ICT RESEARCH INSTITUTE

Optimization of camera arrangement

بهینه سازی چیدمان دوربینها

Farzad Safaei

UNIVERSITY OF WOLLONGONG AUSTRALIA

Multi-Camera Systems

- Multi-camera systems are used widely
  - Coverage of the scene
  - Stereo or multiple cameras for depth estimation
  - Motion capture or 3D reconstruction
  - Video surveillance
  - Free viewpoint video or TV

استفاده از سیستمهای چند دوربینی در حال گسترش است

جهت پوشش کامل صحنه

جهت تخمین عمق

جهت بازسازی مدل سه بعدی صحنه

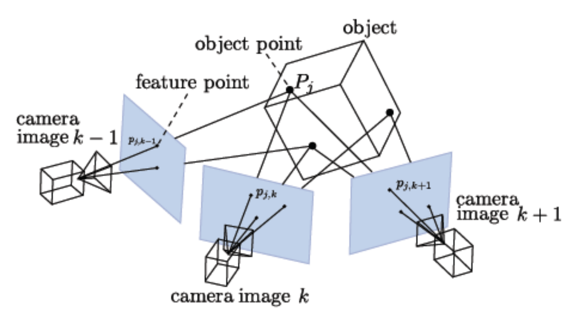
ترکیب تصاویرجهت خلق ویدئو با نقطه دید آزاد

RESEARCH INSTITUTE

Multi-Camera Systems

- Single camera, but acting like multiple cameras
  - 3D reconstruction from video, e.g. SfM (Structure from motion)
  - Simultaneous localization and mapping (SLAM)

استفاده از یک دوربین متحرک که بازدهی مشابه سیستم چند دوربینی دارد



Challenges

- It is still practically impossible to generate
  - Truly navigatable free viewpoint video
  - Photo realistic 3D modelling of scenes
- One of the fundamental Issues
  - Knowing little on how well a multi-camera system captures the *wanted* information

هنوز موارد زیر از نظر تکنیکی میسر نیست

ویدئوی که به معنای واقعی دارای نقطه دید آزاد باشد

بازسازی واقع گرایانه مدل سه بعدی صحنه

یک مشکل اساسی چیدمان دوربینها به نوعی است که بهترین داده ها را از صحنه اخذ کند

Existing research

- Some researchers have examined converged and diverged camera configurations*
- Seleting the most suitable images from multi-cameras has also been investigated %
- Iso-disparity surfaces studied by Pollefeys and Sinha&

*V.V. Petrov and K.A. Grebenyuk, "Improved stereoscopic imaging with converged camera configuration," in Saratov Fall Meeting 2006: Coherent Optics of Ordered and Random Media VII. International Society for Optics and Photonics, 2007, pp. 65 360T–65 360T.
* X. Song, Y. Wu, L. Yang, and Z. Liu, "Object position measuring based on adjustable dual-view camera," in Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on, July 2013, pp. 1–6.
* T. Yoshida and T. Fukao, "Dense 3d reconstruction using a rotational stereo camera," in System Integration (SII), 2011 IEEE/SICE International Symposium on, Dec 2011, pp. 985–990.

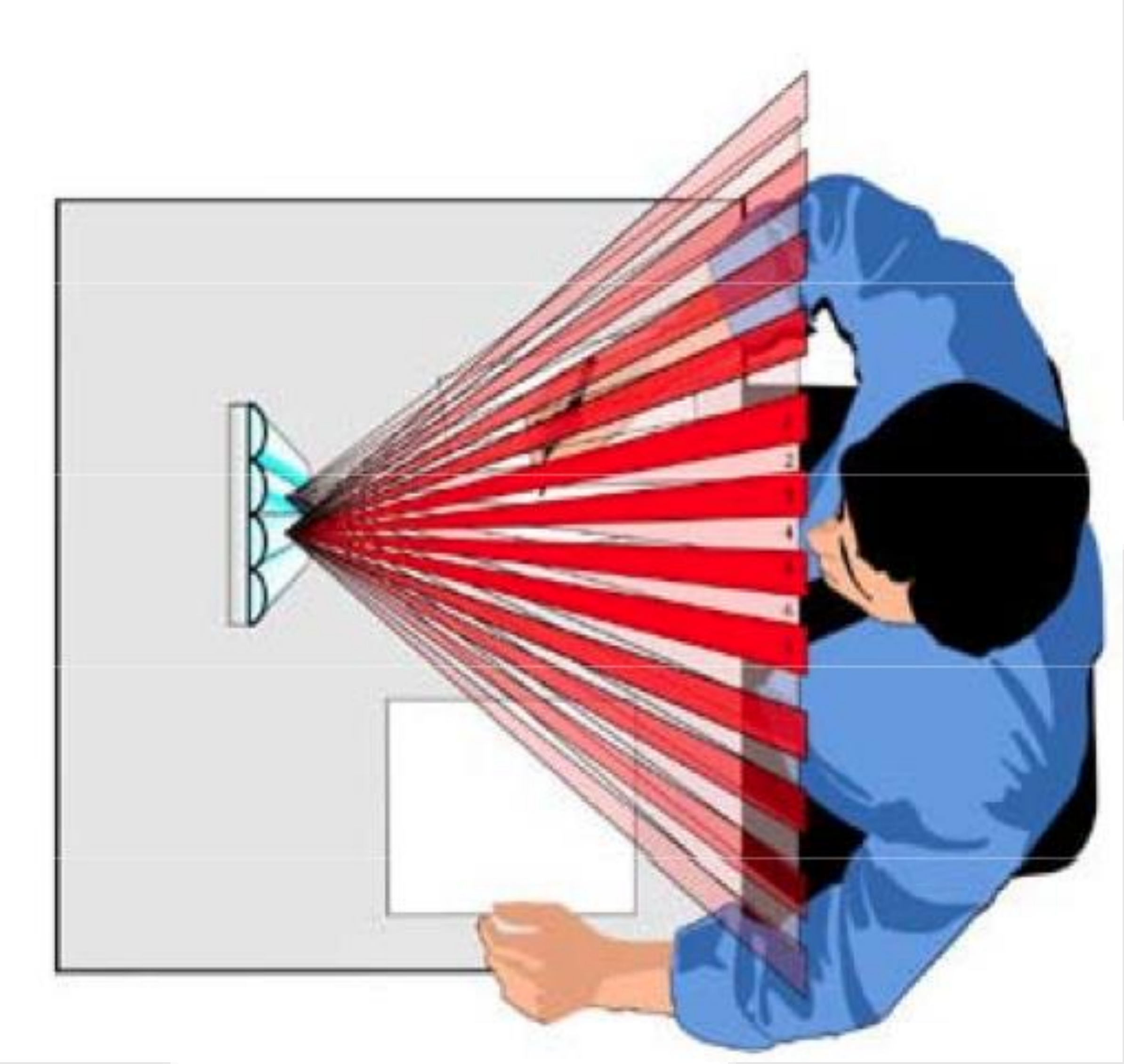
%A. Hornung, B. Zeng, and L. Kobbelt, "Image selection for improved multi-view stereo," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, June 2008, pp. 1–8.

& M. Pollefeys and S. Sinha, "Iso-disparity surfaces for general stereo configurations," in ECCV, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, vol. 3023, pp. 509–520.

Free Viewpoint TV/Video

- Small number of cameras – Unlimited viewpoints

Auto-Stereoscopic 3D Display



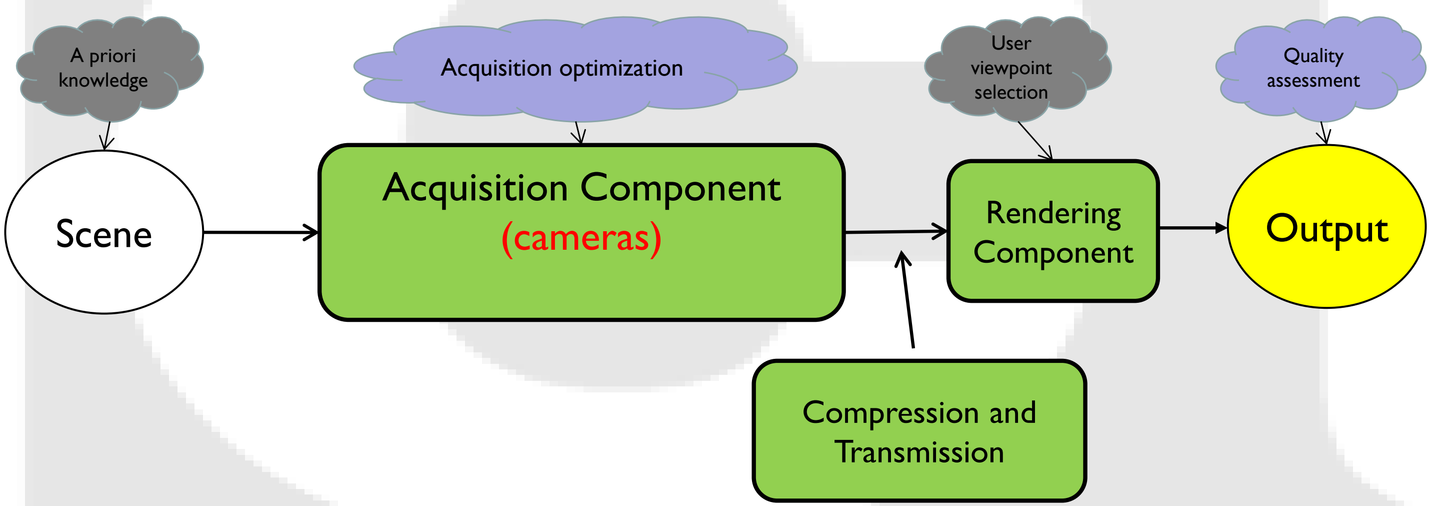
FVV System

- Light field acquisition (sampling),
- Light field compression/transmission
- Viewpoint rendering

گردآوری داده ها از صحنه

کم حجم سازی و ارسال داده ها

پردازش داده ها جهت خلق نقطه دید مورد نظر



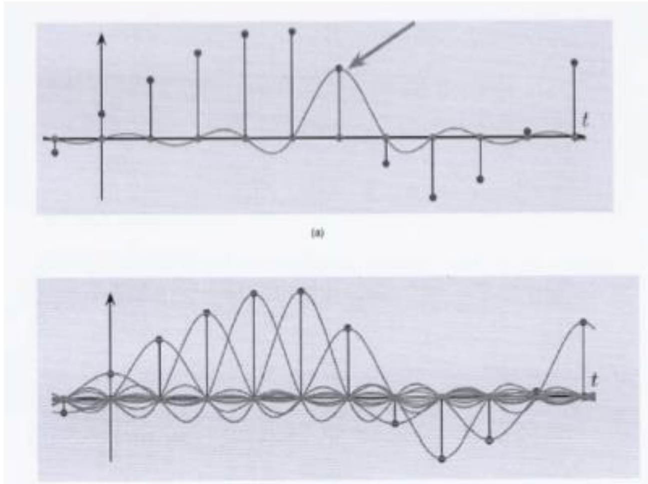
Light Field Acquisition

- How many cameras?
- How to arrange them?


Classical sampling

- Sampling at twice the max frequency
- Assumptions:
  - Band-limited signal
  - Regularly spaced samples
  - Linear interpolation of the samples
- This model results in an impractically large number of cameras*

$$x(t) = \sum_{n=-\infty}^{\infty} x[n]\mathrm{sinc}\left(\frac{t - nT_s}{T_s}\right)$$



RESEARCH INSTITUTE

* e.g. C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 1038-1050, 2003.

