Computer Vision for Autonomous Systems - Part 2: A VMA Perspective

Jiangbo Lu

https://publish.illinois.edu/**visual-modeling-and-analytics/**






ADSC
Illinois at Singapore Pte Ltd

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Our VMA **Vision**

Holistic (Computer/Robotic) Vision

Geometric Reconstruction meets
Semantic Recognition for 3D holistic vision:
Real-time, robust, geometry-centric vision

Where: geometry, location…

What: semantics, action…

Reconstruction

- Real-time camera pose localization
- 3D environment mapping
- Depth and motion estimation
- Large-scale urban reconstruction

Recognition

- Visual place/scene recognition
- Object recognition, localization
- Human re-identification
- Action recognition, tracking

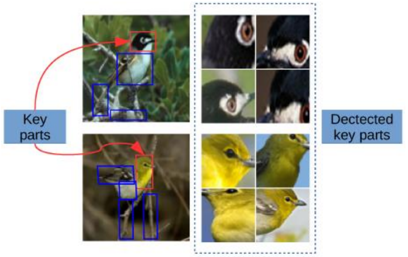
How: model, solve, compute

- Multiple sensors
- Modern vehicles
- Moving robots
- Opportunistic scan

Key techniques

- Geometry-aware filtering
- Fast randomized algorithms
- Efficient inference models
- Deep learning innovation

- Mobile cameras
- Big visual data
- Rich annotations
- Powerful machine



Scene structure

- Depth enhance.
- Slanted stereo
- Optical flow
- Live FG segment
- 3D city recons.

Comp. imaging

- Photo refocus
- Rain removal
- Image stitching
- Multi-scale dec.
- Structure extra.
- Image warping
- Colorization

3D vision & percept.

CODE [ECCV'14, '16]
Feature matching

DSE [CVPR'15]
Camera pose rec.

SLAM-O [TCSVT'16]
3D object prop.

Recogn. & segment.

Fine-grained recog. [TIP'16]

PISA [CVPR'13b]
Pixel-acc. saliency

MHIC [TMM'16]
Recog & Co-seg.

Edge-aware filtering

CLMF [CVPR'12]
Local EAF

FGS [TIP'14]
Global EAF

FGI [ECCV'16]
Sparse to dense

Efficient inference

PMF [CVPR'13]
Local optim.

DFF [CVPR'14]
Generalized

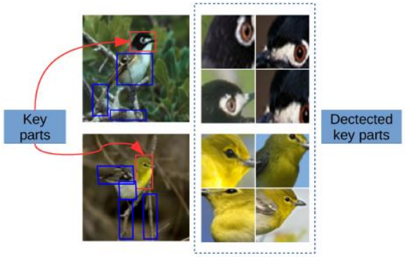
SPM-BP [ICCV'15]
Global optim.

Current topics

- *Scene flow*
- *Robust SLAM*
- *Motion seg.*

- *Text detection*
- *Place recogn.*
- *Scene labeling*
- *Action recog.*

- *Deep learning*



3D vision & percept.

- CODE [ECCV'14, '16] **Feature matching**
- DSE [CVPR'15] **Camera pose rec.**
- SLAM-O [TCSVT'16] **3D object prop.**

Recogn. & segment.

- **Fine-grained recog.** [TIP'16]
- PISA [CVPR'13b] **Pixel-acc. saliency**
- MHIC [TMM'16] **Recog & Co-seg.**

Scene structure

- Depth enhance.
- Slanted stereo
- Optical flow
- Live FG segment
- **3D city recons.**

Comp. imaging

- Photo refocus
- Rain removal
- **Image stitching**
- Multi-scale dec.
- Structure extra.
- Image warping
- Colorization

Edge-aware filtering

- CLMF [CVPR'12] **Local EAF**
- FGS [TIP'14] **Global EAF**
- FGI [ECCV'16] **Sparse to dense**

Efficient inference

- PMF [CVPR'13] **Local optim.**
- DFF [CVPR'14] **Generalized**
- SPM-BP [ICCV'15] **Global optim.**

Current topics

- *Scene flow*
- *Robust SLAM*
- *Motion seg.*

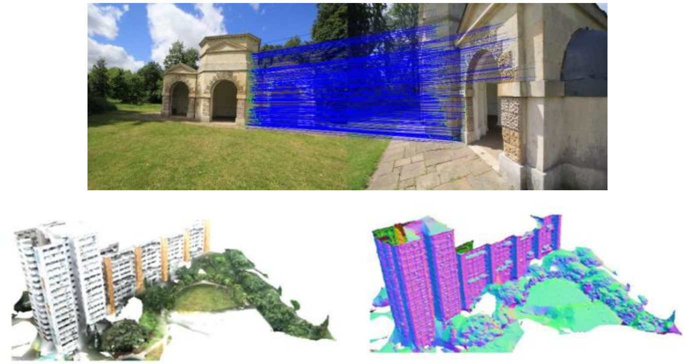
- *Text detection*
- *Place recogn.*
- *Scene labeling*
- ***Action recog.***

