Three-Dimensional Sensors
Lecture 2: Projected-Light Depth Cameras

Radu Horaud
INRIA Grenoble Rhone-Alpes, France
Radu.Horaud@inria.fr
http://perception.inrialpes.fr/

Outline

- The geometry of active stereo.
- Surface reflectivity.
- Camera-projector calibration
- Active stereo matching and reconstruction

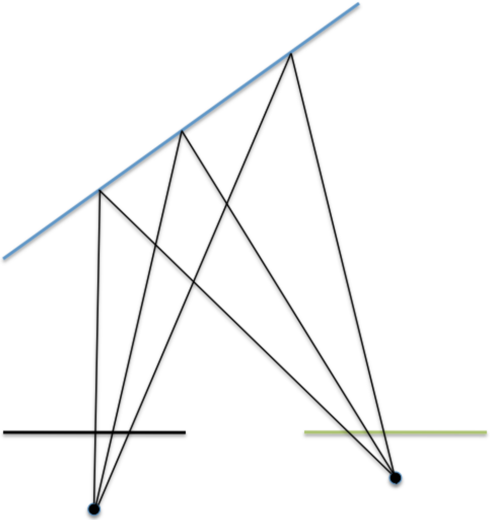
Generalities

- The basic principle is to associate a light projector with a standard camera.
- The projector and the camera share the same geometric model.
- Optical (non-linear distorsions) are present with both devices. They will be treated separately.
- 3D reconstruction is based on the same principle as stereo: sensor calibration and matching.
- Instead of matching two images, we match a pattern with "itself" after being backscattered by scene objects.

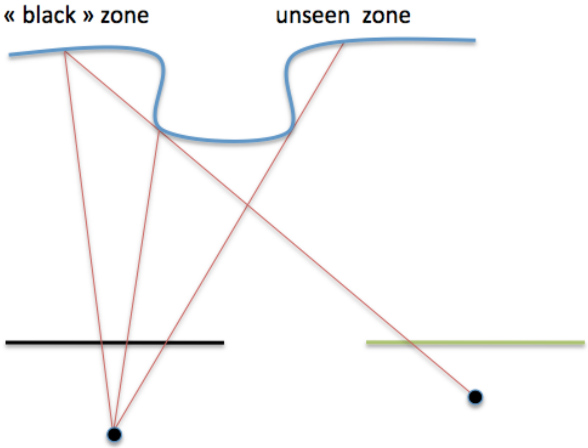
The Geometry of Active Stereo

- Perspective effects (foreshortening).
- Surface discontinuities.
- Cluttered scenes.

