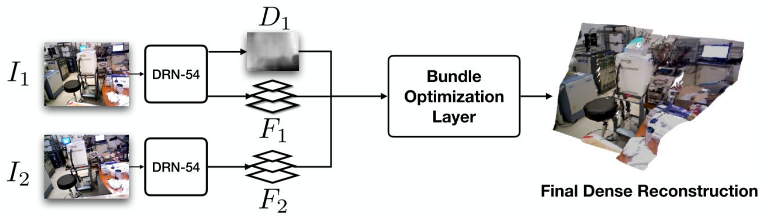
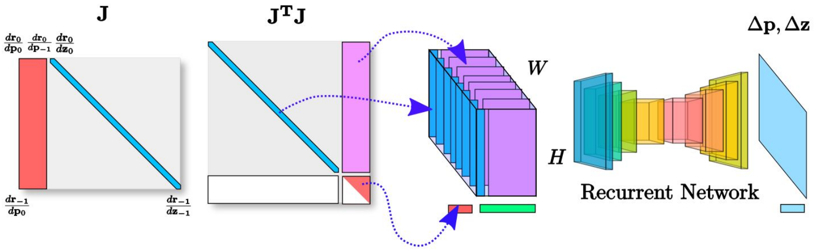
Geometry-Aware Learning Methods for Computer Vision

Ioan Andrei Bârsan
iab@cs.toronto.edu

Source: BA-Net by Tang and Tan





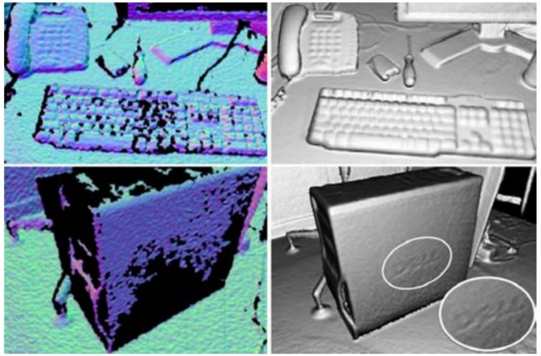
Source: LS-Net by Clark et al.

Supervisor: Prof. Raquel Urtasun

January 18, 2019

Overview and Motivation

- **Scope:** Recent advances in combining **geometry** and **learning** for computer vision tasks.
- **Motivation:** Deep learning is powerful but computer vision is still inherently geometric in nature.
- **Big question:** How can we get the best of both worlds?

TensorFlow mxnet +  = ?

Source: KinectFusion by Newcombe et al.

Agenda

- Area 1: Matching
- Area 2: Optimization and Meta-Optimization
- Area 3: Unsupervised Geometric Learning
- Existing & Future Work
- Conclusions

[1] Matching (for Stereo)

- Fundamental CV problem:
  - Find local correspondences across multiple sensors.
- Basis for numerous tasks:
  - Stereo
  - Flow
  - Structure from Motion
- Classic:
  - Engineered descriptors
  - SSD/MI-based matching
  - Cannot be adapted to specific problem domain
    - Matching is rarely the end-goal!

