

Multi-View and Structured 3D Shape Representations

Richard (Hao) Zhang

CMPT 464/764: Geometric Modeling in Computer Graphics

Lecture 7

Outline

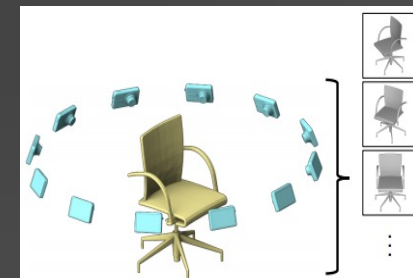
- Implicit representations
- Parametric representations
- Meshes (subdivision)
- Point clouds
- Voxels
- Multi-view representations
- Structured representations

Smooth curves and surfaces

Discrete representations

2D $\rightarrow$ 3D

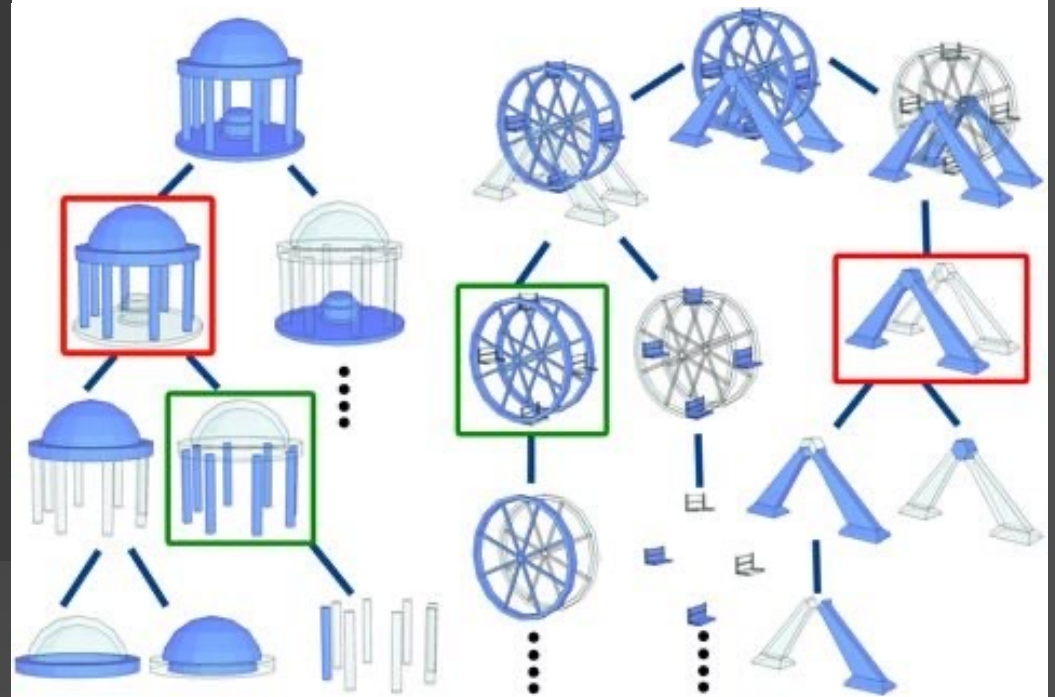
Parts + relations = structures



Outline

- Implicit representations
 - Parametric representations
 - Meshes (subdivision)
 - Point clouds
 - Voxels
-
- Multi-view representations
 - Structured representations

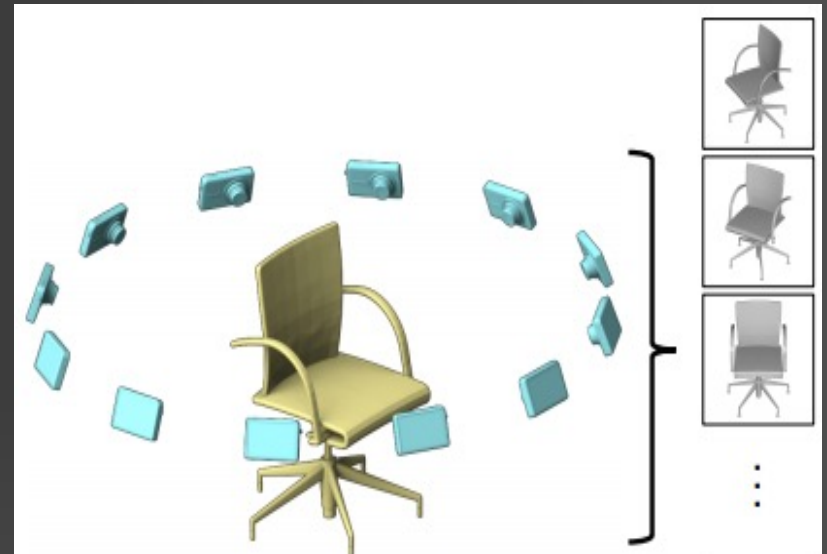
Symmetry hierarchies



Parts + relations = structures
Encompasses all low-level reps

Multi-view representation of 3D shapes

- Represent a 3D shape as a set of **images from one or more views**
- Technically not a 3D representation
- Why would this be a good idea?
 - Images are easy to acquire



