L1a: Introduction to Light Fields

2018 IEEE SPS Summer School on Light Field Data
Representation, Interpretation, and Compression

Donald G. Dansereau, May 2018

Schedule

09:30 - 10:15	Lecture 1 a D Danserau
10:15 - 10:45	Coffee break
10:45 - 11:30	Lecture 1 b D Danserau
11:30 - 11:45	Break
11:45 - 12:30	Lecture 1 c D Danserau
12:30 - 13:00	Lunch
13:00 - 14:00	

14:00 - 15:00	IEEE
15:00 - 15:30	Exercise 1 D Danserau
15:30 - 16:00	Coffee break
16:00 - 16:45	Exercise 1 D Danserau
16:45 - 17:00	
17:00 - 17:30	

Outline

Lecture 1a: Introduction to Light Fields

- Intro
- History
- Parameterizations
- Visualizations

Lecture 1b: Cameras, Sampling, & Calibration

Lecture 1c: Basic Processing

Hands-on: Writing a renderer, handling light fields in matlab

Resources

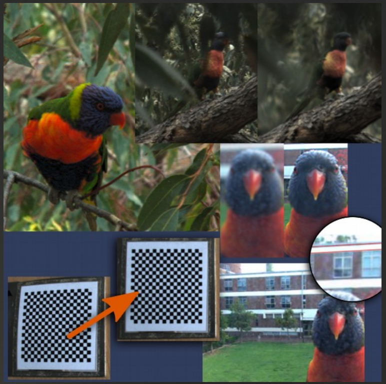
Exercise handouts are available at: http://dgd.vision/LF2018/

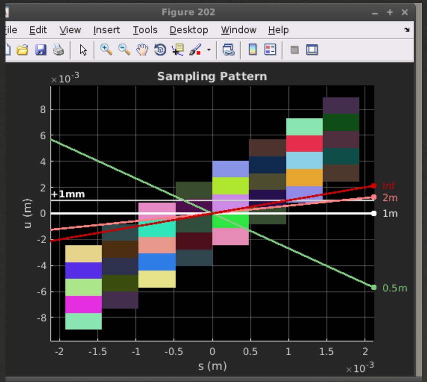
These slides will also be up there soon

Light Field Resources page on GitHub with links to datasets, forums, tools:

https://github.com/lightfield-analysis/resources

Light field toolbox, sampling pattern explorer, LF Synth: http://dgd.vision/Tools

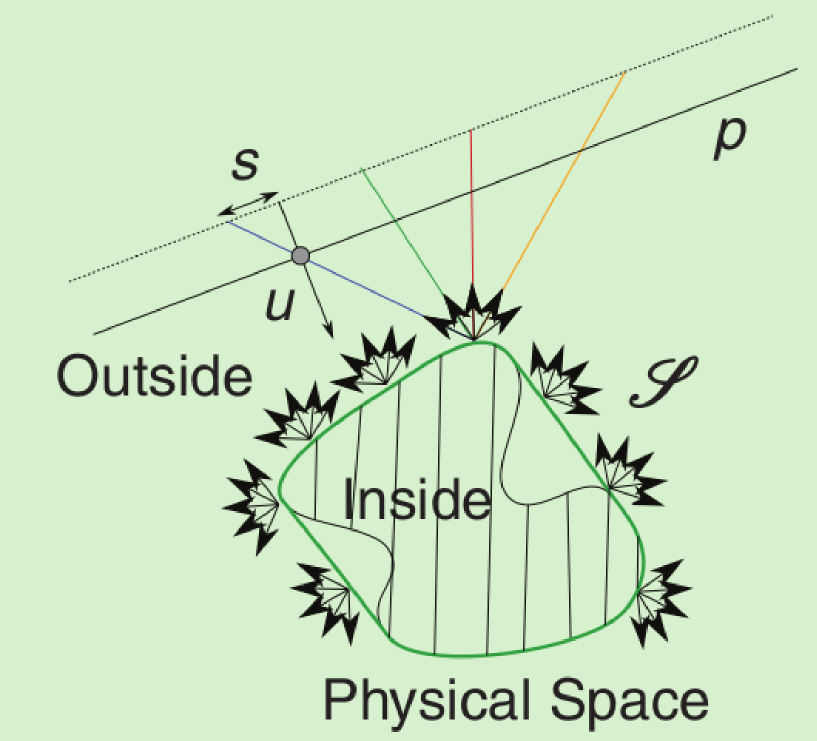




What is a light field?

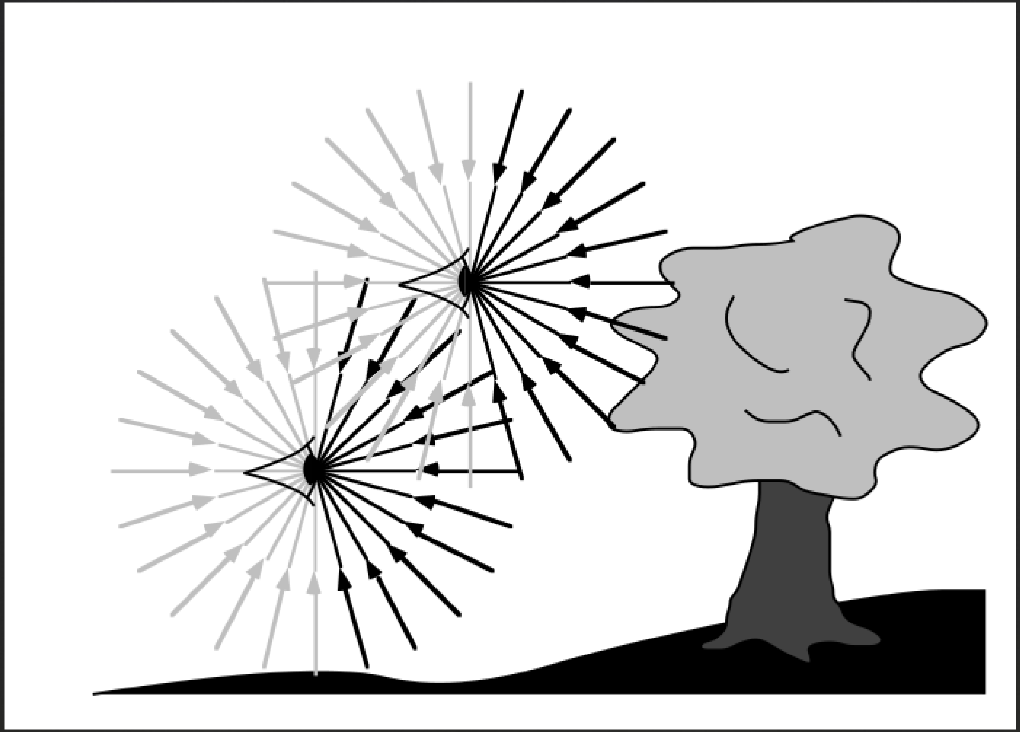
1) Represent scenes as a surface of light

We can represent complex scenes inside a volume by describing the light rays passing through a surface surrounding the volume



[Image c/o Ihrke et al 2016]

2) Represent light as a scalar field



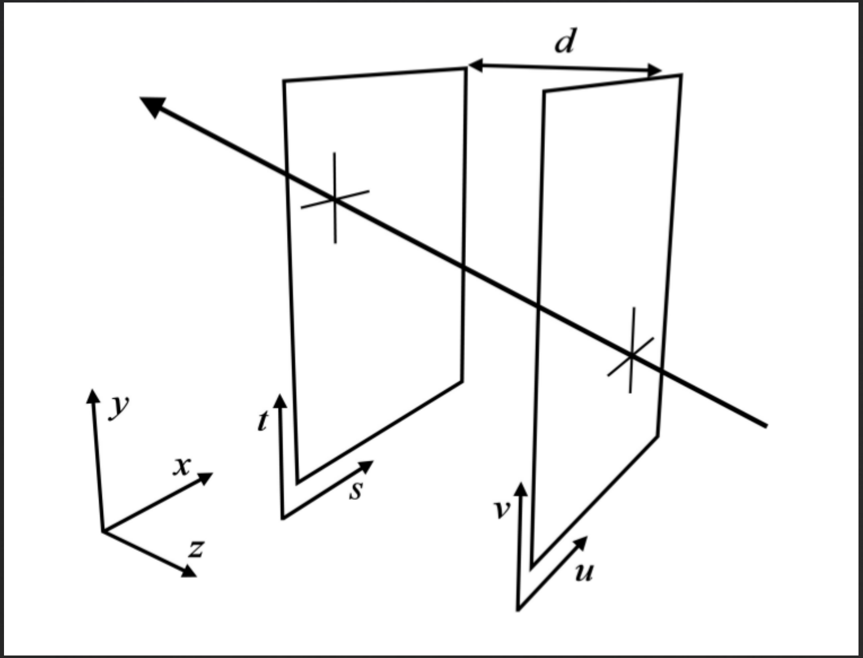
[Adelson and Bergen 1991]

Monochrome Photo: $\mathcal{L}(u, v)$

Plenoptic Function $\mathcal{L}(.)$

- Position (3)
- Direction (2)
- Time (1)
- Wavelength (1)
- *Phase (1)*
- *Polarization (1..3)*

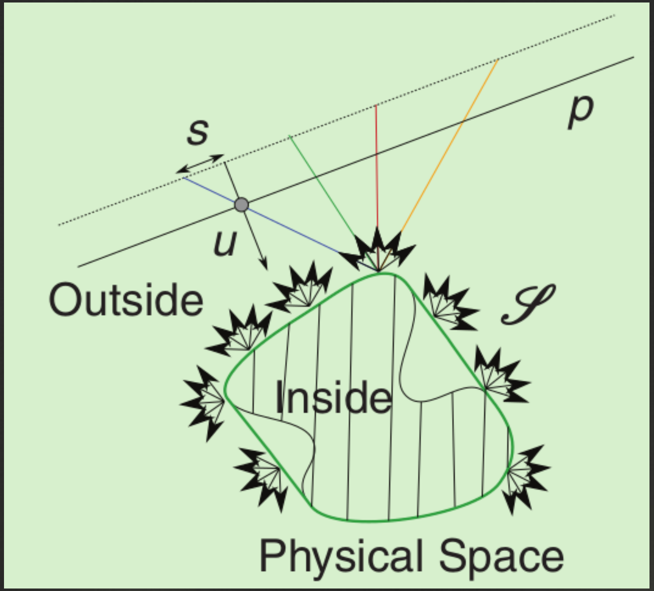
3) 4D is a sweet spot for rays



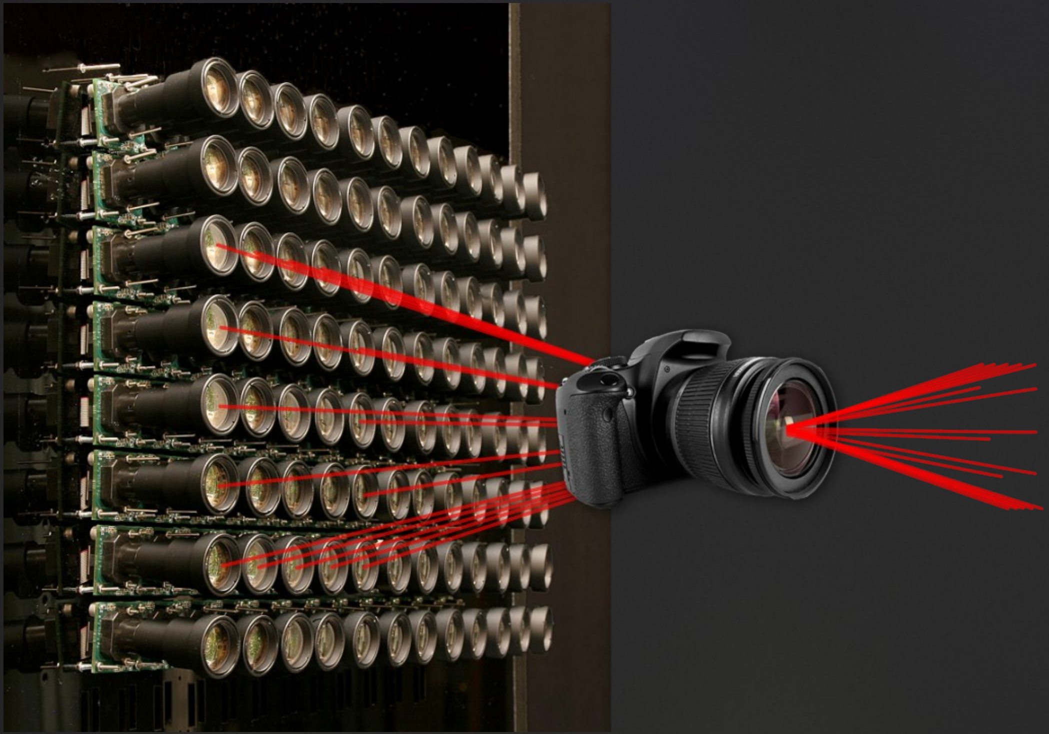
Minimum for describing position & direction

