Mapping Local Image Deformations into Depth

Stephen Benoit
Frank P. Ferrie

Centre for Intelligent Machines (CIM)
McGill University, Montreal.

Introduction: Range and Time to Collision

- Optical flow methods:
  - Range (distance) map depends on global motion.
  - Estimate, adds uncertainty to conclusion.
  - Segmentation of scene into different rigid bodies.
  - Time to collision undefined where optical flow is null.
  - Range (and time to collision) unknown scale factor.
- Direct estimation of dense TTC from image pairs.
  - Recognise image deformations, not optical flow.
  - Automatic segmentation of floor and ceiling.
  - Straightforward factoring of uncertainty.

Overview

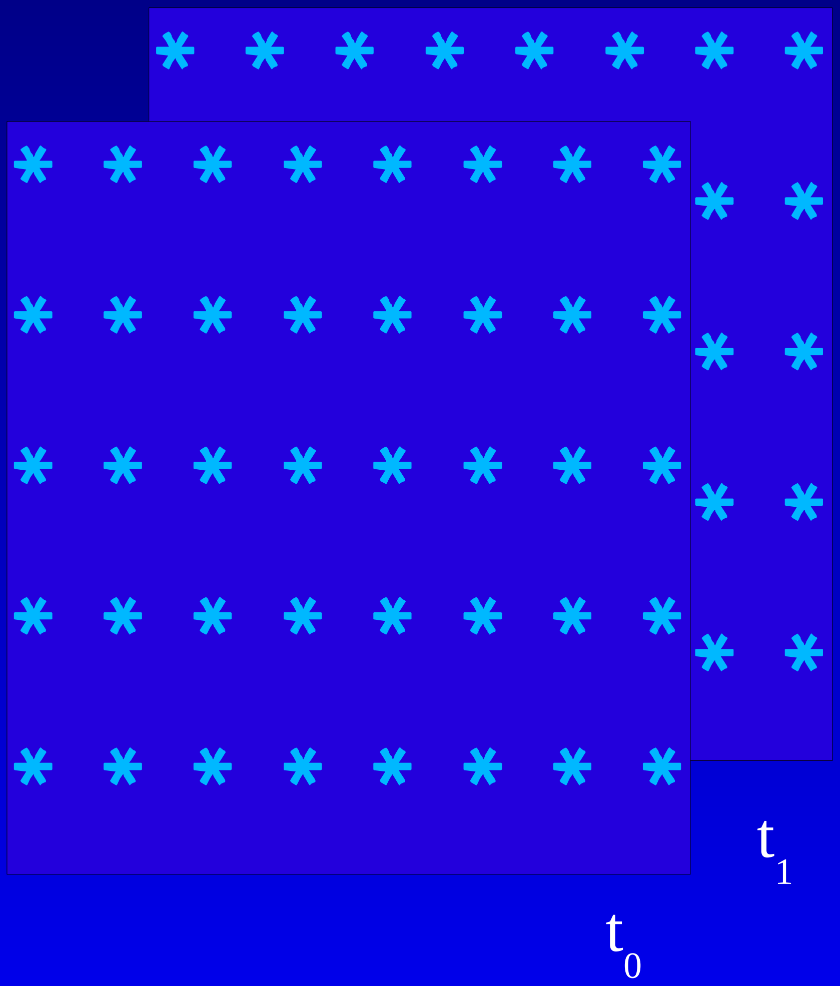
- 2 image frame structure-from-motion
- Model local image deformations
  - Image slits with camera model
  - Surface shape and motion model
- Recognise local image deformations
  - Represent as correspondence matrix H
  - Detectors from Singular Value Decomposition (SVD)
- Estimate Time To Collision (TTC)
- Experimental results

Ensemble of image slits

- Two instants in time
- Multiple orientations
- Distributed over image
- Eulerian (vs. Lagrangian)