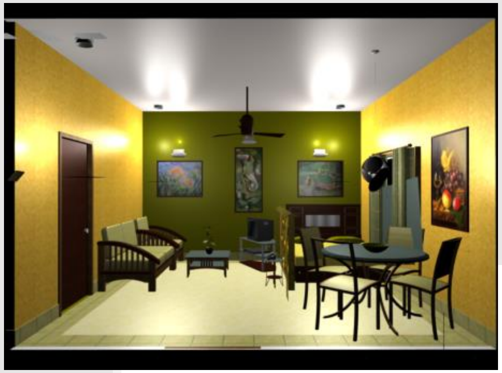
Adaptive sampling

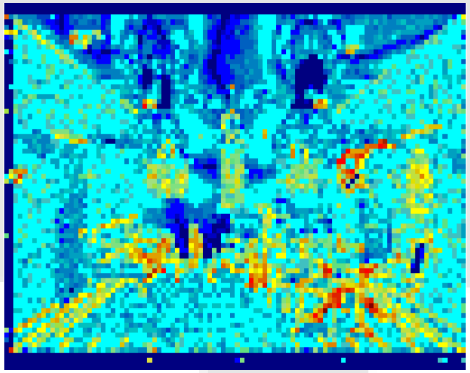
- Frequently used in computer graphics
  - Basic idea: adaptively distribute more samples on pixels with large errors for rendering
  - Many approaches: [Mitchell, 1987], [Bala et al., 2003], [Rousselle, et al., 2011]

ICT RESEARCH INSTITUTE

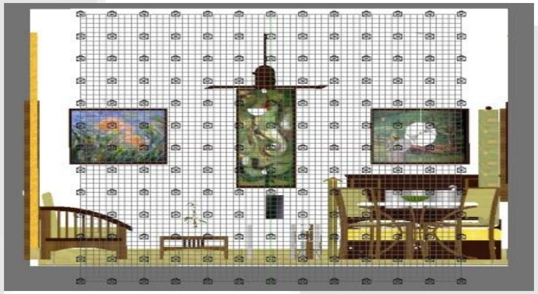
Light Field Acquisition

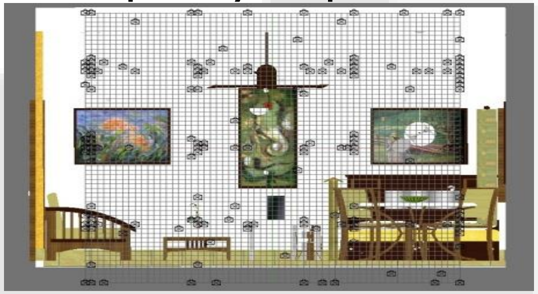
- Uniform sampling (or camera grid) is not optimal





Complexity map



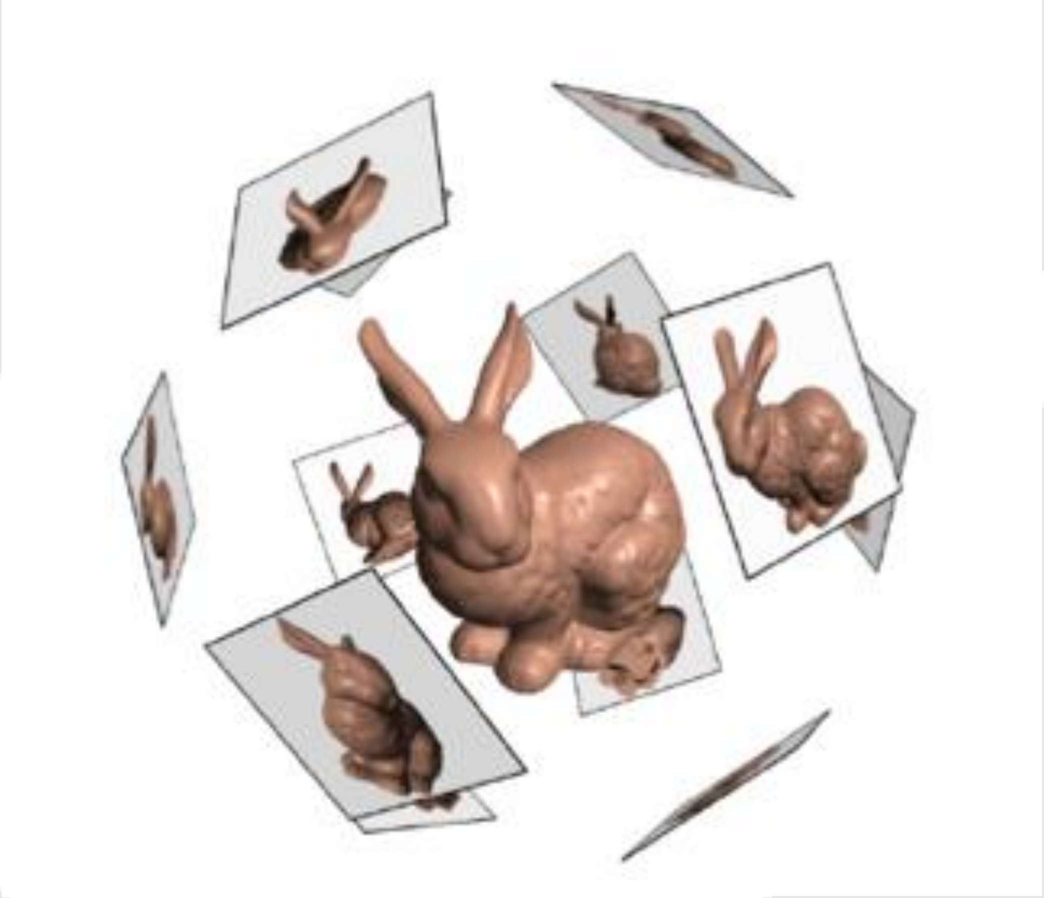


Optimum camera arrangement

H. Shidanshidi, F. Safaei, A. Zamani-Farahani, and L. Wanqing, "Non-uniform sampling of plenoptic signal based on the scene complexity variations for a free viewpoint video system," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, 2013, pp. 3147-3151.

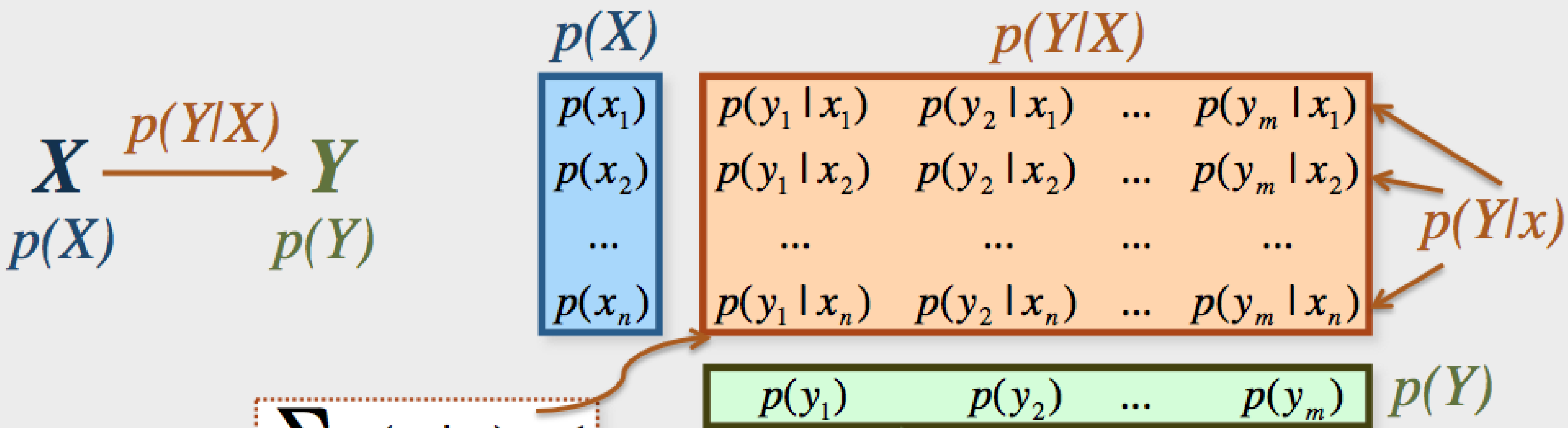
Viewpoint selection problem:

- Which views are the most informative?



Information theory

- Communication channel $X \longrightarrow Y$:



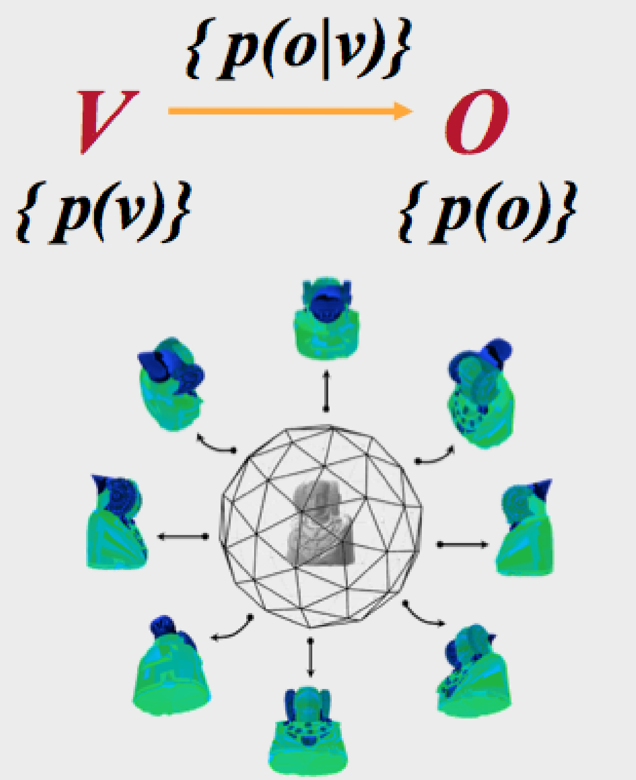
$$\sum_{x \in X} p(y \mid x) = 1$$

$$p(y) = \sum_{x \in X} p(x) p(y \mid x)$$

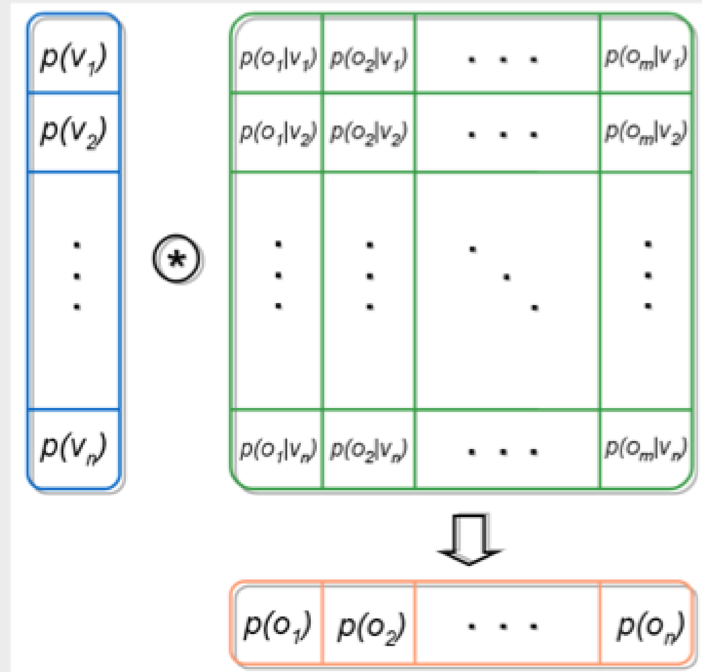
Bayes' rule $p(x,y) = p(x)p(y \mid x) = p(y)p(x \mid y)$

Viewpoint selection

- Model as an information channel [sbert et al 2006]



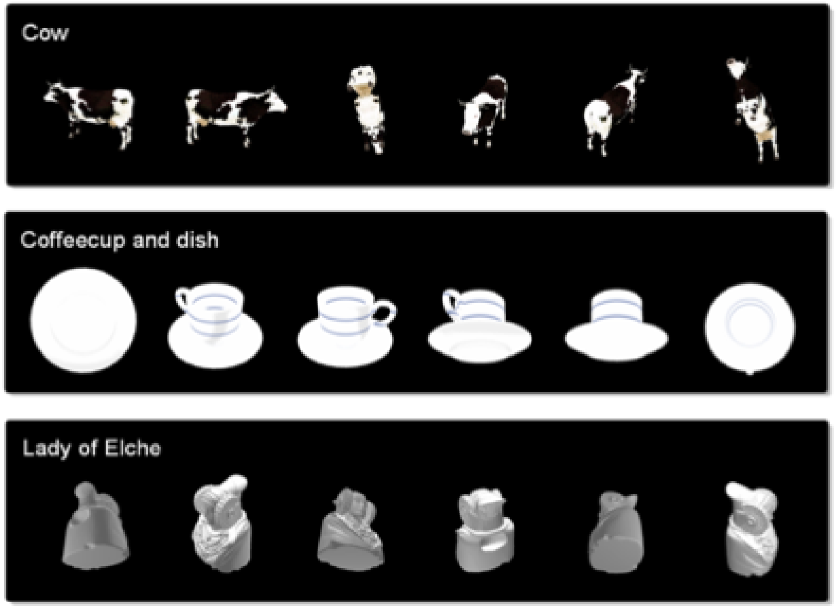
This framework is based on geometric characteristics