- *Deep learning*

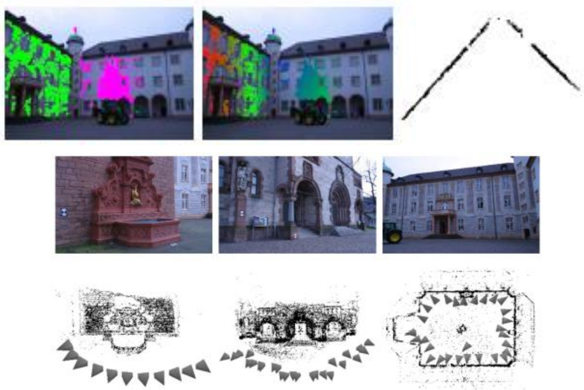
VMA Techs for Autonomous Systems

Dense depth & motion

Feature matching and mapping

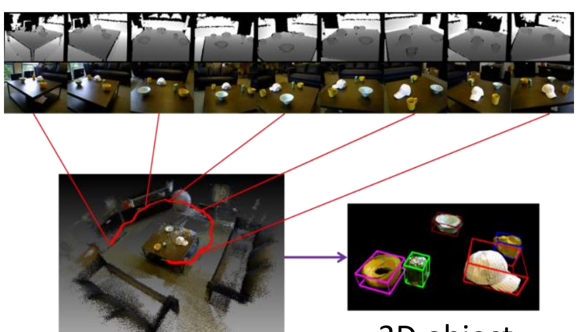


Camera pose localization



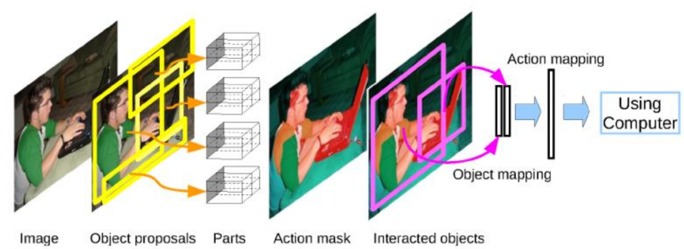
Stitching & visualization

Localizing 3D object proposals



3D object proposals

Action recog. w/ min. labelling

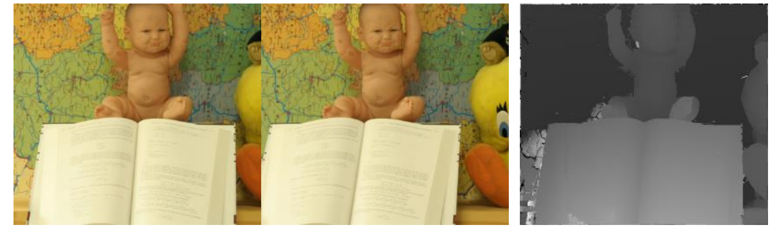


Related VMA **Tech**

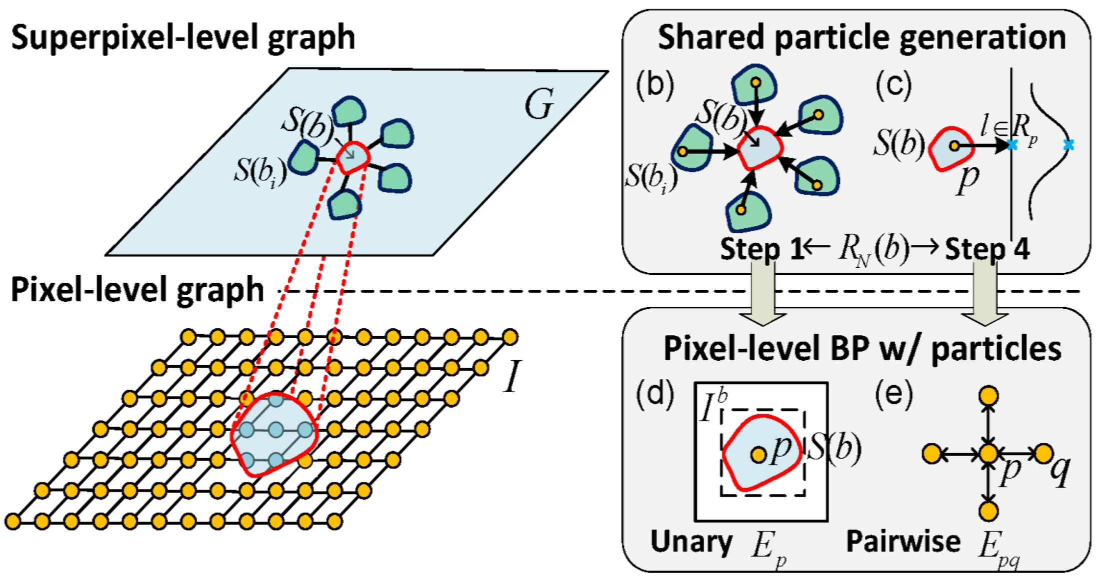
#1 Perceiving Depth & Motion for AS: Matching

A general, efficient discrete optimizer for **stereo** & **flow** & **etc**

$$E = \sum_{p} E_p(l_p; W) + \sum_{p} \sum_{q \in \mathcal{N}_p} E_{pq}(l_p, l_q)$$