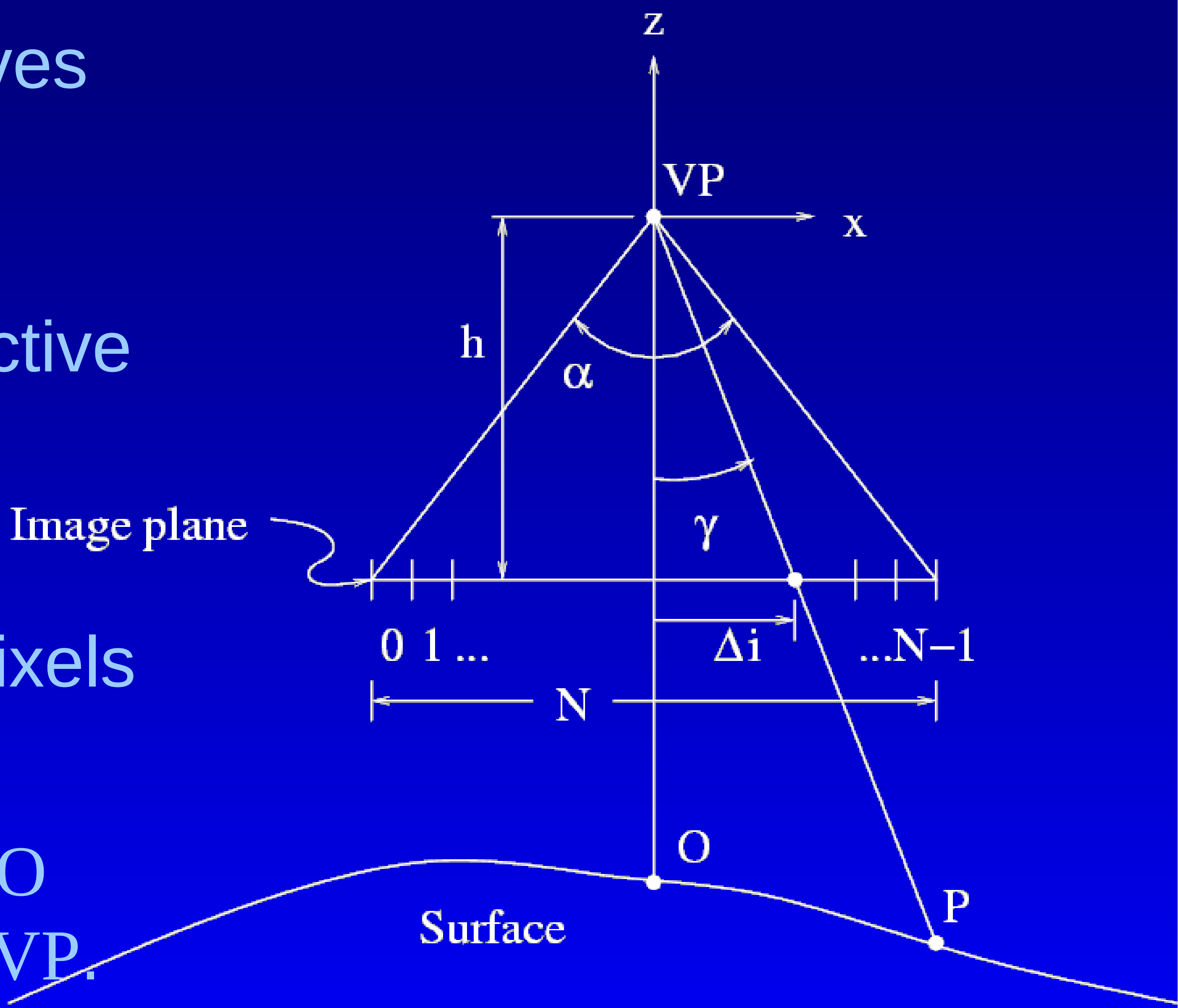
(x_i,y_i,t_0)

(x_i,y_i,t_1)