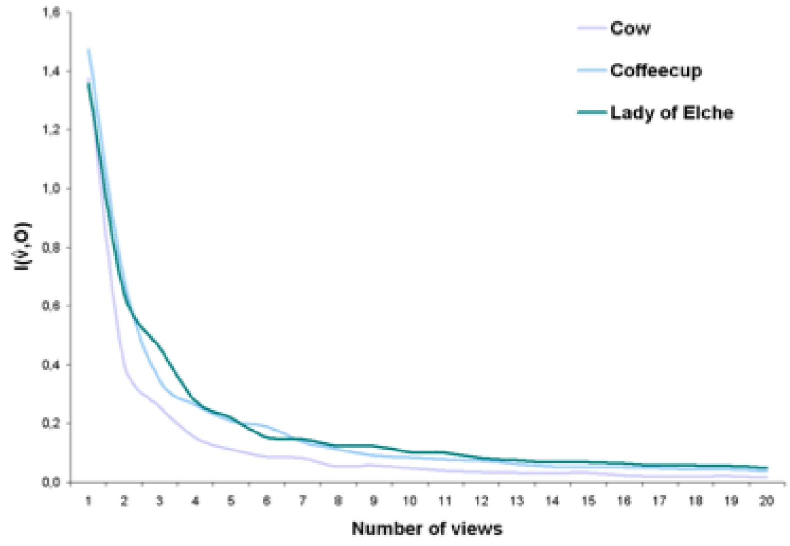


$$p(o) = \sum_{v \in \mathcal{V}} p(v) p(o|v)$$

Viewpoint selection

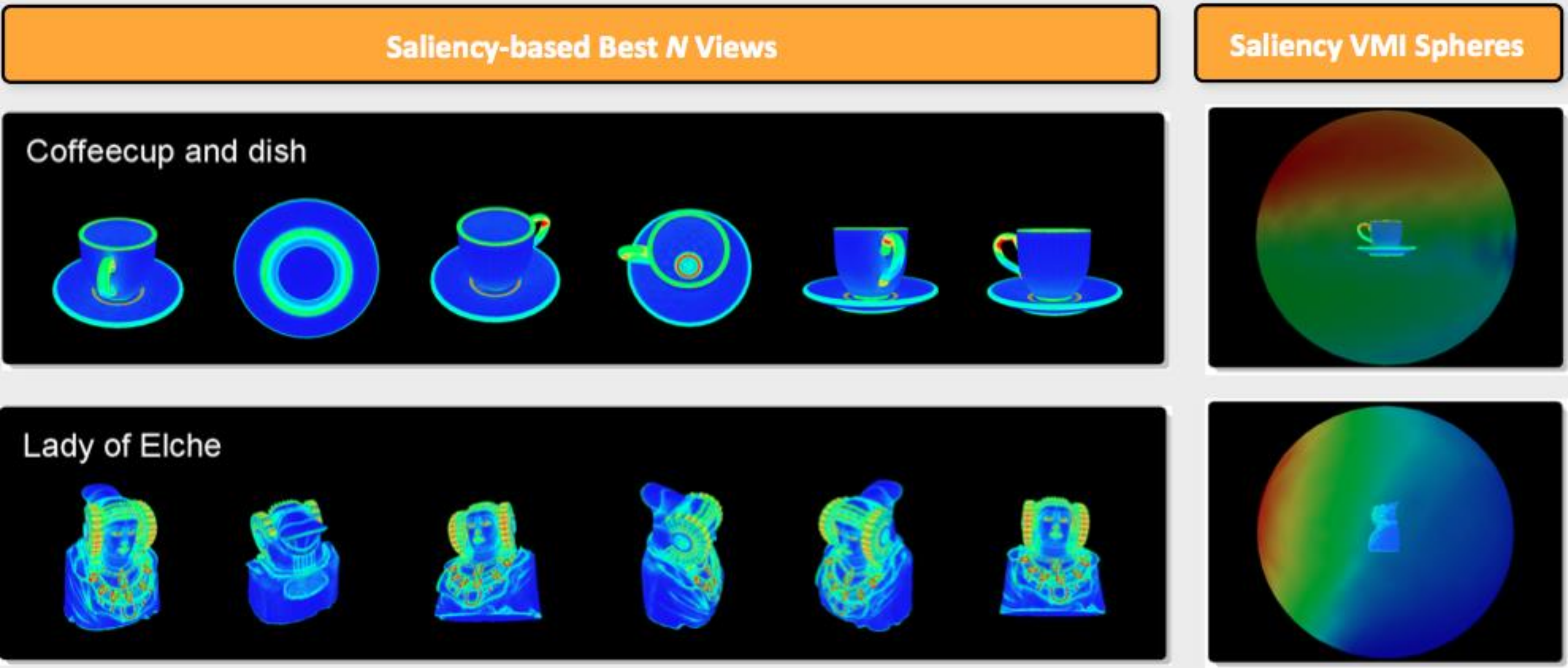
- Select those that have min mutual information





Viewpoint selection

- Similar approaches for saliency and importance



Mathematical representation of cameras

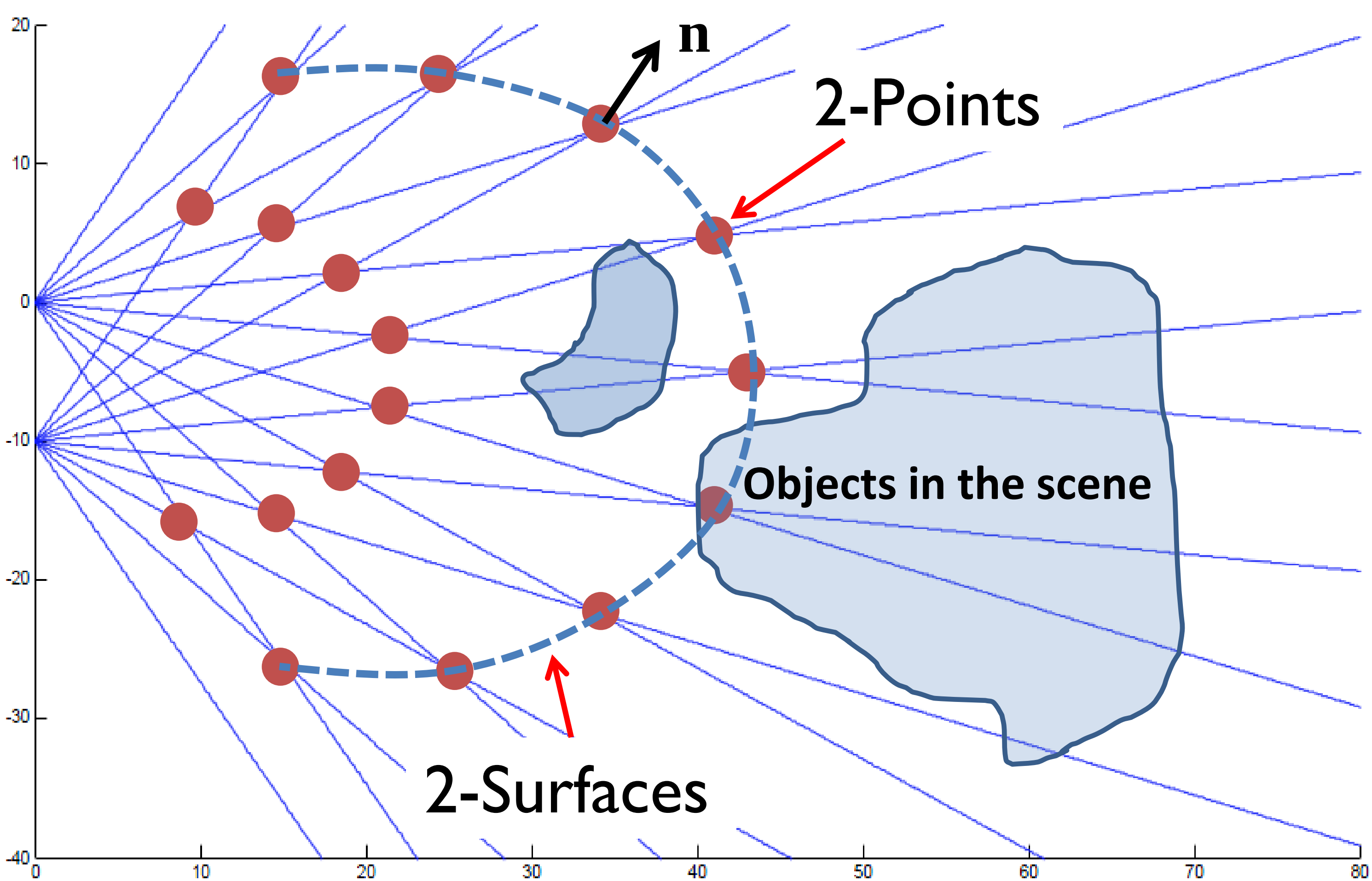
- We recently introduced the concept of Correspondence Field (CF)
- CF determines the spatial relationship between the cameras and the objects in the scene

Correspondence field

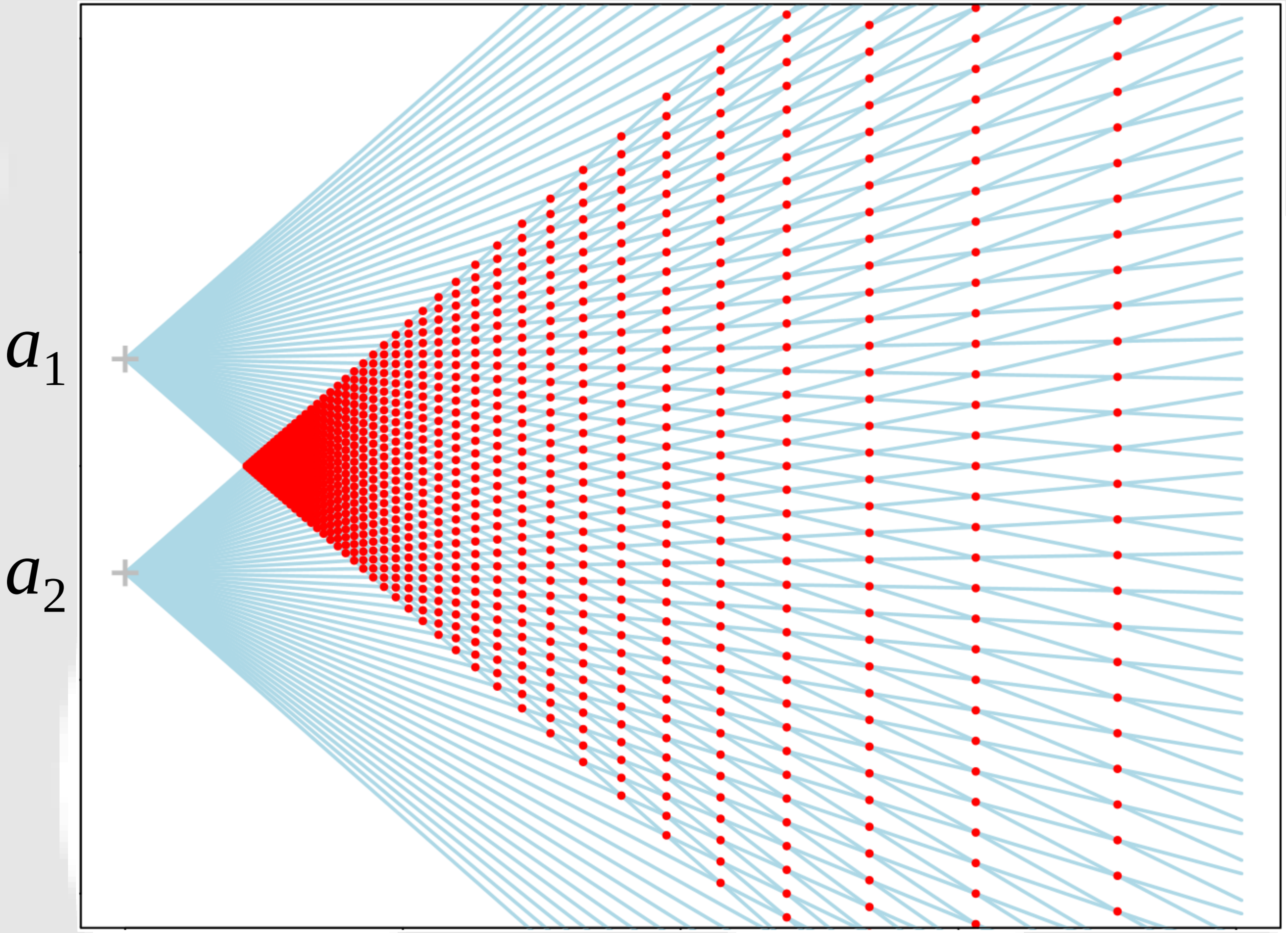
از نظر ریاضی تعیین کننده رابطه بین دوربینها و اشیا موجود در صحنه است