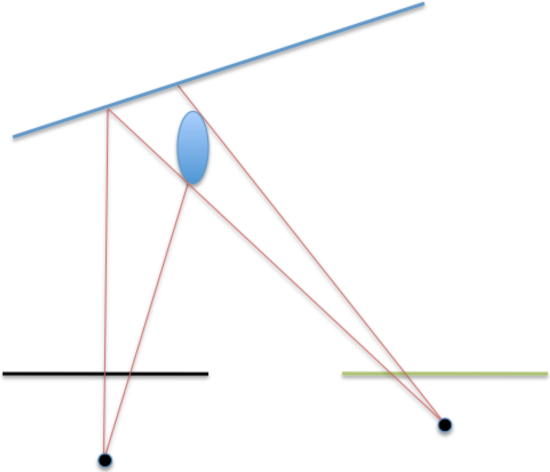
Foreshortening



Discontinuities



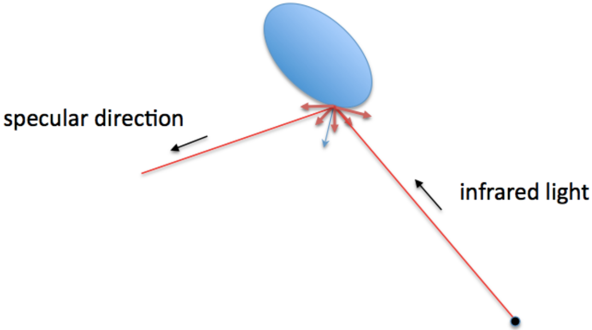
Clutter



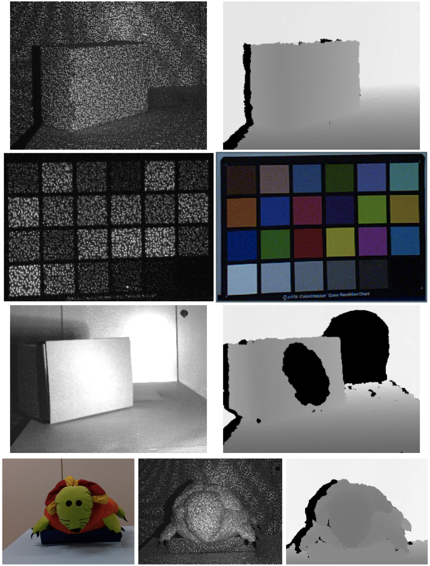
Surface Reflectivity and External Illumination

- This is a scientific topic in its own right.
- Reflectance (or absorption) depends on the material properties and of the wavelength of the incident light.
- A rough way to think of surface reflectivity is to divide surfaces into **diffuse** (matte) and **specular** (mirror).
- In general a surface is a combination of these effects plus other ones: metal, glossy, rough, transparent, translucent, etc. cause problems in practice.
- External illumination (daylight, artificial light, etc.) acts as *noise* added to the infra-red light emitted by the projector.

Reflectivity



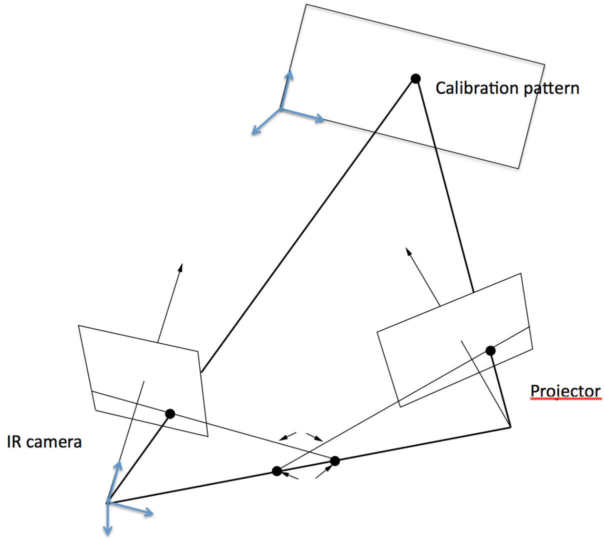
Artifacts with the Kinect Camera



Back to Geometry

- Consider an IR camera and a projector in general positions (before rectification).
- Parameters to be estimated:
  - The internal parameters of the IR camera
  - The external parameters of the IR camera
  - The projection matrix of the light projector

IR-Camera and Projector Geometry and Calibration



External IR-Camera Calibration

- The basic equation is $\boldsymbol{m} \approx \left[\begin{array}{cc} \mathbf{R} & \boldsymbol{t} \end{array}\right] \boldsymbol{M}'$ where $\boldsymbol{M}'$ lies on the plane $Z' = 0$, hence we obtain:

$$\left(\begin{array}{c} sx \\ sy \\ s \end{array}\right) = \left[\begin{array}{ccc} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t} \end{array}\right] \left(\begin{array}{c} X' \\ Y' \\ 1 \end{array}\right)$$

- Each correspondence between an image point $\boldsymbol{m}_i$ and a calibration point $\boldsymbol{M}'_i$ yields two homogeneous linear equations in the unknown extrinsics:

$$\left\{ \begin{array}{lcl} x_i(X'_i r_{13} + Y'_i r_{23} + t_3) & = & X'_i r_{11} + Y'_i r_{12} + t_1 \\ y_i(X'_i r_{13} + Y'_i r_{23} + t_3) & = & X'_i r_{21} + Y'_i r_{22} + t_2 \end{array} \right.$$

- In practice it is useful to constrain the parameters of the rotation matrix, i.e., $\|\boldsymbol{r}_1\| = \|\boldsymbol{r}_2\| = 1$, $\boldsymbol{r}_1^\top \boldsymbol{r}_2 = 0$. Various optimization methods can be used in practice.