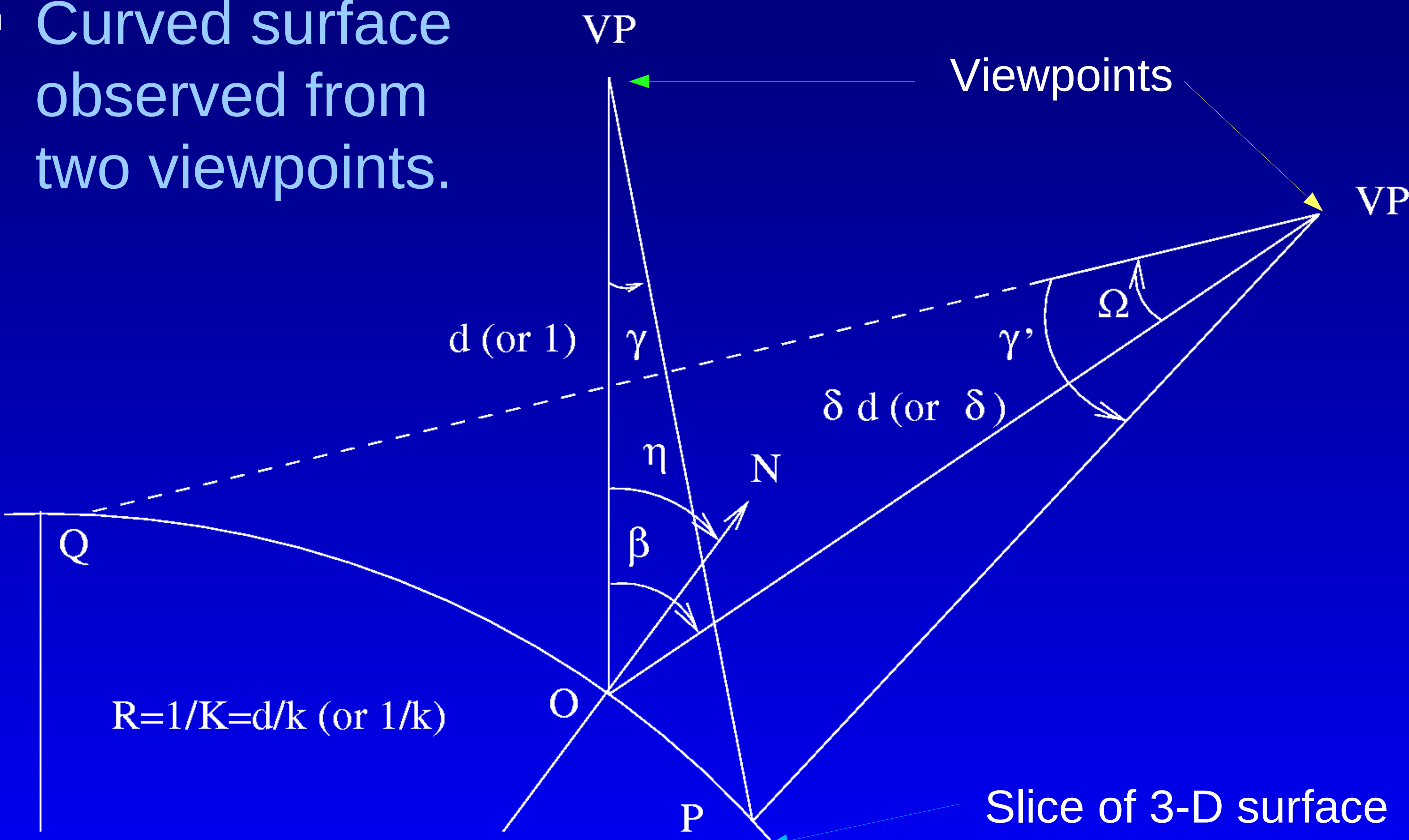
Camera model of slit

- Each slit observes small piece of surface.
- Pinhole perspective camera model.
- Aperture of N pixels in angle α.
- Fixate on point O from viewpoint VP.

