Towards Visual SLAM in Dynamic Environments

MIT 16.412J Spring 2004

Vikash K. Mansinghka

Key Points

- Visual metric SLAM in environments with motion is an open problem
- SIFT SLAM error handling makes odd choices:
  - Cheats on position estimation (by trusting vision) - see report
  - Sets arbitrary matching thresholds for vision
- These choices impact SLAM performance:
  - Imposes a limit on motion error (odometry vs. actual)
  - Slow-moving objects can arbitrarily foul localization
  - Matching thresholds may not be best (though “good enough”)
- Partial solution: Minimize uncertainty and use optical flow

SIFT Landmark Tracking

- Predict where landmarks should appear (reliability, speed)
- Note: Robot moves in xz plane
- Given $[p, q, \delta]$ and old relative position $[X, Y, Z]$, find expected position $[X', Y', Z']$ by:

  $X' = (X - p)cos(\delta) - (Z - q)sin(\delta)$

  $Y' = Y$

  $Z' = (X - p)sin(\delta) - (Z - q)cos(\delta)$

- By pinhole camera model ((u_0, v_0) image center coords, I interocular distance, f focal length):

  $r' = v_0 - f\frac{Y'}{Z'}$

  $c' = u_0 + f\frac{X'}{Z'}$

  $d' = f\frac{I}{Z'}$

  $\sigma' = \sigma\frac{Z}{Z'}$

SIFT Landmark Tracking

- V is camera field of view angle (60 degrees)
- A landmark is expected to be in view if:

  $Z' > 0$

  $tan^{-1}(\frac{|X'|}{Z'}) < \frac{V}{2}$

  $tan^{-1}(\frac{|Y'|}{Z'}) < \frac{V}{2}$

- An expected landmark matches an observed landmark if:
  - Obs. center within a 10x10 region around expected
  - Obs. scale within 20% of expected
  - Obs. orientation within 20 degrees of expected
  - Obs. disparity within 20% of expected

Impact of Loose Matching Tolerance

- Mario does not know landmarks on Wario come from same source
- Robot (Mario) confused as long as Wario moves slowly

Possible Solution

- Add landmarks only when sensor error very low (robot stationary)
- Do not accept landmarks from image regions with motion
- Approximate motion either via SIFT matching or other optical flow scheme
- Alternately: Only trust largest set of consistent errors (how?)

