Basics for Enhanced Visualization: 3D/Data Stereopsis

Rodrigo Cabral

Polytech Nice - Data Science

cabral@unice.fr

Source: `https://www.lhup.edu/~dsimanek/3d/stereo/3dgallery3.htm`.

Outline

1. Introduction:
   - How to recover the 3D structure?
   - Depth from two views: stereo principle
   - The two problems of stereo
2. Point correspondences
   - Point matches and matching criteria
   - Epipolar geometry
3. General triangulation with calibrated cameras
4. Conclusions

How to recover 3D structure?

- Problem: from camera matrix $\mathbf{M}$ and image point $\mathbf{u}_{\text{im}}$ we cannot do it!

$$\lambda \mathbf{u}_{\text{im}} = \mathbf{M}\,\mathbf{u}_w$$

Matrix $\mathbf{M}$ is a 3×4 matrix: the system of equations does not have a unique solution.

Introduction

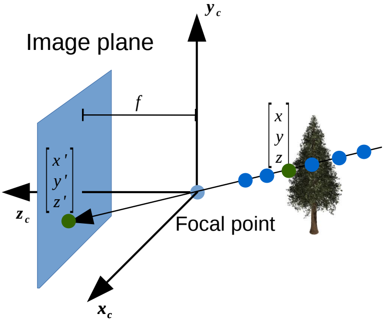
How to recover 3D structure?

- Problem: from camera matrix $\mathbf{M}$ and image point $\mathbf{u}_{\text{im}}$ we cannot do it!

$$\lambda \mathbf{u}_{\text{im}} \quad = \quad \mathbf{M}\, \mathbf{u}_w$$

Matrix $\mathbf{M}$ is a 3×4 matrix: the system of equations does not have a unique solution.

- From one image, we know the incoming ray, but not the point:



Introduction

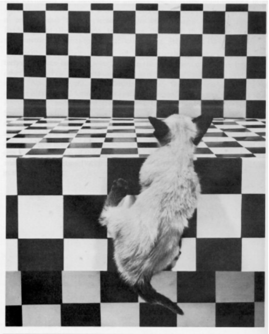
How to recover 3D structure?

- Solution: trivial - find further non redundant information that will allow to retrieve depth.
- Multiple options...
  - Known size of patterns: depth from texture.
  - Known circle of confusion: depth from focusing.
  - Known structured lighting.
  - and the subject of this class.

Introduction

How to recover 3D structure?

Depth from texture: texture with known structure is deformed by depth



Source: The visual cliff, by William Vandivert, 1960.
http://www.csd.uwo.ca/faculty/olga/Courses//Winter2010//CS4442_9542b/L14-CV-stereo.pdf.

Introduction

How to recover 3D structure?

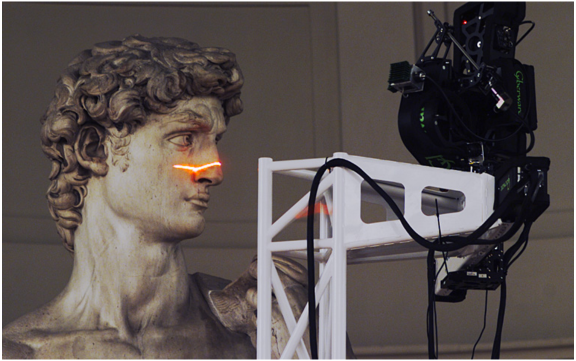
Depth from focus: blur gives information on depth

Source: from *The Art of Photography*, Canon.

Introduction

How to recover 3D structure?

Known structured light: projected light patterns are deformed by 3D structure.



- This is an active method, since we emit the light.

Source: https://graphics.stanford.edu/projects/mich/more-david/scanner-head-and-david-head-s.jpg.

How to recover 3D structure?

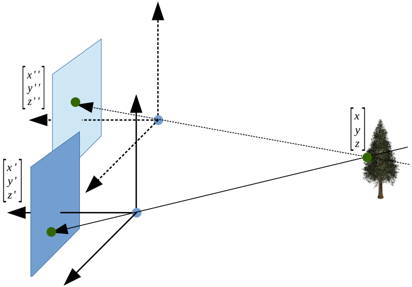
Known structured light: projected light patterns are deformed by 3D structure.

- It is similar to depth from texture, but it is active.
- This is one of the ways that Kinect uses to retrieve depth information.
  - The dots are actually ellipses with angles depending on depth.
  - Kinect also uses the **stereo principle** (next slides).

Source: http://www.youtube.com/watch?v=-gbzXjdHfJA.