RESEARCH INSTITUTE

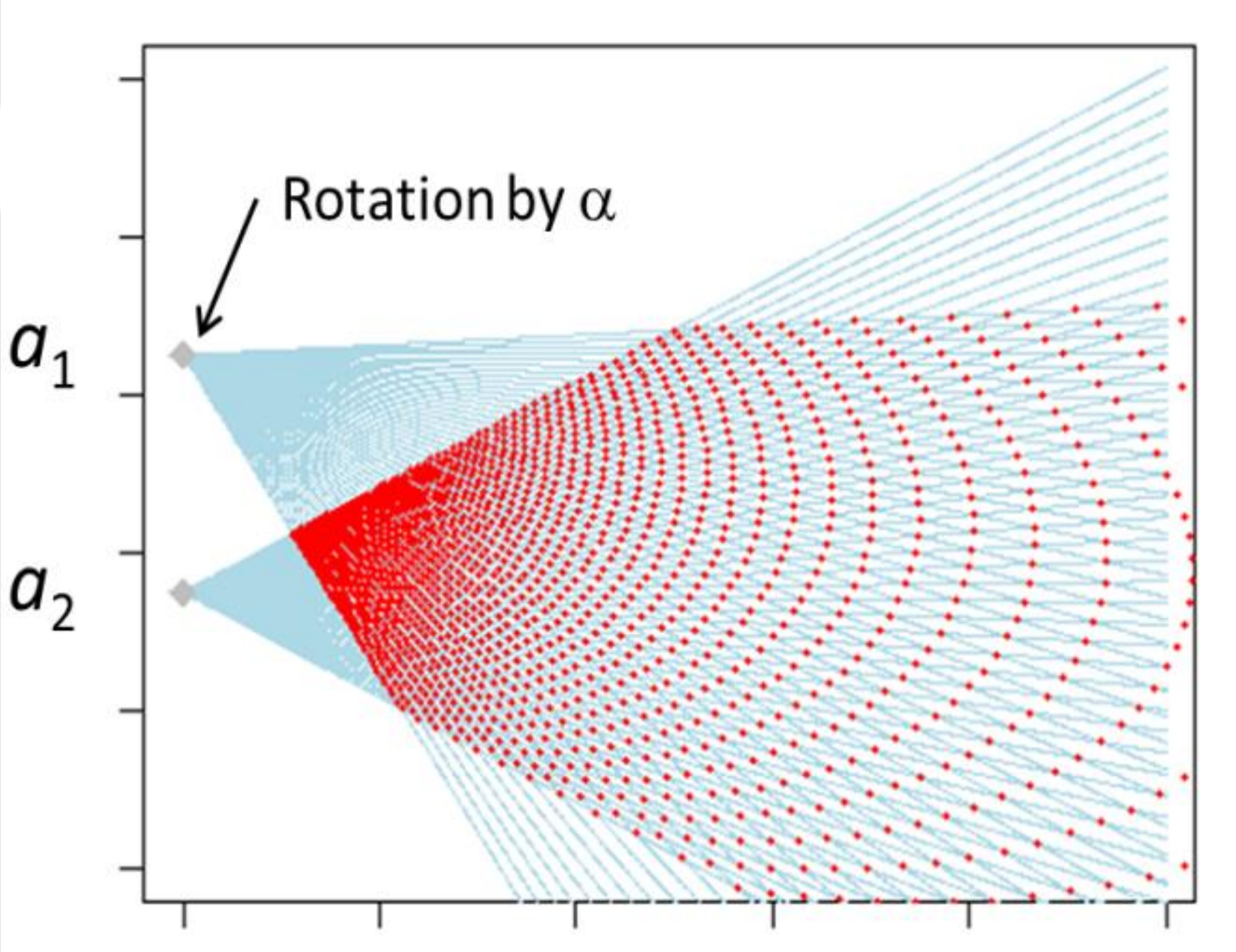
Correspondence field



Correspondence Field

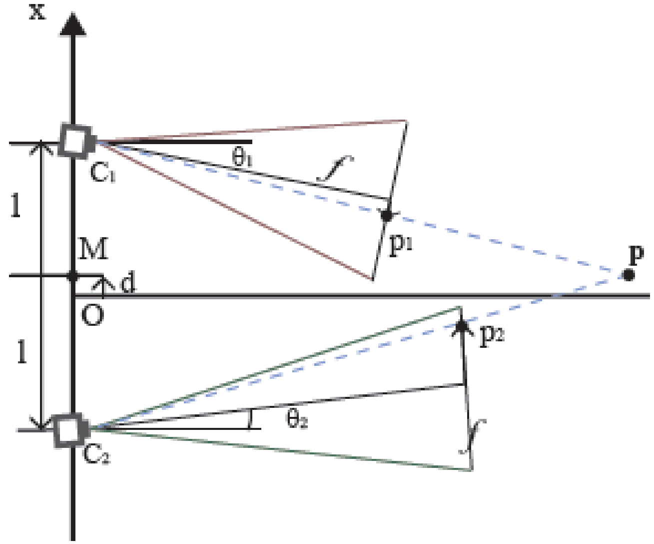


Correspondence Field



Correspondence Field

- Analytic form derived for 2-Surfaces



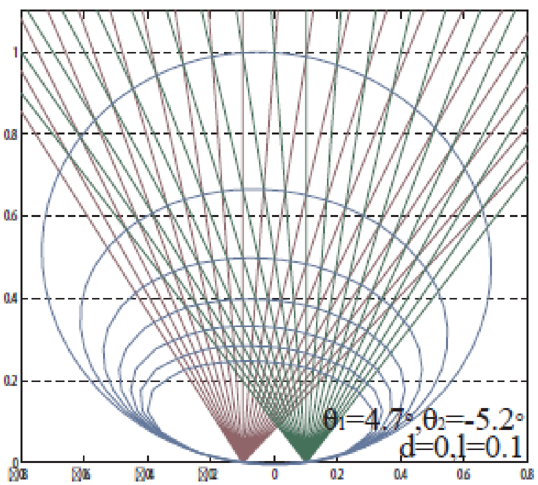
$$A(x-d)^2+2B(x-d)z+Cz^2+2E(x-d)+2Fz+G=0 \quad (4)$$

where

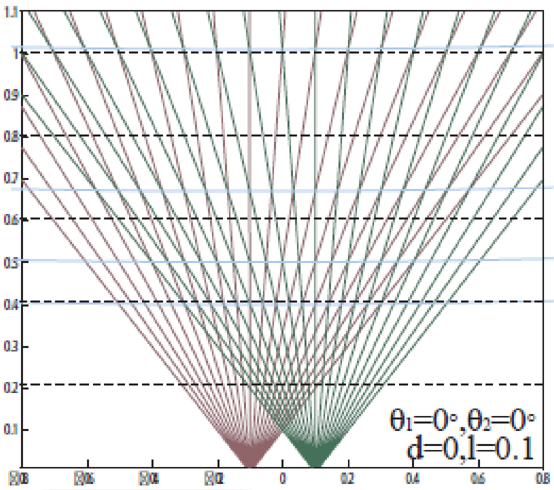
$$A = \sin(\theta_1-\theta_2)-\lambda\sin\theta_1\sin\theta_2 \quad (5)$$
$$B = \frac{\lambda}{2}\sin(\theta_1+\theta_2) \quad (6)$$
$$C = \sin(\theta_1-\theta_2)-\lambda\cos\theta_1\cos\theta_2 \quad (7)$$
$$E = 0 \quad (8)$$
$$F = \frac{l}{2}(\lambda\sin(\theta_1-\theta_2)+2\cos(\theta_1-\theta_2)) \quad (9)$$
$$G = l^2(\lambda\sin\theta_1\sin\theta_2-\sin(\theta_1-\theta_2)) \quad (10)$$

$$\lambda = k/f$$

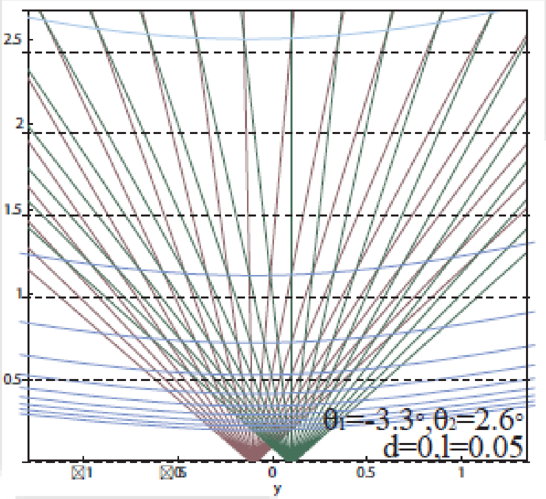
2-surfaces are conic curves



Converged



Parallel



Diverged

Important measures

- 2-point density: This represents the sampling density at each location of space

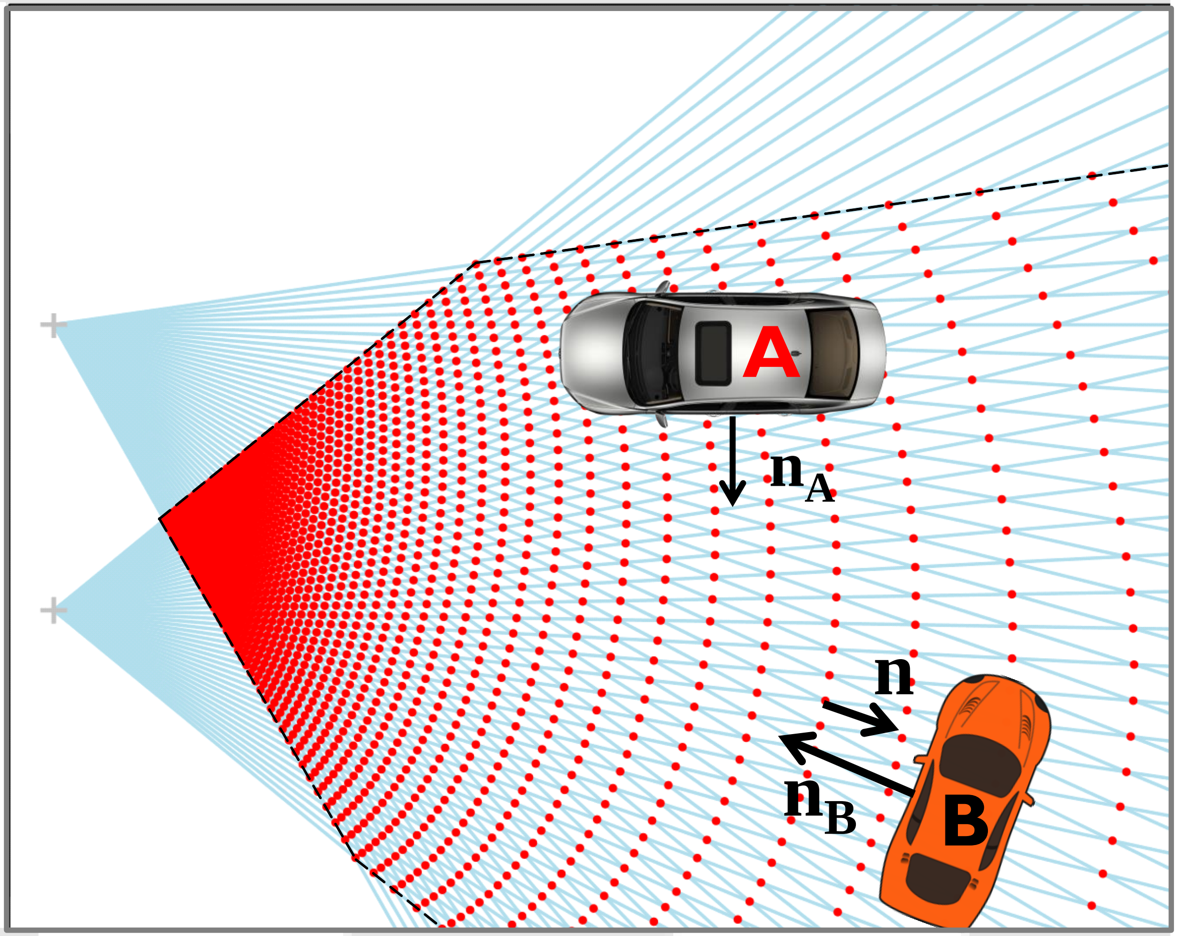
2-point density
نشان دهنده تراکم داده هایی است
که از هر نقطه صحنه اخذ میشود.