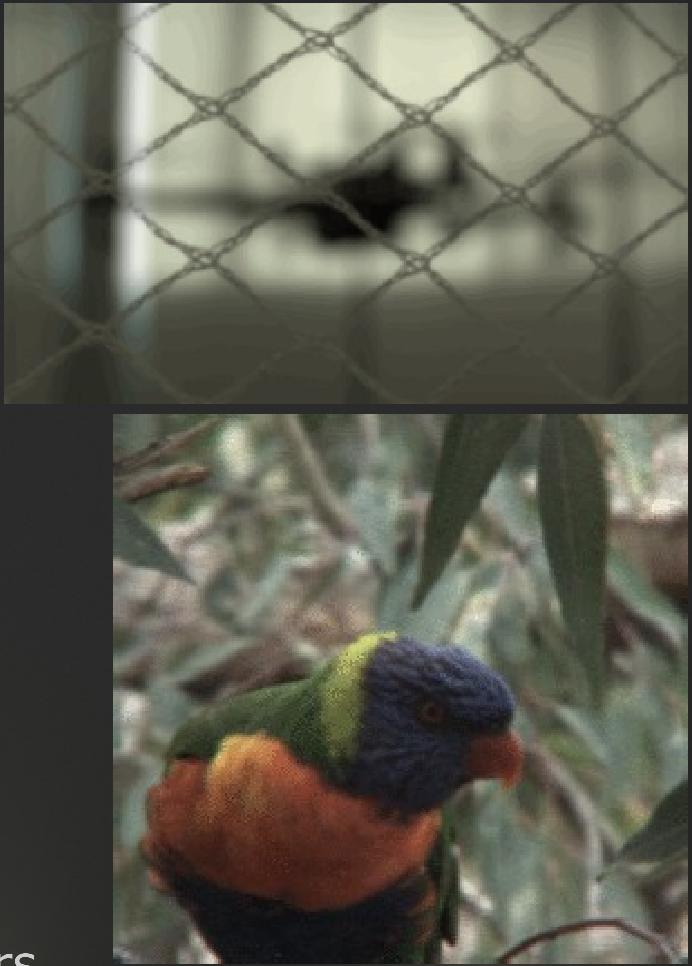
1 less than the obvious 5

- Assumes non-participating medium
- Restricted to outside a volume



A Versatile Representation



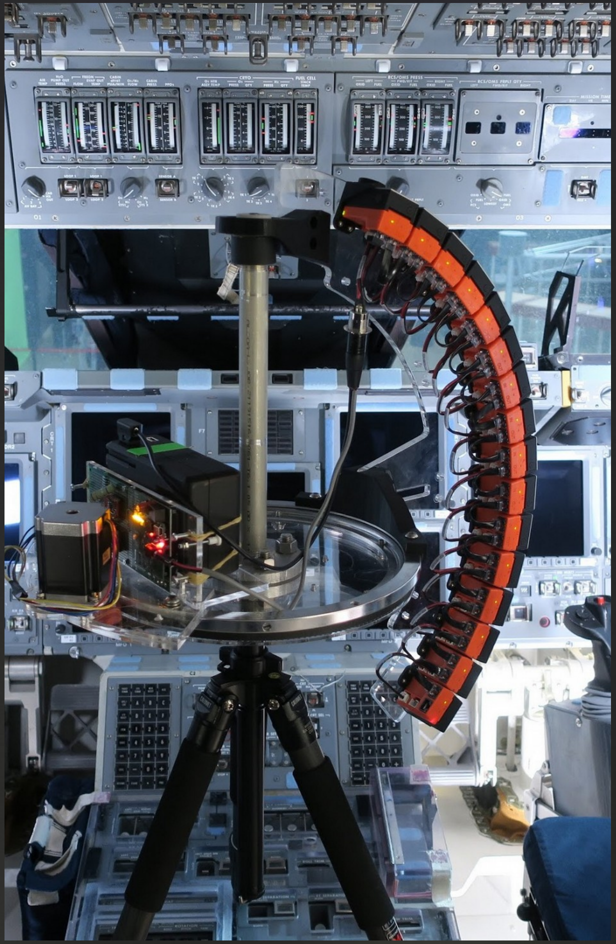


- Regular, densely sampled light
- Emulate virtual cameras – a meta-camera
- Fundamental structure capturing complex behaviours

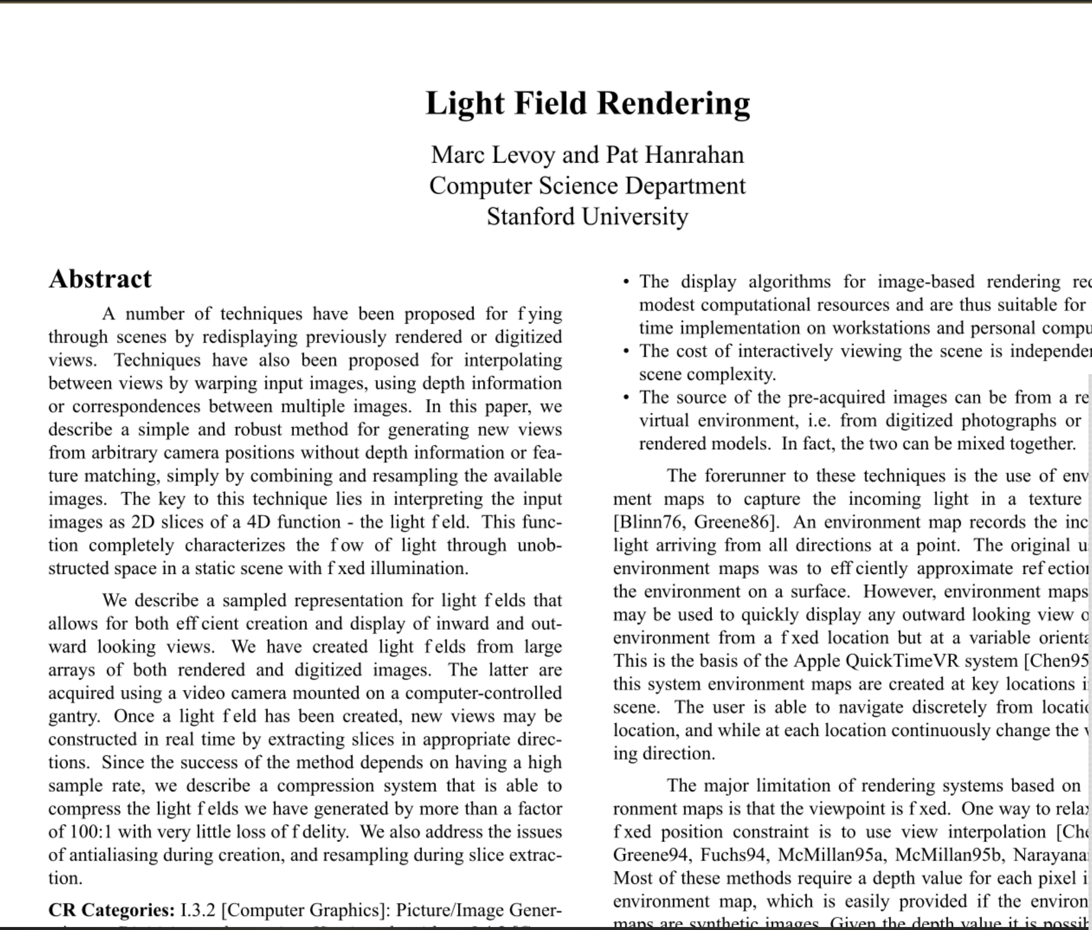
History – How old are these ideas?

Welcome to Lights Fields, Google 2018

Video: Google’s Welcome to Light Fields Demo
https://www.youtube.com/watch?v=OUU2yGHgPQY



1996 The Light Field Representation



Light Field Rendering

Marc Levoy and Pat Hanrahan
Computer Science Department
Stanford University

Abstract

A number of techniques have been proposed for flying through scenes by redisplaying previously rendered or digitized views. Techniques have also been proposed for interpolating between views by warping input images, using depth information or correspondences between multiple images. In this paper, we describe a simple and robust method for generating new views from arbitrary camera positions without depth information or feature matching, simply by combining and resampling the available images. The key to this technique lies in interpreting the input images as 2D slices of a 4D function - the light field. This function completely characterizes the flow of light through unobstructed space in a static scene with fixed illumination.

We describe a sampled representation for light fields that allows for both efficient creation and display of inward and outward looking views. We have created light fields from large arrays of both rendered and digitized images. The latter are acquired using a video camera mounted on a computer-controlled gantry. Once a light field has been created, new views may be constructed in real time by extracting slices in appropriate directions. Since the success of the method depends on having a high sample rate, we describe a compression system that is able to compress the light fields we have generated by more than a factor of 100:1 with very little loss of fidelity. We also address the issues of antialiasing during creation, and resampling during slice extraction.

CR Categories: I.3.2 [Computer Graphics]: Picture/Image Gener-

- The display algorithms for image-based rendering require modest computational resources and are thus suitable for real-time implementation on workstations and personal computers.
- The cost of interactively viewing the scene is independent of scene complexity.
- The source of the pre-acquired images can be from a re virtual environment, i.e. from digitized photographs or rendered models. In fact, the two can be mixed together.

The forerunner to these techniques is the use of env ment maps to capture the incoming light in a texture [Blinn76, Greene86]. An environment map records the inc light arriving from all directions at a point. The original u environment maps was to efficiently approximate reflection the environment on a surface. However, environment maps may be used to quickly display any outward looking view o environment from a fixed location but at a variable orienta This is the basis of the Apple QuickTimeVR system [Chen95 this system environment maps are created at key locations i scene. The user is able to navigate discretely from locatic location, and while at each location continuously change the ing direction.

The major limitation of rendering systems based on ronment maps is that the viewpoint is fixed. One way to relax fixed position constraint is to use view interpolation [Che Greene94, Fuchs94, McMillan95a, McMillan95b, Narayana Most of these methods require a depth value for each pixel i environment map, which is easily provided if the environ maps are synthetic images. Given the depth value it is possi

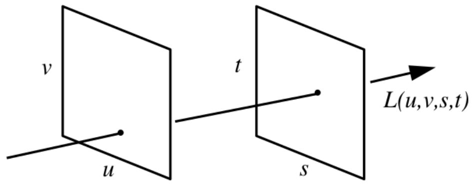
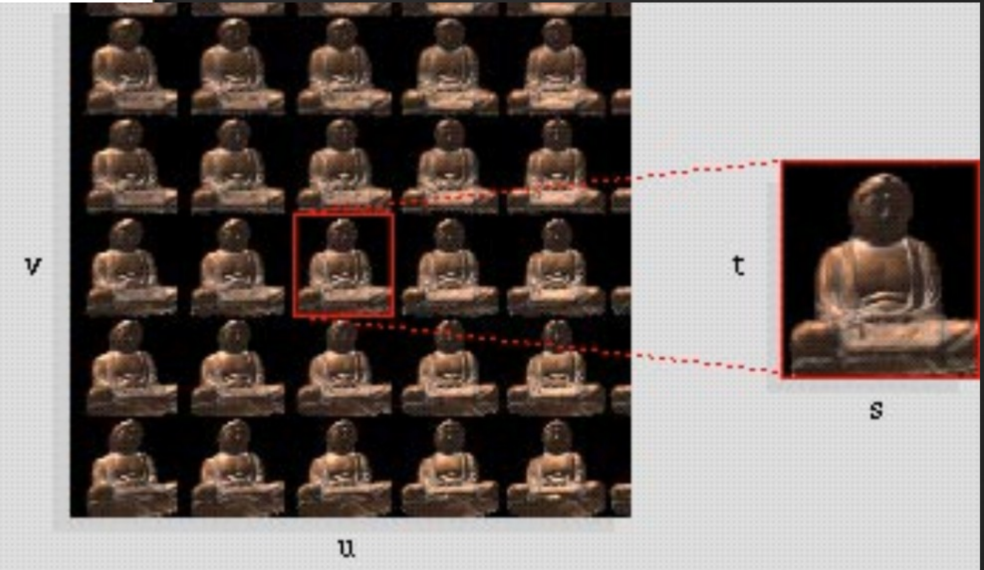
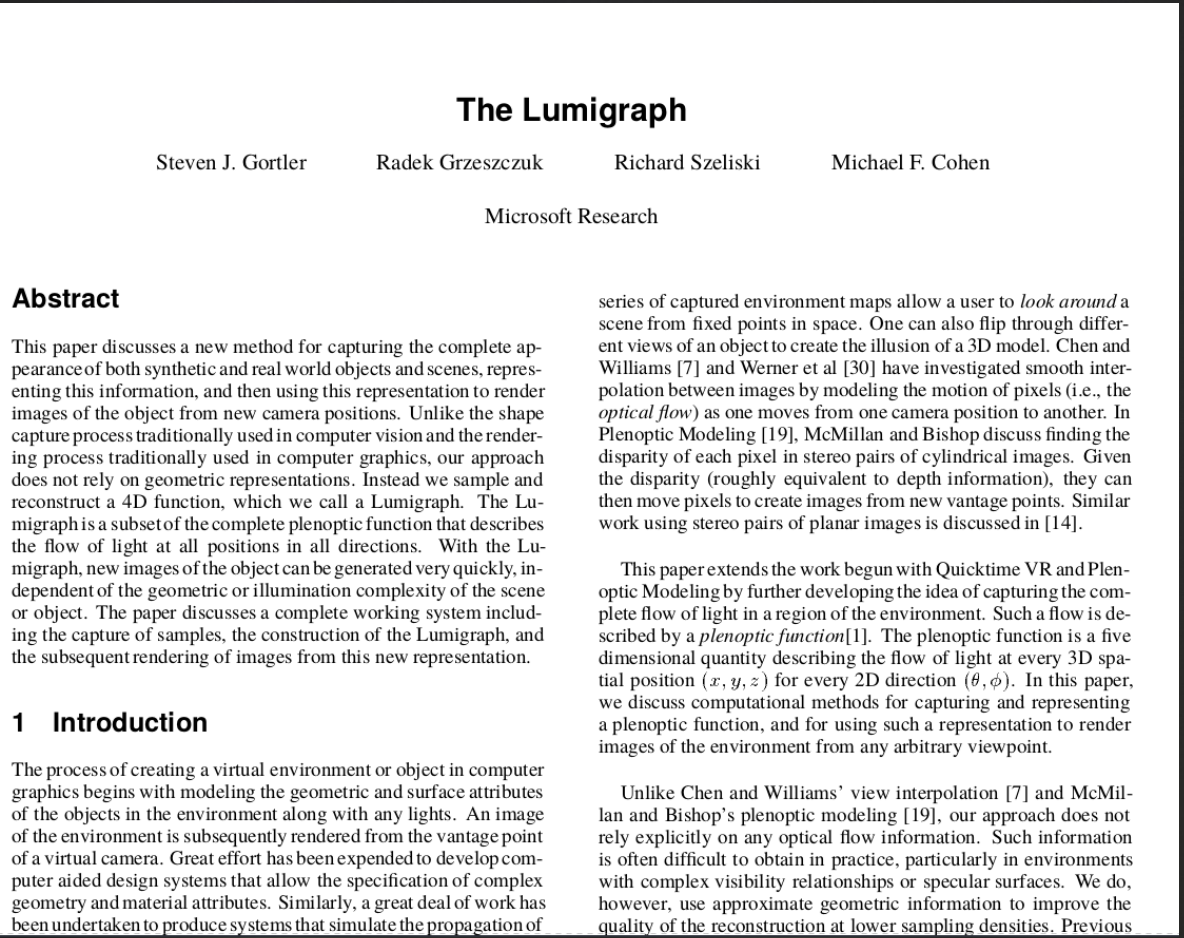