Depth from two views: stereo principles

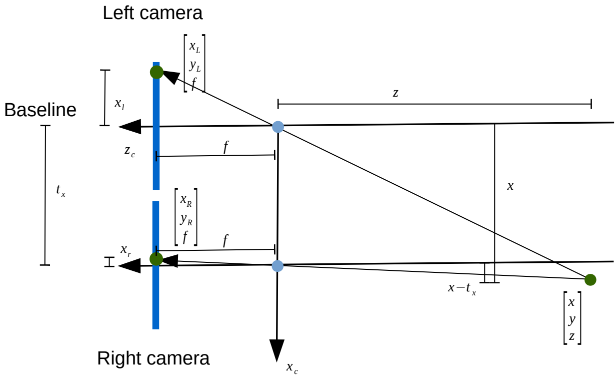
Another solution: use a second camera.



- The rays intersect at the true depth $\Longrightarrow$ We can retrieve the depth with two cameras!

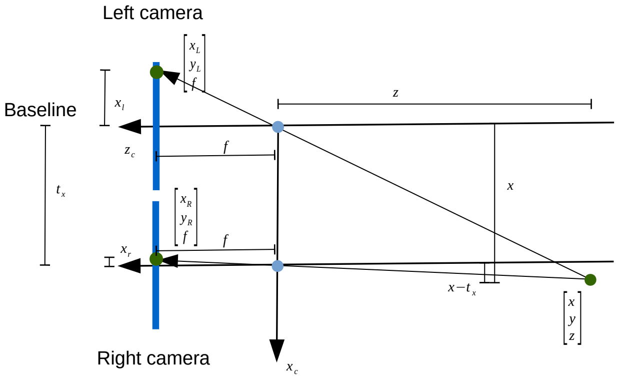
Depth from two views: stereo principles

- Simplified model: two cameras with parallel optical axis and same focal length f. Reference frame is left camera frame.



Depth from two views: stereo principles

- Simplified model: two cameras with parallel optical axis and same focal length f. Reference frame is left camera frame.



- From the two rays we get: $\frac{x}{z} = \frac{x_l}{f}$ and $\frac{x-t_x}{z} = \frac{x_r}{f}$.

Depth from two views: stereo principles

- From the two rays we get: $\frac{x}{z} = \frac{x_l}{f}$ and $\frac{x-t_x}{z} = \frac{x_r}{f}$.
- Solving for z we get:

$$Z = \frac{f\ t_x}{d}$$

where $d = x_l - x_r$ is called **disparity**.

Depth from two views: stereo principles

- From the two rays we get: $\frac{x}{z} = \frac{x_l}{f}$ and $\frac{x - t_x}{z} = \frac{x_r}{f}$.
- Solving for z we get:

$$Z = \frac{f\ t_x}{d}$$

where $d = x_l - x_r$ is called **disparity**.

Depth from two views: stereo principles

- From the two rays we get: $\frac{x}{z} = \frac{x_l}{f}$ and $\frac{x-t_x}{z} = \frac{x_r}{f}$.
- Solving for z we get:

$$z = \frac{f\,t_x}{d}$$

  where $d = x_l - x_r$ is called **disparity**.
- x and y follow trivially:

$$x = \frac{x_l\,t_x}{d} = \frac{x_r\,t_x}{d} \quad \text{and } y = \frac{y_l\,t_x}{d} = \frac{y_r\,t_x}{d}$$

- This is also used in our vision. The brain is constantly evaluating disparity.
- Note that disparity is inversely proportional to depth. Remember when you are on a train: far away mountains move slower than houses close to the railway.

Depth from two views: stereo principles

Some remarks

- *stereopsis* from Greek: "stereo" means solid and "opsis" means appearance
- Therefore, stereopsis should be the name of any vision method which retrieves depth.

Depth from two views: stereo principles

Some remarks

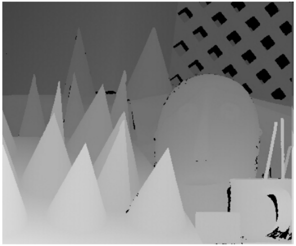
- *stereopsis* from Greek: "stereo" means solid and "opsis" means appearance
- Therefore, stereopsis should be the name of any vision method which retrieves depth.
- Unfortunately this is not the case. Stereopsis is the name for two-camera systems.
- Fundamental principle behind stereo: retrieve the depth from the disparity between corresponding projections in two different image planes.

Depth from two views: stereo principles

Left image

Right image

Disparity image



Source: `http://web.eecs.umich.edu/~silvio/teaching/lectures/lecture7.pdf`

The two problems of stereo

Stereo principle: reconstruct 3D scene from the pixel disparities between two image projections.

- Problem 1: correspondences

  How can you retrieve automatically and efficiently pixel correspondences on both images?

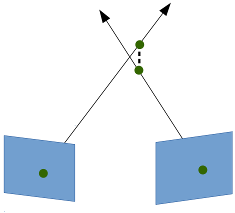
The two problems of stereo

Stereo principle: reconstruct 3D scene from the pixel disparities between two image projections.

- Problem 2: general reconstruction

  What to do when...

  1. image planes are not side to side and intrinsic parameters are different?
     They may be also rotated and focal lengths may be different.
  2. pixel correspondences are noisy?
     Image is noisy and/or algorithm to find correspondences is not perfect.