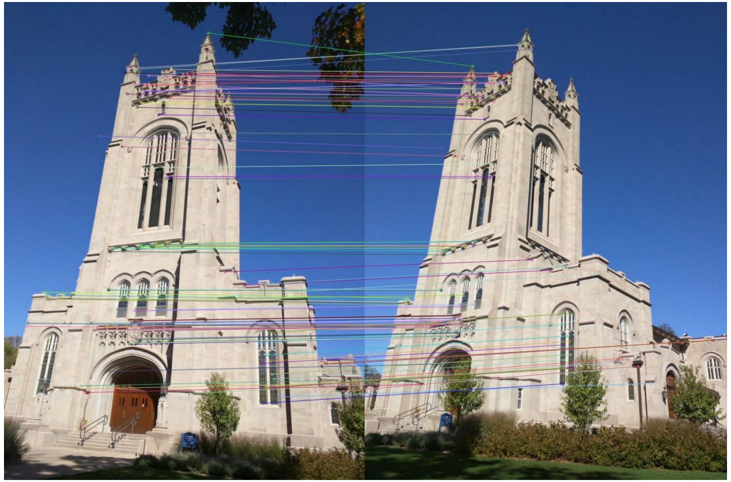


Image credit: http://cs.carleton.edu

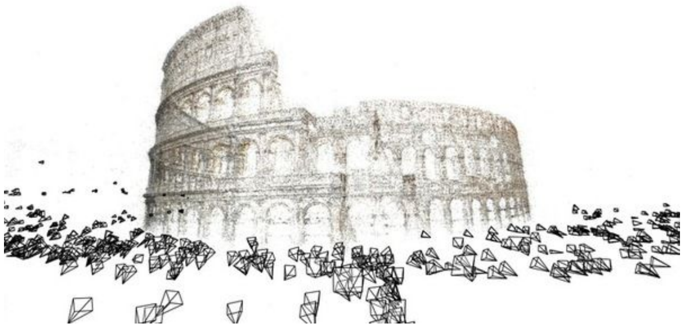
[1] Matching: Geometry + Learning

- **Problem:** Limitations of classic cost functions for stereo matching.
  - Limited context and no semantics
  - Unable to adapt to specific domain: flow / stereo / SfM / etc.
- **Solution:** (one of many) Replace engineered matcher with learned one.
  - Žbontar and LeCun pioneered this (learned cost + SGM)
  - Luo et al. made this more efficient (+ output uncertainty!)

[2] Optimization in Vision

- Many non-learning tasks vision require geometric optimization:
  - Visual odometry
  - Structure from motion
  - Image alignment and stitching

- Example: Bundle adjustment typically solved with **Gauss-Newton** and **Levenberg-Marquardt**. Main challenges?



Example task: Bundle Adjustment
Image credit: Building Rome in a Day by Agarwal et al.

[2] Optimization in Vision Continued

- Example: Bundle adjustment typically solved with **Gauss-Newton** and **Levenberg-Marquardt**. Main challenges?

$$\boxed{\sum_i \sum_j} \boxed{\|x_{ij} - \pi\left(C_i X_j\right)\|^2}$$

What if there are millions of parameters?

↓

📈 **Scalability**

What if a match is incorrect?

↓

💪 **Robustness**

[2A] Improving Scalability

- **Problem:** Dense depth is difficult to optimize over in **bundle adjustment**.
- **Solution:** Reduce effective number of unknowns.

Option 1: CodeSLAM
- Train autoencoder to predict dense depth conditioned on intensity image.
- Optimize over low-dimensional (128D) **latent code** in BA. instead of dense depth.

Option 2: BA-Net
- Train network to predict monocular depth as **linear combination** of “depth bases”.
- Optimize over weights (128) in BA instead of dense depth.

Joint BA over dense depth and camera poses becomes easier to solve!

[2B] Improving Robustness

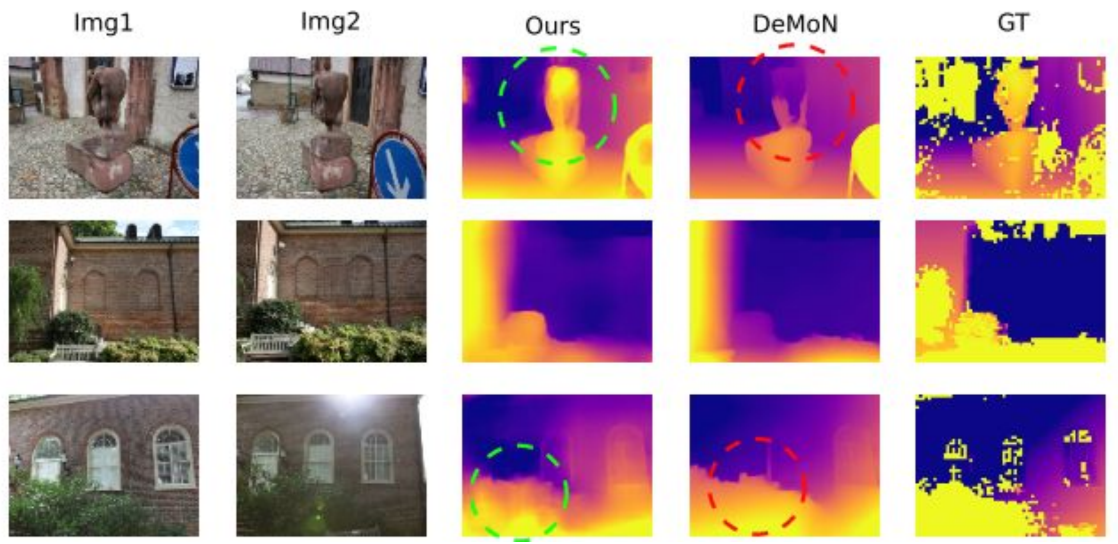
- **Problem:** Optimizing photometric-like error terms requires good initialization.
    - Limited robustness to dynamic object and complex lighting changes.
- **Solution:** Introduce learning in the optimizer itself as an implicit regularizer.