Figure 1: The light slab representation.



1996 The Lumigraph



The Lumigraph

Steven J. Gortler Radek Grzeszczuk Richard Szeliski Michael F. Cohen

Microsoft Research

Abstract

This paper discusses a new method for capturing the complete appearance of both synthetic and real world objects and scenes, representing this information, and then using this representation to render images of the object from new camera positions. Unlike the shape capture process traditionally used in computer vision and the rendering process traditionally used in computer graphics, our approach does not rely on geometric representations. Instead we sample and reconstruct a 4D function, which we call a Lumigraph. The Lumigraph is a subset of the complete plenoptic function that describes the flow of light at all positions in all directions. With the Lumigraph, new images of the object can be generated very quickly, independent of the geometric or illumination complexity of the scene or object. The paper discusses a complete working system including the capture of samples, the construction of the Lumigraph, and the subsequent rendering of images from this new representation.

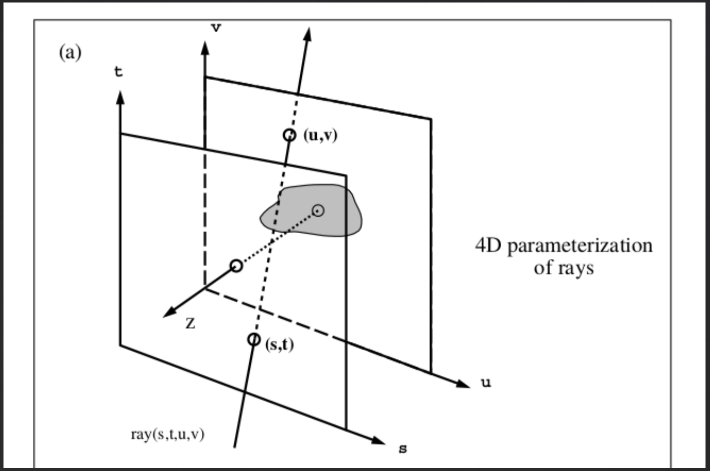
1 Introduction

The process of creating a virtual environment or object in computer graphics begins with modeling the geometric and surface attributes of the objects in the environment along with any lights. An image of the environment is subsequently rendered from the vantage point of a virtual camera. Great effort has been expended to develop computer aided design systems that allow the specification of complex geometry and material attributes. Similarly, a great deal of work has been undertaken to produce systems that simulate the propagation of

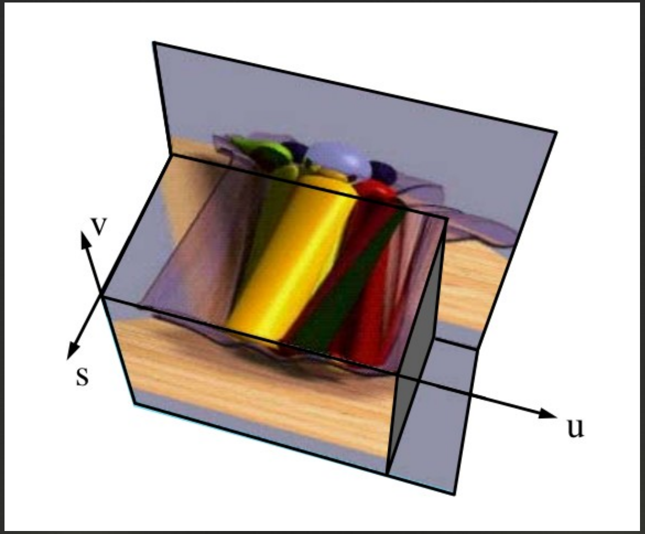
series of captured environment maps allow a user to *look around* a scene from fixed points in space. One can also flip through different views of an object to create the illusion of a 3D model. Chen and Williams [7] and Werner et al [30] have investigated smooth interpolation between images by modeling the motion of pixels (i.e., the *optical flow*) as one moves from one camera position to another. In Plenoptic Modeling [19], McMillan and Bishop discuss finding the disparity of each pixel in stereo pairs of cylindrical images. Given the disparity (roughly equivalent to depth information), they can then move pixels to create images from new vantage points. Similar work using stereo pairs of planar images is discussed in [14].

This paper extends the work begun with Quicktime VR and Plenoptic Modeling by further developing the idea of capturing the complete flow of light in a region of the environment. Such a flow is described by a *plenoptic function*[1]. The plenoptic function is a five dimensional quantity describing the flow of light at every 3D spatial position (x, y, z) for every 2D direction (θ, ϕ). In this paper, we discuss computational methods for capturing and representing a plenoptic function, and for using such a representation to render images of the environment from any arbitrary viewpoint.

Unlike Chen and Williams' view interpolation [7] and McMillan and Bishop's plenoptic modeling [19], our approach does not rely explicitly on any optical flow information. Such information is often difficult to obtain in practice, particularly in environments with complex visibility relationships or specular surfaces. We do, however, use approximate geometric information to improve the quality of the reconstruction at lower sampling densities. Previous



4D parameterization of rays



1996

Nintento 64

Deep blue beats Garry Kasparov

MP3 patented

IPv6 introduced

Tom's Hardware starts

IMDB

KDE

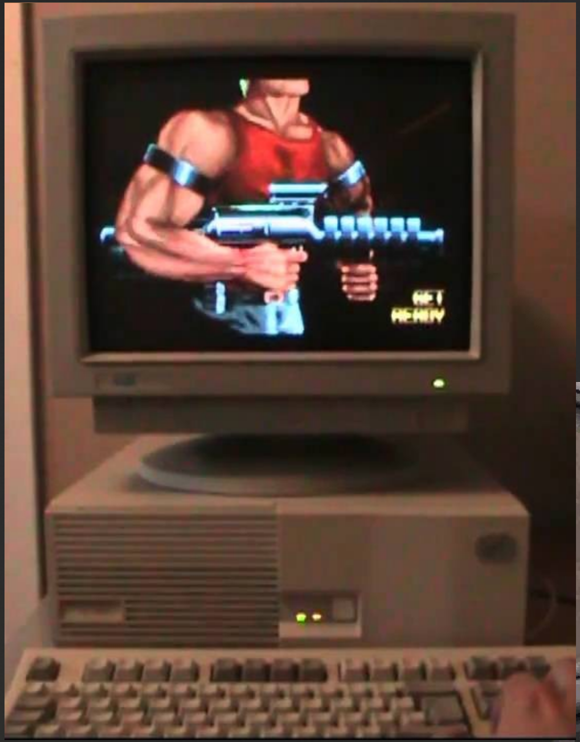
MySpace.Com

JDK 1.0

The Internet Archive

The wheel mouse hits mainstream

~~Pentium II~~ 486DX!



1996

Video: Levoy et al Light Field Rendering - Siggraph '96 video
https://www.youtube.com/watch?v=dMcZpeGOBPI

1992 Depth from Epipolar Planes

Single Lens Stereo with a Plenoptic Camera

Edward H. Adelson and John Y.A. Wang

Abstract—Ordinary cameras gather light across the area of their lens aperture, and the light striking a given subregion of the aperture is structured somewhat differently than the light striking an adjacent subregion. By analyzing this optical structure, one can infer the depths of objects in the scene, i.e., one can achieve "single lens stereo." We describe a novel camera for performing this analysis. It incorporates a single main lens along with a lenticular array placed at the sensor plane. The resulting "plenoptic camera" provides information about how the scene would look when viewed from a continuum of possible viewpoints bounded by the main lens aperture. Deriving depth information is simpler than in a binocular stereo system because the correspondence problem is minimized. The camera extracts information about both horizontal and vertical parallax, which improves the reliability of the depth estimates.

I. INTRODUCTION

"EVERY BODY in the light and shade fills the surrounding air with infinite images of itself; and these, by infinite pyramids diffused in the air, represent this body

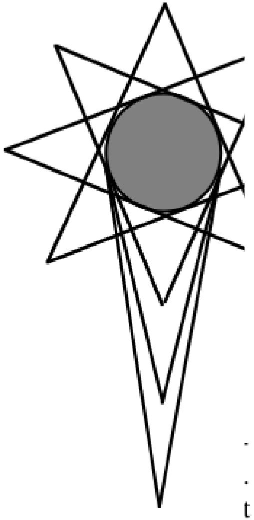


Fig. 1. Diagram from Leonardo's notebooks illu
rays leaving an object's surface may be consi
cones (which Leonardo calls "pyramids"), eac
that would be seen by a pinhole camera at a give

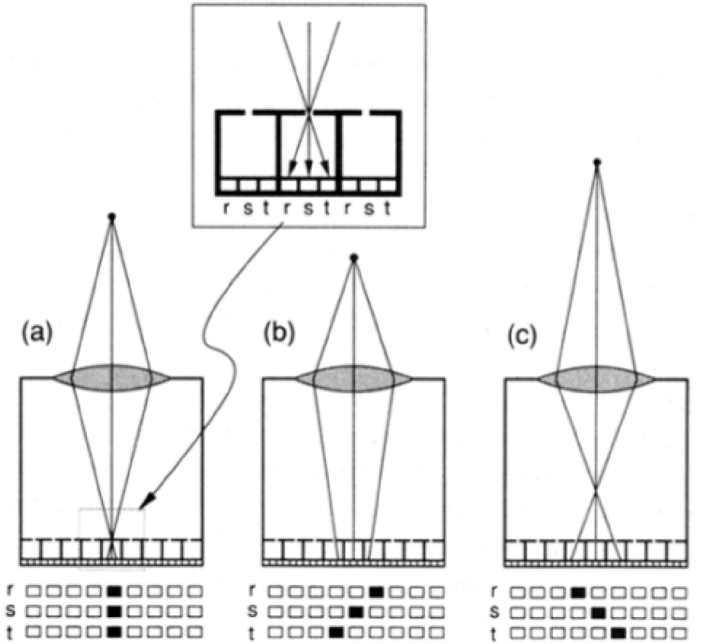
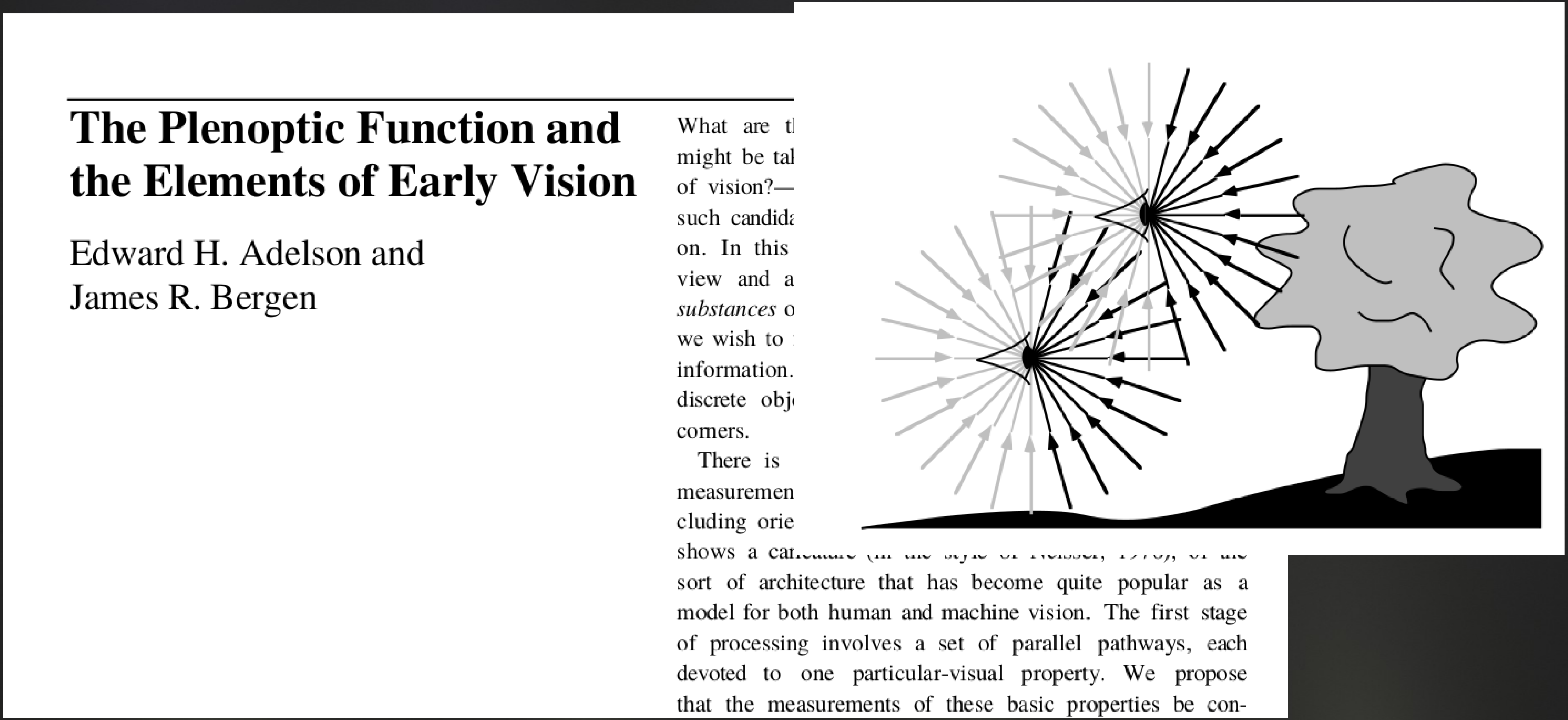


