Lecture 10: Dense 3D Reconstruction

1 Dense Reconstruction

For the 3D reconstruction from multiple views we assume that cameras are calibrated

- **intrinsically** (K is known for each camera), and
- **extrinsically** (T and R between cameras are known, for instance, from Structure From Motion).

Considering multi-view stereo, one has:

- **Input**: calibrated images from several viewpoints.
- **Output**: 3D object dense reconstruction.

Remark. Recall: The two camera centers and the image point p determine the epipolar plane, which intersects each camera image plane in the epipolar lines. Since we use the epipolar constraint, corresponding points only need to be searched along epipolar lines.

We want to estimate the structure from a *dense* region of pixels (hence not only from corners). The workflow consists of:

1. **Local methods**: estimate depth for every pixel independently.
2. **Global methods**: refine the depth surface as a whole by enforcing smoothness constraint.

1.1 Photometric Error

We use the **photometric error** (SSD): this is derived for every combination of the reference image and any further image. **Idea**: optimal depth minimizes the photometric error in all images as a function of the depth in the first image.

1.1.1 Aggregated Photometric Error

The Dense reconstruction requires establishing dense correspondences. These are computed based on the photometric error (SSD between corresponding patches of intensity values (min patch size: 1×1 pixels). What are the pros and cons of large and small patches?

- Small window: Pro: more detail, Cons: more noise.
- Large window: Pro: smoother disparity maps, Cons: less detail.

Not all the pixels can be matched reliably, due to viewpoint changes, occlusions. We take advantage of *many* small baseline views, where *high quality* matching is possible. Important facts:

- Non distinctive features (repetitive texture) show multiple minima.

- The aggregated photometric error for *flat regions* and *edges* parallel to the epipolar line show **flat valleys** (noise!).
- For distinctive features, the aggregated photometric error has one clear minimum.

1.2 Disparity Space Image (DSI)

For a given image point (u, v) and for discrete depth hypotheses d, the aggregate photometric error $C(u, v, d)$ with respect to the reference image I_r can be stored in a volumetric 3D grid called the *Disparity Space Image (DSI)*, where each voxel (group of u, v, d) has value

$$C(u, v, d) = \sum_k \rho \left(\tilde{I}_k(u', v', d) - I_r(u, v) \right), \tag{1.1}$$

where $\tilde{I}_k(u', v', d)$ is the patch of intensity values in the k-th image centered on the pixel (u', v') corresponding to the patch $I_r(u, v)$ in the reference image I_r an depth hypothesis d. Furthermore ρ is the photometric error (SSD).

1.2.1 Solution to Depth Estimation Problem

The solution to the depth estimation problem is a function $d(u, v)$ in the DSI that satisfies:

$$\begin{aligned} &\text{Minimum aggregated photometric error (i.e. } argmin_d\ C) \\ &\text{AND} \\ &\text{Piecewise smooth (global methods)} \end{aligned} \tag{1.2}$$

Interpolating while not overfitting!

Global Methods:

We formulate them in terms of energy minimization. The objective is to find a surface $d(u, v)$ that minimizes a global energy

$$E(d) = \underbrace{E_d(d)}_{\text{data term}} + \underbrace{\lambda \cdot E_s(d)}_{\text{regularization term}}, \tag{1.3}$$

where

$$E_d(d) = \sum_{(u,v)} C(u, v, d(u, v)) \tag{1.4}$$

and

$$E_s(d) = \sum_{(u,v)} \rho_d(d(u, v) - d(u + 1, v)) + \rho_d(d(u, v) - d(u, v + 1)). \tag{1.5}$$

ρ_d is a norm (e.g. the $L_{1,2}$ or Huber norm), while λ controls the tradeoff data (regularization). What happens as λ increases? Higher **smoothing**!

Regularized Depth Maps

- **The regularization term** $E_s(d)$
  - **Smooths** non smooth surfaces (result of noisy measurements) as well as discontinuities.
  - Fills the holes.
- Popular assumption: discontinuities in intensity **coincide** with discontinuities in depth.
- We control **smoothness penalties** according to image gradient (discrete)

$$\rho_d(d(u,v) - d(u+1,v)) \cdot \rho_I(\|I(u,v) - I(u+1,v)\|) \tag{1.6}$$

- ρ_I is some monotically *decreasing* function of intensity differences: **lower** smoothness cost for **high intensity gradients** (if there are high intensity gradients, you don't want to smooth them as they are a crucial information in your image).

Choosing the stereo baseline

What is the optimal stereo baseline? As we have previously introduced:

- Too small: large depth error.
- Too large: difficult search problem.

A possible solution is to obtain depth maps from small baselines: when the baseline becomes too large, create a new reference frame and start a new depth computation (**depth map fusion**, different depth maps with different perspectives gives a complete image).

GPGPU for Dense Reconstruction

General Purpose Computing on Graphics Processing Unit. Perform demanding calculations on the GPU instead of the CPU. We can run processes in **parallel** on thousands of cores (CPU is optimized for serial processing). More transistors for data processing.

- Fast pixel processing (ray tracing, draw textures, shaded triangles,..)
- Fast matrix/vector operations (transform vertices)
- Programmable (shading, bump mapping)
- Floating-point support (accurate computations)
- Deep learning.

And

- **Image processing**
  - Filtering and feature extractions (e.g. convolutions)
  - Warping (e.g. epipolar rectification, homography).

- **Multiple-view geometry**
  - Search for dense correspondences (pixel wise operations, matrix and vector operations (epipolar geometry).
  - Aggregated photometric error.
- **Global Optimation**
- Variational methods (i.e. regularization (smoothing)) (divergence computation)

Typically on consumer hardware: 1024 threads per multiprocessor, 30 multiprocessors: 30000 threads. CPU with 4 cores which supports 32 threads. High arithmetic intensity.

1.3 Understanding Check

Are you able to answer the following questions?

- *Are you able to describe the multi-view stereo working principle? (aggregated photometric error)*
- *What are the differences in the behavior of the aggregated photometric error for corners, flat regions and edges?*
- *What is the disparity space image (DSI) and how is it built in practice?*
- *How do we extract the depth from the DSI?*
- *How do we enforce smoothness (regularization) and how do we incorporate depth discontinuities (mathematical expressions)?*
- *What happens if we increase lambda (the regularization term)? What if lambda is 0? And if lambda is too big?*
- *What is the optimal baseline for multi-view stereo?*
- *What are the advantages of GPUs?*