- 2-surface direction: This represents the orientation of constant depth surface

2-Surface Direction
نشان دهنده جهت سطوحی است
که از منظر دوربینها عمق ثابت
دارند

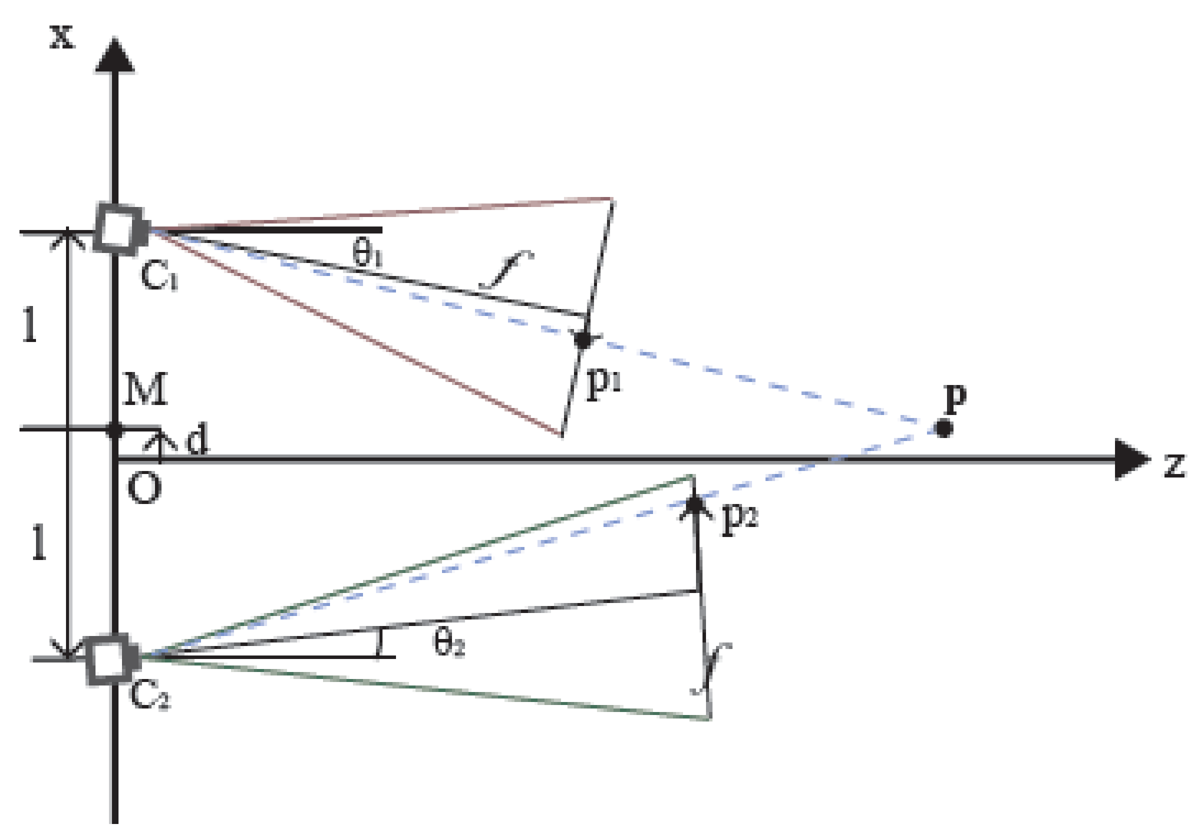
RESEARCH INSTITUTE

2-Surface Direction



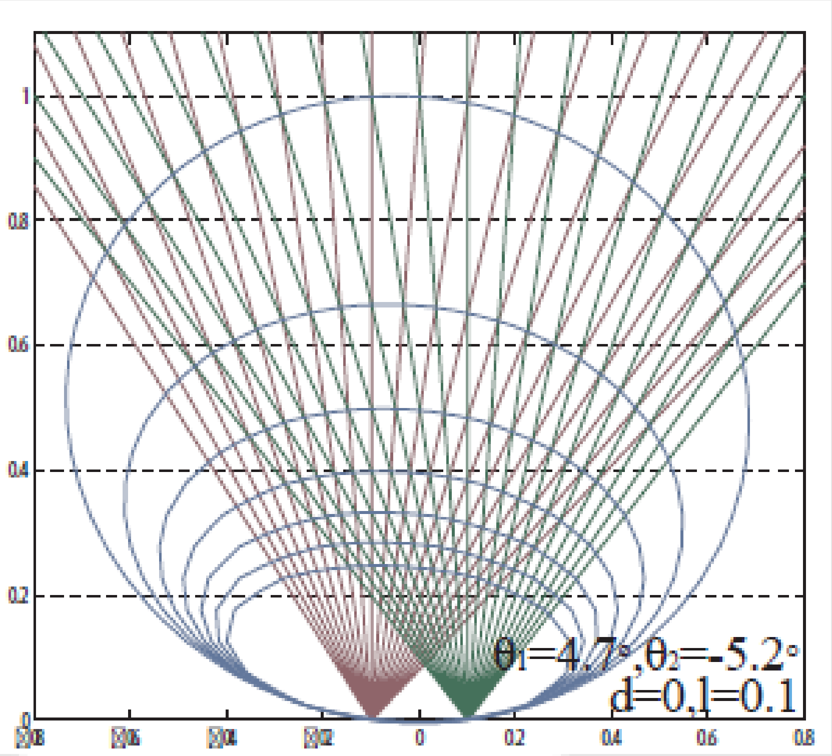
Optimisation approach for depth estimation using two cameras

- Let $\mathbf{a} = (\theta_1, \theta_2, l, d)$ represent the arrangement of two cameras



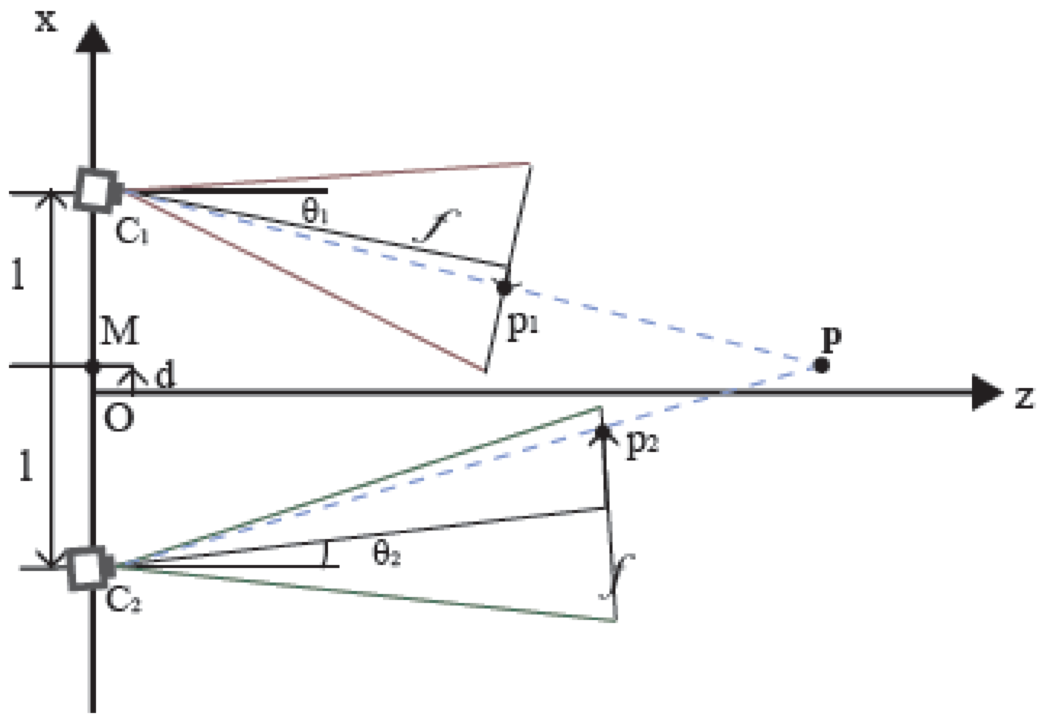
Optimisation approach for depth estimation

- Let F represent the CF of two cameras
- Then $F(k, \mathbf{a})$ is the 2-surface at disparity k



Optimisation approach for depth estimation

- Consider point p somewhere in the scene with the position vector $\mathbf{r}$
- The gradient of F at this point is $\nabla F(\mathbf{r}, \mathbf{a})$



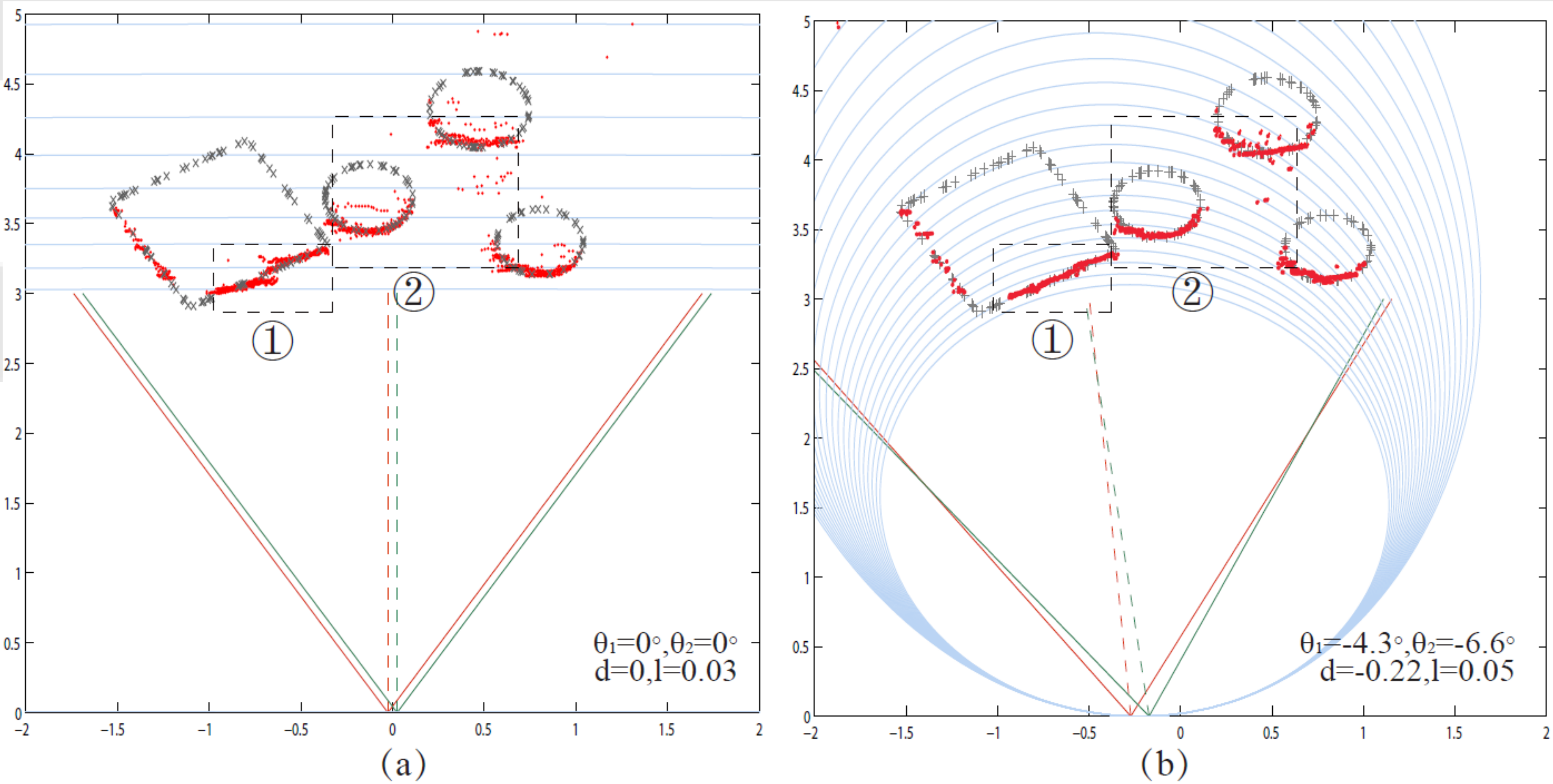
Optimisation approach for depth estimation

- The sampling density at this point is the magnitude of the gradient $||\nabla F(\mathbf{r}, \mathbf{a})||$
- For a small region Ω in the scene space, one possible optimisation is to maximise the sum of densities in that region

$$\underset{\mathbf{a}}{\operatorname{argmax}} \int_{\Omega} ||\nabla F(\mathbf{r}, \mathbf{a})|| \, d\mathbf{r} \qquad (1)$$

RESEARCH INSTITUTE

Example: CF density optimisation



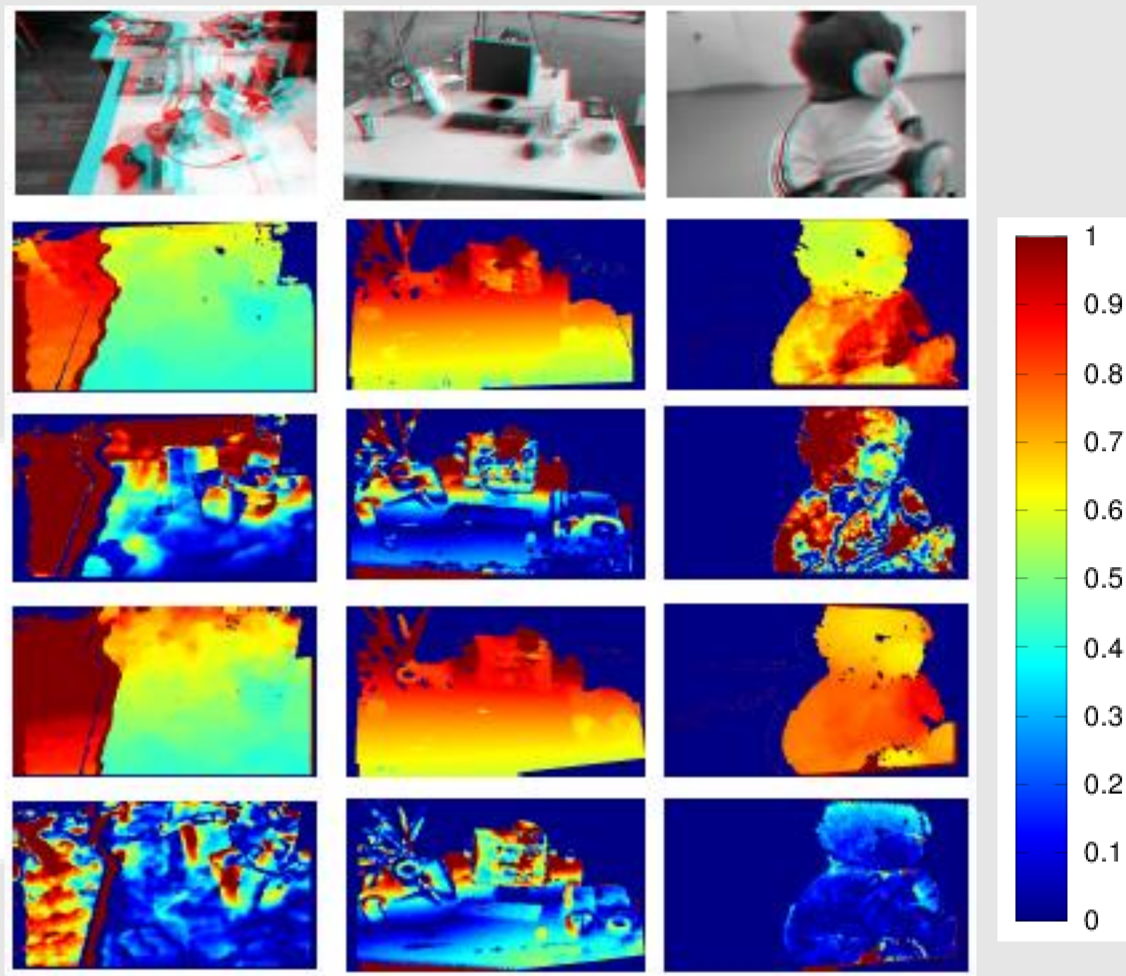
Optimisation approach for depth estimation