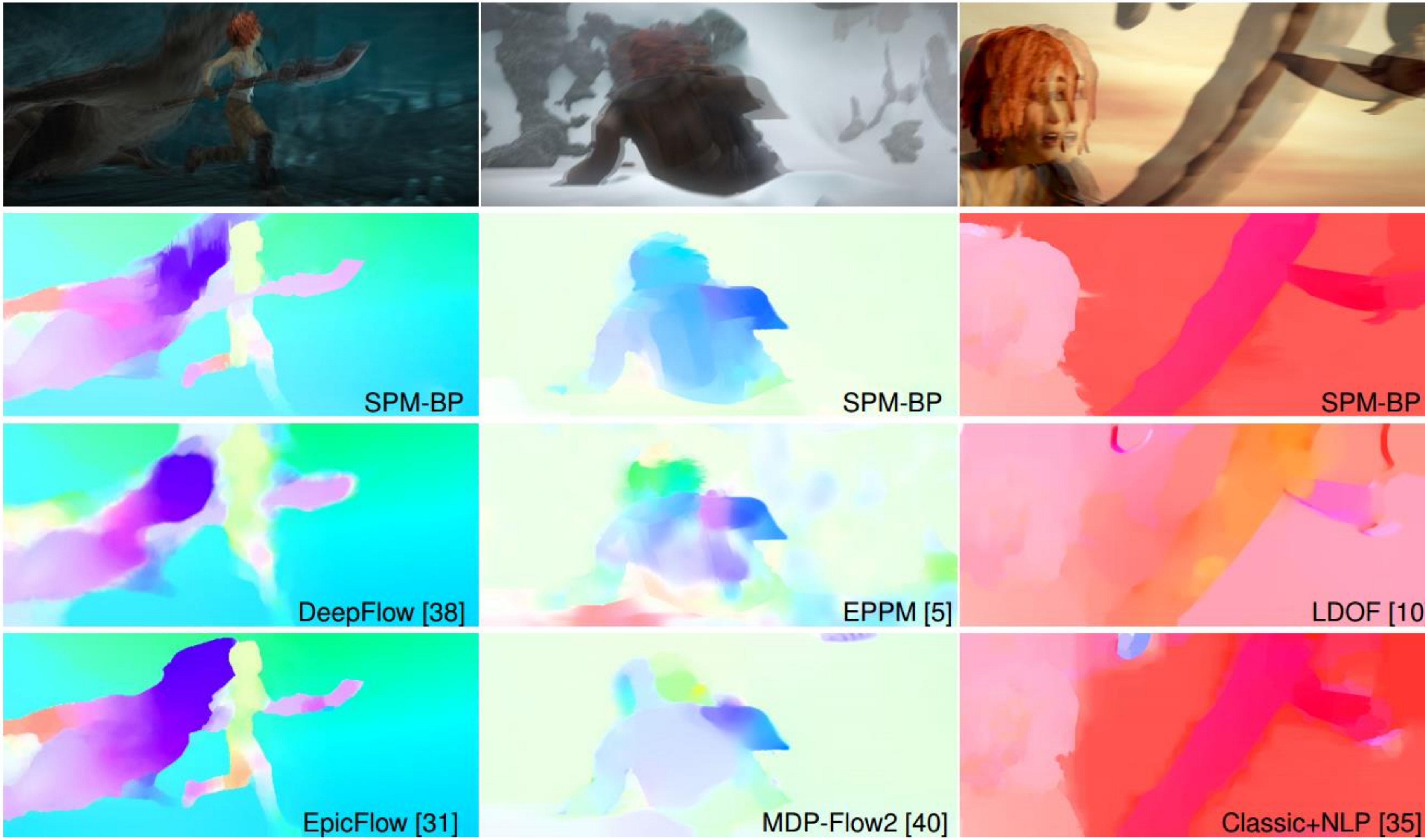


- ✓ *General MRF for labeling problems*
- ✓ *Superior for huge label spaces*
- ✓ *50-100x faster than [PMBP]*
- ✓ *Edge-aware filtering + PM + BP*



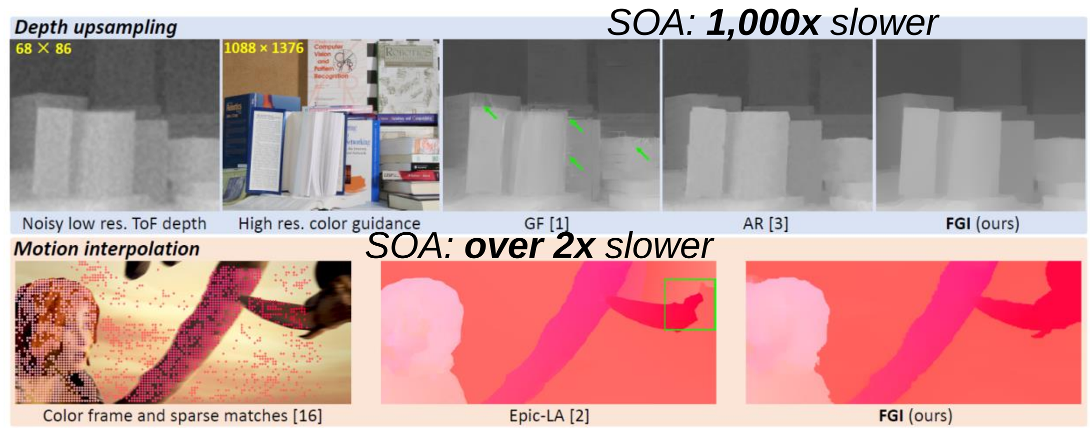
- A simple formulation w/o
  - *complex* energy terms
  - a separate *initialization*
- Achieved top-tier perform., even when compared to *task-specific* techniques
- Applied on the full pixel grid, *w/o coarse-to-fine*

* Y. Li et al., "SPM-BP: Sped-up PatchMatch Belief Propagation for Continuous MRFs," ICCV 2015 (Oral)

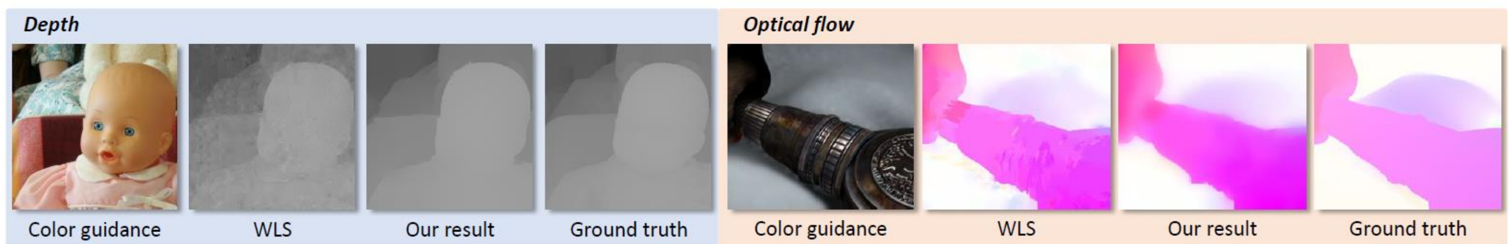

SPM-BP
SPM-BP
SPM-BP
DeepFlow [38]
EPPM [5]
LDOF [10]
EpicFlow [31]
MDP-Flow2 [40]
Classic+NLP [35]