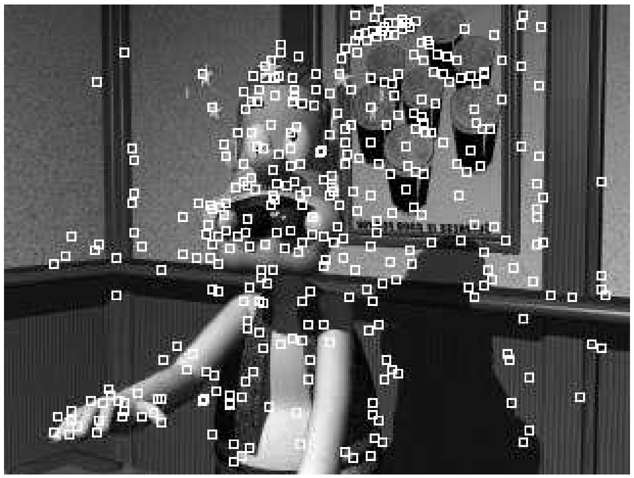
Horn-Schunck Optical Flow

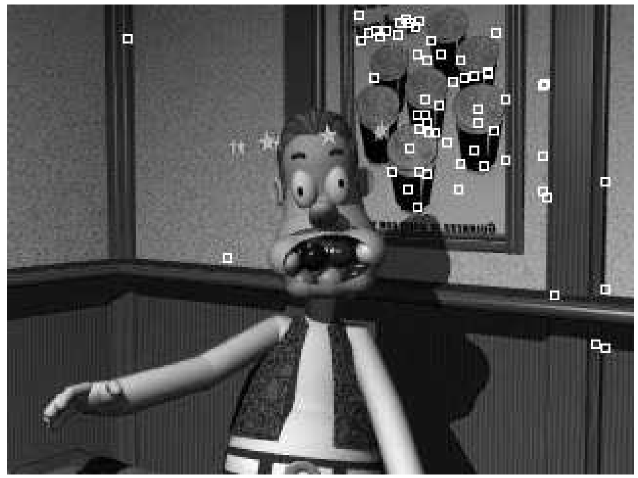
- Let $E(x, y, t)$ be image brightness at pixel (x, y) at time t
- Goal: find $u = \frac{dx}{dt}$ and $v = \frac{dy}{dt}$ (horiz, vert flow)
- Assume $\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0$
- Also minimize measures of the departure from smoothness and above assumption
- One set of choices yields coupled PDEs:
  $\nabla^2 u = \lambda(E_x u + E_y v + E_t)E_x$ and $\nabla^2 v = \lambda(E_x u + E_y v + E_t)E_y$

Proof of Concept for Optical Flow

- Uses Horn-Schunck optical flow; could use SIFT matching
- First demo prunes landmarks which move/are missed too much
- Second demo attempts place recognition by ignoring features on moving image regions

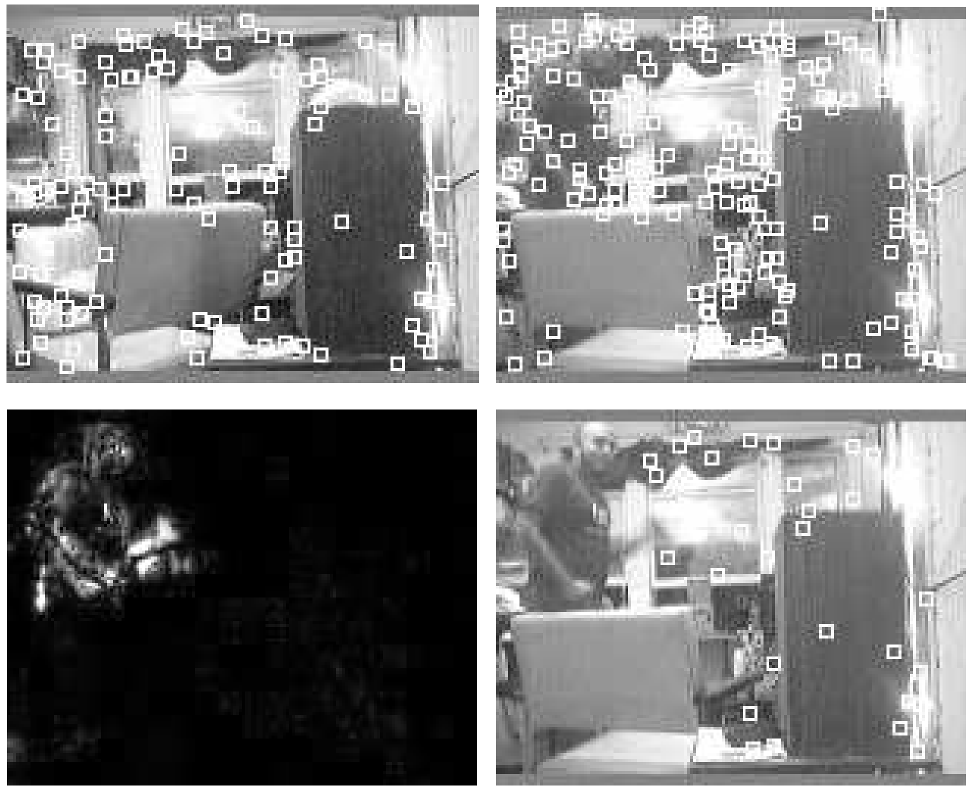
Simple Results





- A restrictive optical flow threshold (0.1) reliably prunes features on moving objects

Place Recognition Results



- Matching threshold of 20%
- Lowering the optical flow threshold from 30 to 1 raised performance from 43% to 56% in an image with motion
- Low performance due primarily to naive place recognition approach

Contributions

- Identified limitation in SIFT error handling; probably present in other SLAM systems
- Suggested simple optical flow technique for enabling dynamic visual SLAM
- Suggested other techniques for improving visual SLAM performance (see report)