Problem: How do we determine
- Our position/orientation
- The identity/position/orientation of other objects

Solutions:
- Mechanical, ultrasonic, electromagnetic, GNSS, earth-relative, inertial, optical/vision, …

http://www.youtube.com/watch?v=6AKjH4On65A
**Optical and Vision-Based Tracking**

- Can use existing environmental light or special-purpose sources
- Light sensors are becoming increasingly better, smaller, cheaper, commonplace (e.g., smartphone cameras)
- Can benefit from computer vision research
  - Computer vision used for navigation, robotics, medical, security, manufacturing, etc.
- Cameras are already used to capture the real world in many (i.e., video see-through) AR applications
  - No extra hardware needed
  - “Why augment what you can’t see?”
    - Note: There are reasons! E.g., Important objects may be outside your FOV or hidden behind others!
- When using a video see-through HWD or handheld device, pixel-level accuracy is achievable today using optical tracking

---

**A Simple Vision-based Object Identification Pipeline**

1. Original
   - 256-levels
2. Threshold
   - Bi-modal: objects vs. background
3. Binary
   - High-pass filter
4. Blob detection
   - Connected pixels
5. Boundaries
   - Region growing
6. Identification
   - Statistical differences

Vision-Based Approaches for Unprepared Environments

- **Unprepared environment = not modified for tracking**
- Look within image for:
  - **Region-based tracking**
    - Connected regions (e.g., characterized by color or intensity)
  - **Contour-based tracking**
    - Boundaries
  - **Model-based tracking**
    - Instances of known 3D models
  - **Feature-based tracking**
    - Points, lines, …
    - SIFT (Scale-Invariant Feature Transform) features are invariant to image scale and rotation:
      "Key points" determined to be distinctive, with associated descriptors that make them robust to changes in illumination and viewpoint. [D. Lowe, Proc. ICCV 1999]
    - SURF (Speeded Up Robust Features) [Bay et al., ECCV 2006]
- In general, these methods work well in 2D, but while determining object identity and pose in 3D is hard, they can work well for feature sets on known geometry

Optical Approaches for Prepared Environments

- **Prepared environment = modified for tracking**
- Multiple cameras look within image for point features:

  [Diagram of epipolar geometry] (en.wikipedia.org/wiki/Epipolar_geometry)
Optical Approaches for Prepared Environments

- Prepared environment = modified for tracking
- Multiple cameras look within image for point features:
  - Active: LEDs
    - Can use IR LEDs to allow use of IR pass filter to view invisible (to humans) targets
    - Can illuminate LEDs in rapid sequence (e.g., 4600Hz) to allow each to be identified by synchronized cameras
    - Small LEDs form well-defined targets (e.g., 0.1mm accuracy)
  - Passive: Retroreflective spheres or patches
    - Use IR source around camera lens (e.g., ring of LEDs)
    - Retroreflective target minimizes response to spurious illumination, maximizes return from source
    - Identity established by unique 3D configuration of targets

Vision-Based Approaches for Prepared Environments

- Marker tracking (fiducial tracking)
  - Camera looks within image for markers (predefined patterns with known characteristics)
    - Typically planar
    - Encode identity
    - Allow recovery of camera pose relative to marker
**Example Marker Tracking Systems**

- **Data only**
- **Tracking (ID and pose)**

**Marker Tracking Approach**

- **Find all objects that might be markers**
  - ARToolKit: Threshold image to bilevel, find connected blobs, select blobs with quadrilateral boundary
  - ARTag: Find edge pixels, threshold and link to form segments, group segments into quadrilateral boundaries
- **Verification/Identification**
  - Verify that the object is a marker and identify it
  - Establish homography between square and quadrilateral boundary
    - ARToolKit: Sample marker interior with 16×16 grid and correlate with each known marker; discard if below confidence cutoff
    - ARTag: Sample marker interior with 6×6 grid to determine 36-bit redundant encoding of 10-bit marker ID; discard if ID invalid (Similar approach now also used in NyARToolKit)
**ARTag Approach**

- (a) Original image
- (b) Find edge pixels and link to form segments
- (c) Group segments to form quadrilaterals
- (d) Retain only quadrilaterals whose interiors contain a valid code

**Marker Tracking Variations**

- Invisible markers
  - IR-reflective markers [Y. Nakazato et al., ISWC 2005]
- Hidden markers
  - Generate texture over markers to hide them [S. Siltanen, ISMAR 2006]
- Buttons
  - Cover marker to act as a button press (need to be careful with false negatives)
- Tangible AR
  - Markers and objects to which markers are attached become a first-class part of the UI
Natural Feature Tracking
G. Klein and D. Murray, ISMAR 2007

- PTAM: Parallel tracking and mapping
- Previously unknown environment
- Fast, robust alternative to sequential, frame-by-frame SLAM (Simultaneous Localization and Mapping)
  - SLAM: A technique used by robots and autonomous vehicles to build up a map within an unknown environment while at the same time keeping track of the current location
  - Classic chicken or egg problem
- PTAM
  - Tries to determine single dominant plane in small, (mostly) static environment
  - Uses dense map of low-quality features, instead of a sparse map of high-quality features.
  - Has separate mapping and tracking threads that run asynchronously
  - Tracking can focus on robustness and real-time performance
- See also later work on single-camera SLAM
  - http://www.youtube.com/watch?feature=player_embedded&v=Df9WhgibCQA
  - http://www.doc.ic.ac.uk/~rfs09/slampp.html
  - G. Klein, PTAM. http://www.robots.ox.ac.uk/~gk/PTAM/

Tracking Nonrigid Objects
J. Pilet, V. Lepetit and P. Fua (EPFL)

- Real-time technique to detect deformable surfaces
  - 10 fps
- Problem Statement
  - Looking for the transformation that maps the known, undeformed model surface (i.e., mesh) into the deformed target one such that (squared) error is minimized
- Also models lighting and detects occlusions
  - Model image
  - Virtual character
  - Augmented image
  - http://cvlab.epfl.ch/research/detect/deformable/