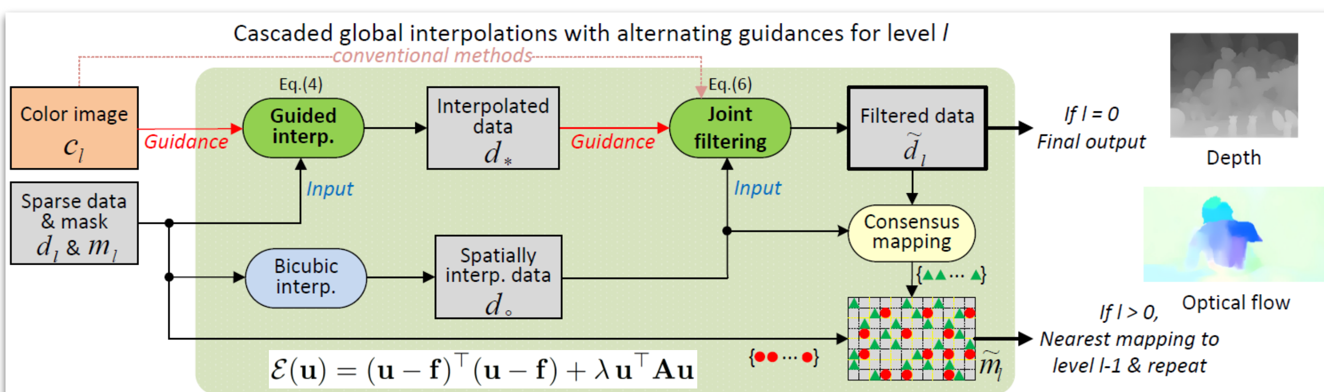
#1 Perceiving Depth & Motion for AS: Interpolation

A unifying framework for fast **guided** (global) **interpolation**



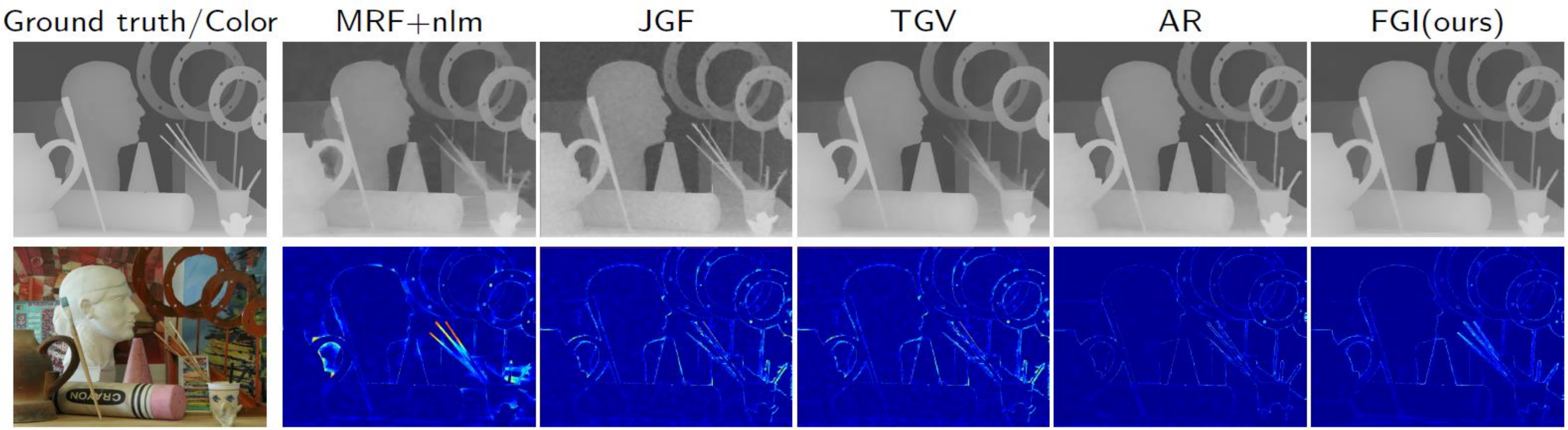
Texture-copying artifacts due to inconsistent structures across modalities

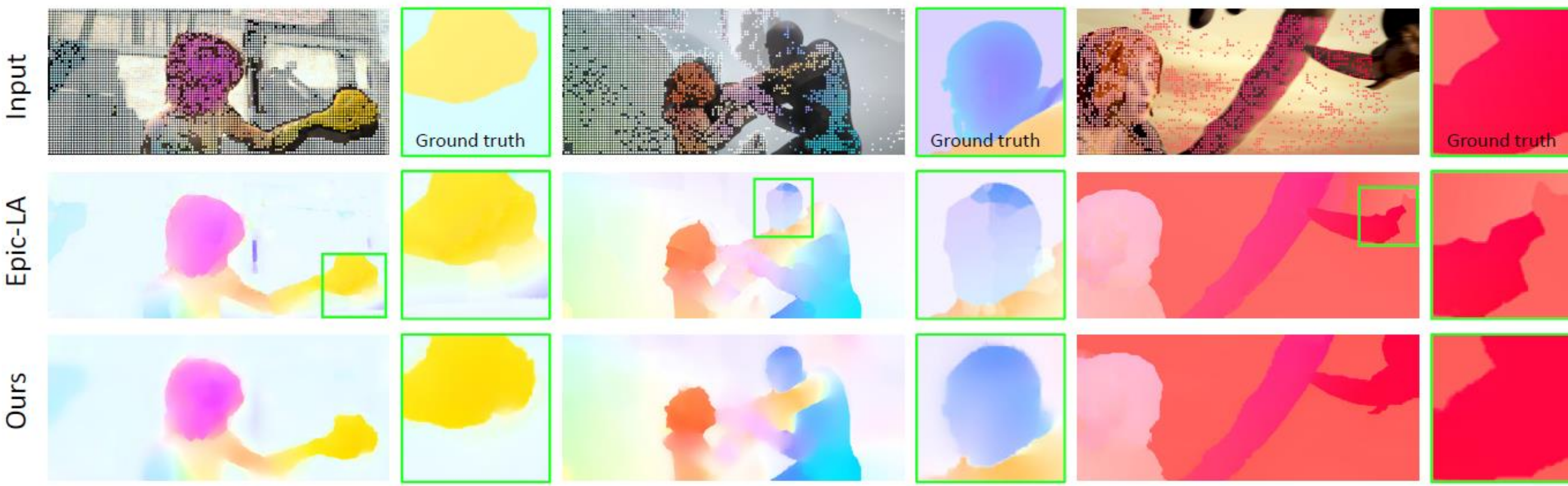




- General, versatile for various tasks
- Simple & effective
- Fast computation
- [FGS] in OpenCV

* Y. Li et al., "Fast Guided Global Interpolation for Depth and Motion," ECCV 2016 (spotlight)
* D. Min et al., "Fast Global Image Smoothing Based on Weighted Least Squares," TIP 2014