Point matches and matching criteria

Problem: how to obtain point correspondences?

Solution: **region based matching.**

Point matches and matching criteria

Problem: how to obtain point correspondences?

Solution: **region based matching.**

1. Choose a comparison criterion for image regions $\mathbf{R}^L$, $\mathbf{R}^R$: $\mathcal{C}(\mathbf{R}^{\mathbf{L}}, \mathbf{R}^R)$ (to be defined later).

Point matches and matching criteria

Problem: how to obtain point correspondences?

Solution: **region based matching.**

1. Choose a comparison criterion for image regions $\mathbf{R}^L$, $\mathbf{R}^R$: $\mathcal{C}(\mathbf{R}^{\mathbf{L}}, \mathbf{R}^R)$ (to be defined later).
2. Choose a size w to define square regions.

Point matches and matching criteria

Problem: how to obtain point correspondences?

Solution: **region based matching.**

1. Choose a comparison criterion for image regions $\mathbf{R}^L$, $\mathbf{R}^R$: $\mathcal{C}(\mathbf{R}^\mathbf{L}, \mathbf{R}^R)$ (to be defined later).
2. Choose a size w to define square regions.
3. For all $[i\ j]^\mathsf{T}$ in the left image buffer.
   3.1 Pick a region $R^L_{i,j}$ centered on $[i\ j]^\mathsf{T}$.

Point matches and matching criteria

Problem: how to obtain point correspondences?

Solution: **region based matching.**

1. Choose a comparison criterion for image regions $\mathbf{R}^L$, $\mathbf{R}^R$: $\mathcal{C}(\mathbf{R}^\mathbf{L}, \mathbf{R}^R)$ (to be defined later).
2. Choose a size w to define square regions.
3. For all $[i\ j]^\mathsf{T}$ in the left image buffer.
   3.1 Pick a region $R^L_{i,j}$ centered on $[i\ j]^\mathsf{T}$.
   3.2 For all $[d_x\ d_y]^\mathsf{T}$ such that the region R^R_{i,j,d_x,d_y} centered on point $[i - d_x\ j - d_y]^\mathsf{T}$ in the right image buffer is a valid region:
      - Evaluate $\mathcal{C}(R^L_{i,j}, R^R_{i,j,d_x,d_y})$.

Point matches and matching criteria

Problem: how to obtain point correspondences?

Solution: **region based matching.**

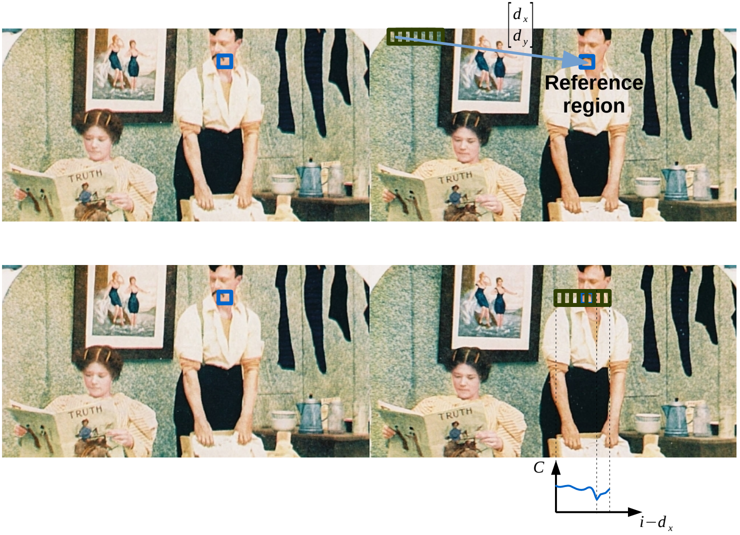
1. Choose a comparison criterion for image regions $\mathbf{R}^L$, $\mathbf{R}^R$: $\mathcal{C}(\mathbf{R}^\mathbf{L}, \mathbf{R}^R)$ (to be defined later).
2. Choose a size w to define square regions.
3. For all $[i\ j]^\mathsf{T}$ in the left image buffer.
   3.1 Pick a region $R^L_{i,j}$ centered on $[i\ j]^\mathsf{T}$.
   3.2 For all $[d_x\ d_y]^\mathsf{T}$ such that the region R^R_{i,j,d_x,d_y} centered on point $[i - d_x\ j - d_y]^\mathsf{T}$ in the right image buffer is a valid region:
      - Evaluate $\mathcal{C}(R^L_{i,j}, R^R_{i,j,d_x,d_y})$.
   3.3 The corresponding point to $[i\ j]^\mathsf{T}$ in the right image buffer is $\left[i - d^\star_x\ j - d^\star_y\right]^\mathsf{T}$ where

$$\left[d^\star_x\ d^\star_y\right]^\mathsf{T} = \underset{d_x,\, d_y}{\arg\min}\ (\text{or}\ \underset{d_x,\, d_y}{\arg\max})\ \mathcal{C}(R^L_{i,j}, R^R_{i,j,d_x,d_y})$$

Correspondences

Point matches and matching criteria

Region based matching



Correspondences

Point matches and matching criteria

- **Sum of squared differences (SSD):**

$$\mathcal{C}(\mathbf{R}^L, \mathbf{R}^R) = \sum_{i,j} (I^L_{i,j} - I^R_{i,j})^2$$

where $I^L_{i,j}$ are the pixel values in region $\mathbf{R}^L$,
$I^L_{i,j}$ are the pixel values in region $\mathbf{R}^R$.