Fig. 5. Array of miniature pinhole cameras placed at the image plane can be used to analyze the structure of the light striking each macropixel.

1991 The Plenoptic Function



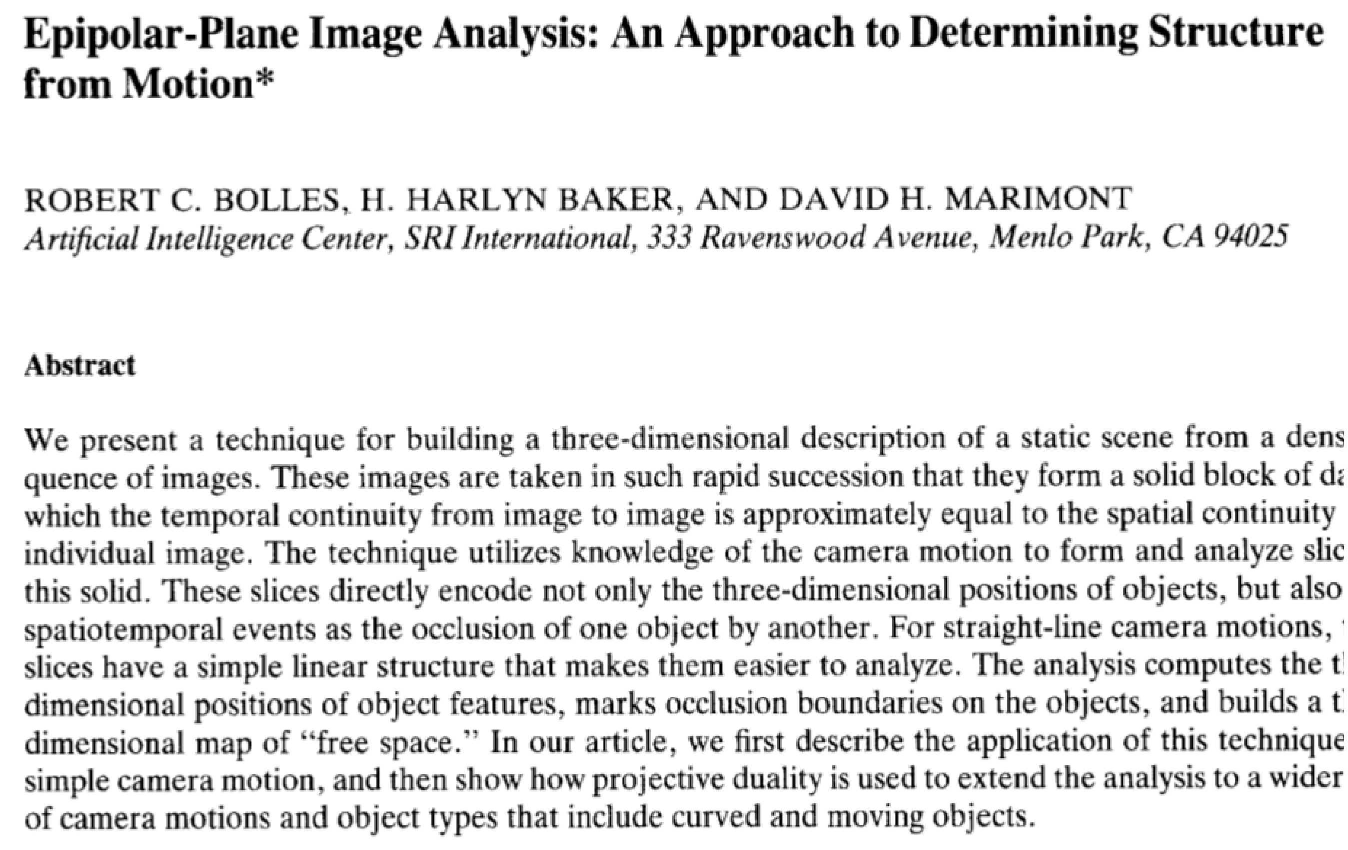
The Plenoptic Function and the Elements of Early Vision

Edward H. Adelson and
James R. Bergen

What are tl
might be tal
of vision?—
such candida
on. In this
view and a
substances o
we wish to
information.
discrete obj
corners.

There is
measuremen
cluding orie
shows a car
sort of architecture that has become quite popular as a model for both human and machine vision. The first stage of processing involves a set of parallel pathways, each devoted to one particular-visual property. We propose that the measurements of these basic properties be con-

1985/1987 Epipolar Plane Images

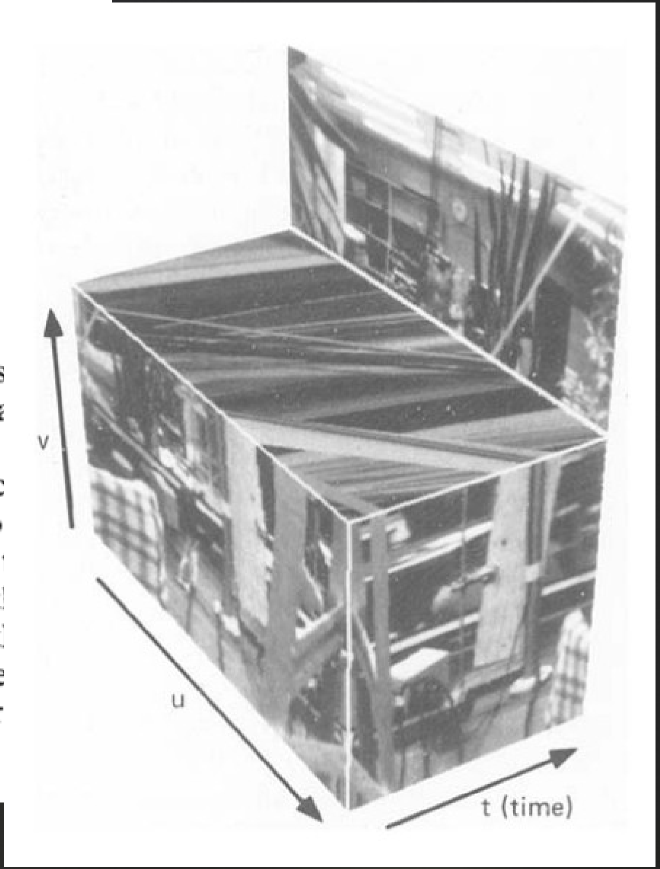


Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion*

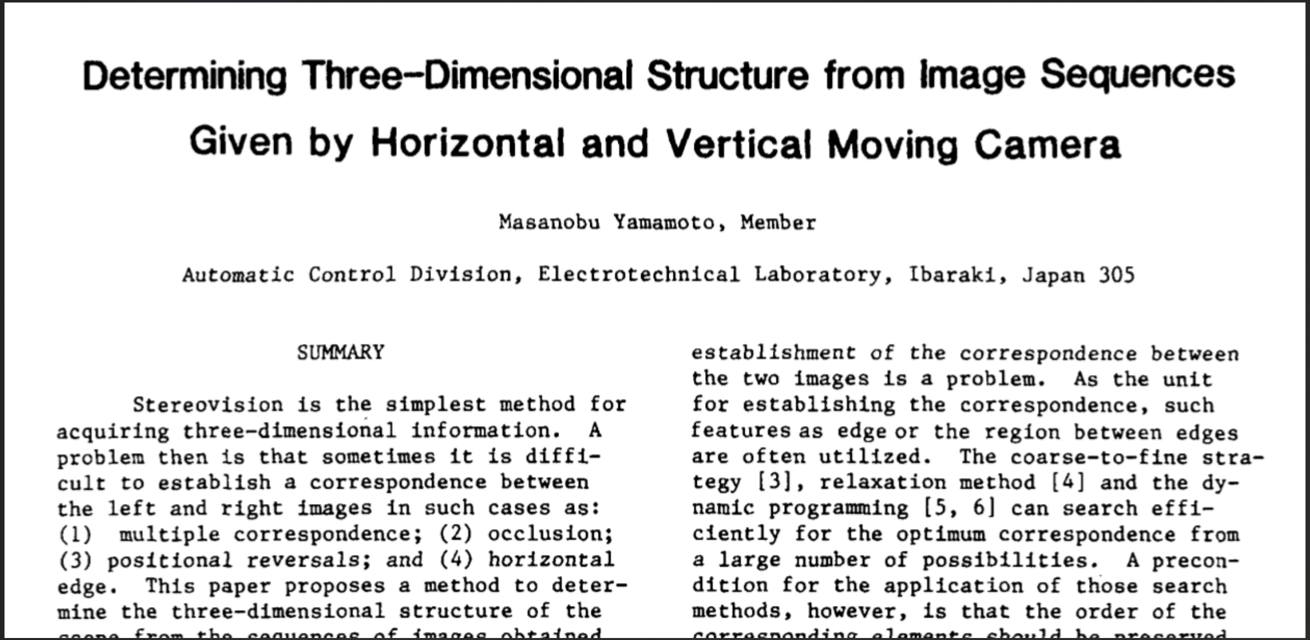
ROBERT C. BOLLES, H. HARLYN BAKER, AND DAVID H. MARIMONT
Artificial Intelligence Center, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025

Abstract

We present a technique for building a three-dimensional description of a static scene from a dens
quence of images. These images are taken in such rapid succession that they form a solid block of da
which the temporal continuity from image to image is approximately equal to the spatial continuity
individual image. The technique utilizes knowledge of the camera motion to form and analyze slic
this solid. These slices directly encode not only the three-dimensional positions of objects, but also
spatiotemporal events as the occlusion of one object by another. For straight-line camera motions,
slices have a simple linear structure that makes them easier to analyze. The analysis computes the t
dimensional positions of object features, marks occlusion boundaries on the objects, and builds a t
dimensional map of "free space." In our article, we first describe the application of this technique
simple camera motion, and then show how projective duality is used to extend the analysis to a wider
of camera motions and object types that include curved and moving objects.