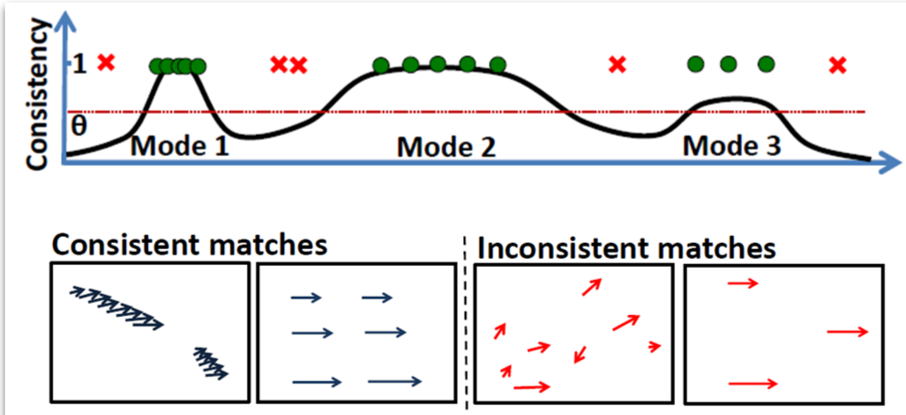




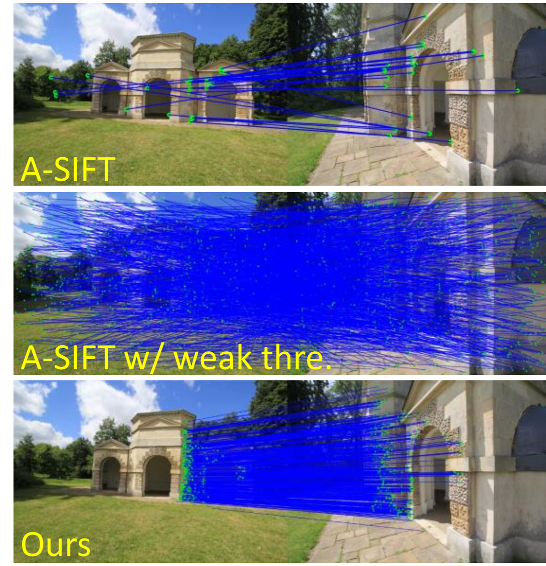
(a) Weak edge in color guidance (b) Sparsely scattered points (c) Extrapolation

#2 Wide-Baseline Matching & 3D Mapping for AS

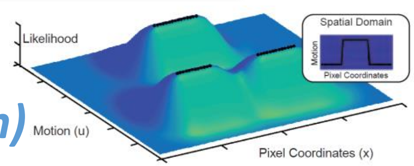
A reliable **feature matcher** for **pose** and **3D reconstruction**



$$\arg\min_{\mathbf{w}} \sum_{i=1}^{N} C(1 - f(\mathbf{m}_i)) + \lambda \mathbf{w}^T G \mathbf{w}$$



- ✓ ***Motion coherence modeling***
- ✓ ***Bilateral domain (spatial + motion)***



- Providing a number of matches, while having almost *no* outliers
- w/ RANSAC to handle *repetitive structures*
- Highly reliable 2-view pose for SfM, mapping

(a) Visual SfM (b) Visual SfM with our matches (c) Dense reconstruction

* W.-Y. Lin et al., "RepMatch: Robust Feature Matching and Pose for Reconstructing Modern Cities," ECCV 2016

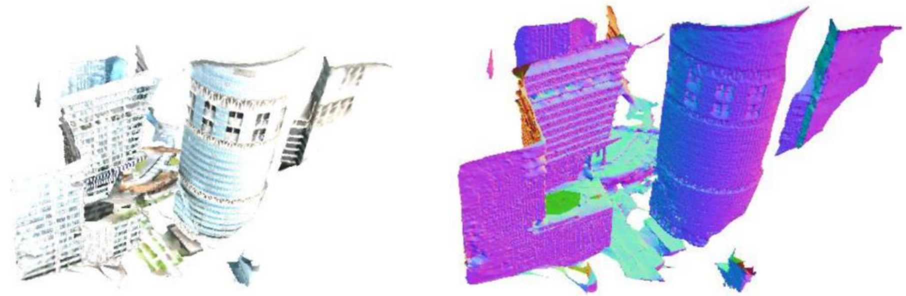
* W.-Y. Lin et al., "Bilateral functions for global motion modeling," ECCV 2014

A set of multi-view images [42]

Agisoft [47]: A commercial 3D reconstruction software

Visual SfM [3], [43], [44], [45], [46]

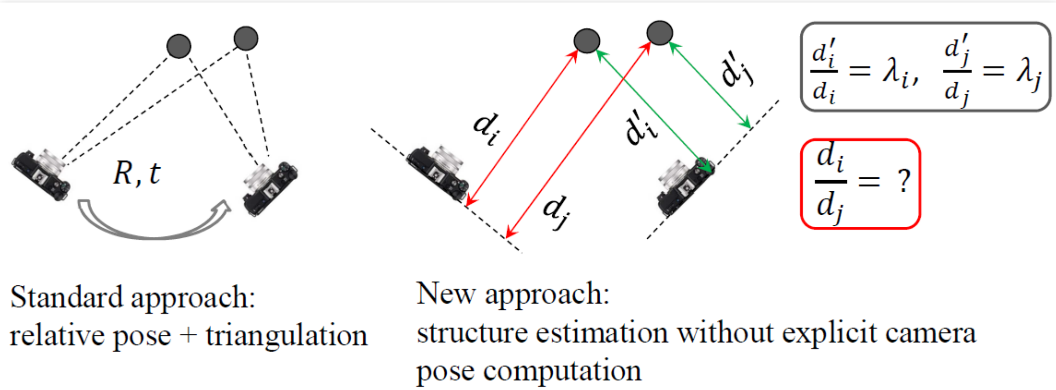
Visual SfM using feature matches returned by *A-SIFT w CODE*