- The CF direction at this point is the direction of the gradient
- Assume there are object surfaces in the region Ω with their normals being $\mathbf{n_r}$
- One possible optimisation is to align the CF surfaces as best as possible with these surfaces

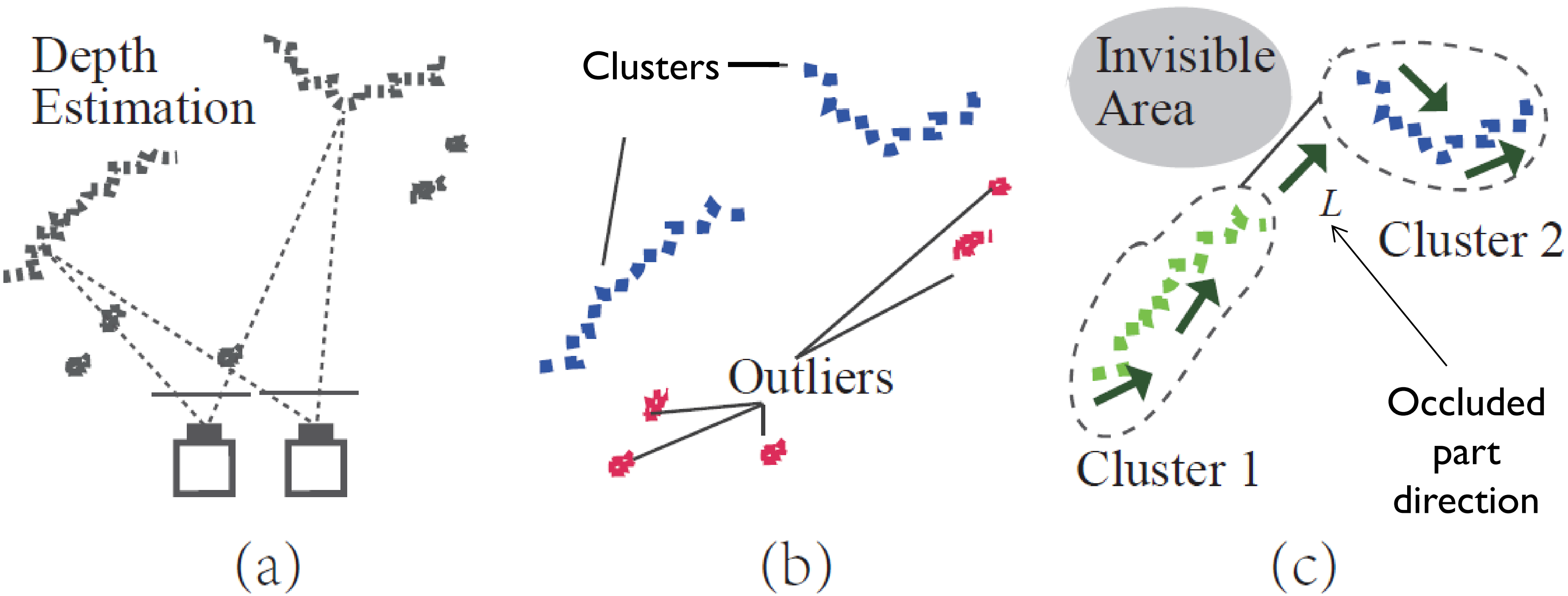
$$\underset{\mathbf{a}}{\operatorname{argmax}} \int_{\Omega} |\boldsymbol{\nabla} F(\mathbf{r}, \mathbf{a}) . \mathbf{n_r}| \, d\mathbf{r} \qquad (2)$$

Example: depth estimation optimization

- Improving depth estimation by 30% for the same stereo matching algorithm



Iterations and dealing with occlusions

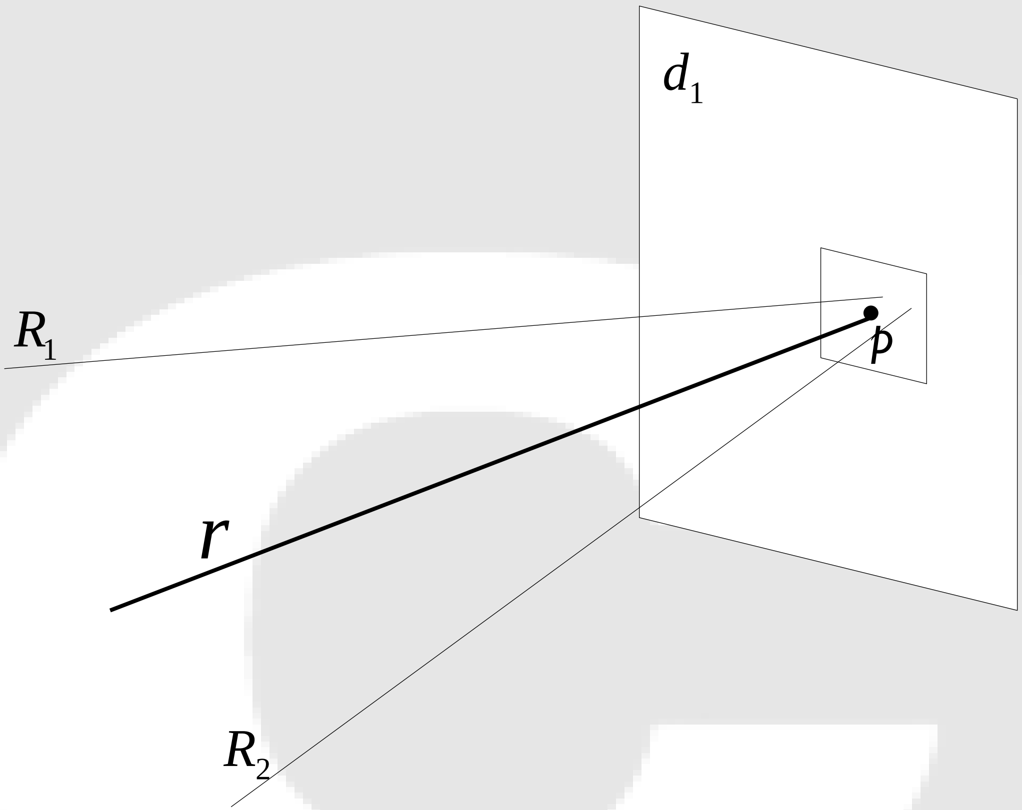


Imperfections of rendering

- Even when we have enough samples, rendering may not know how to use them

Effective Sampling Density (ESD)

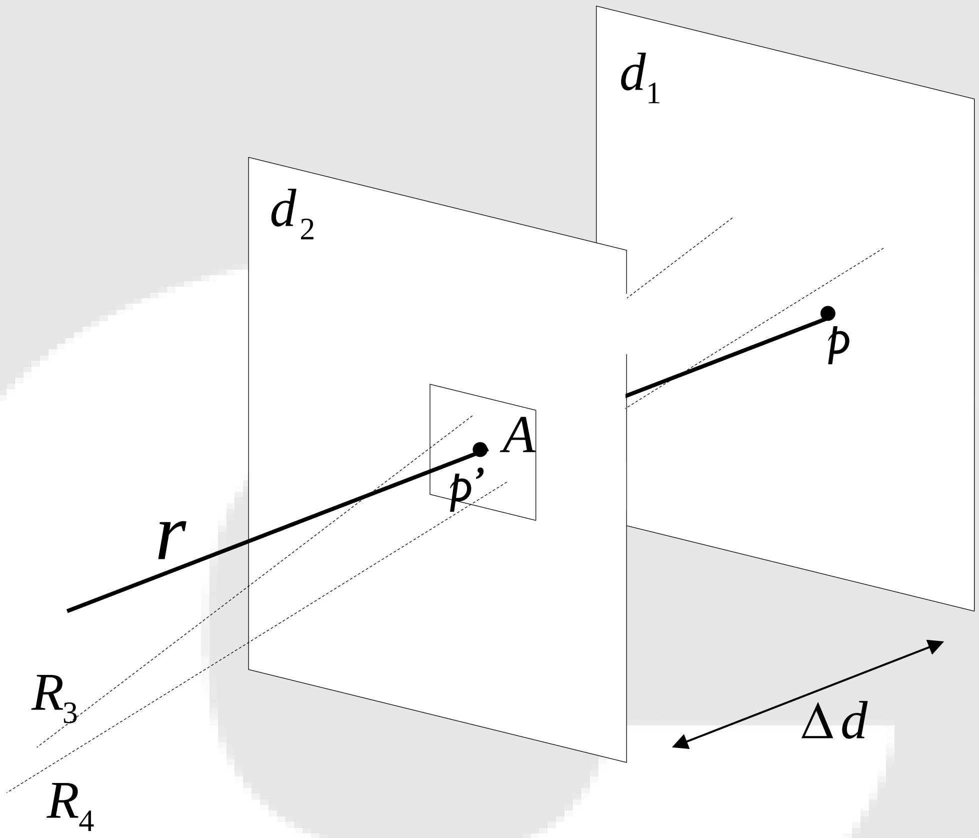
- Introduced by us in 2011 and discussed in *



* H. Shidanshidi, F. Safaei, and L. Wanqing, "Estimation of Signal Distortion Using Effective Sampling Density for Light Field-Based Free Viewpoint Video," *Multimedia, IEEE Transactions on*, vol. 17, pp. 1677-1693, 2015.

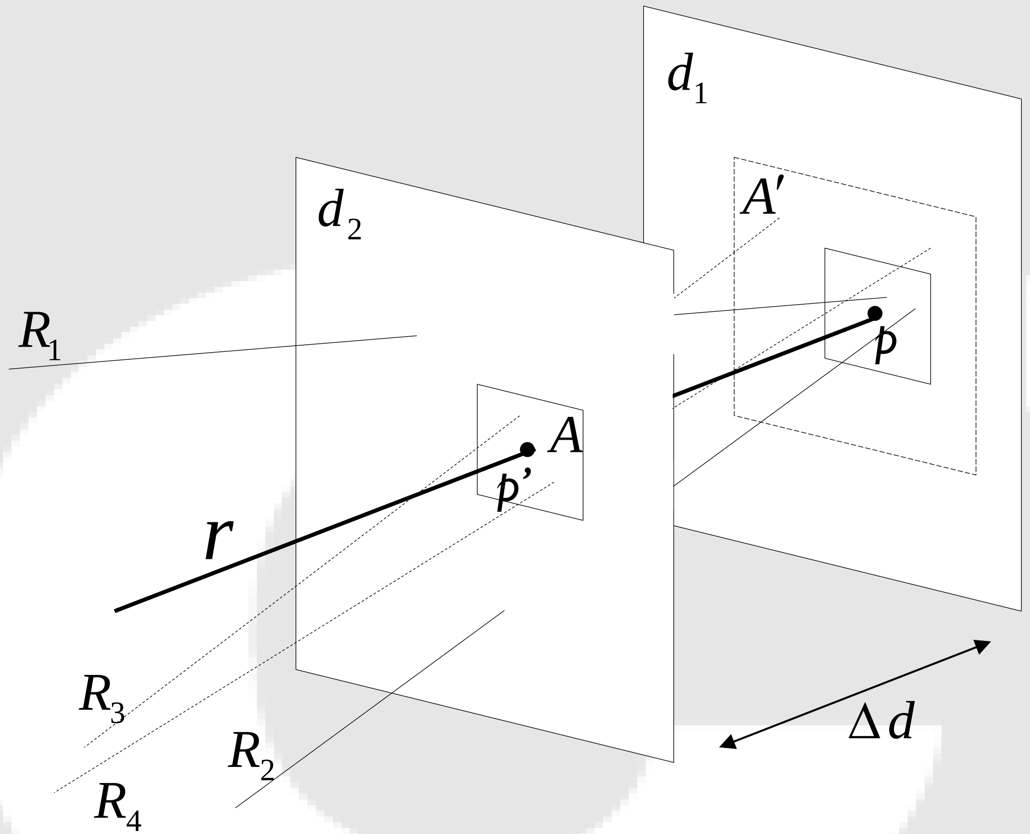
Effective Sampling Density (ESD)

