Sensing Deforming and Moving Objects with Commercial Off the Shelf Hardware

Philip Fong
Florian Buron
Stanford University

This work supported by:

[Logos of National Science Foundation, National Institutes of Health, and Stanford University]
Motivational Applications

Human tissue modeling for surgical simulation

Robotic manipulation of fabric or rope
Motivational Applications

Need to sense 3D geometry of moving/deforming objects

Deformable object modeling
Goal

• Generate rangemaps (2.5D) from single images
  – No temporal coherence assumption
  – No restriction on motion
• Use commercially available hardware
Existing 3D Sensing Methods

- **Time of flight**
  - Lidar, radar, sonar, shuttered light pulse

- **Triangulation**
  - Laser stripe scanner
  - Stereo
  - **Structured Light**
    - Can be implemented with commercial cameras and projectors
    - Sinusoids / Moiré gratings (Takeda and Kitoh; Tang and Hung, Sansoni *et al*).
    - Stripe patterns (Koninckx, Griesser, and Van Gool; Zhang, Curless, and Seitz; Caspi, Kiryati, and Shamir; Liu, Mu, and Fang)
Limitations of Existing Methods

- Requires scene be rigid and not moving
  - Laser scanner
- Requires non-repeating texture
  - Stereo vision
    - Applied to deforming cloth (Pritchard and Heidrich)
- Requires known scene topology and known anchor points
  - Sinusoids / Moiré gratings
- Requires multiple frames
  - Restricts movement
  - Spacetime Stereo (Davis, Ramamoothi, and Rusinkiewicz)
  - Stripes
    - In single frames, spatial resolution does not scale due to fixed number of encodings
System Overview

- Idea: Combine colored stripes with sinusoid pattern
- Use sinusoidal pattern to produce dense rangemap
- Colored stripe transitions give sparse absolute depths
  - Use as known anchor points
System Geometry

- Camera projection center at (0,0,0)
- Projector at \((p_x, p_y, p_z)\)
- Parallel optical axes
- Pinhole projection model for camera and projector
System Overview

Diagram showing the process:

1. Image
2. Demodulate
3. Segment
4. Phase Unwrap
5. Rangemap
6. Label colors
7. Find color transitions
8. Label transitions
9. Generate guesses
Depth from Sinusoid

- Projected sinusoid:
  \[ v^{\text{PI}}(x, y) = A \cos(\omega^{\text{PI}} y) + C \]

- Camera sees deformed sinusoid:
  \[ v^{\text{CI}}(x, y) = A \cos(\omega^{\text{CI}} y - \theta^{\text{CI}}(x, y)) \]
  \[ \theta^{\text{CI}}(x, y) = \omega \left( \frac{-p_z^{\text{CI}} y + p_Y}{w z^{\text{CI}}(x, y) - w p_z} \right) \]

- Demodulate to get wrapped phase (Tang and Hung):
  \[ \theta_w = \arctan 2(\sin(\theta), \cos(\theta)) \]
Segmenting

- Phase unwrapping assumes $\theta$ changes by less than $2\pi$ between pixels
- Segment image into regions based on phase variance (Ghiglia and Pritt) using snakes
Labeling Colors

- Label and score pixels with colors using Bayesian classifier
- PDFs of colors in hue space are approximated with a gaussian distribution
Labeling Color Transitions

- Threshold change in hue between pixels along X direction
- Label each detected transition according to the projected pattern
  - Based on color label scores in pixel windows to the left and right of transition
  - Ignore transitions:
    - Not consistent with projected pattern
    - Over a max width
Generating Guesses

• In projected pattern
  – Each transition appears only once
  – Identifies unique vertical plane
• Intersect with ray corresponding to transition location in camera to compute depth
• Use as guesses in phase unwrapping
Phase Unwrapping

- Compute phase gradient assuming no jumps greater than $2\pi$
- Integrate to get $\theta_u$
  \[ \theta(x, y) = \theta_u(x, y) + 2\pi k \]
- For each region compute $k$ from median of the difference between guesses and $\theta_u$
- Compute rangemap from $\theta$
Calibrating for Chromatic Aberration

- Chromatic aberration in camera and projector lenses distort pattern
- Model distortion with:
  - Focal length ratio of red and green in camera and projector ($\alpha$, $\beta$)
  - Red optical center shift ($\gamma$) in projector
- Minimize:
  $$\sum_{Q} (\theta_s(x\alpha, y\alpha) - \theta_g(x, y))^2$$
  $$\theta_s(x, y) = \beta(\theta_r(x, y) - \omega(\frac{1}{\alpha} - 1)y - \gamma)$$
Results: Plastic Duck

- 0.48mm (0.1%) RMS error compared to Cyberware 3030MS laser scanner
Results: Popcorn Tin
Results: Water Balloon

Bouncing Water Balloon
Captured at 100 fps @ ~52 cm Sequence
Advantages / Limitations

• Spatial resolution scales with camera and projector resolution
• Temporal resolution scales with camera speed
  – Not limited by projector speed

• Each segmented region must contain at least one recognized color transition
• Objects with many saturated colors are hard to sense
  – Mitigated by choosing right set of colors in pattern
• Single viewpoint only senses one side of objects
  – No straight forward way to project and sense from multiple viewpoints simultaneously
Conclusions / Results

- Combined sinusoidal and colored stripe pattern is effective
  - Produces good quality dense range maps of moving and deforming objects
- Chromatic aberration calibration allows use of commercial cameras and projectors
Questions?

These and more at:
http://www.stanford.edu/~fongpwf/research.html