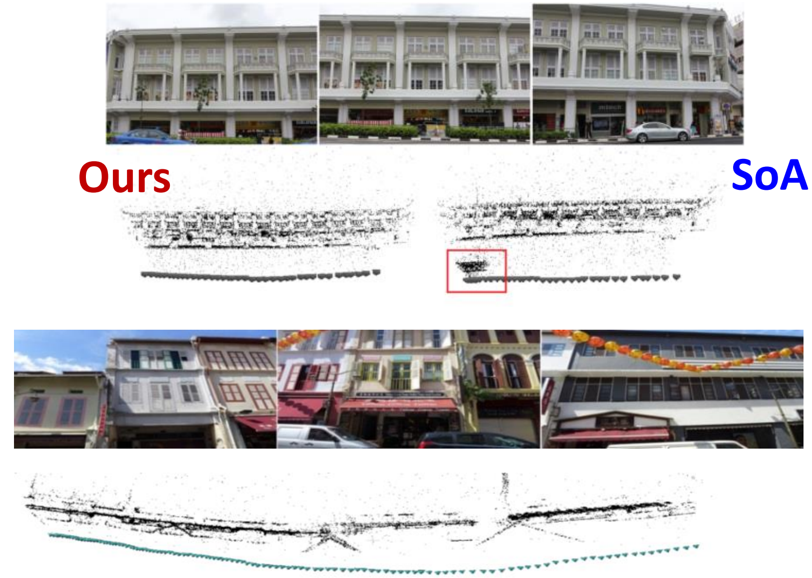


Our ADSC lab reconstructed w/ only color images

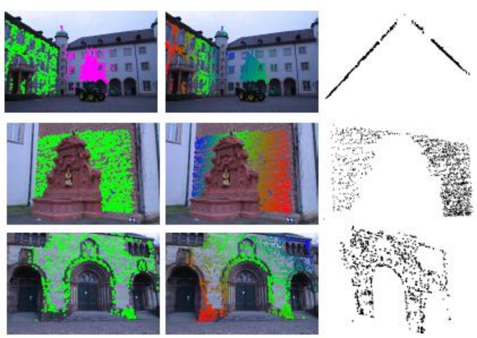
#3 Structure-First Camera Pose Estimation for AS

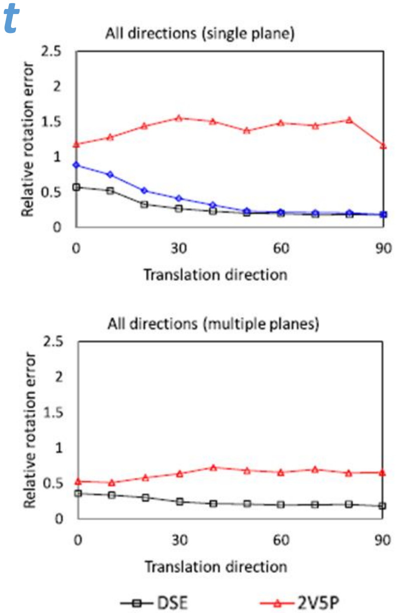
A unified **pose estimation** approach to **man-made scenes**

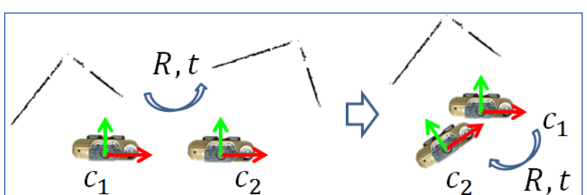




- ✓ ***Using homography estimation***
- ✓ ***Euclidean rigidity constraint***





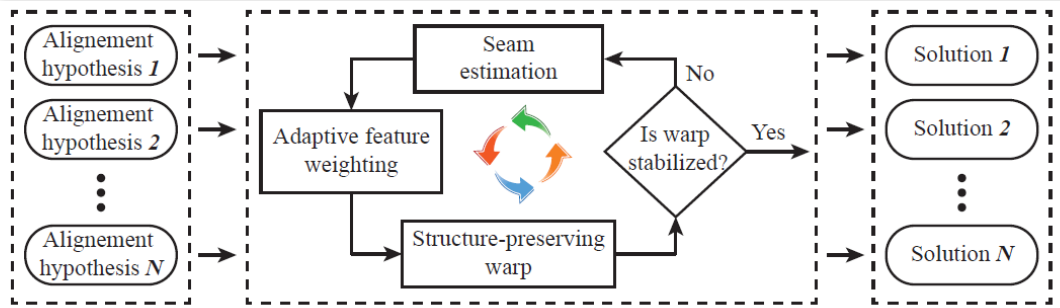


- Reversing the std. pipeline
- *No* model selection regardless of #planes
- Significant outperformance in stability & accuracy
- Esp. for *sideway* motions

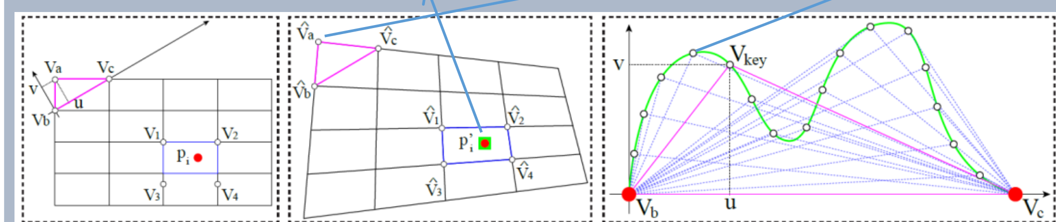
* N. Jiang et al., "Direct structure estimation for 3D reconstruction," CVPR 2015