- Introduced by us in 2011



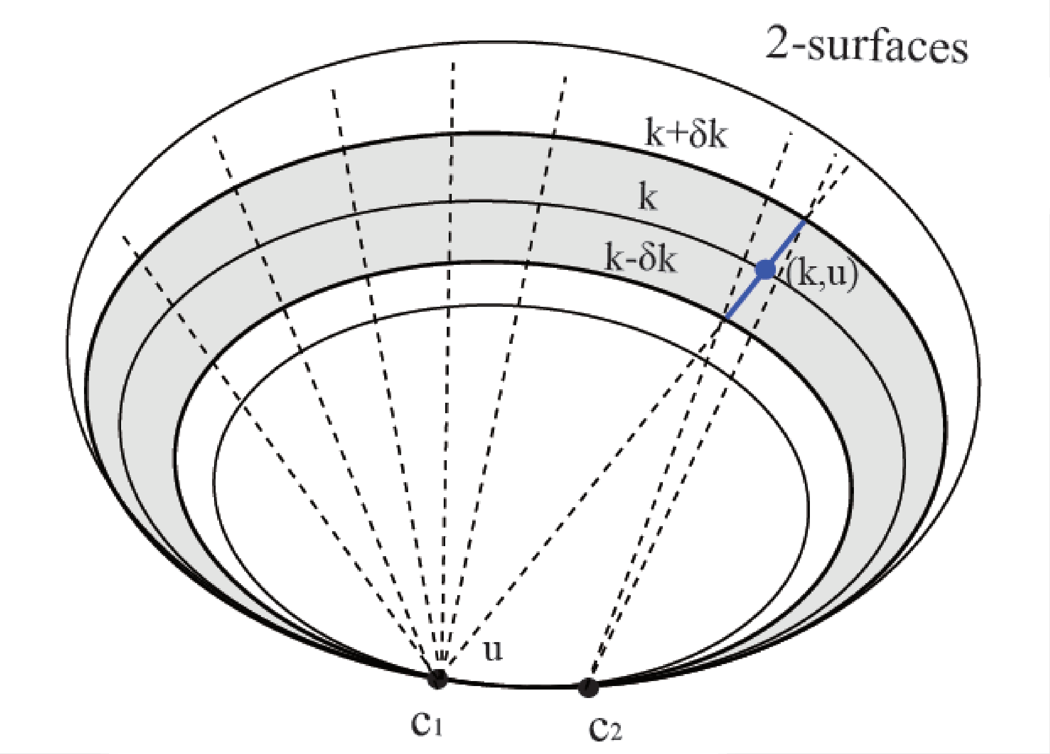
Effective Sampling Density (ESD)

A' > A, Hence the sampling density is effectively reduced



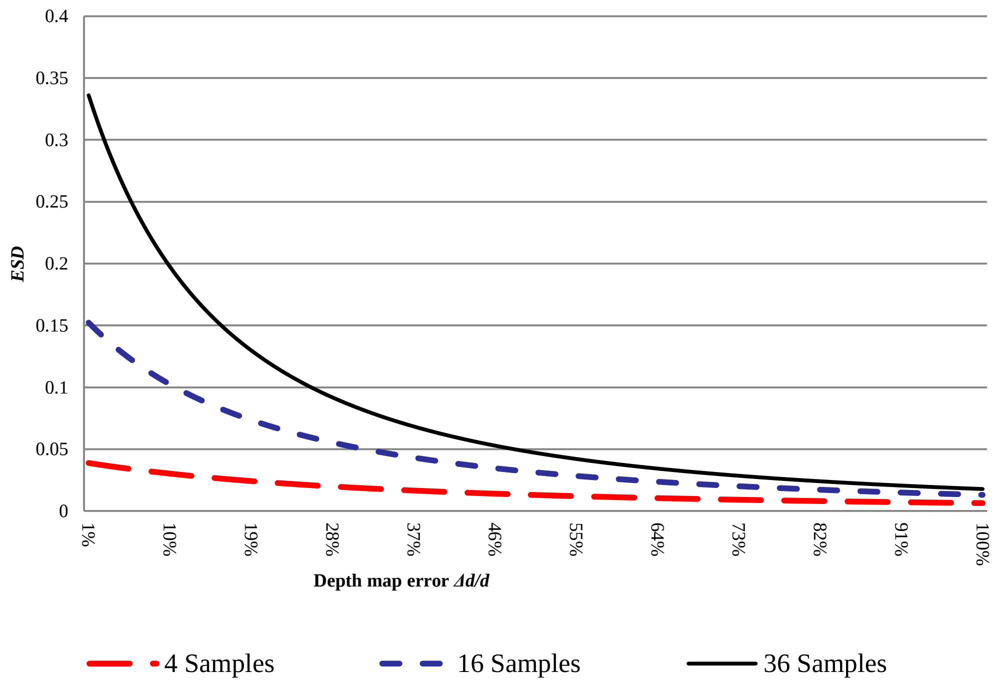
The error bound of depth

- The depth error Δd is bounded by the distance between 2-surfaces
  - Measured by CF density

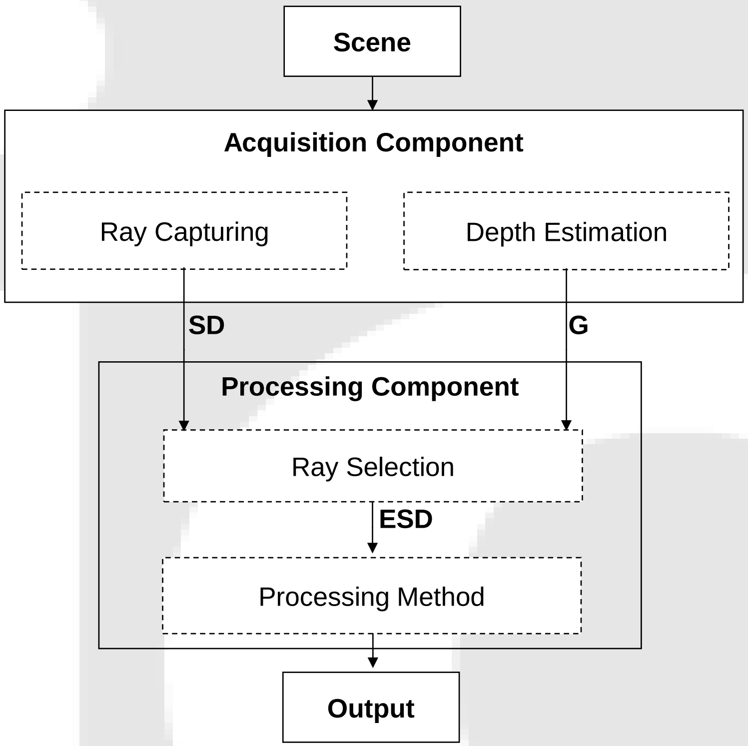


ESD vs Error in depth estimation

- ESD decreases rapidly with Δd even for powerful rendering algorithms



Effective Sampling Density (ESD)



θ : captured
Ω : captured rays passing through A
ω : rays selected by M & processed by F

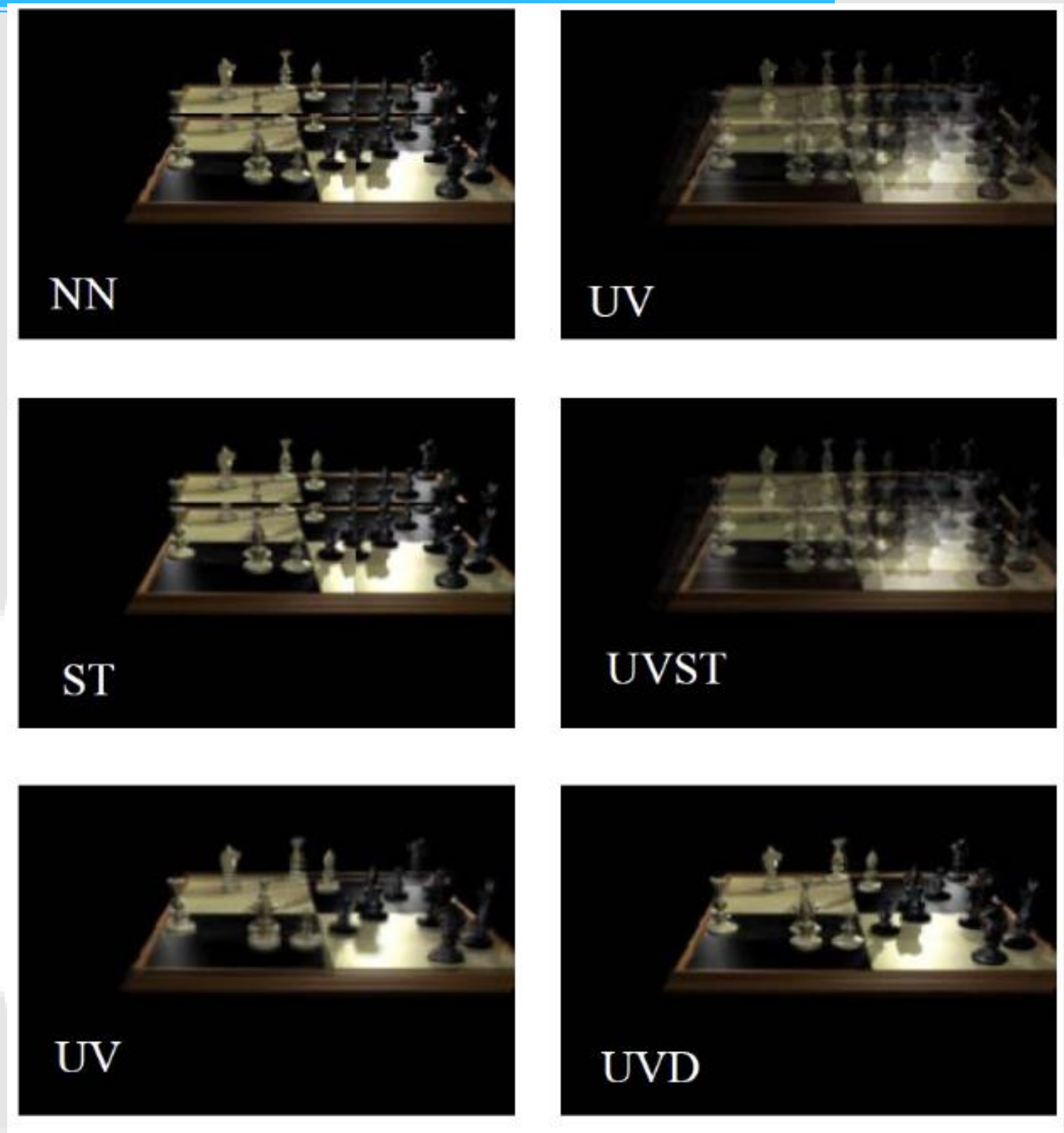
$$\omega \subseteq \Omega \subseteq \theta \,;\, \omega = M(\theta, G),\; r = F(\omega, G)$$

$$SD = \frac{|\Omega|}{A},$$

$$\mathrm{ESD} = \frac{|\omega|}{A} = \frac{|M(\theta, G)|}{A}$$

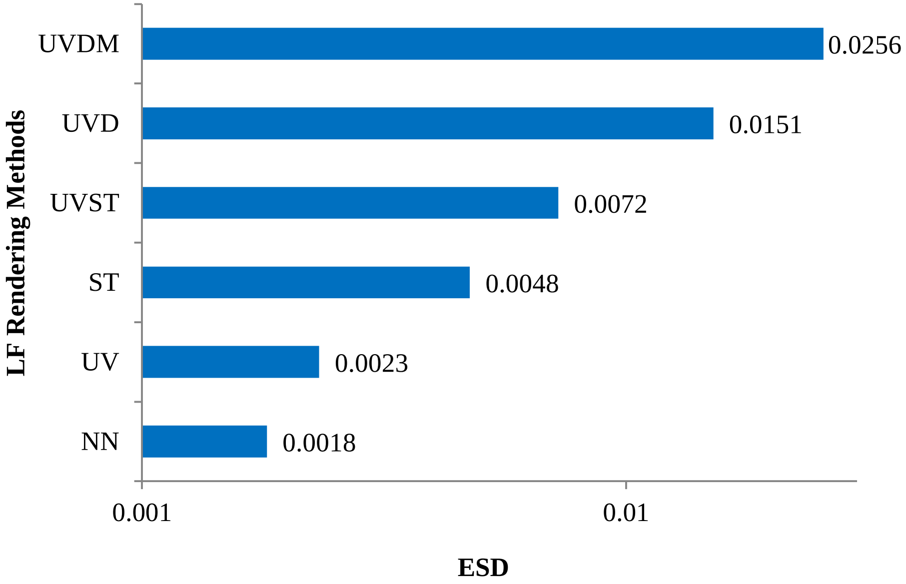
H. Shidanshidi, F. Safaei and W. Li, Estimation of Signal Distortion Using Effective Sampling Density for Light Field-based Free Viewpoint Video, *IEEE Trans Multimedia*, 17(10), 2015, pp. 1677-1693