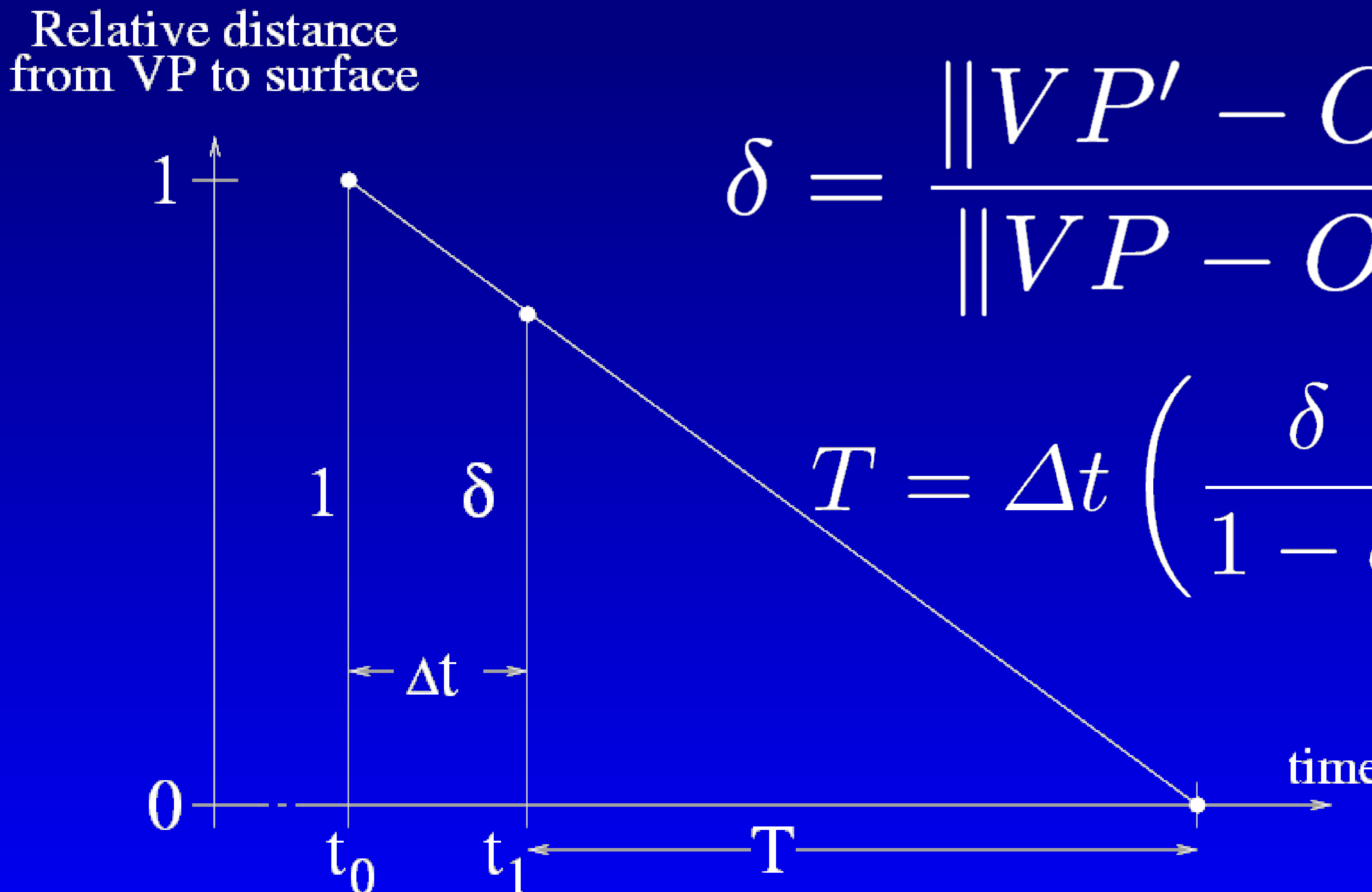


Surface model of slit

- Curved surface observed from two viewpoints.



Time To Collision (TTC) of slit

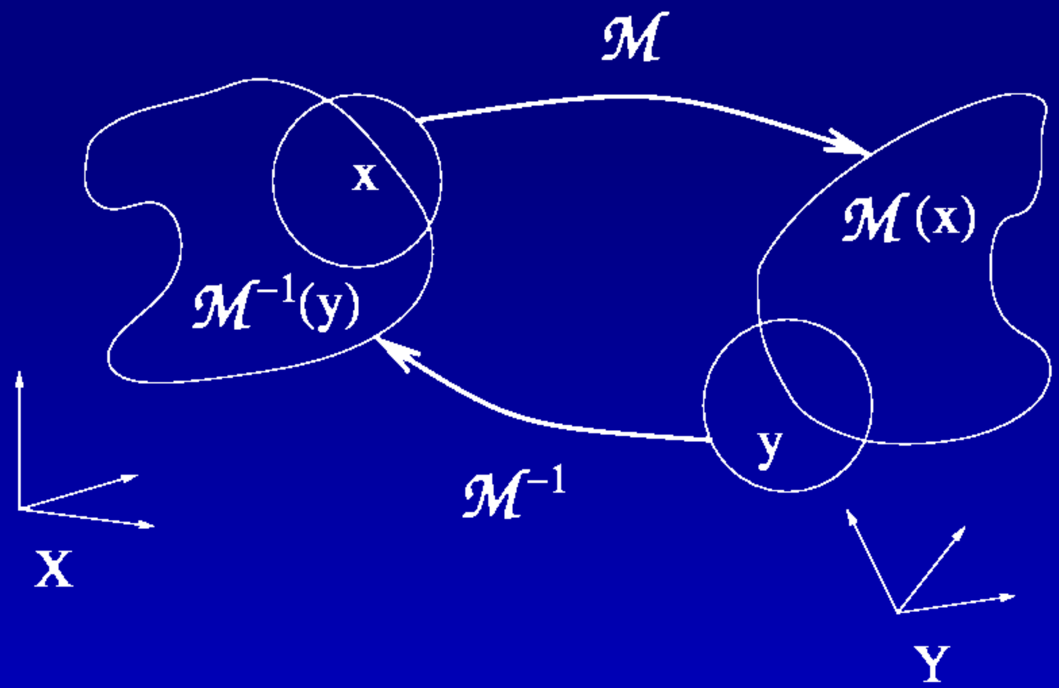


$$\delta = \frac{\|VP' - O\|}{\|VP - O\|}$$

$$T = \Delta t \left(\frac{\delta}{1 - \delta} \right)$$

Correspondence matrix $\mathbf{H}$

Encode mapping between each pixel in first view $\langle I \rangle_i$ and each pixel in second view $\langle I' \rangle_j$.