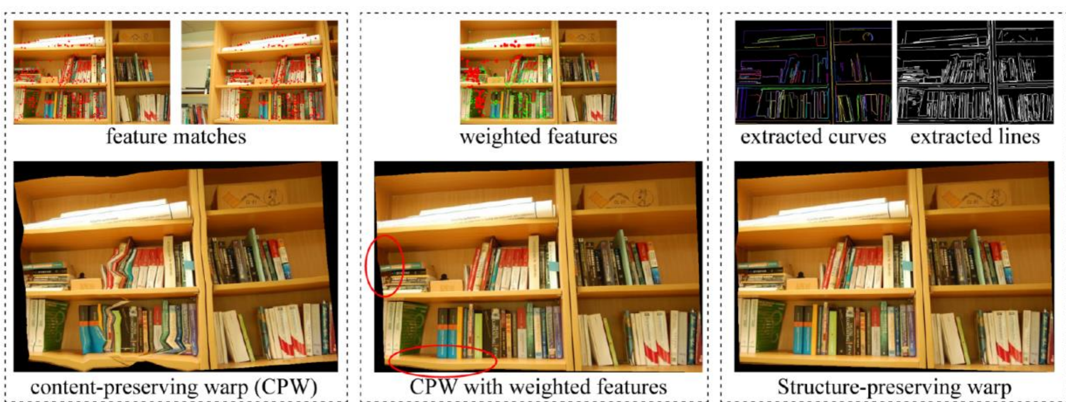
#4 Large Parallax Stitching & Visualization for AS

A high-quality image **stitching** method to handle **large parallax**



$$E(\hat{V}) = \lambda_1 E_f(\hat{V}) + \lambda_2 E_{ls}(\hat{V}) + \lambda_3 E_{cs}(\hat{V})$$





- Coupled local alignment and seam estimation
- #alignment hypotheses significantly reduced
- Curve structure constraint
- State-of-the-art quality for challenging scenes

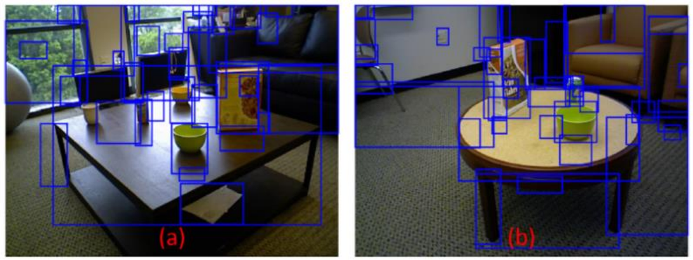
* K. Lin et al., “SEAGULL: Seam-guided local alignment for large parallax image stitching,” ECCV 2016


APAP
Ours

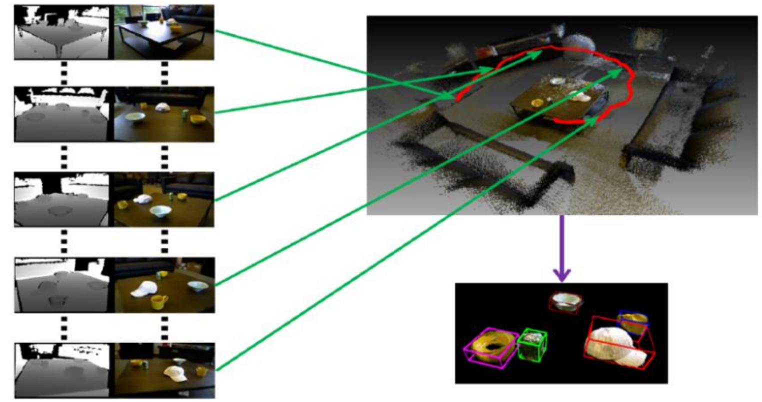

sentosa
KFC

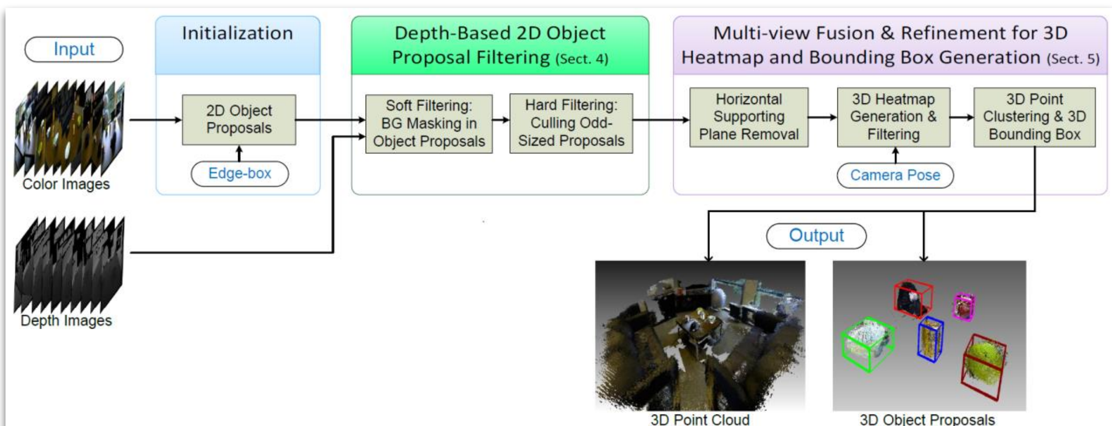
#5 Localizing 3D Object Proposals for AS

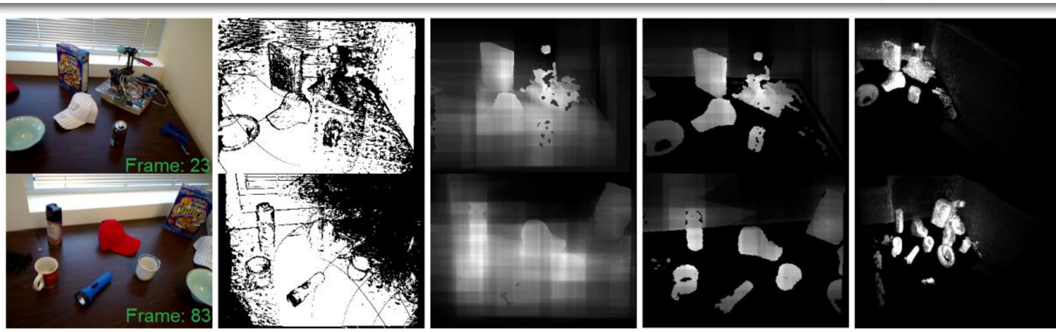
An online method for **3D object proposals** using **RGB-D videos**



- Allow any 2D object prop., e.g. Edge-box
- *No* good bbox out of 1,000 for yellow bowl







