Visual Information Processing

Menandro Roxas

Computer Vision (Oishi) Laboratory
Today’s Topic

- Spatio-Temporal Image Processing / Video Processing
  - 3D reconstruction
  - Video Modification/Restoration
3D Reconstruction

• Model Based Methods
  • Pose Estimation
  • PTAM
  • DTAM
  • LSD-SLAM

• Learning Based Methods
  • CNN-SLAM
  • DeMoN
Camera Pose Estimation

• Two cases

2D-2D

3D-2D
2D-2D Pose Estimation

• Find correspondences \( x, x' \)
• Solve for the fundamental matrix \( F(3 \times 3) \)
• Extract the essential matrix \( E \)
• Decompose \( E \) to find relative camera rotation \( R \) and translation \( t \)

\[ x' = K' T F K \]

\( E = [t] \times R \)

• Scale is relative \( |t| = 1 \)
3D-2D Pose Estimation

- (PnP or Perspective n-point problem)
- Camera projection model

\[
\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}
\]

\[
\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}
\]

- Calibrated (K = known) and Uncalibrated (K = unknown)
- Minimal case: solution for P3P or n=3 (n correspondences)
PTAM
(Parallel Tracking and Mapping)
PTAM

Frame-by-frame SLAM
- Find features
- Draw graphics
- Update camera pose and entire map (Many DOF)

Parallel Tracking and Mapping
- Find features
- Draw graphics
- Update camera pose (6-DOF)

A comparison between PTAM and frame-by-frame algorithms (2).
PTAM

On the train to Kyoto
Parallel Tracking and Mapping for Small AR Workspaces

ISMAR 2007 video results

Georg Klein and David Murray
Active Vision Laboratory
University of Oxford
Key Framing
DTAM:
Dense Tracking and Mapping in Real-Time
DTAM
(Dense Tracking and Mapping)

Live Dense Reconstruction with a Single Moving Camera
LSD-SLAM
(Large-Scale Direct Monocular SLAM)
Feature-Based Alignment vs. Direct Alignment

Uses point correspondence and 3D point differences

Failure: Non-textured and repeating patterns

Uses intensity difference

Failure: Large homography
LSD-SLAM

We present LSD-SLAM: a Large-Scale Direct monocular SLAM system, which allows to build large-scale semi-dense maps in real-time on a CPU.
Scale Ambiguity
CNN-SLAM
(Dense SLAM with learned depth prediction)

Figure 2. CNN-SLAM Overview.
CNN-SLAM
DeMoN
(Learning Monocular Stereo)
DeMoN

Figure 2. Overview of the architecture. DeMoN takes an image pair as input and predicts the depth map of the first image and the relative pose of the second camera. The network consists of a chain of encoder-decoder networks that iterate over optical flow, depth, and egomotion estimation; see Fig. 3 for details. The refinement network increases the resolution of the final depth map.

Figure 3. Schematic representation of the encoder-decoder pair used in the bootstrapping and iterative network. Inputs with gray font are only available for the iterative network. The first encoder-decoder predicts optical flow and its confidence from an image pair and previous estimates. The second encoder-decoder predicts the depth map and surface normals. A fully connected network appended to the encoder estimates the camera motion $r, t$ and a depth scale factor $s$. The scale factor $s$ relates the scale of the depth values to the camera motion.
DeMoN

DeMoN: Depth and Motion Network for Learning Monocular Stereo

Benjamin Ummenhofer*,1 Huizhong Zhou*,1 Jonas Uhrig1,2
Nikolaus Mayer1 Eddy Ilg1 Alexey Dosovitskiy1 Thomas Brox1

1University of Freiburg 2Daimler AG R&D

*equal contribution
Video Modification/Restoration

• Video Completion
  • Wexler, Irani
  • Shiratori
  • Roxas

• Video Stabilization
  • Matsushita
Video Completion, Wexler
Video Completion, Wexler
Video Completion, Wexler
Video Completion, Wexler
Video Completion, Wexler
Video Completion, Wexler
Video Completion, Shiratori
Video Completion, Shiratori
Video Completion, Shiratori
Remaining Issue

• Temporal aliasing or “ghost” exists

• Why do ‘ghosts’ exist?
  • Discontinuity in motion of objects
  • Inconsistent between the hole and its boundary
Video Completion, Roxas

- Input
- Image pyramid
- Trajectory prior estimation
- Motion estimation and inpainting
- Iterative step
- Color propagation
- Update mask function
- Output
Three-frame Optical Flow

\[
\min_{u_f, u_b} E_{\text{data}} + E_{\text{spatial}} + E_{\text{trajectory}}
\]

\[
E_{\text{data}} = \mu(x)f(I, u_f, u_b)
\]

\[
E_{\text{spatial}} = \lambda_s TV(u_f, u_b)
\]

\[
E_{\text{trajectory}} = \lambda_t \varphi(u_f, u_b)
\]

\(f = \text{brightness constancy}\)

\(\mu = \text{mask}\)

\(TV = \text{total variation}\)

\(\varphi = \text{trajectory}\)

backward flow, \(u_b\)

forward flow, \(u_f\)
Color Propagation

\[ \mu(x_1) = 1 \]
\[ \mu(x_2) = 2 \]

distance of the source to frame 0
Color Propagation

- $\mu(x)$:

source frame distance from **one optical flow direction**
Color Propagation

input frame

using one direction

INCONSISTENCY!
Color Propagation

- Issue: two optical flow direction (forward and backward)
- How to choose between the two? -> Blend both directions using $\mu(x)$
Color Propagation

- Using one direction:
  - Input frame
  - Using one direction

- Using both directions without blending:
  - Using both directions without blending

- Blending both directions:
  - Blending both directions

Inconsistency!
Results

Input Sequence

Ground Truth Optical Flow

Inpainted Video

Solved and Inpainted Optical Flow
Results
Video Completion, Roxas 2014
Video Stabilization, Matsushita

Stabilized sequence

Trimming  Mosaicing  Our method
Video Stabilization, Matsushita
Video Stabilization, Matsushita
Video Stabilization, Matsushita
Removing Raindrops from Videos

**Fig. 1.** An example of the results of our proposed detection and removal method. (a) Scene with raindrop. (b) Dense long trajectories. (c) Matching of trajectories occluded by raindrop. (d) Trajectory based video completion.

**Fig. 2.** The pipeline of our method.
Removing Raindrops from Videos
Papers

• **PTAM**
  - Georg Klein and David Murray, "Improving the Agility of Keyframe-based SLAM", Proc. ECCV 2008

• **DTAM**

• **LSD-SLAM**

• **CNN-SLAM**

• **DeMoN**
Papers

• Video Completion, Wexler
  • Y. Wexler, et al. “Space-time video completion”, CVPR 2004

• Video Completion, Shiratori

• Video Completion, Roxas

• Video Stabilization, Matsushita

• Removing Raindrops, You