- **Correlation:**

$$\mathcal{C}(\mathbf{R}^L, \mathbf{R}^R) = \sum_{i,j} \frac{(I^L_{i,j} - \bar{I}_L)(I^R_{i,j} - \bar{I}_R)}{\sigma(I_L)\sigma(I_R)}$$

where $\bar{I}_L$ and $\bar{I}_R$ are the mean intensities,
$\sigma(I_L)$ and $\sigma(I_R)$ are the standard deviations.

Point matches and matching criteria

- In practice, SSD is most commonly used.

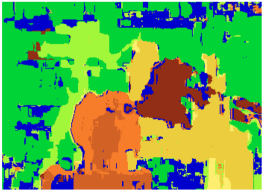
Source: https://courses.cs.washington.edu/courses/cse455/09wi/Lects/lect16.pdf

Point matches and matching criteria

- In practice, SSD is most commonly used.
- Local matching is never perfect: noise, occlusions, ambiguity...

Image

Ground truth disparity



Region based disparity

Source: https://courses.cs.washington.edu/courses/cse455/09wi/Lects/lect16.pdf

Point matches and matching criteria

- We can also match specific features: lines, circles, ellipses...
- We can develop a global cost function: sum of local matches costs + smoothness of matchings + penalty for repeated matches + sum of local matches costs with right image reference.

Point matches and matching criteria

- We can also match specific features: lines, circles, ellipses...
- We can develop a global cost function: sum of local matches costs + smoothness of matchings + penalty for repeated matches + sum of local matches costs with right image reference.
- Note that region based matching gives directly the disparities d_x and d_y.
- The price to pay with this procedure: **huge complexity! We look for a matching pixel on the entire image!**

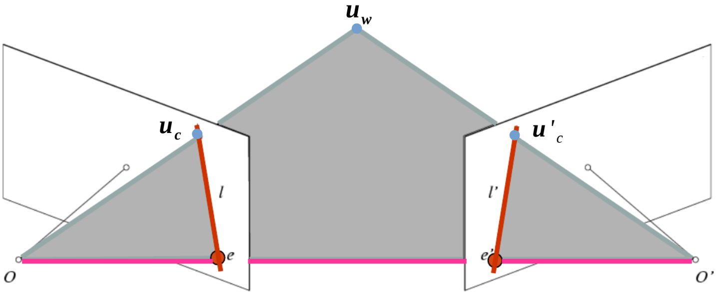
Epipolar geometry

- The price to pay with this procedure: **huge complexity! We look for a matching pixel on the entire image!**
- Problem: how do we reduce the search region for a correspondence?
- Solution: use the geometry of stereo - **epipolar geometry**.

Correspondences

Epipolar geometry

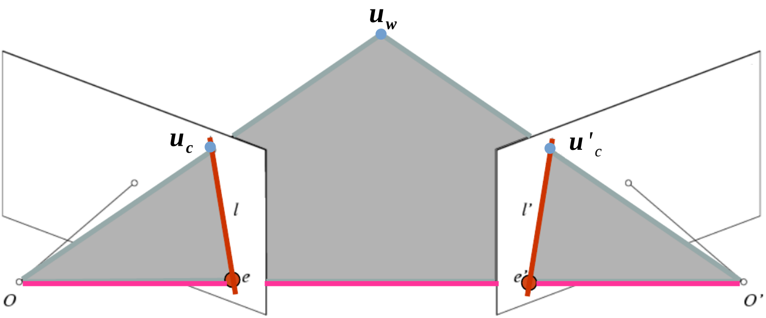
- For simplicity we consider the image planes in front of the focal point $\Longrightarrow$ reversed x_c and y_c.
- We also consider a normalized focal length $f = 1$.



Left image plane

Right image plane

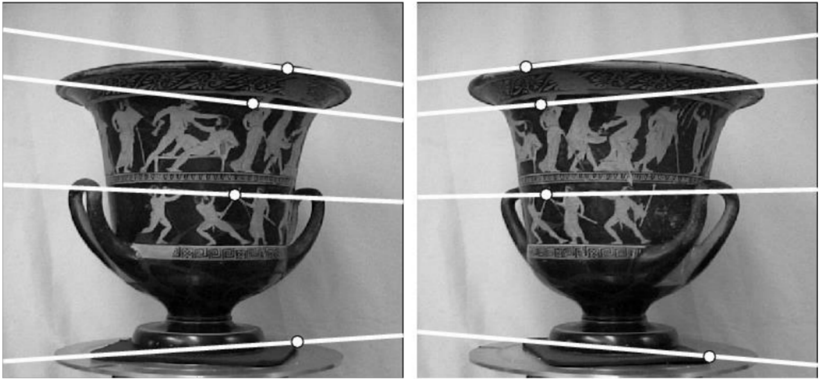
Epipolar geometry



- Baseline: line (OO') connecting the two camera centers.
- Epipolar plane: plane containing baseline and a point (O,O', $\mathbf{u}_w$).
- Epipoles $\mathbf{e}$ and $\mathbf{e}'$: intersection of baseline and image planes.
- Epipolar lines l and l': intersection of epipolar plane with image planes.