(a) Color image (b) Matched pixels (c) Weighted 2D heatmap H_{2D} (d) Heatmap after plane removal $\hat{H}_{2D}$ (e) Global heatmap H_{3D} projected onto image plane

- RGB-D video for 3D object proposals w/o detectors
- High precision, much more accurate than state-of-art
- Class-independent for *new* data collection on the fly
- Good for improved SLAM, navigation, object search

* R. Pahwa et al., “Locating 3D Object Proposals: A Depth-Based Online Approach,” TCSVT 2016 (minor revision)

UW-RGBD dataset

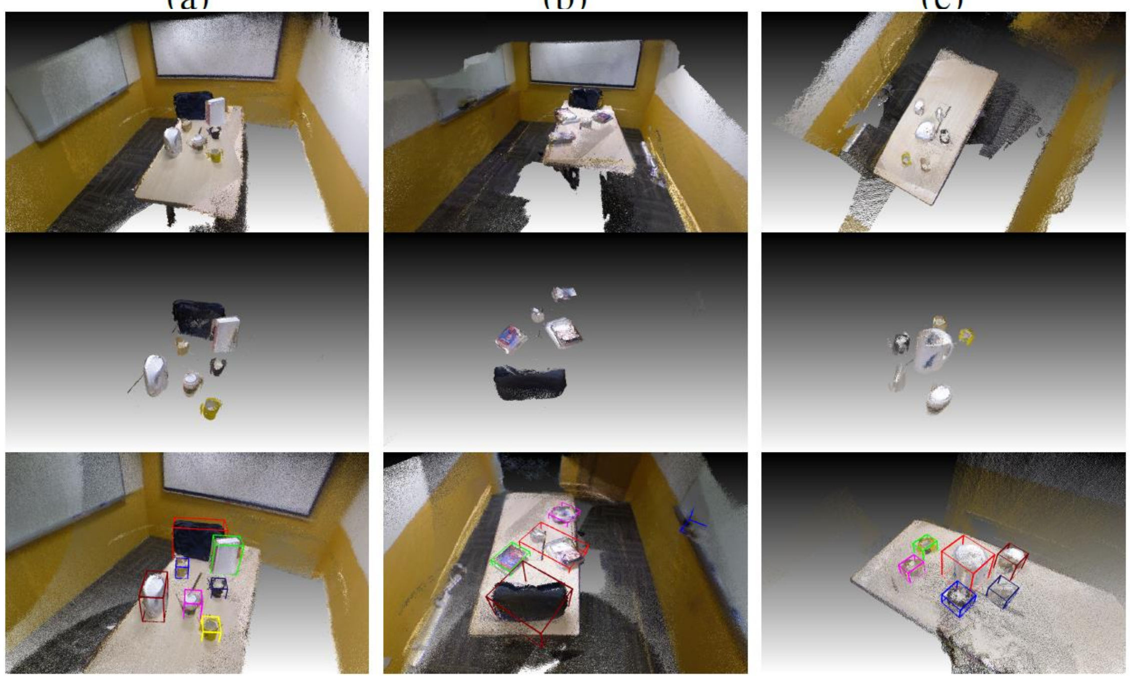
Point cloud

Top-ranked filtered points

Resulting 3D object proposals

(a) (b) (c)

Our dataset



Point cloud

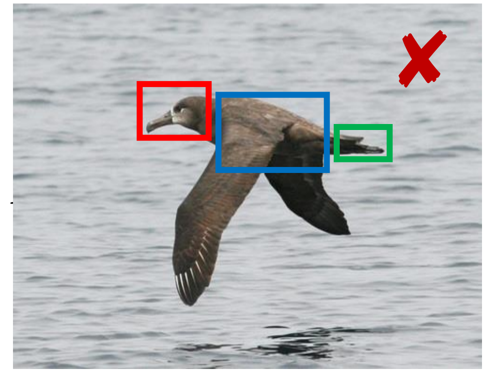
Top-ranked filtered points

Resulting 3D object proposals

(d) (e) (f)

#6 DNN Recognition w/ Minimum Annotation Efforts

Red-bellied Woodpecker vs. Red-headed Woodpecker [Zhang et al. TIP'16]



- Without human-annotated bounding-boxes for parts
- 78.92% on CUB 200-2011 datasets of 200 bird species

- Without requiring human bounding-boxes
- On a single-image
- 83.23% on PASCAL VOC 2012 of 10 action classes

[Zhang et al. TIP'16-2]

https://publish.illinois.edu/visual-modeling-and-analytics/research/

ADSC
Illinois at Singapore Pte Ltd

ILLINOIS

Visual Modeling and Analytics of Dynamic Environments for the Mass

HOME | NEWS | RESEARCH | PUBLICATIONS | DEMOS | PEOPLE | CONTACT US

Research

This project will focus on addressing the following fundamental challenges key to a multitude of visual data analytics applications: 1) raw visual data cleaning; 2) visual data registration and fusion; and 3) visual data analytics and management. The existing efforts, either from the academia or the industry, are not capable of robustly and efficiently modeling, analyzing, and fusing continuous or discrete visual data captured by individuals or big companies.

Grounded on the recent novel and exciting developments described in this project, we plan to extend, generalize, and optimize them to address the aforementioned key challenges of visual modeling and analytics for the masses using the following research directions:

1. Localization – recovering the geometric locations of the user, the camera viewpoint, or the objects in the environment around the user;
2. Registration – aligning and modeling dynamically captured images and measurements of the scene over different time and viewpoints together;
3. Inference – estimating and analyzing the semantic information of the scene from the registered visual information and recovered geometric information.

We aim at achieving both high robustness and accuracy for the above tasks at unprecedented processing speeds on commodity computing devices and mobile cameras, often producing more than one or two orders of magnitude of speedup over the existing state-of-the-art solutions.

We have been working on the following clusters of research topics, and now are actively innovating in a broader scope.

- Edge-aware filtering and joint filtering
- Dense stereo, optical flow and view synthesis
- Dense correspondences across scenes
- Motion coherence and wide-baseline matching
- Structure from motion, 3D reconstruction
- Computational photography, image enhancement
- Efficient inference for continuous MRFs
- Fast guided global interpolation
- Saliency, recognition, cosegmentation
- Hash techniques

Thanks