$$
\begin{aligned}
\langle \mathbf{H} \rangle_{ij} &\triangleq \mathrm{P}\left([\mathbf{X}]_i \overset{\mathcal{M}}{\Longleftrightarrow} [\mathbf{Y}]_j \right) \subset [0,1] \\
&= \frac{\left(\begin{array}{c} \mathcal{A}\left(\mathcal{M}([\mathbf{X}]_i) \cap [\mathbf{Y}]_j \right) \\ + \mathcal{A}\left([\mathbf{X}]_i \cap \mathcal{M}^{-1}([\mathbf{Y}]_j) \right) \end{array} \right)}{\mathcal{A}([\mathbf{X}]_i) + \mathcal{A}\left([\mathbf{Y}]_j \right)}
\end{aligned}
$$

Image slit normalization

- Remove influence of absolute image intensity.
- Normalize image vectors I, I'
- 0-mean, unit contrast.

$$\mu_{\mathcal{I}} \triangleq \frac{\sum_i \mathcal{I}_i + \sum_i \mathcal{I}'_i}{2N} ,$$

$$\Delta_{\mathcal{I}} \triangleq \frac{\max_i(|\mathcal{I}_i - \mu_{\mathcal{I}}|, |\mathcal{I}'_i - \mu_{\mathcal{I}}|)}{\mu_{\mathcal{I}}} \subset (0,1)$$

$$\tilde{\mathcal{I}} = \frac{\mathcal{I} - \mu_{\mathcal{I}}}{\mu_{\mathcal{I}} \Delta_{\mathcal{I}}} , \tilde{\mathcal{I}}' = \frac{\mathcal{I}' - \mu_{\mathcal{I}}}{\mu_{\mathcal{I}} \Delta_{\mathcal{I}}}$$

Role of SVD in synthesizing detectors

- U, V are optimal image-domain detectors for deformation represented by H.

$$\underset{N\times N}{\mathbf{H}} = \underset{N\times N}{\mathbf{U}}\,\underset{N\times N}{\mathbf{\Sigma}}\,\underset{N\times N}{\mathbf{V}^T}$$

$$\vec{w} = \mathbf{U}^T\tilde{\mathcal{I}} \qquad \vec{w}' = \mathbf{V}^T\tilde{\mathcal{I}}'$$

S. Benoit and F. P. Ferrie. Towards direct recovery of shape and motion parameters from image sequences. In Proceedings of ICCV, pages 1395 1402, Nice, France, October 2003.

Finding maximum likelihood H of slit

Find H_i that minimizes residual error r_i.

$$\hat{w}_i = \left[\mathbf{U}_{\mathbf{k}i}^{T}/\sqrt{2} \vdots \mathbf{V}_{\mathbf{k}i}^{T}/\sqrt{2} \right] \begin{bmatrix} \tilde{\mathcal{I}} \\ \cdots \\ \tilde{\mathcal{I}}' \end{bmatrix}$$