1981/1986 Epipolar Plane Images



Determining Three-Dimensional Structure from Image Sequences Given by Horizontal and Vertical Moving Camera

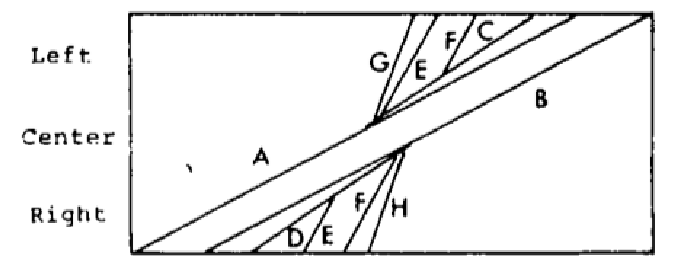
Masanobu Yamamoto, Member

Automatic Control Division, Electrotechnical Laboratory, Ibaraki, Japan 305

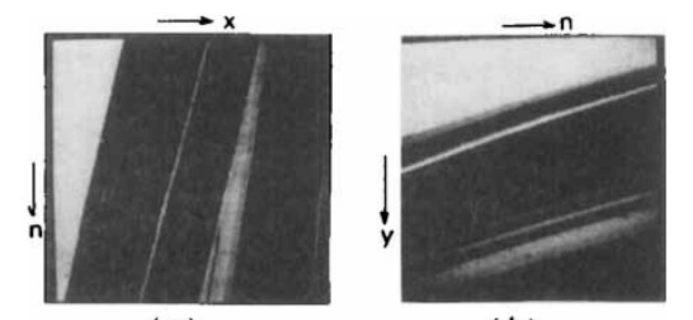
SUMMARY

Stereovision is the simplest method for acquiring three-dimensional information. A problem then is that sometimes it is difficult to establish a correspondence between the left and right images in such cases as: (1) multiple correspondence; (2) occlusion; (3) positional reversals; and (4) horizontal edge. This paper proposes a method to determine the three-dimensional structure of the

establishment of the correspondence between the two images is a problem. As the unit for establishing the correspondence, such features as edge or the region between edges are often utilized. The coarse-to-fine strategy [3], relaxation method [4] and the dynamic programming [5, 6] can search efficiently for the optimum correspondence from a large number of possibilities. A precondition for the application of those search methods, however, is that the order of the



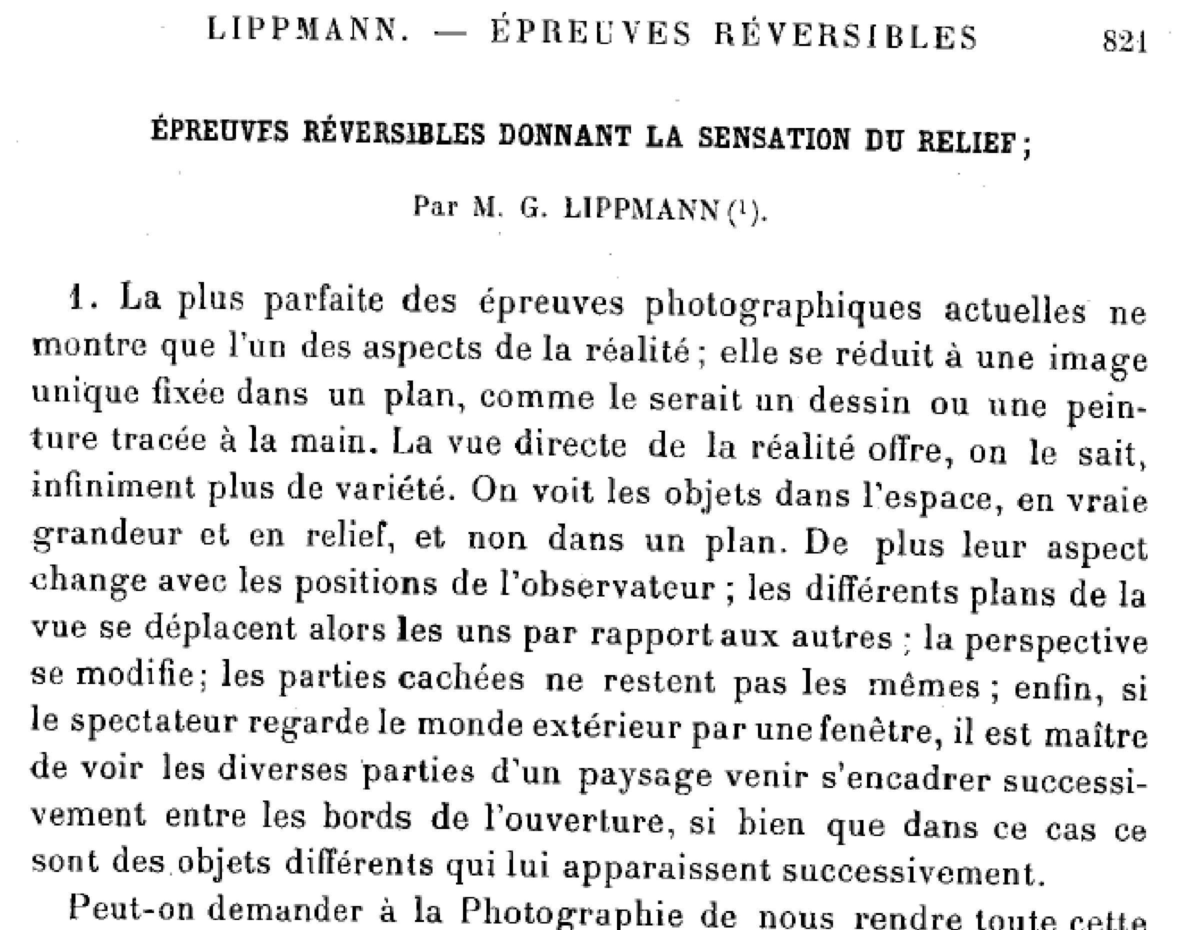
(c) Cross section



(a) (b)

M. Yamamoto, "Motion analysis using the visualized locus method," untranslated Japanese articles, 1981.

1908 Light Field Capture and Display



LIPPMANN. — ÉPREUVES RÉVERSIBLES 821

ÉPREUVES RÉVERSIBLES DONNANT LA SENSATION DU RELIEF;

Par M. G. LIPPMANN (¹).

1. La plus parfaite des épreuves photographiques actuelles ne montre que l'un des aspects de la réalité; elle se réduit à une image unique fixée dans un plan, comme le serait un dessin ou une peinture tracée à la main. La vue directe de la réalité offre, on le sait, infiniment plus de variété. On voit les objets dans l'espace, en vraie grandeur et en relief, et non dans un plan. De plus leur aspect change avec les positions de l'observateur; les différents plans de la vue se déplacent alors les uns par rapport aux autres; la perspective se modifie; les parties cachées ne restent pas les mêmes; enfin, si le spectateur regarde le monde extérieur par une fenêtre, il est maître de voir les diverses parties d'un paysage venir s'encadrer successivement entre les bords de l'ouverture, si bien que dans ce cas ce sont des objets différents qui lui apparaissent successivement.

Peut-on demander à la Photographie de nous rendre toute cette

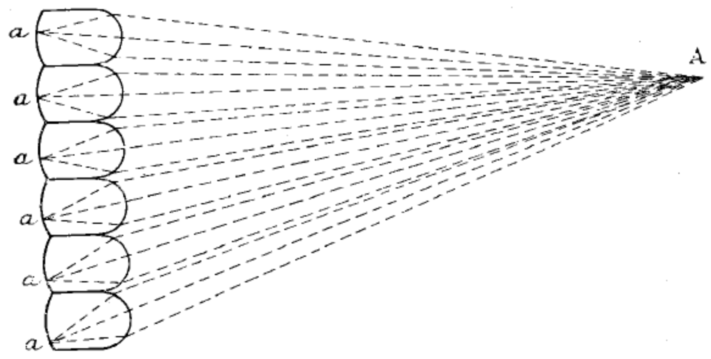
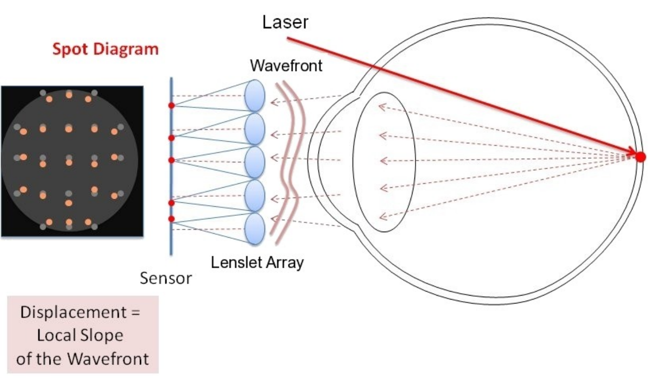


FIG. 1.

1900 Hartmann Mask for Astronomy

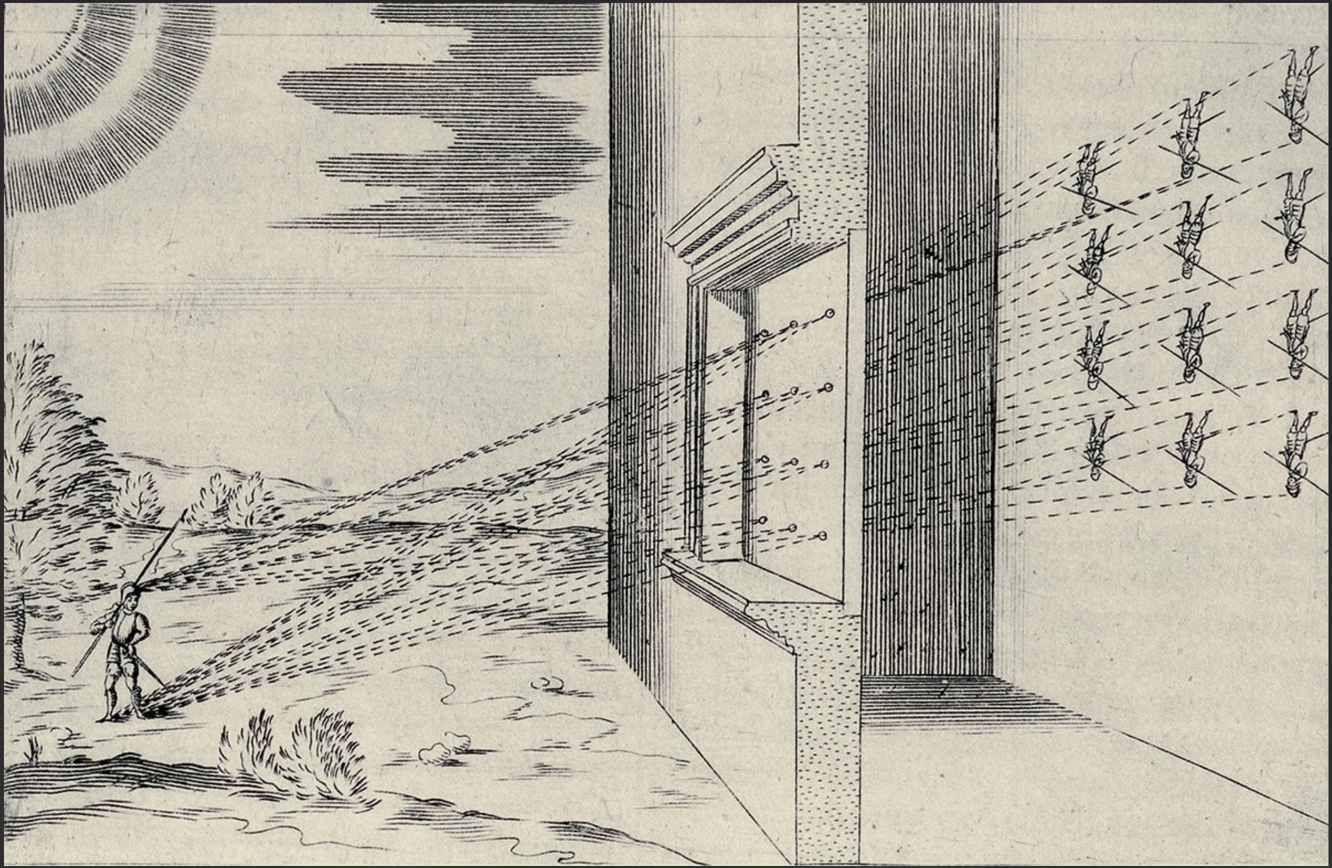
J. Hartmann, “Bemerkungen uber den bau und die justirung von spektrographen,” Z. Instrumentenkd, vol. 20, no. 47, p. 2, 1900.

...Evolved into Shack-Hartmann sensors for adaptive optics



R. V. Shack and B. C. Platt, “Production and use of a lenticular Hartmann screen,” Journal of the Optical Society of America, vol. 61, no. 5, p. 656, 1971.