External IR-Camera Parameters

- The external camera matrix is:

$$\left[\begin{array}{cc} \mathbf{R} & \boldsymbol{t} \end{array} \right] = \left[\begin{array}{cccc} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{r}_1 \times \boldsymbol{r}_2 & \boldsymbol{t} \end{array} \right]$$

- The coordinates of the planar set of points in the IR-camera frame:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \left[\begin{array}{ccc} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t} \end{array} \right] \begin{pmatrix} X' \\ Y' \\ 1 \end{pmatrix}$$

- Recall the simple relationship between $\boldsymbol{m}$ and $\boldsymbol{M}$ when the latter is in camera frame:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = Z \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Point/Plane Duality in Projective Space

- The point $\boldsymbol{M}$ belongs to the plane $\boldsymbol{\mu}$:

$$\alpha X + \beta Y + \gamma Z + \delta H = 0 \text{ or } \boldsymbol{\mu}^\top \tilde{\boldsymbol{M}} = 0$$

  with the notation $\tilde{\boldsymbol{M}} = \begin{pmatrix} X & Y & Z & H \end{pmatrix}$ and $H = 1$.
- The 4×4 point-transformation $\mathbf{D}$ (a rotation followed by a translation) maps the coordinates of a point from the calibration frame to the camera frame: $\tilde{\boldsymbol{M}} = \mathbf{D}\tilde{\boldsymbol{M}}'$
- Let $\boldsymbol{\mu}'^\top \tilde{\boldsymbol{M}}' = 0$ with $\boldsymbol{\mu}' = \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}$ be the plane of the calibration pattern. Substituting $\tilde{\boldsymbol{M}}'$ we obtain $\left(\mathbf{D}^{-\top}\boldsymbol{\mu}'\right)^\top \tilde{\boldsymbol{M}} = 0$ with:

$$\mathbf{D}^{-\top} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ -\boldsymbol{t}^\top \mathbf{R} & 1 \end{bmatrix} \text{ and } \boldsymbol{\mu} = \mathbf{D}^{-\top}\boldsymbol{\mu}'$$

- The latter is the plane-transformation or the dual of the point-transformation.

Projector Calibration

- We assume that the planar calibration pattern is in the **same** position as before. A ray of light through the projector image-point $\boldsymbol{p} = (u \; v \; 1)^\top$ is $\boldsymbol{p} \approx \mathbf{P}\tilde{\boldsymbol{M}}$.
- This ray of light is observed in the IR-camera image at $\boldsymbol{m}$. Recall that $\boldsymbol{M} = Z\boldsymbol{m}$ and that $\boldsymbol{\mu}^\top \tilde{\boldsymbol{M}} = 0$. Hence the depth Z is:

$$Z = -\frac{\delta}{\alpha x + \beta y + \gamma}$$

- For each camera-projector point correspondence $\boldsymbol{m}_i \leftrightarrow \boldsymbol{p}_i$ we obtain two equations:

$$\left\{ \begin{array}{lcl} u_i[Z_i(x_i p_{31} + y_i p_{32} + p_{33}) + p_{34}] & = & Z_i(x_i p_{11} + y_i p_{12} + p_{13}) + p_{14} \\ v_i[Z_i(x_i p_{31} + y_i p_{32} + p_{33}) + p_{34}] & = & Z_i(x_i p_{21} + y_i p_{22} + p_{23}) + p_{24} \end{array} \right.$$

- This requires several such correspondences and several parallel planes.

Projector's Intrinsics and Extrinsics

- If enough correspondences are available, one can estimate the projection matrix $\mathbf{P}$ either linearly or with the help of a non-linear optimization method (bundle adjustment).
- Intrinsic and extrinsic parameter can be made explicit by identification:

$$\mathbf{P} \approx \mathbf{K} \begin{bmatrix} \mathbf{R} & \boldsymbol{t} \end{bmatrix}$$

or

$$\begin{bmatrix} \mathbf{P}_{3\times 3} & \boldsymbol{p}_4 \end{bmatrix} \approx \begin{bmatrix} \mathbf{KR} & \mathbf{K}\boldsymbol{t} \end{bmatrix}$$

- Making use of $\mathbf{R}\mathbf{R}^\top = \mathbf{I}$ one can easily extract the intrinsic and extrinsic parameters from $\mathbf{P}$.