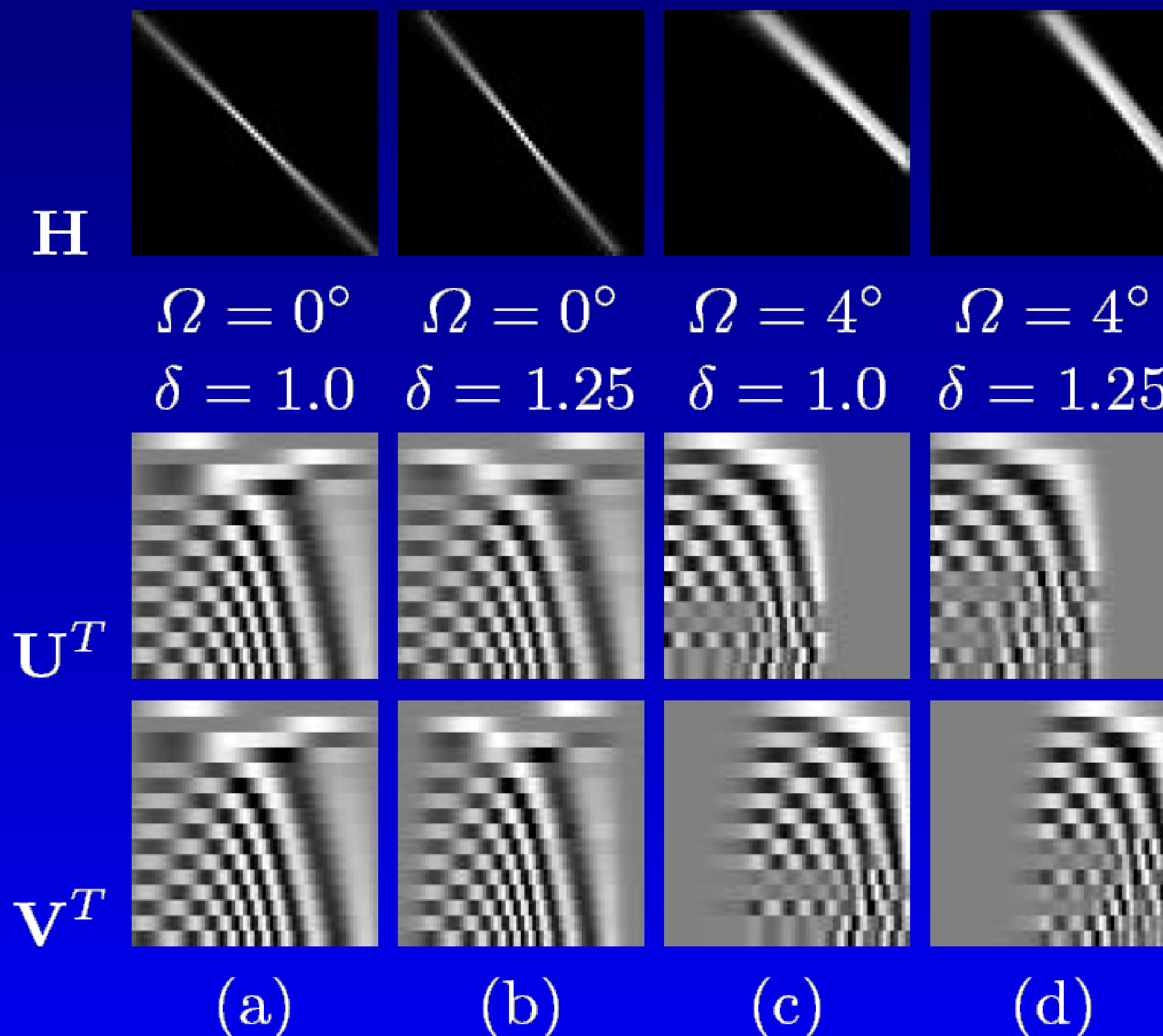
$$\boldsymbol{r}_i = \left(\begin{bmatrix} \tilde{\mathcal{I}} \\ \cdots \\ \tilde{\mathcal{I}}' \end{bmatrix} - \begin{bmatrix} \mathbf{U}_{\mathbf{k}i} \\ \cdots \\ \mathbf{V}_{\mathbf{k}i} \end{bmatrix} \hat{w}_i \right) \Bigg/ \left\| \begin{bmatrix} \tilde{\mathcal{I}} \\ \cdots \\ \tilde{\mathcal{I}}' \end{bmatrix} \right\|$$

$$\mathcal{L}\left(\mathbf{H}_i | \tilde{\mathcal{I}}, \tilde{\mathcal{I}}' \right) \triangleq e^{-\|\boldsymbol{r}_i\|} \subset (0,1]$$

$$h_i = \frac{-\sum_{i=1}^{n} \mathcal{L}\left(\mathbf{H}_i | \tilde{\mathcal{I}}, \tilde{\mathcal{I}}' \right) \log\left(\mathcal{L}\left(\mathbf{H}_i | \tilde{\mathcal{I}}, \tilde{\mathcal{I}}' \right) \right)}{\log(n)} \subset (0,1]$$

Example correspondence matrices

- Correspondence
  - Black: 0
  - White: 1
  - Row: position t_0
  - Col: position t_1
- SVD determines detectors for
  - U: view image I
  - V: view image I'