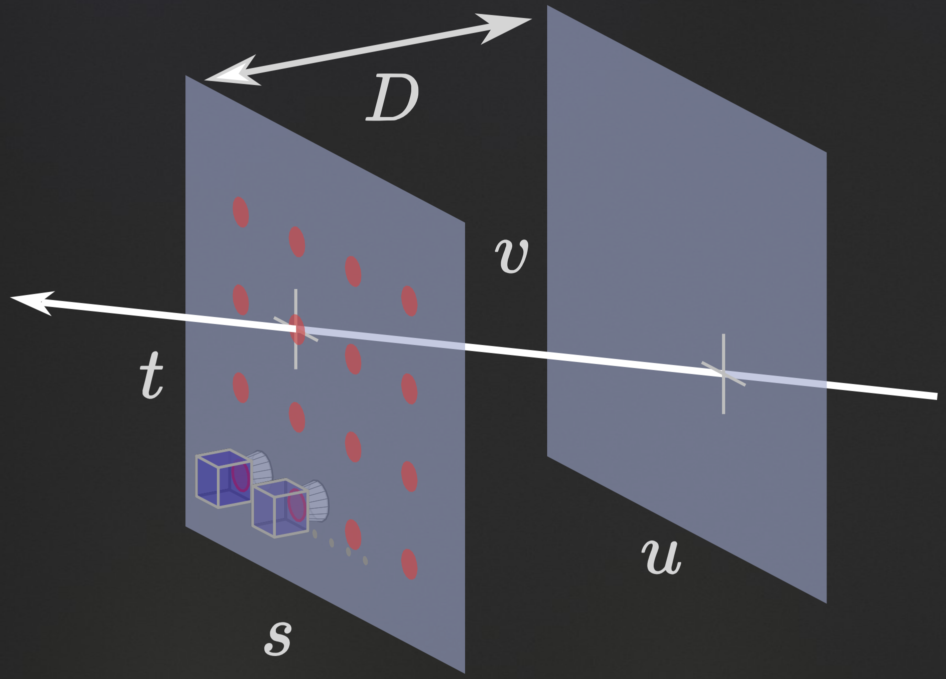
1642 A Light Field Camera Obscura

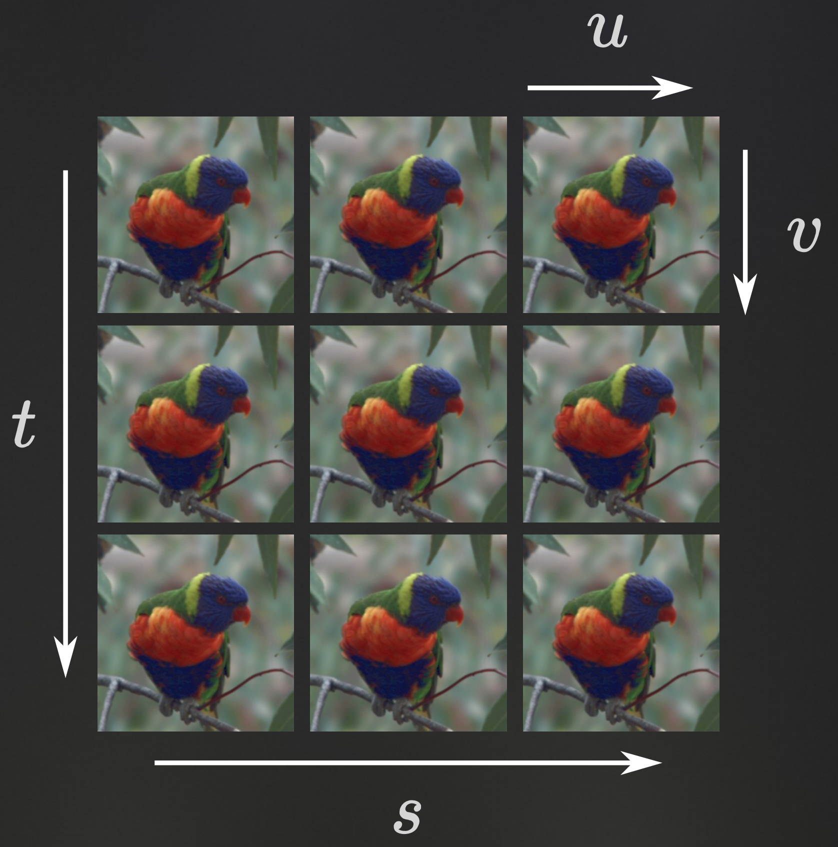


Mario Bettini "Apiaria universae philosophiae mathematicae", 1642

Parameterizations

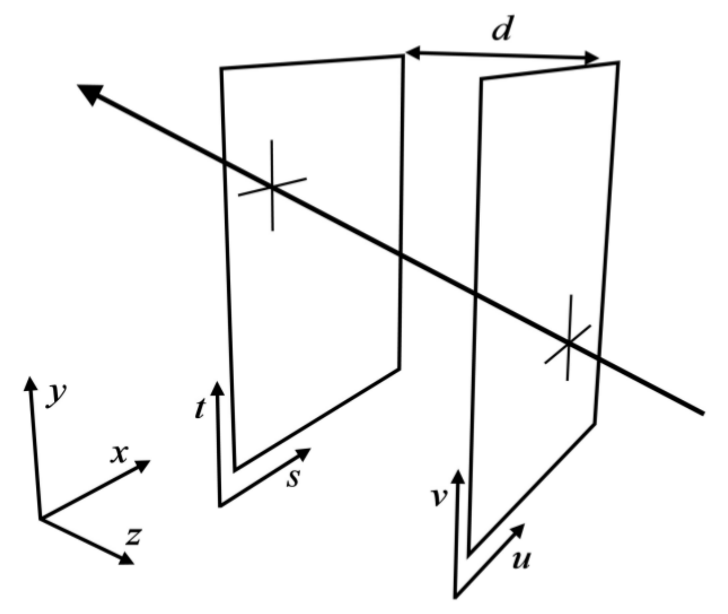
2-Plane Parameterization (2pp)



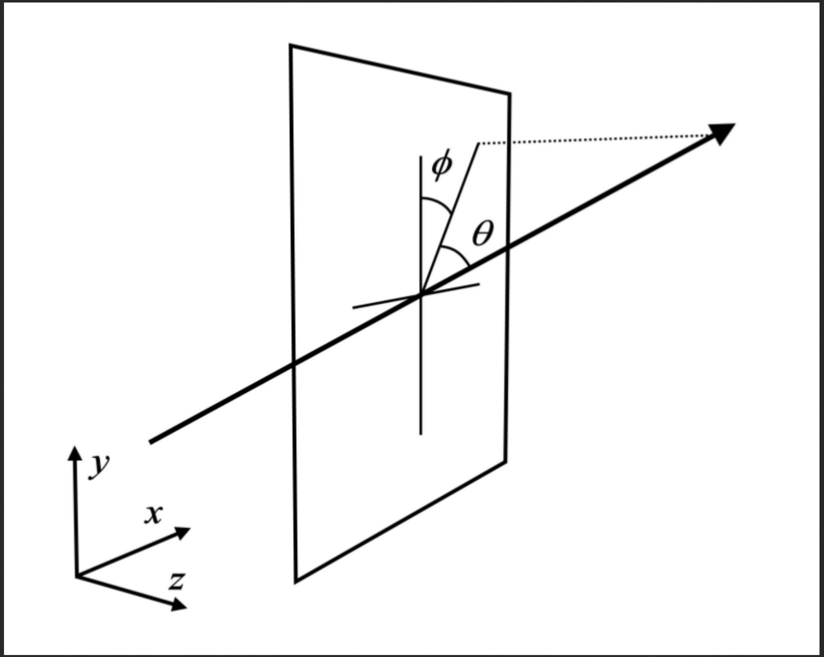


4D Image $\mathcal{L}(s, t, u, v)$

Planar vs Planar/Spherical



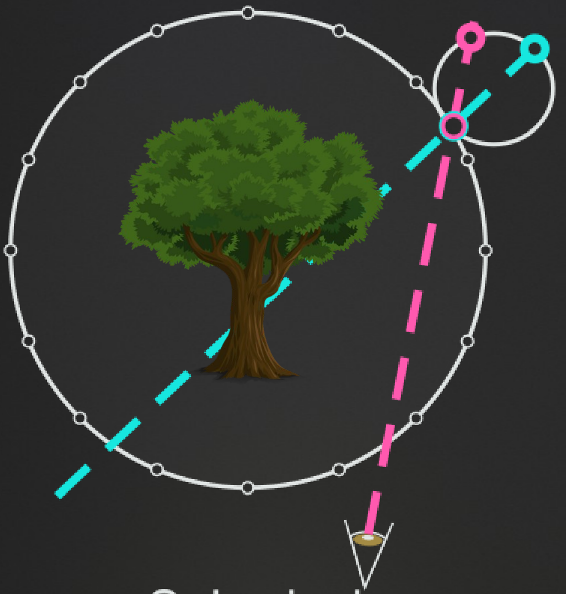
Two-Plane



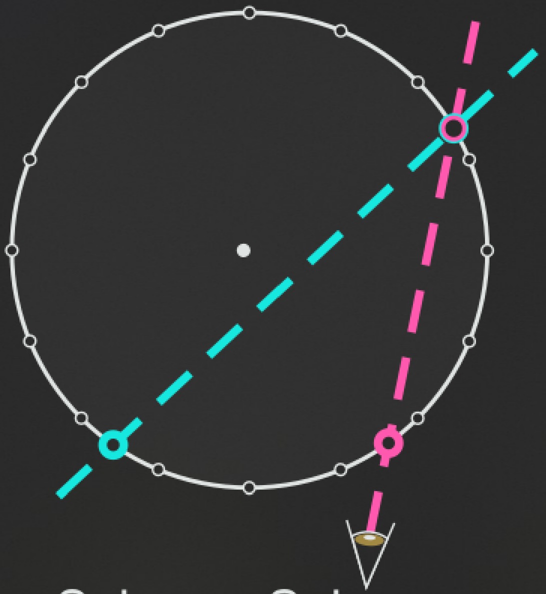
Position + Direction

Spherical

[Todt 2007]

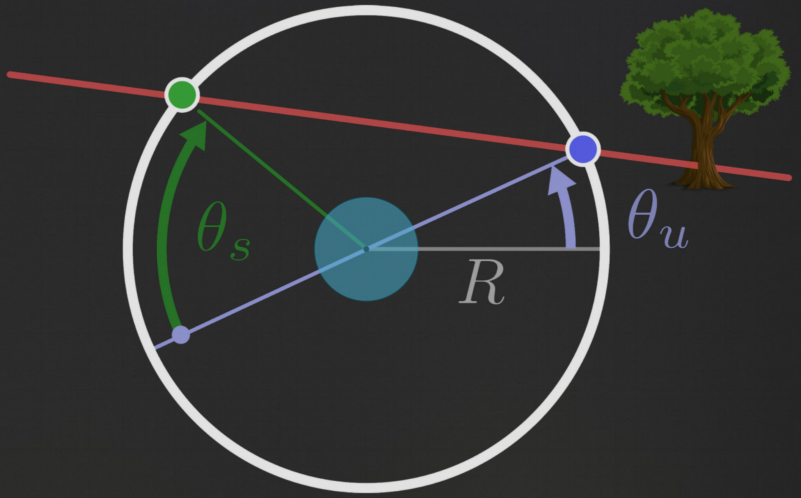


Spherical



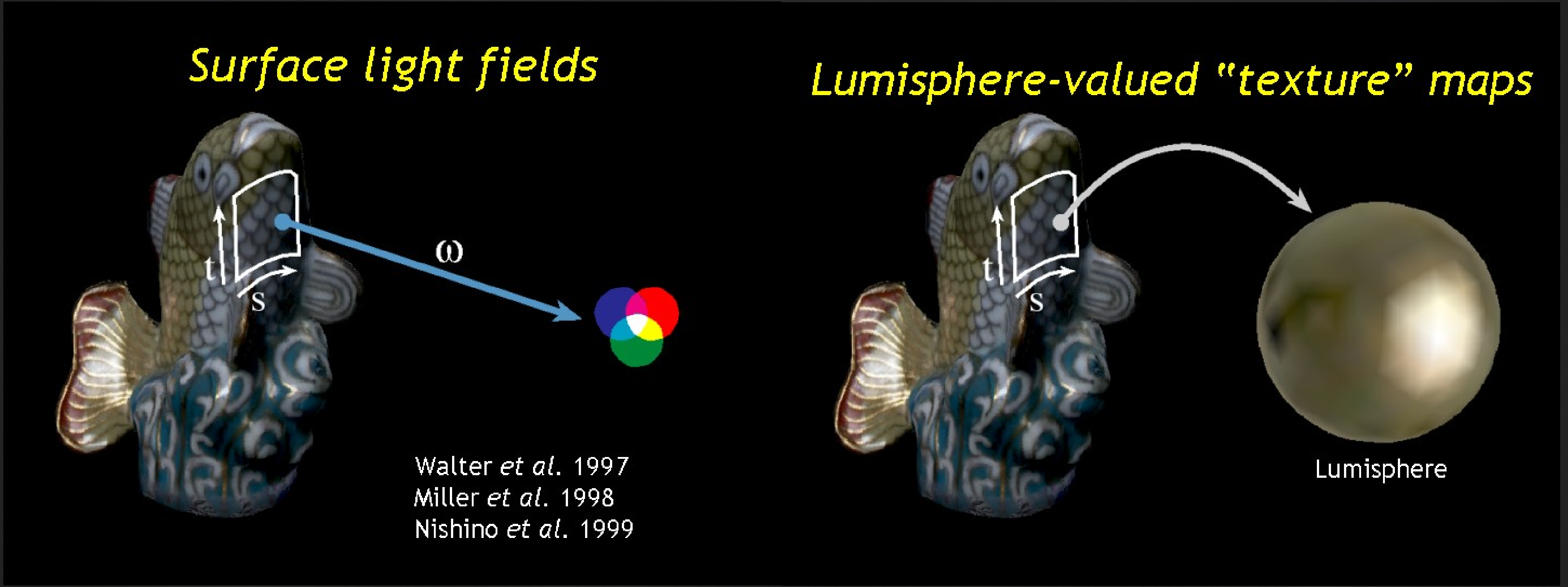
Sphere - Sphere

[Dansereau 2017]