Correspondences

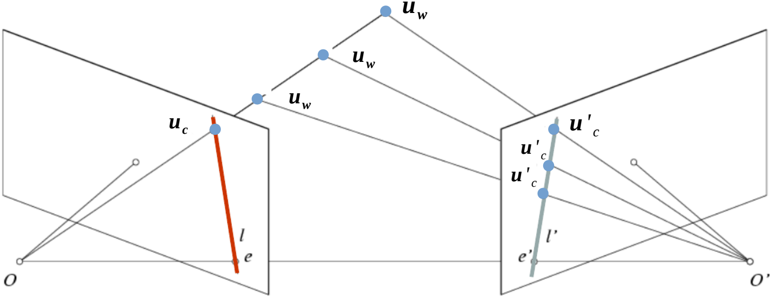
Epipolar geometry



Converging cameras

Epipolar geometry

If we have a point $\mathbf{u}_c$ on the left image plane, what can be the corresponding points on the right image?



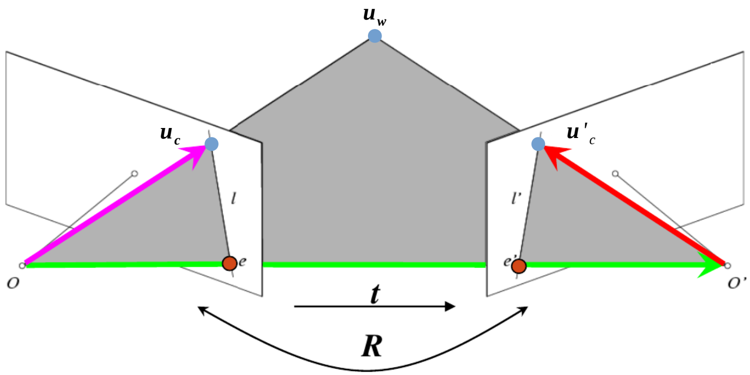
- Candidate matches for $\mathbf{u}_c$ must lie on the epipolar line l'.
- Similarly, candidate matches for $\mathbf{u}'_c$ lie on the epipolar line l.

Epipolar geometry

- The epipolar geometry constrains the search region to a line!
- A much smaller region than the entire image.

Epipolar geometry

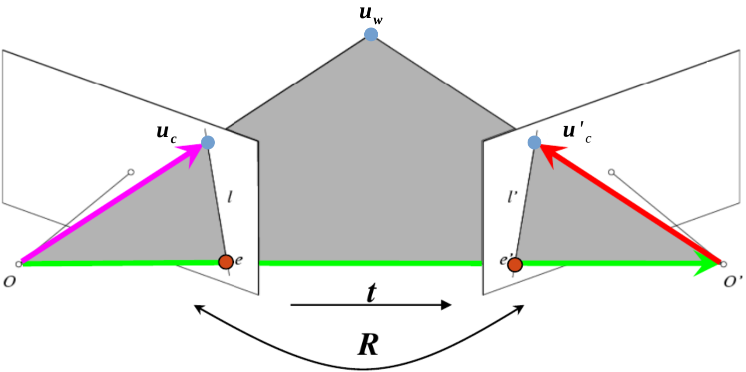
- How do we describe the epipolar constraint mathematically?



- Vector $\mathbf{u}'_c$ in left camera reference is $\mathbf{R}\mathbf{u}'_c$.
- Vectors $\mathbf{u}_c$, $\mathbf{t}$ and $\mathbf{R}\mathbf{u}'_c$ are coplanar.

Epipolar geometry

- How do we describe the epipolar constraint mathematically?



- Co planarity: $\mathbf{u}_c \cdot (\mathbf{t} \times \mathbf{R}\mathbf{u}'_c) = 0 \quad \longrightarrow \mathbf{u}_c^{\mathsf{T}}\mathbf{E}\mathbf{u}'_c = 0$
- $\mathbf{E} = \mathbf{T}_\times \mathbf{R}$ is called **essential matrix** (see second class for $\mathbf{T}_\times$).

Epipolar geometry

Essential matrix $\mathbf{E}$:

- Epipolar line l associated with a fixed $\mathbf{u}_c'$: $\mathbf{u}_c^\top(\mathbf{E}\mathbf{u}_c') = 0$.
- Epipolar line associated l' associated with $\mathbf{u}_c$: $(\mathbf{u}_c^\top\mathbf{E})\mathbf{u}_c' = 0$.