Camera-Projector Calibration Step-by-Step

1. Calibrate the internal parameters of the IR camera (OpenCV).
2. Calibrate the external parameters with respect to a single planar calibration pattern (OpenCV).
3. Transform the calibration plane in the camera coordinate frame.
4. Using the same position of the calibration pattern to obtain $\boldsymbol{m}_i \leftrightarrow \boldsymbol{p}_i$ correspondences.
5. Move the calibration pattern in several parallel positions.
6. Estimate the calibration parameters of the projector in the camera frame.
7. Rectify the camera-projector setup to align the camera image with the projector image.

Active Stereo Reconstruction

- With a calibrated & rectified camera-projector setup, the relationship between the depth Z and the disparity d is:

$$Z = \frac{b}{d} \text{ with } d = x_p - x_c$$

where b is the baseline, x_p and x_c are the horizontal coordinates of a projector image-point and a camera image-point.

- The coordinates of a point are:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = Z \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix}$$

Active Stereo Matching

- The problem is to find $x_c \leftrightarrow x_p$ correspondences along epipolar lines (along the rows).
- Each pixel of the projector image must be **signed** such that it can be easily **recognized** in the IR-camera image:
  - Temporal signature;
  - Spatial signature, or
  - a combination of both
- Define a $K \times K \times T$ window around each pixel.
- Arrange the pixels in a vector of size $N = K \times K \times T$

Cross-Correlation

- Let the entries of two N-dimensional vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ correspond to N realizations of two random variables v and w.
- Expectation: $E[v] = \frac{1}{N}\sum_{n=1}^{N} p_n v_n$ with $\sum_{n=1}^{N} p_n = 1$.
- Variance: $var[v] = E[v^2] - E[v]^2$
- Standard deviation: $\sigma_v = \sqrt{var[v]}$
- Standardization: $\hat{v} = \frac{v - E[v]}{\sigma_v}$
- Cross-correlation: $\rho_{vw} = corr[\hat{v}, \hat{w}] = E\left[\frac{v-E[v]}{\sigma_v}\frac{w-E[w]}{\sigma_w}\right]$

$$\rho_{vw} = \sum_{n=1}^{N} p_n \frac{(v_n - E[v])(w_n - E[w])}{\sigma_v \sigma_w}$$

Matching Two Pixels

- Vector $\boldsymbol{v}$ contains pixel information associated with the projected light: $\boldsymbol{v} = (L(n))_{n=1}^{N}$.
- Vector $\boldsymbol{w}$ contains pixel information associated with the IR image: $\boldsymbol{w} = (I(n))_{n=1}^{N}$.
- The entries of these two vectors are non-null, hence:
  - $0 \leq \rho_{vw} \leq 1$.
  - $\rho_{vw} = 1 \Leftrightarrow \boldsymbol{v} = \boldsymbol{w}$.
  - $\rho_{vw} = 0$ if v and w are uncorrelated.
- $\boldsymbol{v}$ is associated with IR image pixel (x_c, y)
- $\boldsymbol{w}$ is associated with projector image pixel (x_p, y)
- A match is accepted if $\rho_{vw} > t$.

Matching Two Images

- Based on the cross-correlation coefficient it is possible to implement various global image-matching techniques.
- Because of the epipolar geometry holds and if the images are rectified, the matching can be done row-wise.
- The most straightforward way to match two corresponding row is to use a sequence-alignment algorithm.
- Sequence-alignment algorithms are generally based on dynamic programming (DP).

Conclusions

- To calibrate a Kinect-like camera, it is necessary to have access to the projector image, but this is not the case with Kinect because the manufacturer (Primesense) does not provide this access.
- The principles behind calibration allow to understand the weaknesses and the current limitations of this type of cameras: limited range, limited to indoor lighting, **extremely sensitive to saturated light**, glossy surfaces, occlusions, transparent and translucent objects, scattering effects (wool, hair, etc.).
- Projected-light depth cameras have good resolution and good depth accuracy, both are better than time-of-flight cameras.