$$\Delta \mathbf{x}_i = -(\mathbf{J}_i^T \mathbf{J}_i + \lambda \, \mathrm{diag}(\mathbf{J}_i^T \mathbf{J}_i))^{-1} \mathbf{J}_i^T \mathbf{r}_i$$

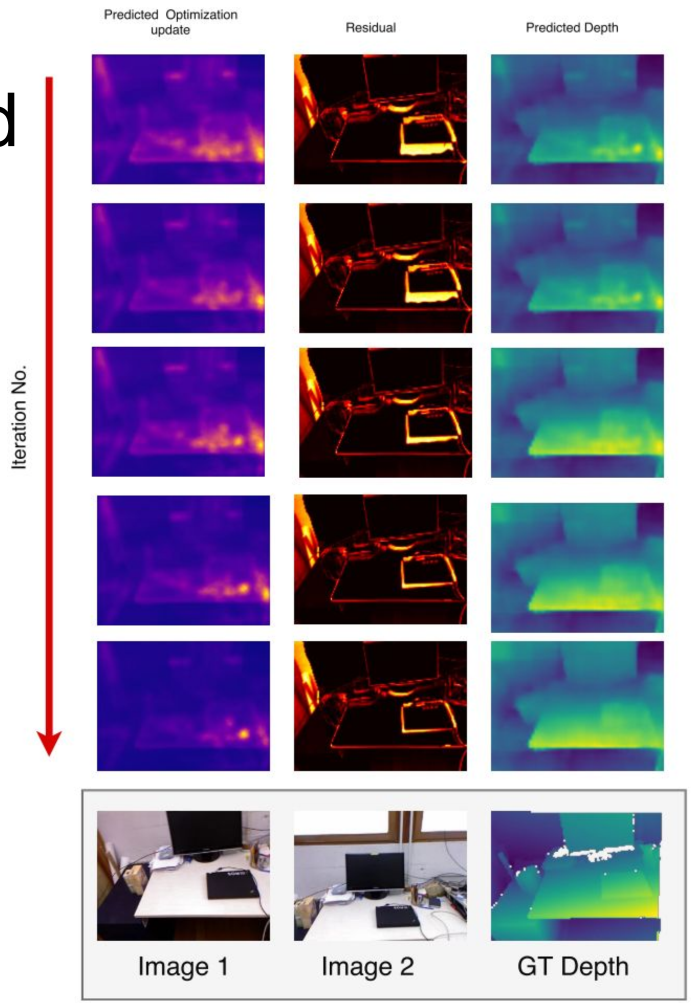
LS-Net (Clark et al.)

$$\Delta \mathbf{x}_i, \mathbf{h}_{i+1} \leftarrow f_\theta \left(\mathit{\Phi}(\mathbf{J}_i, \mathbf{r}_i), \mathbf{h}_i \right)$$

[2B] Improving Robustness Cont'd



- Results from NYU dataset.



[3] Unsupervised Geometric Learning

- **Problem:** Learn low-level vision on real data without supervision.
- **Challenges:** How do we get the network to learn what we want?
- **Solution:**
  - Given monocular frames (t) and (t - 1), try to estimate each from the other.
  - Use **photometric** and **geometric** loss terms.

- **Problem 2.0:** Photometric / 2D Geometric loss terms are biased towards areas closer to the camera!

