Probabilistic 3D Occupancy Flow with Latent Silhouette Cues

Li Guan
UNC-Chapel Hill, U.S.A.
liguan@cs.unc.edu

Jean-Sébastien Franco
University of Bordeaux
LaBRI - INRIA Sud-Ouest, France
jean-sebastien.franco@labri.fr

Edmond Boyer
Grenoble Universities
INRIA Grenoble, France
edmond.boyer@inrialpes.fr

Marc Pollefeys
ETH Zürich, Switzerland
UNC-Chapel Hill, U.S.A.
marc.pollefeys@inf.ethz.ch

Objective:
- to estimate both the shape and motion information from multiple views in a probabilistic framework.

Motivation:
- knowing the motion between time steps can help refine the shapes
- 3D probabilistic shape reconstruction methods exist, but not for motion
- 3D motion field methods exist, but only for surface representations

Difficulty: motion is not well-defined for non-deterministic shapes

Objective & Motivations

Example of 3D Scene Flow methods, figure from [Pons et al., 2007]

Previous work: probabilistic occupancy grid from silhouettes cues [Franco and Boyer, 2005]

Variables
- $T$ = view images
- $p$ = motion field
- $G$ = grid at time $t - 1$
- $G'$ = known grid at time $t$
- $G_X$ = binary state at location $X \in \{0,1\}$

Intuition: each voxel occupancy at $t$ should be explained by a voxel occupancy moving from $t-1$ as well as by the image cues at $t$.

Modeling & Inference

Joint probability: dependence and distributions

$$p(G_{t-1}, G_t, D_t) = \prod_{X \in \Omega} \left( p(G_X \g T_{t-1}) p(G_X \g T_t) \prod_{X' \in \Omega} p(D_{X'} \g D_{X}) \right)$$

E step: Estimate $p(G_{t-1} \g T_t)$, i.e. the distribution of the grid occupancy knowing the previously estimated displacement

$$E_{G_{t-1}} [\ln p(T_t, G_{t-1}, G_t)] = \sum_{G_{t-1}} p(G_{t-1} \g T_t, G_t) \ln p(T_t, G_{t-1}, G_t)$$

M step: $d^{th+1} = \text{argmax}_d Q(D_t | d^{th})$, where

$$Q(D_t | d^{th}) = E_{G_{t-1}} [\ln p(T_t, G_{t-1}, G_t)] = \sum_{G_{t-1}} p(G_{t-1} \g T_t, G_t) \ln p(T_t, G_{t-1}, G_t)$$

Discretizing the space of possibilities, we reduce M to a discrete minimization, solved with Fast-PD [Fornodaki et al. 2007] on multi-scale.

Eq. (1) allows to calculate all the above distributions (Bayes’ Rule).

Contributions
- formalization and modeling of the problem
- EM algorithm solution to make the interdependence manageable
- first results on indoor and outdoor datasets
- no assumption on the shape and motion, besides simple continuity.

Background models
- Observed videos
- Inference
- How to manage the spatiotemporal complexity and interdependence between shape and motion estimation?

Principle

Method

Result

Test on synthetic ellipses
- Estimated flow vs. the ground truth
- Error plots with different velocity, shape, noise

All above datasets have 6 ~ 9 views. The volume size is $128^3$. Three levels of control grid with control points 11, 7 and 3 voxels apart. The EM converges in less than three iterations for all datasets. <1 minute per frame on an 8-core CPU.

Applications
- 3D Motion Segmentation
- Occupancy refinement (outdoor shadow removal)

This work was partially supported by David & Lucille Packard Foundation Fellowship, the European Research Council grant 4DVIDEO and ANR-06-MOCA-003-01 DALIA project.

Synthetic Validation

CVPR 2010, San Francisco, U.S.A.