Epipolar geometry

Essential matrix $\mathbf{E}$:

- Epipolar line l associated with a fixed $\mathbf{u}_c'$: $\mathbf{u}_c^{\top}(\mathbf{E}\mathbf{u}_c') = 0$.
- Epipolar line associated l' associated with $\mathbf{u}_c$: $(\mathbf{u}_c^{\top}\mathbf{E})\mathbf{u}_c' = 0$.
- $\mathbf{E}\mathbf{e}' = \mathbf{0}$ and $\mathbf{e}^{\top}\mathbf{E} = \mathbf{0}^{\top}$
- $\mathrm{rank}(\mathbf{E}) = 2$, it is a singular matrix.

Epipolar geometry

Essential matrix $\mathbf{E}$:

- Epipolar line l associated with a fixed $\mathbf{u}'_c$: $\mathbf{u}_c^\top(\mathbf{E}\mathbf{u}'_c) = 0$.
- Epipolar line associated l' associated with $\mathbf{u}_c$: $(\mathbf{u}_c^\top\mathbf{E})\mathbf{u}'_c = 0$.

- $\mathbf{E}\mathbf{e}' = \mathbf{0}$ and $\mathbf{e}^\top\mathbf{E} = \mathbf{0}^\top$
- $\text{rank}(\mathbf{E}) = 2$, it is a singular matrix.
- $\mathbf{E}$ can be obtained from the extrinsic parameters of the cameras.
- $\mathbf{E}$ only relate points on the image planes, not on the image buffers.

Epipolar geometry

How do we relate points on the image buffers?

- Note that $\mathbf{u}_c = \mathbf{M}_{\text{int}}^{-1}\mathbf{u}_{\text{im}}$ and $\mathbf{u}'_c = (\mathbf{M}'_{\text{int}})^{-1}\mathbf{u}'_{\text{im}}$,
  where $\mathbf{M}_{\text{int}}$ and $\mathbf{M}'_{\text{int}}$ are the intrinsic matrices of both cameras.
- Using the epipolar constraint:
  $$\mathbf{u}_c^{\mathsf{T}}\mathbf{E}\mathbf{u}'_c = \mathbf{u}_{\text{im}}^{\mathsf{T}}(\mathbf{M}_{\text{int}}^{-1})^{\mathsf{T}}\mathbf{E}(\mathbf{M}'_{\text{int}})^{-1}\mathbf{u}'_{\text{im}} = \mathbf{u}_{\text{im}}^{\mathsf{T}}\mathbf{F}\mathbf{u}'_{\text{im}} = 0$$
- $\mathbf{F} = (\mathbf{M}_{\text{int}}^{-1})^{\mathsf{T}}\mathbf{E}(\mathbf{M}'_{\text{int}})^{-1}$ is called **fundamental matrix**.

Epipolar geometry

Fundamental matrix $\mathbf{F}$:

- Epipolar line l associated with a fixed $\mathbf{u}'_{\text{im}}$: $\mathbf{u}_{\text{im}}^{\mathsf{T}}(\mathbf{F}\mathbf{u}'_{\text{im}}) = 0$.
- Epipolar line associated l' associated with $\mathbf{u}_{\text{im}}$: $(\mathbf{u}_{\text{im}}^{\mathsf{T}}\mathbf{F})\mathbf{u}'_{\text{im}} = 0$.

Epipolar geometry

Fundamental matrix $\mathbf{F}$:

- Epipolar line l associated with a fixed $\mathbf{u}'_{\text{im}}$: $\mathbf{u}_{\text{im}}^{\mathsf{T}}(\mathbf{F}\mathbf{u}'_{\text{im}}) = 0$.
- Epipolar line associated l' associated with $\mathbf{u}_{\text{im}}$: $(\mathbf{u}_{\text{im}}^{\mathsf{T}}\mathbf{F})\mathbf{u}'_{\text{im}} = 0$.
- $\text{rank}(\mathbf{F}) = 2$, it is a singular matrix (why?).

Epipolar geometry

Fundamental matrix **F**:

- Epipolar line l associated with a fixed $\mathbf{u}'_{\text{im}}$: $\mathbf{u}^{\top}_{\text{im}}(\mathbf{F}\mathbf{u}'_{\text{im}}) = 0$.
- Epipolar line associated l' associated with $\mathbf{u}_{\text{im}}$: $(\mathbf{u}^{\top}_{\text{im}}\mathbf{F})\mathbf{u}'_{\text{im}} = 0$.
- $\text{rank}(\mathbf{F}) = 2$, it is a singular matrix (why?).
- **F** needs both the extrinsic and intrinsic parameters of the cameras.
- **F** has 7 free parameters.