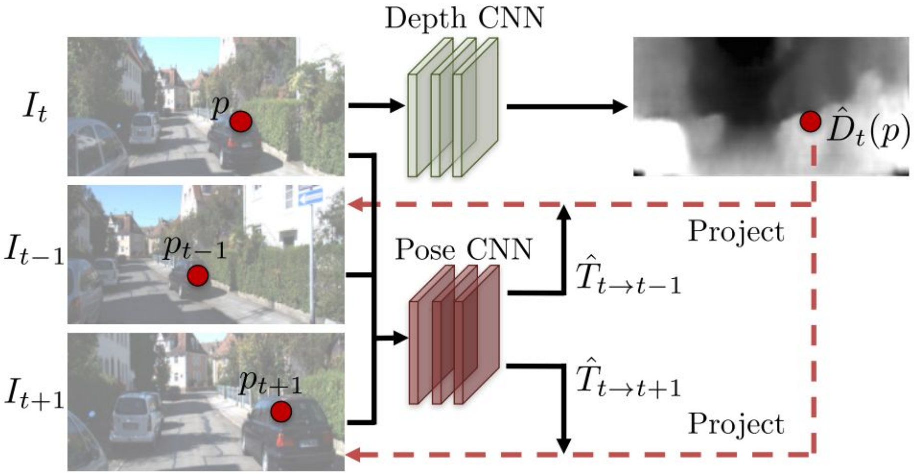
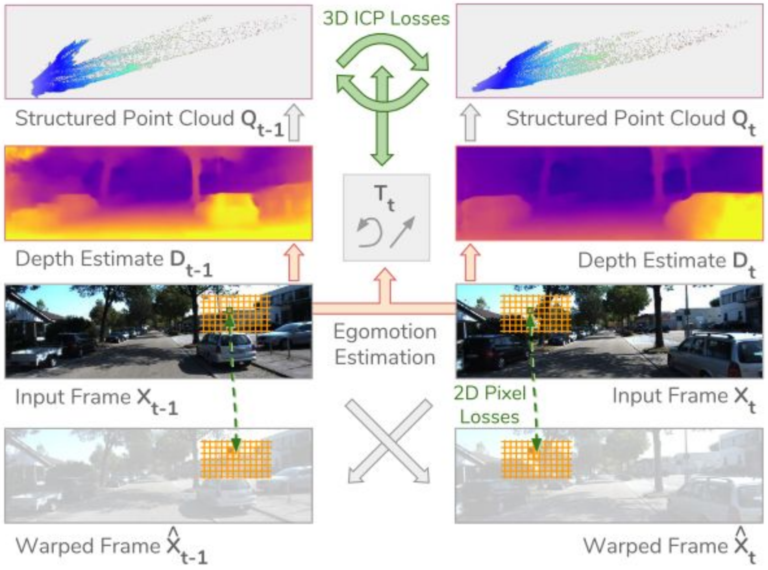


Image credit: Zhou et al., 2017

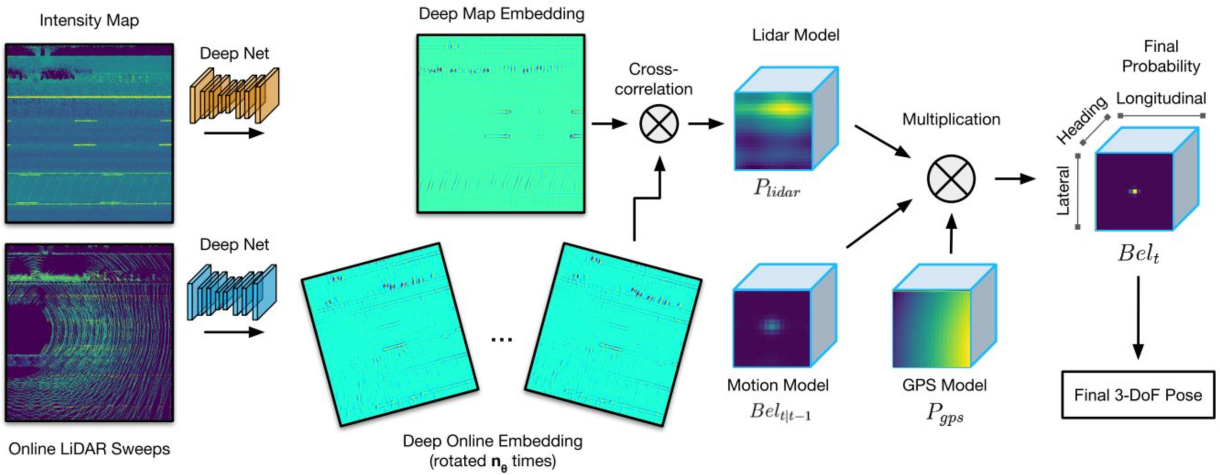
[3] Unsupervised Geometric Learning Continued

- **Problem 2.0:** Photometric / 2D Geometric loss terms are biased towards areas closer to the camera!
- **Solution 2.0:** Compute geometric loss in 3D!
    - Unsupervised Learning [...] Using 3D Geometric Constraints by Mahjourian et al., 2018
- Sharper and more accurate results than methods using just 2.5D losses.



Existing Work

- Learning to Localize using a LiDAR Intensity Map
- **Motivation:** Match LiDAR observation to a map for high-precision localization.
- **Challenges:** Dynamic objects, calibration mismatch.
- **Idea:** Replace matching heuristics with “learning to match”.



Future Work

- More sophisticated geometry + more learning = more power
- Relaxing nondifferentiable methods to be diffable and integrating them in structured prediction pipelines trained E2E is very powerful!
  - Example: Differentiable semantic point cloud registration
- Experiments: lots of challenging datasets available!
  - KITTI, SUN3D, etc.

Conclusions

- Adding geometric structure allows a neural net to focus on questions that geometry can’t answer!
- We saw examples in **stereo**, **bundle adjustment**, and **unsupervised learning**.
- Paradigm not restricted to cameras.
- Fast-growing field with lots of opportunities.

Thank you!