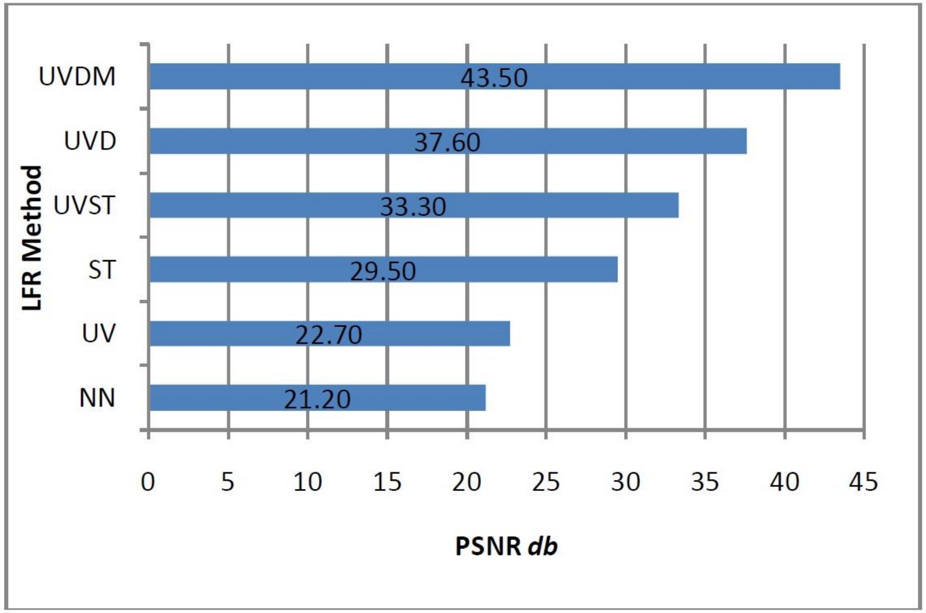
Output comparison



ESD (*not* SD) is the indicator of output quality

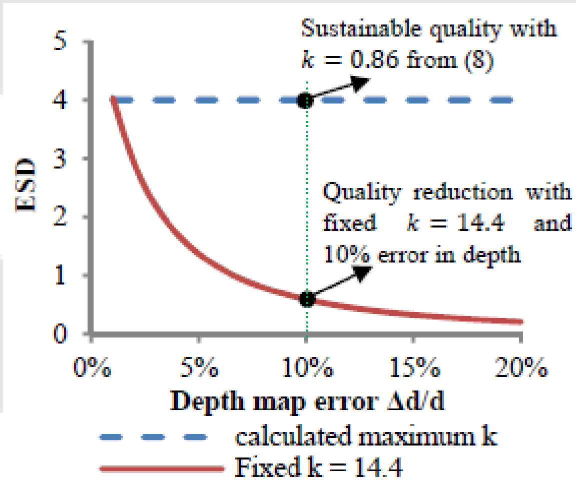
- Comparison of ESD and output quality for six rendering methods

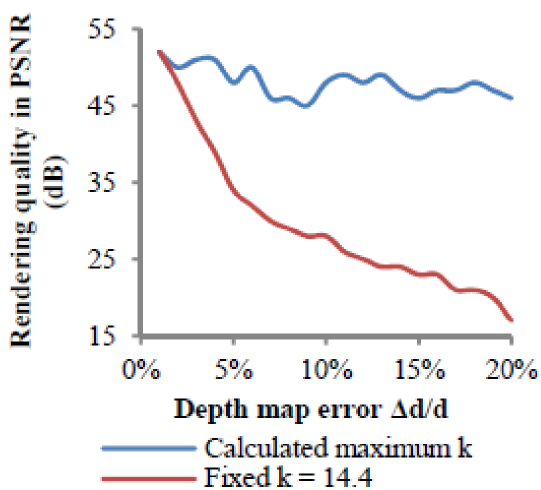


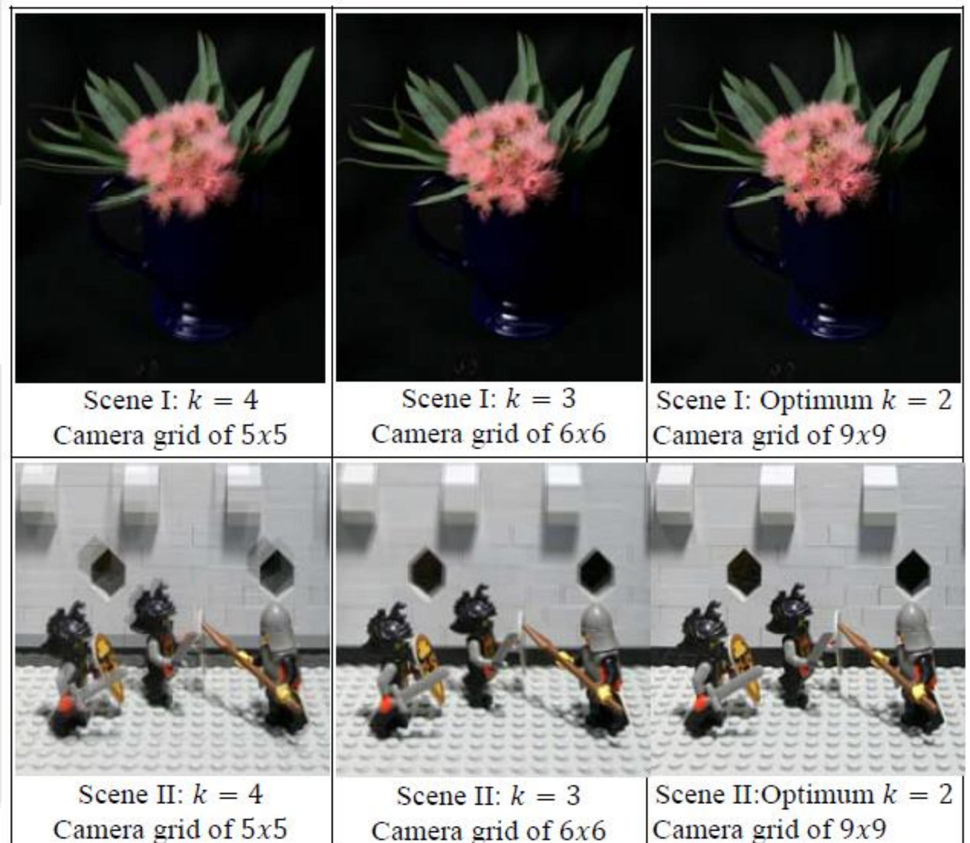


Optimization of the number of cameras

$$k = \frac{l\left(\left(\sqrt{\frac{|\omega|}{n}} - 1\right) d^2 - d\Delta d\right)}{\Delta d\left(\sqrt{|\omega|} - 1\right)}$$



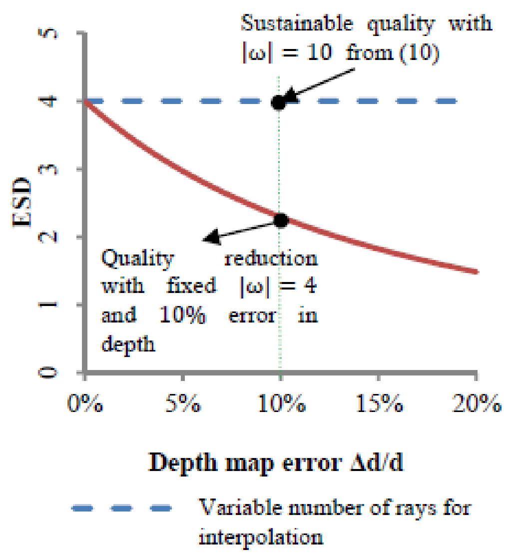


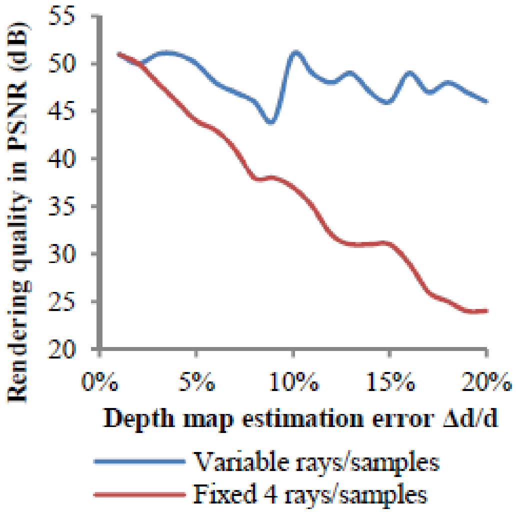


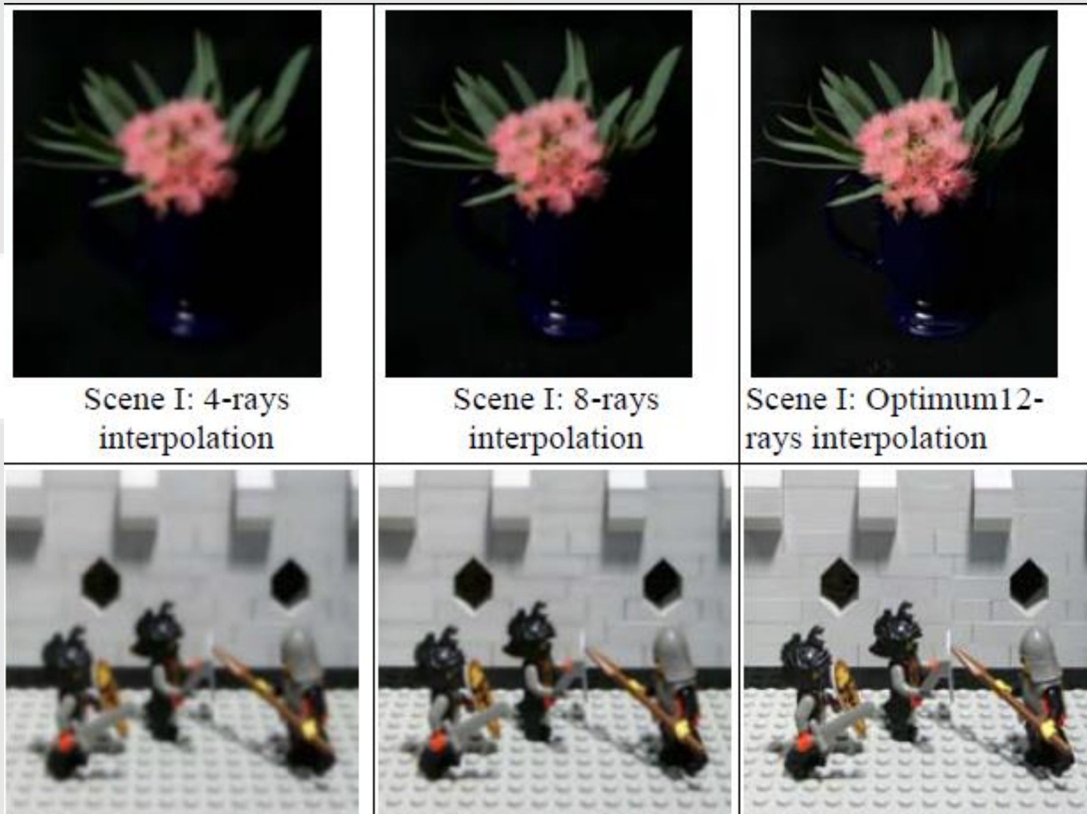
ict RESEARCH INSTITUTE

Optimizing Ray Selection

$$|\omega| = \left(\frac{l(d+\Delta d) - \frac{\Delta d.k}{d}}{\frac{ld}{\sqrt{n}} - \frac{\Delta d.k}{d}}\right)^2$$







Scene I: 4-rays interpolation	Scene I: 8-rays interpolation	Scene I: Optimum12-rays interpolation
Scene II: 4-rays interpolation	Scene II: 8-rays interpolation	Scene II: Optimum 14-rays interpolation

Conclusion

- Acquisition in most cases has been “ad-hoc” or experienced based
- ESD provides a tractable metric to quantify the joint impact of how cameras sampling a scene and how the samples are used
- CF offers an effective representation to numerically calculate ESD and to study the interaction between the cameras and scenes
- ESD + CF opens a new approach to optimizing multiple cameras

Acknowledgement

- This presentation is based on the joint work with
  - A/Professor Wanqing Li
  - Dr Hooman Shidanshidi
  - Mr Shichao Fu (PhD)

- Partially supported by Smart Services Cooperative Research Centre, Australia