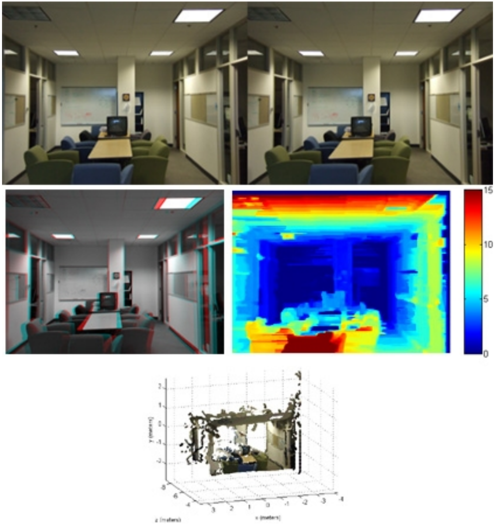
Epipolar geometry

Fundamental matrix $\mathbf{F}$:

- Epipolar line l associated with a fixed $\mathbf{u}'_{\text{im}}$: $\mathbf{u}_{\text{im}}^{\text{T}}(\mathbf{F}\mathbf{u}'_{\text{im}}) = 0$.
- Epipolar line associated l' associated with $\mathbf{u}_{\text{im}}$: $(\mathbf{u}_{\text{im}}^{\text{T}}\mathbf{F})\mathbf{u}'_{\text{im}} = 0$.
- $\text{rank}(\mathbf{F}) = 2$, it is a singular matrix (why?).
- $\mathbf{F}$ needs both the extrinsic and intrinsic parameters of the cameras.
- $\mathbf{F}$ has 7 free parameters.
- It is possible to estimate matrix $\mathbf{F}$ from a set of point correspondences with a modification of the DLT method.
- The estimation procedure is known as the **8-point algorithm** since you need at least 8 point correspondences to retrieve $\mathbf{F}$ (see bibliographic references).

General triangulation with calibrated cameras

3D reconstruction



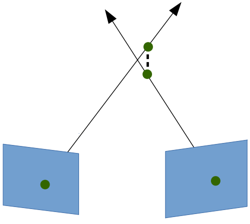
Source: `https://fr.mathworks.com/discovery/stereo-vision.html`

General triangulation with calibrated cameras

What do we do in the general case?

Many problems:

- Right camera may be not only translated but also rotated with respect to left camera.
- Intrinsic parameters may be different.
- There is noise on the point correspondences: **the rays never intersect in practice.**



General triangulation with calibrated cameras

What do we do in the general case?

General solution:

- Suppose that the two camera matrices are known:

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \quad \mathbf{M}' = \begin{bmatrix} m'_{11} & m'_{12} & m'_{13} & m'_{14} \\ m'_{21} & m'_{22} & m'_{23} & m'_{24} \\ m'_{31} & m'_{32} & m'_{33} & m'_{34} \end{bmatrix}$$

- **If there was no noise:** for each point correspondence

$$\mathbf{u}_{\text{im}} = \begin{bmatrix} x_{\text{im}} \\ y_{\text{im}} \\ 1 \end{bmatrix} \text{ and } \mathbf{u}'_{\text{im}} = \begin{bmatrix} x'_{\text{im}} \\ y'_{\text{im}} \\ 1 \end{bmatrix}$$

  we solve the 4 following equations for the 3D point coordinates x_w, y_w and z_w:

$$x_{\text{im}} = \frac{m_{11}x_w + m_{12}y_w + m_{13}z_w + m_{14}}{m_{31}x_w + m_{32}y_w + m_{33}z_w + m_{34}} \qquad y_{\text{im}} = \frac{m_{21}x_w + m_{22}y_w + m_{23}z_w + m_{24}}{m_{31}x_w + m_{32}y_w + m_{33}z_w + m_{34}}$$

$$x'_{\text{im}} = \frac{m'_{11}x_w + m'_{12}y_w + m'_{13}z_w + m_{14}}{m'_{31}x_w + m'_{32}y_w + m'_{33}z_w + m'_{34}} \qquad y'_{\text{im}} = \frac{m'_{21}x_w + m'_{22}y_w + m'_{23}z_w + m'_{24}}{m'_{31}x_w + m'_{32}y_w + m'_{33}z_w + m'_{34}}$$

General triangulation with calibrated cameras

What do we do in the general case?

