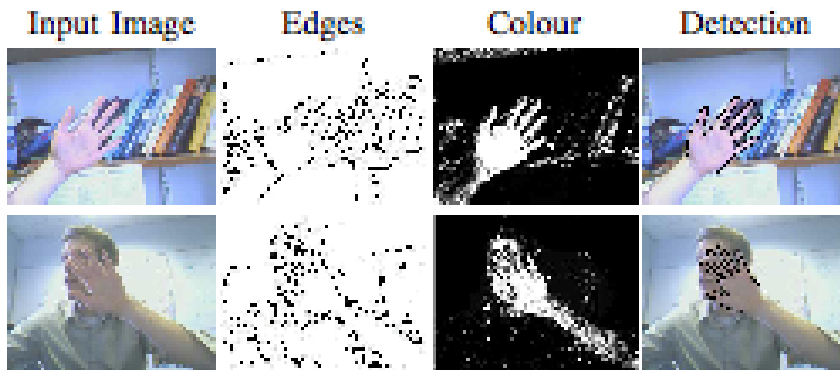
Dense reconstruction

- Allows occlusion and interaction between physical and real world
- Newcombe & Davison 2010



Hand Tracking / Interaction

- Real-time unassisted monocular hand tracking is still unsolved



Stenger et al 2006



Wang and Popovic 2009

Visual Tracking Methods for Augmented Reality

Georg Klein

Scientist

Microsoft Corporation