Camera-centered
R = focal distance

Surface



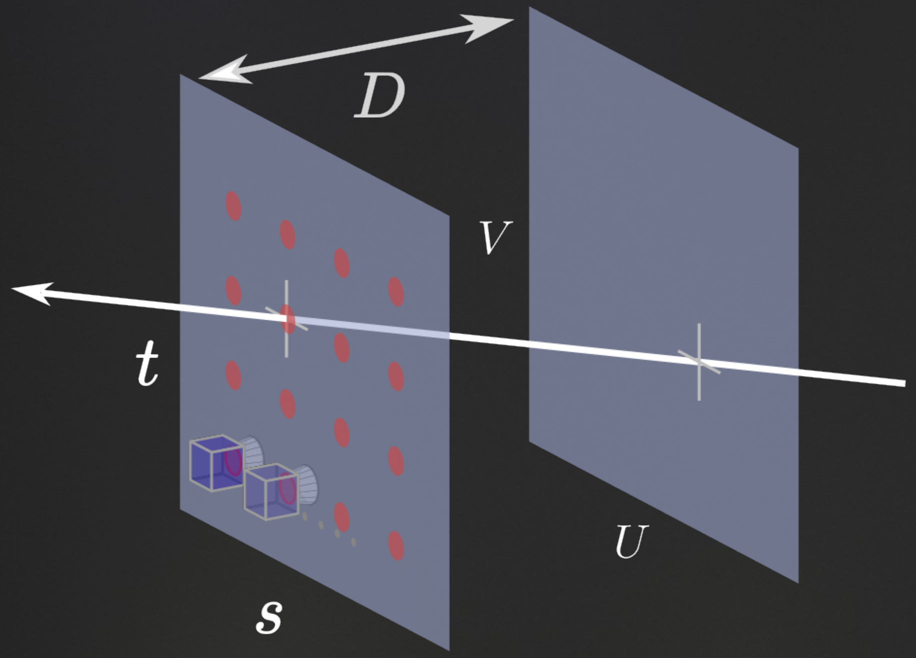
Surface light fields

Lumisphere-valued "texture" maps

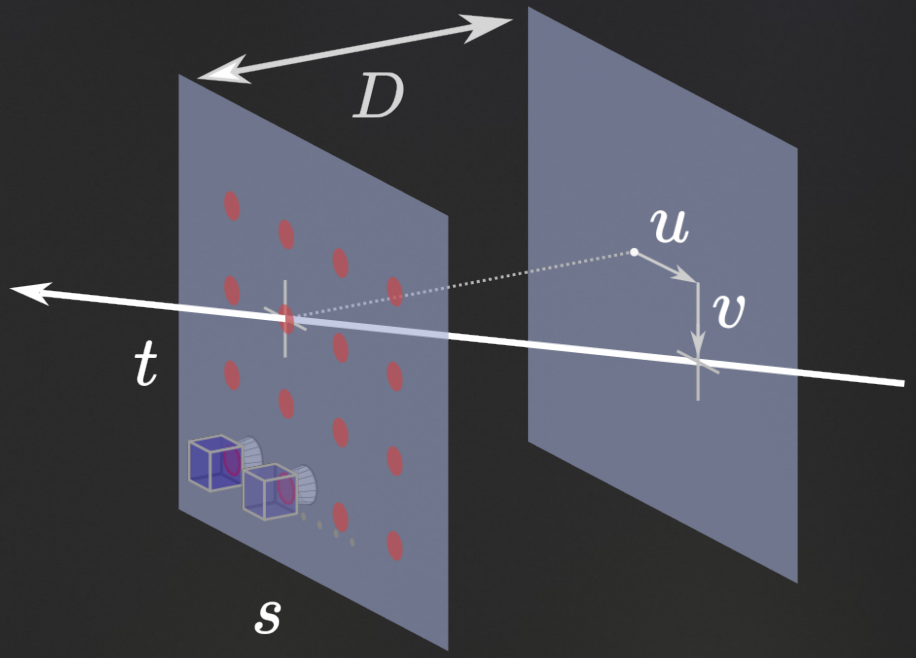
Walter *et al.* 1997
Miller *et al.* 1998
Nishino *et al.* 1999

[images c/o Wood et al, UWashington, http://grail.cs.washington.edu/projects/slf/papers/siggraph2000/talk/]

Absolute vs Relative 2pp

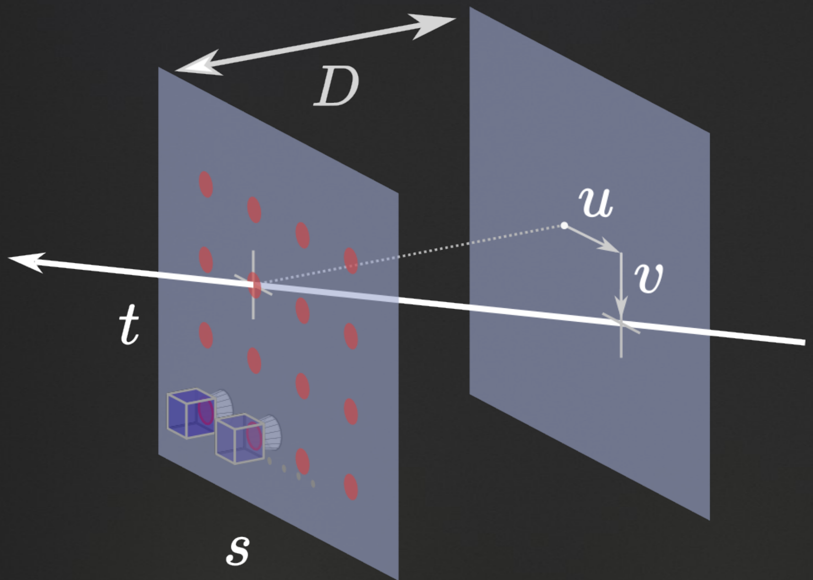


Absolute 2pp:
U,V relative to fixed point



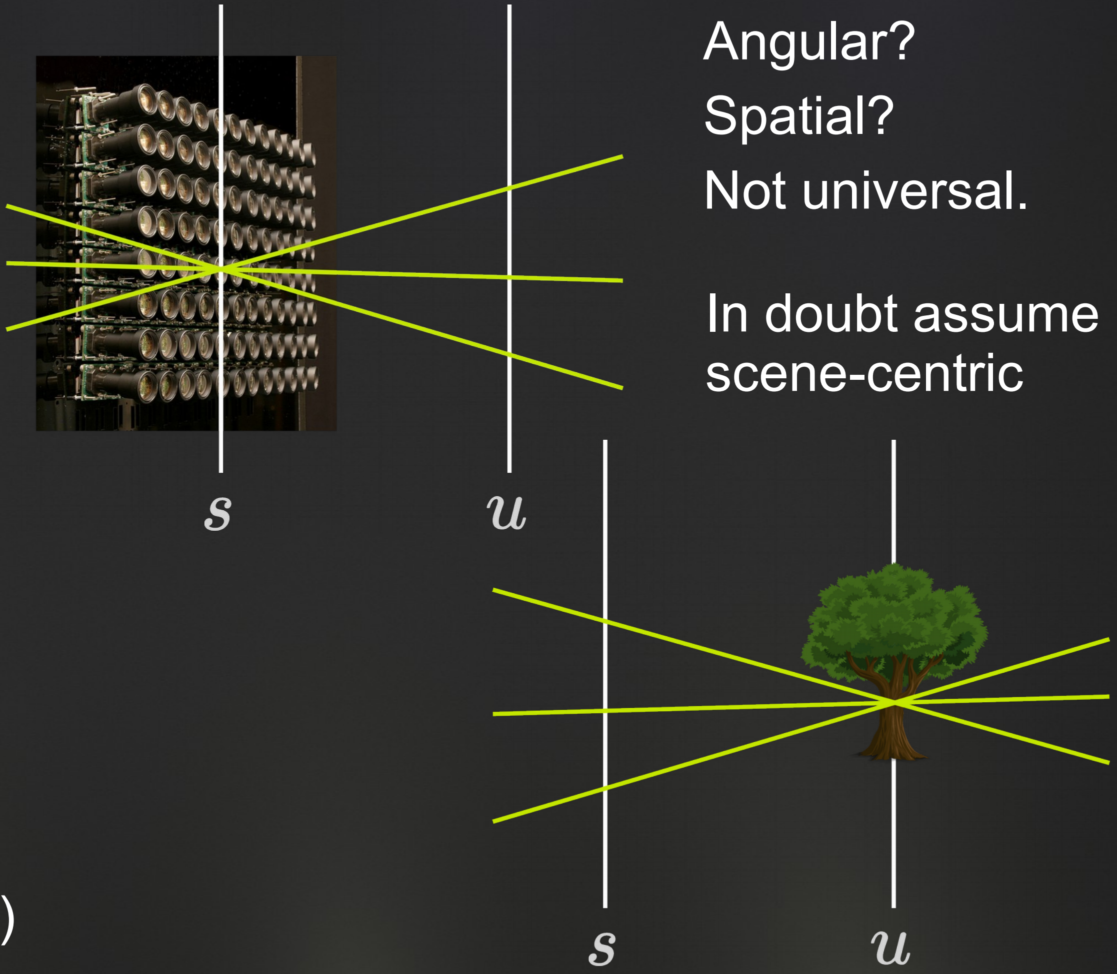
Relative 2pp:
u,v relative to s,t

A Caution on Terminology



s,t,u,v not universal, careful of order!

Mix of continuous (m) / sampled (index)



Angular?

Spatial?

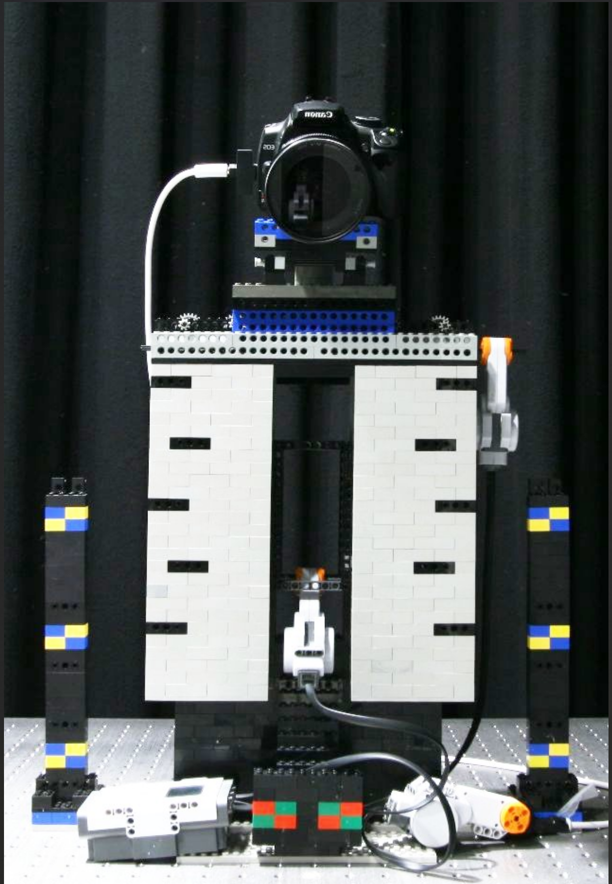
Not universal.

In doubt assume scene-centric

Visualizations

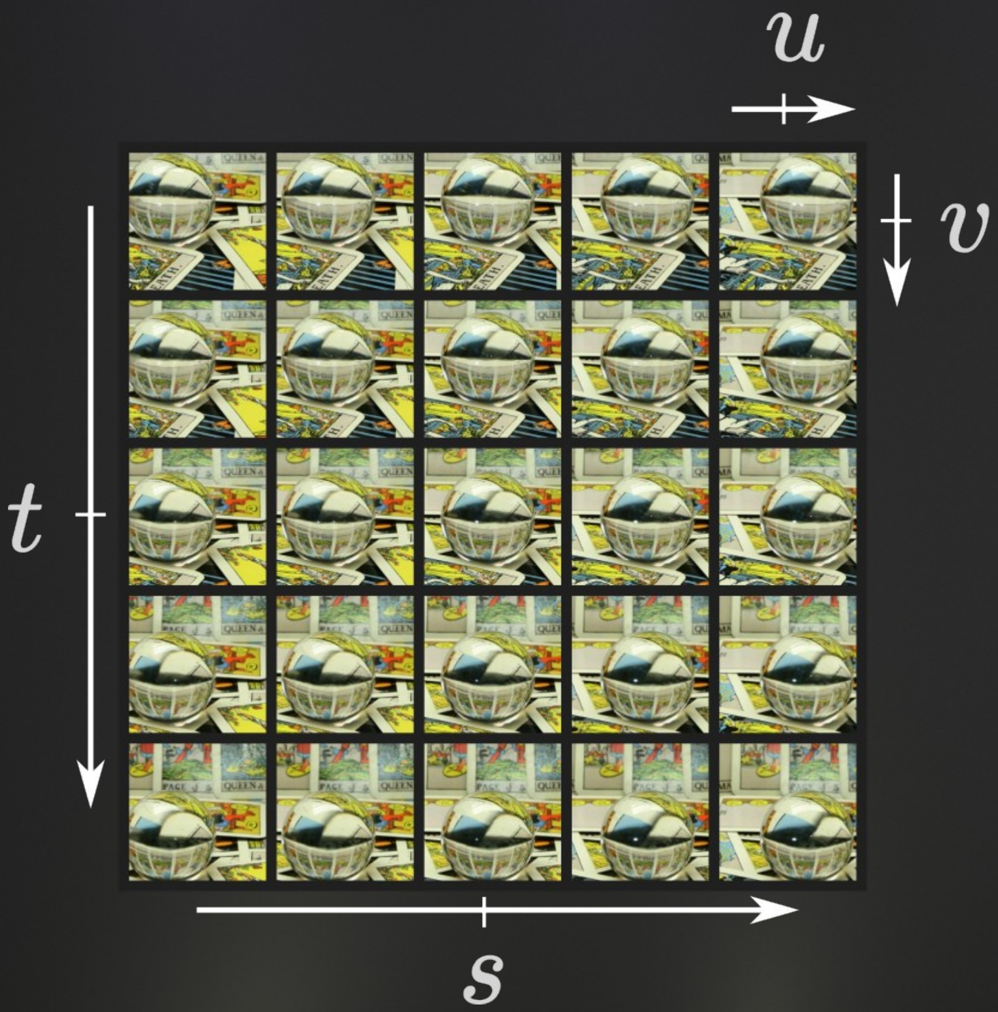
How many ways can you slice a 4D function into 2D slices?