General solution - Bundle adjustment:

- **But there is noise:** therefore we can only try to minimize the residuals $\Longrightarrow$ use least squares approach.

$$\text{minimize} \qquad (\varepsilon_x)^2 + (\varepsilon_y)^2 + (\varepsilon'_x)^2 + \left(\varepsilon'_y\right)^2$$

$$\text{with respect to} \qquad x_w,\, y_w,\, z_w$$

where the residuals are

$$\varepsilon_x = x_{\text{im}} - \frac{m_{11}x_w + m_{12}y_w + m_{13}z_w + m_{14}}{m_{31}x_w + m_{32}y_w + m_{33}z_w + m_{34}} \qquad \varepsilon_y = y_{\text{im}} - \frac{m_{21}x_w + m_{22}y_w + m_{23}z_w + m_{24}}{m_{31}x_w + m_{32}y_w + m_{33}z_w + m_{34}}$$

$$\varepsilon'_x = x'_{\text{im}} - \frac{m'_{11}x_w + m'_{12}y_w + m'_{13}z_w + m_{14}}{m'_{31}x_w + m'_{32}y_w + m'_{33}z_w + m'_{34}} \qquad \varepsilon'_y = y'_{\text{im}} - \frac{m'_{21}x_w + m'_{22}y_w + m'_{23}z_w + m'_{24}}{m'_{31}x_w + m'_{32}y_w + m'_{33}z_w + m'_{34}}$$

General triangulation with calibrated cameras

What do we do in the general case?

Linear approach:

- In homogeneous coordinates, we have two linear systems of equations:

$$\lambda \mathbf{u}_{\text{im}} = \mathbf{M}\mathbf{u}_w \quad \lambda' \mathbf{u}'_{\text{im}} = \mathbf{M}'\mathbf{u}_w$$

where λ and λ' are two unknown scaling factors.

General triangulation with calibrated cameras

What do we do in the general case?

Linear approach:

- In homogeneous coordinates, we have two linear systems of equations:

$$\lambda \mathbf{u}_{\text{im}} = \mathbf{M}\mathbf{u}_w \quad \lambda' \mathbf{u}'_{\text{im}} = \mathbf{M}'\mathbf{u}_w$$

  where λ and λ' are two unknown scaling factors.

- These systems of equation can be put together as follows:

$$\begin{bmatrix} \mathbf{M} & -\mathbf{u}_{\text{im}} & \mathbf{0} \\ \mathbf{M}' & \mathbf{0} & -\mathbf{u}'_{\text{im}} \end{bmatrix} \begin{bmatrix} \mathbf{u}_w \\ \lambda \\ \lambda' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0} \implies \mathbf{Cs} = \mathbf{0}$$

General triangulation with calibrated cameras

What do we do in the general case?

Linear approach:

- These systems of equation can be put together as follows:

$$\begin{bmatrix} \mathbf{M} & -\mathbf{u}_{\text{im}} & \mathbf{0} \\ \mathbf{M}' & \mathbf{0} & -\mathbf{u}'_{\text{im}} \end{bmatrix} \begin{bmatrix} \mathbf{u}_w \\ \lambda \\ \lambda' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0} \implies \mathbf{Cs} = \mathbf{0}$$

- Due to noise we use the optimization approach

$$\begin{aligned} &\text{minimize} && \|\mathbf{Cs}\|_2^2 \\ &\text{with respect to} && \mathbf{s} \\ &\text{subject to} && \|\mathbf{s}\|_2^2 = 1 \end{aligned}$$

whose solution is again given by the SVD.

General triangulation with calibrated cameras

What do we do in the general case?

Linear approach:

- Due to noise we use the optimization approach

$$\begin{aligned} &\text{minimize} && \|\mathbf{Cs}\|_2^2 \\ &\text{with respect to} && \mathbf{s} \\ &\text{subject to} && \|\mathbf{s}\|_2^2 = 1 \end{aligned}$$

  whose solution is again given by the SVD.

- The 3D point is then obtained from $\mathbf{u}_w$ in $\mathbf{s}$ by transforming it to Cartesian coordinates (dividing the first three coordinates by the last one).

Stereo 3D reconstruction using the linear approach

1. Construct fundamental matrix **F** from the parameters obtained from camera calibration (slides 40 and 44).
2. Find a correspondence on right image for each pixel on the left image using the epipolar lines defined by **F** and a matching criterion $\mathcal{C}$.
3. Construct matrix **C** for each pixel correspondence (slide 54).
4. Retrieve $\mathbf{s} = \mathbf{v}_{\text{min}}$ with SVD of **C** for each pixel.
5. Retrieve the 3D point for each pixel correspondence from **s**.

Conclusions

- Stereopsis is one out of many methods to recover the 3D structure of a scene.
- We can easily modify the linear approach to use more than two cameras (how?).
- Other methods exist for doing triangulation without complete camera information:
  - Euclidean reconstruction: known intrinsic parameters, an unknown scale is the price to pay.
  - Projective reconstruction: all parameters are unknown, an unknown projective transformation is the price to pay.

Conclusions

- In the nonlinear approach, we can easily introduce other vision modalities (structure light, texture, time-of-flight) we just need to add residual terms corresponding to each new modality.
  - **Kinect 1**: Camera 1 + Camera 2 + Structure light texture size + Structured light texture rotation.
  - **Kinect 2**: Kinect 1 + Time-of-flight
- Fundamental difficulties in stereo: complexity and robustness
  - Complexity : to recover the scene we may not need to reconstruct all pixels.
  - Robustness : we may pick special regions in the image that we known that they will be well-matched.

$\Longrightarrow$ Automatic feature detection approaches (SIFT, ORB, BRIEF)