TTC using available information

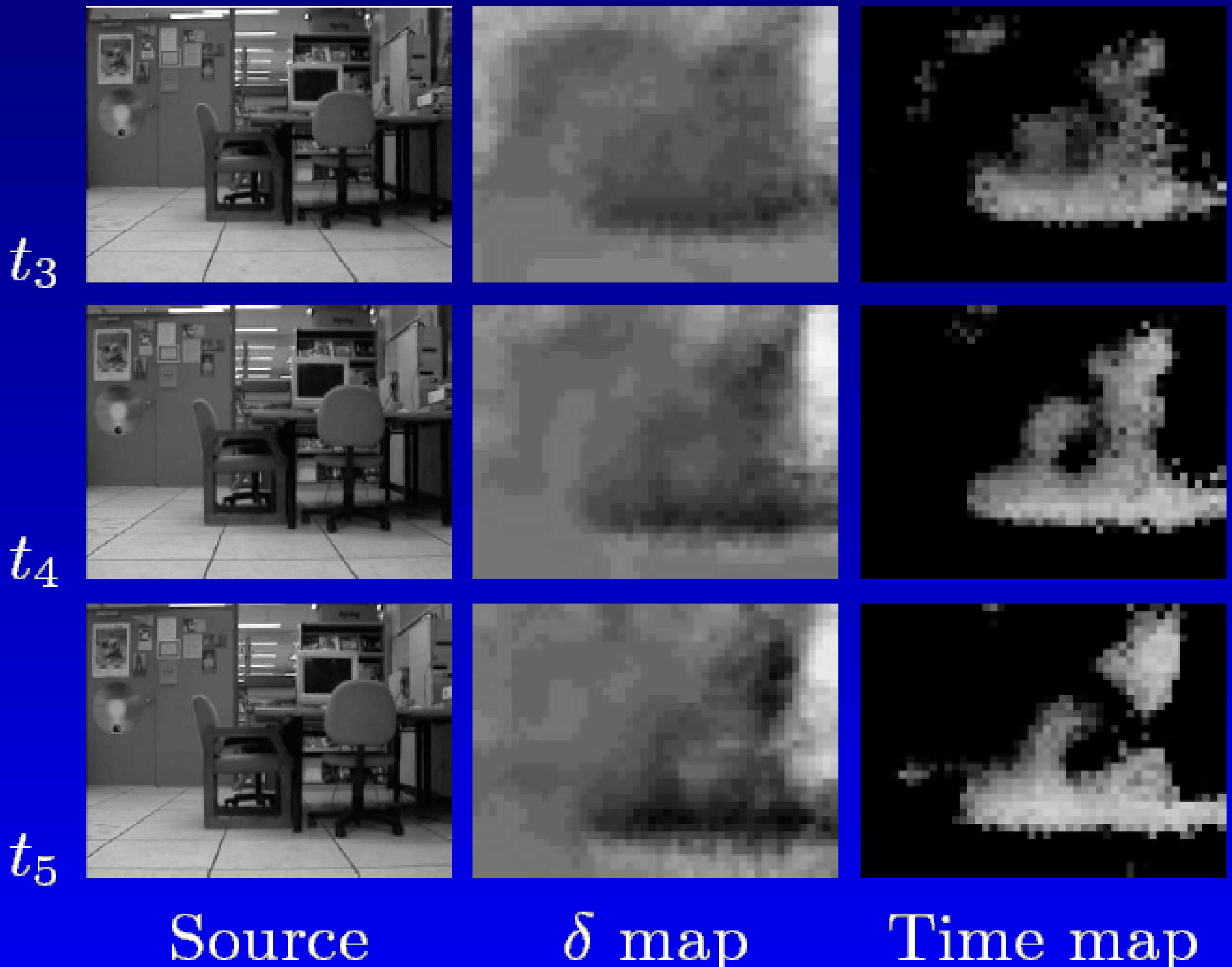
- Using forward and reverse estimates:

$$\tilde{T} = \frac{\Delta t}{2}\left(\frac{\delta}{1-\delta} + \frac{1}{\delta' - 1}\right)$$

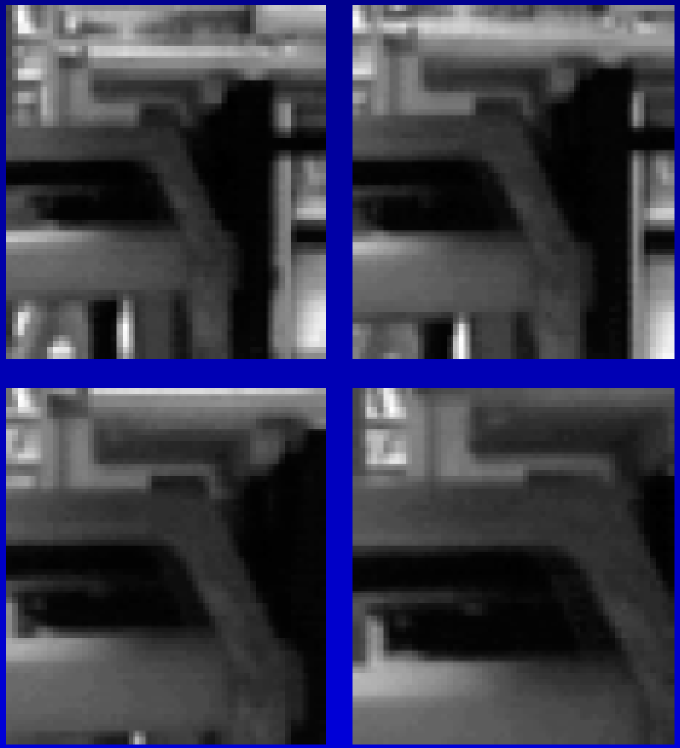
- Using forward and reverse estimates
  - Multiple orientations θ
  - Weight by confidence (1-*h*), where *h* is entropy.