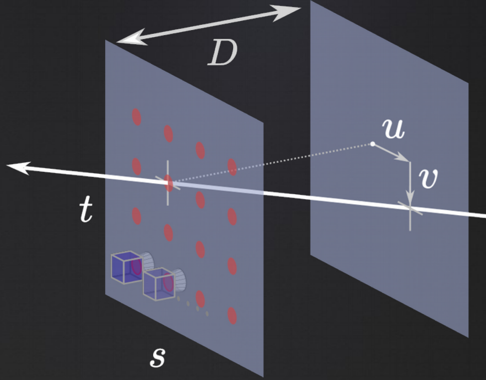
Visualizations



LF c/o Stanford Light Field Archive
http://lightfield.stanford.edu

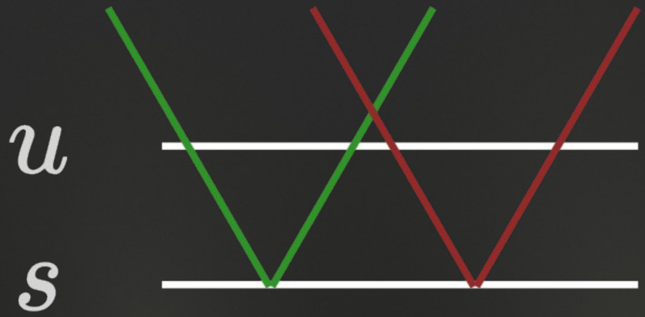
2D images in *u*,*v*



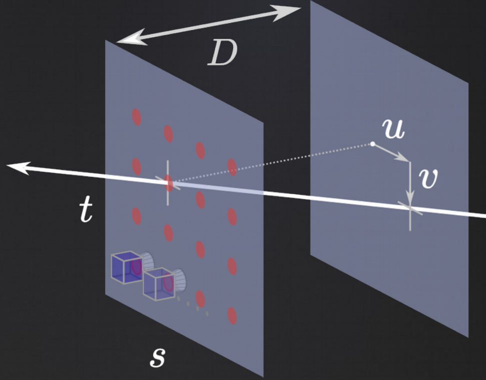


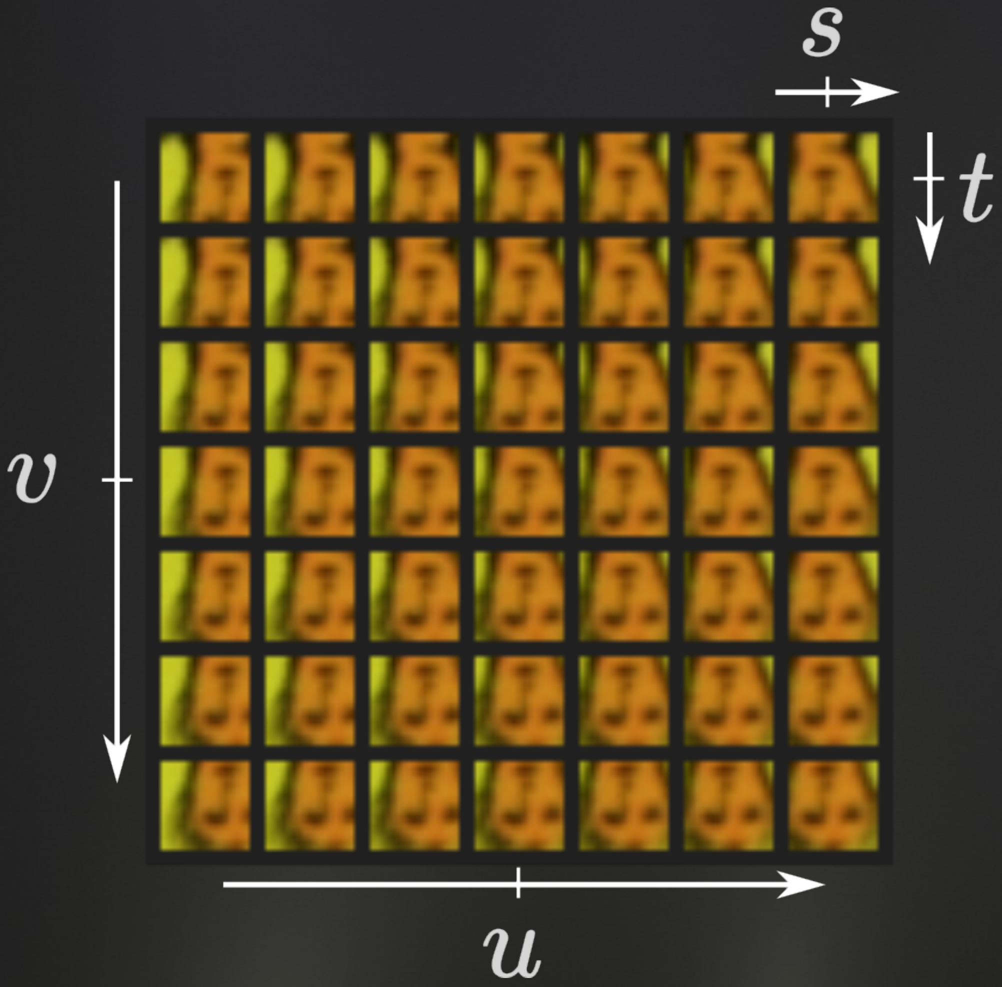
Each image:
$\mathcal{L}(u,v) \mid s,t$ fixed

For relative 2pp these are perspective images



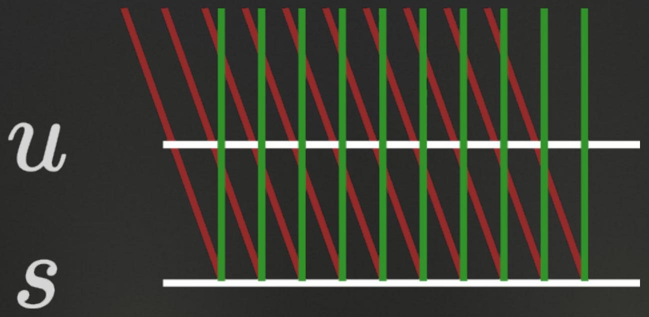
2D images in *s*,*t*



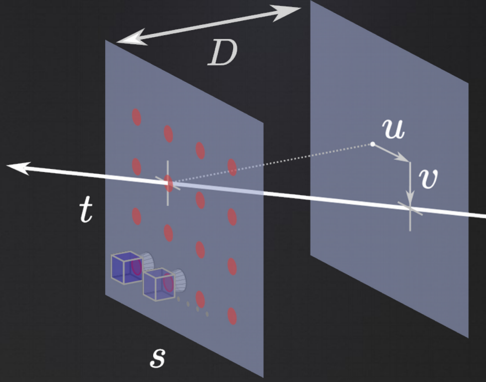


Each image:
$\mathcal{L}(\, s,t \,) \mid u,v$ fixed

For relative 2pp these
are orthographic images

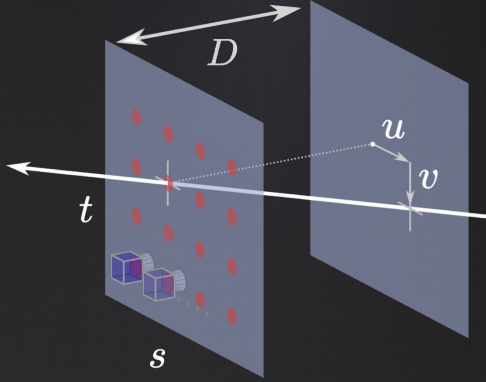


Still perspective image



$\mathcal{L}(u, v) \mid s,t$ fixed

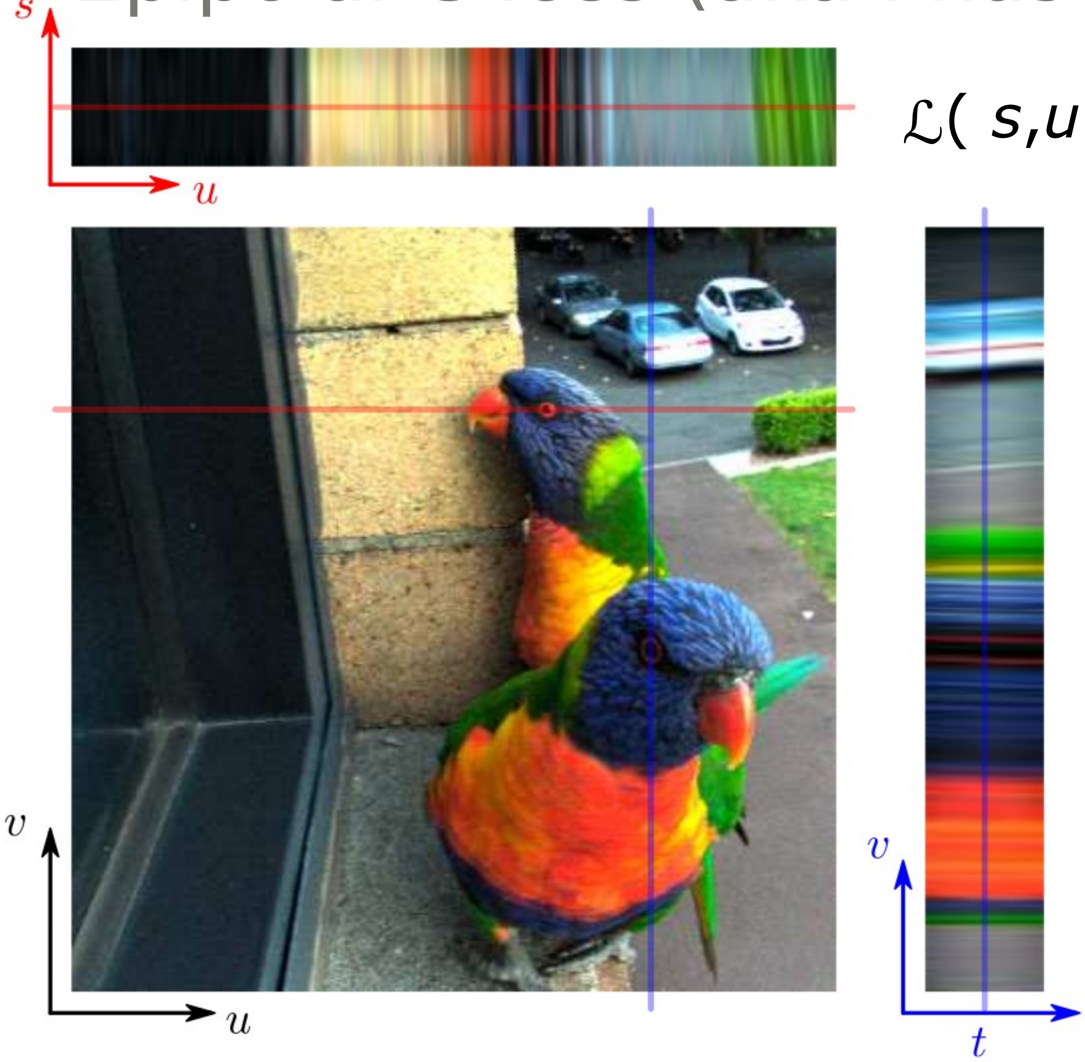
Animated perspective image



Video: panning around the s,t plane

$\mathcal{L}(\, u, v \,) \mid s,t$ animated

Epipolar Slices (aka Phase Space, EPIs)



$\mathcal{L}(s,u) \mid t,v$ fixed

$\mathcal{L}(t,v) \mid s,u$ fixed

More on these later...

Points to Ponder

We saw slices s,t; u,v; s,u; and t,v. What about the other combinations?

What should we call a 3D subset L(*s,u,v*)? Or a 2D subset L(*s,v*)?
When might each of these arise?

What sorts of tasks might be simpler in s,t slices than in u,v slices?
Simpler in s,u or t,v slices?

With what LF parameterizations could you represent all four walls, ceiling, and floor of a room? How about all views surrounding an object?