CSE 252A: Stereo Vision II

Lecturer: David Kriegman

Scribed by Owen Jow on November 08, 2018

1 Stereo: Establishing Correspondence

- *Sparse, feature-based correspondence*: detect interest points in one image, then look for their correspondences in the other image
  - Stereo works well around interest points, as they're easily localized during triangulation.
- *Dense, area-based correspondence*: directly compare window intensities between the two images
  - This is what humans do (as evidenced by Julesz's random dot stereogram experiments).
  - This is implemented more often than feature-based correspondence.

1.1 Dense Correspondence

The core assumption is that corresponding points have the same intensity in both images (*brightness constancy constraint*). Fundamentally, this means the scene is Lambertian.

We also base window matching on the idea that the photographed surfaces are parallel to the image plane post-rectification (otherwise windows will be warped between views and won't have exactly the same contents); this is not generally true, but the hope is that it serves as a good approximation.

To compare windows, we often use an SSD or NCC[1] metric.

1.1.1 Extension: 1-to-"At Most 1" Matching

Naively, we'd take every point in the left image and match it independently to the best point in the right image. But maybe two points in the left image match with the same point in the right image; this is no good. So one improvement over the naive greedy approach is to enforce 1-to-"at most 1" matching through a global optimization.

To rephrase, we'd like to have at most one match for each point. The reason we might not have a match, even in the middle of an image, is that there might be an object occluding things in just one view. We call the occluded region a **half-occluded region**, i.e. an area occluded in one image but not the other. *Where we don't have a match, there will be a depth discontinuity.*

[1] **Normalized cross-correlation**: just a dot product between two normalized vectors.

1.1.2 Extension: Epipolar Ordering

Advanced methods might also account for image information in a way that goes beyond "windows should look the same" and searching along epipolar lines. For example, to help resolve ambiguity, one could search only over a minimum and maximum disparity range, or enforce consistency in the ordering of matches across the epipolar lines (although this ordering doesn't always hold).

1.1.3 Extension: Window Size and Shape

The window size can also make a significant difference. At first, it might seem as though a bigger window would mean a higher accuracy, but we don't want to make the window too big or we will confound matching for small or narrow detail in the image. A solution: **adaptive window size**.

Also, squares might not map to squares. Image patches will probably be transformed between views in a way that extends beyond a pure translation. The exception is when the images are rectified (s.t. the image planes are parallel translations of each other) and the imaged surfaces are parallel to the image plane. To achieve these conditions, we can rectify and apply a per-patch homography so that a window maps to another window exactly.

Note: to get the homography, we need an estimate of the surface normal (not trivial to estimate!).

1.1.4 Extension: Lighting and Specularity

Lighting variation has a big impact. To alleviate this (to some degree), we can perform normalization, e.g. subtracting the mean value. Of course, this still won't help with specularity.

- To handle specularity, one thing we can try is performing matching over the UV channels from the SUV color space, assuming we know the specular lighting color(s).

2 Variations on Binocular Stereo

2.1 Trinocular Stereopsis

If we have three cameras, then between every pair of images there's an epipolar geometry.

$$\mathbf{p_1}^T \mathcal{E}_{12} \mathbf{p_2} = 0$$
$$\mathbf{p_2}^T \mathcal{E}_{23} \mathbf{p_3} = 0$$
$$\mathbf{p_3}^T \mathcal{E}_{31} \mathbf{p_1} = 0$$

The constraints are not independent. If we match a point in image 1 with a point in image 2, we already have a 3D point and by extension a match in image 3. So to help eliminate ambiguity, we can verify that the window around the corresponding point in image 3 looks right.

2.2 Helmholtz Reciprocity Stereopsis

Another problem in stereo matching is non-Lambertian surfaces. To handle completely arbitrary BRDFs, we can take advantage of Helmholtz reciprocity. Again, **Helmholtz reciprocity** means that the BRDF is the same whether you go in at direction θ_i, ϕ_i and come out in direction θ_o, ϕ_o or go in at direction θ_o, ϕ_o and come out in direction θ_i, ϕ_i.

$$\rho(\theta_i, \phi_i; \theta_o, \phi_o) = \rho(\theta_o, \phi_o; \theta_i, \phi_i)$$

So we can swap the camera and the lighting positions, and the BRDF is the same.

If we build an active stereo system in which this is done (switch the camera and the light positions, taking a photo in each configuration), we can use the fact that the BRDF is the same in both cases.

- For a nearby point source, the irradiance is *the cosine of the normal and the lighting direction* times *the BRDF for the incoming and outgoing directions* divided by *the attenuation factor*.
- After taking a photo in each setting, we can equate the irradiances and eliminate the BRDFs from the equation. So the two stereo views are from switching the light and camera positions.
- The normal is part of these equations, so we can estimate surface normals for matched points.
- Since we have multiple views, we can also produce disparities as usual.
  - Note: photometric stereo requires that surfaces be continuous (since we have to integrate the normals); there is no such problem with this method since we have disparities as well.