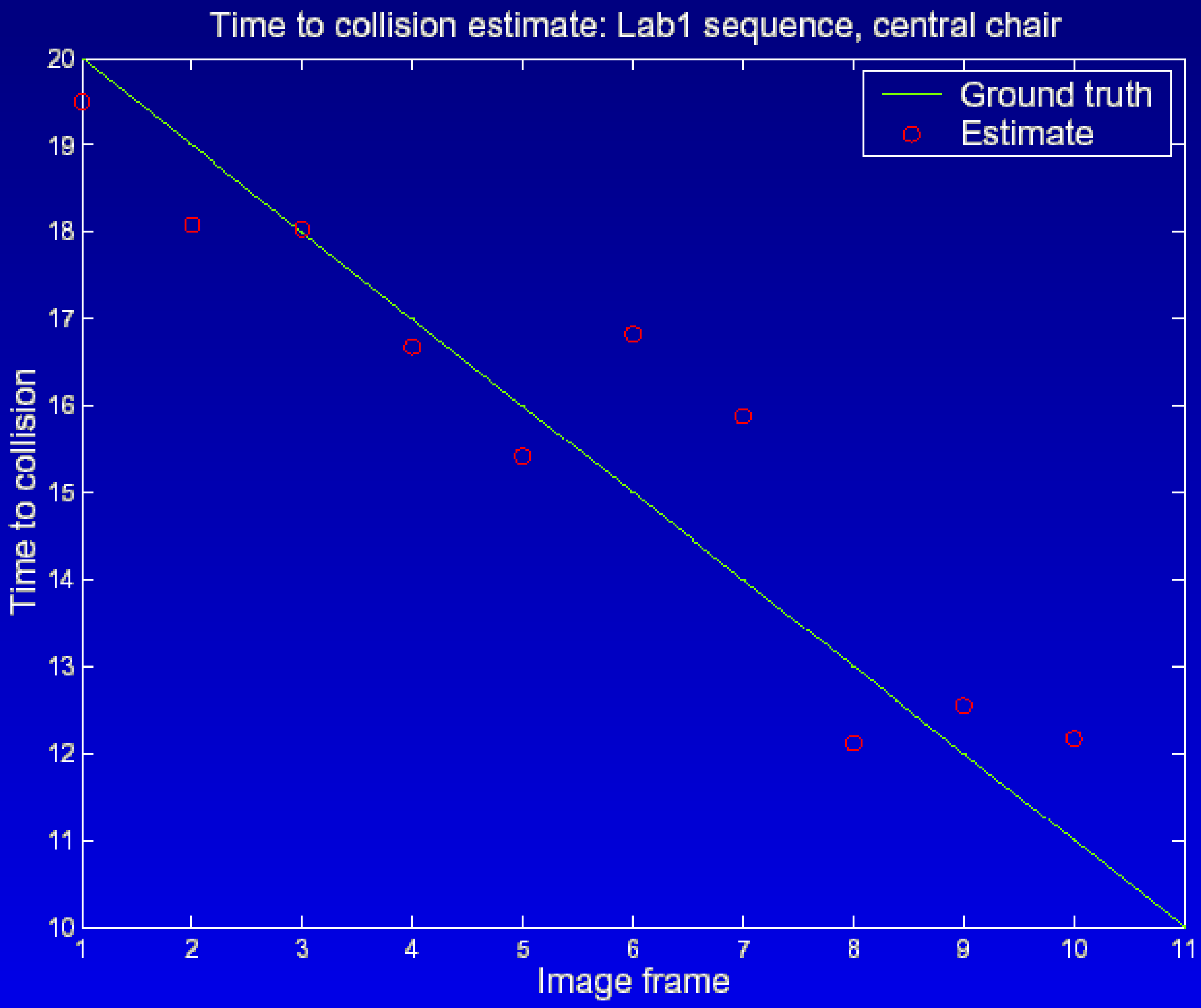
$$\tilde{T} = \frac{\Delta t}{2n}\sum_{\theta=1}^{n}\left(\frac{\delta_\theta h_\theta}{1-\delta_\theta h_\theta} + \frac{1}{\delta'_\theta h'_\theta - 1}\right)$$

Result: Lab1 delta and time maps



Result: Lab1 central chair





Result: Lab1 sequence

Future work

- Fast implementation in graphics card (OpenGL).
- Kalman filtering for more robust TTC estimates.
- Reduce search space for recognition of H.

Conclusion

- Automatic figure-ground separation.
- No optical flow or feature points required.
- TTC more “relevant” than range in egomotion.
- Simple enough model that uncertainties are:
  - easily combined from multiple measurements
  - to form more robust measurements.

Thanks and Acknowledgements.

- Centre for Intelligent Machines (CIM)

  McGill University, Montreal.
- Robert Sim
  - Mobile Robotics Lab (MRL) at CIM.
  - Mobile robot image sequences.