Multi-view representation of 3D shape

- Represent a 3D shape as a set of 2D images from one or more cameras
- Technically not a 3D representation
- Why would this be a good representation?
 - Images are easy to acquire
 - For **visualization**, all we care about is to view the 3D shape from another angle — **novel view synthesis (NVS)**
 - Do we still need 3D then? 😊



Projective shape processing and analysis

[Wang et al. SIGGRAPH Asia 2013]

Projective Analysis for 3D Shape Segmentation

Yunhai Wang^{**} Minglun Gong^{†*} Tianhua Wang^{‡*} Daniel Cohen-Or^{*} Hao Zhang[§] Baoquan Chen^{*#}

^{*}Shenzhen VisuCA Key Lab/SIAT [†]Memorial University of Newfoundland [‡]Jilin University

^{*}Tel-Aviv University [§]Simon Fraser University [#]Shandong University

- Perform analysis in image space and then “back-project” into 3D

Projective shape processing and analysis

[Wang et al. SIGGRAPH Asia 2013]

Projective Analysis for 3D Shape Segmentation

Yunhai Wang^{**} Minglun Gong^{†*} Tianhua Wang^{‡*} Daniel Cohen-Or^{*} Hao Zhang[§] Baoquan Chen^{*#}

^{*}Shenzhen VisuCA Key Lab/SIAT [†]Memorial University of Newfoundland [‡]Jilin University

^{*}Tel-Aviv University [§]Simon Fraser University [#]Shandong University



- Perform analysis in image space and then “back-project” into 3D

Projective shape processing and analysis

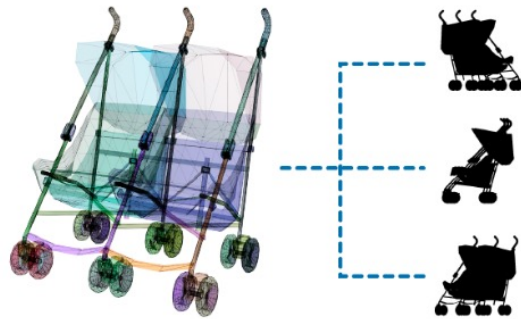
[Wang et al. SIGGRAPH Asia 2013]

Projective Analysis for 3D Shape Segmentation

Yunhai Wang^{**} Minglun Gong^{†*} Tianhua Wang^{‡*} Daniel Cohen-Or^{*} Hao Zhang[§] Baoquan Chen^{*##}

^{*}Shenzhen VisuCA Key Lab/SIAT [†]Memorial University of Newfoundland [‡]Jilin University

^{*}Tel-Aviv University [§]Simon Fraser University [#]Shandong University



- Perform analysis in multi-view image space ...

Projective shape processing and analysis

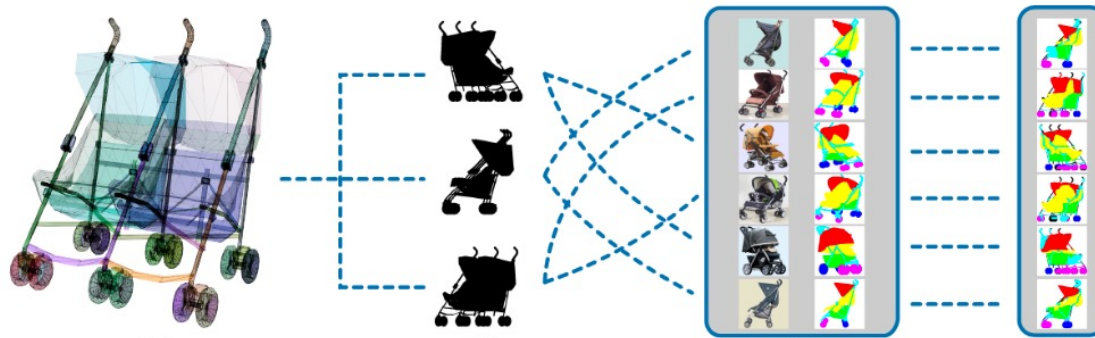
[Wang et al. SIGGRAPH Asia 2013]

Projective Analysis for 3D Shape Segmentation

Yunhai Wang^{**} Minglun Gong^{†*} Tianhua Wang^{‡*} Daniel Cohen-Or^{*} Hao Zhang[§] Baoquan Chen^{*#*}

^{*}Shenzhen VisuCA Key Lab/SIAT [†]Memorial University of Newfoundland [‡]Jilin University

^{*}Tel-Aviv University [§]Simon Fraser University [#]Shandong University



- Perform analysis in image space, leveraging **large volumes of 2D segmentation training data**

Projective shape processing and analysis

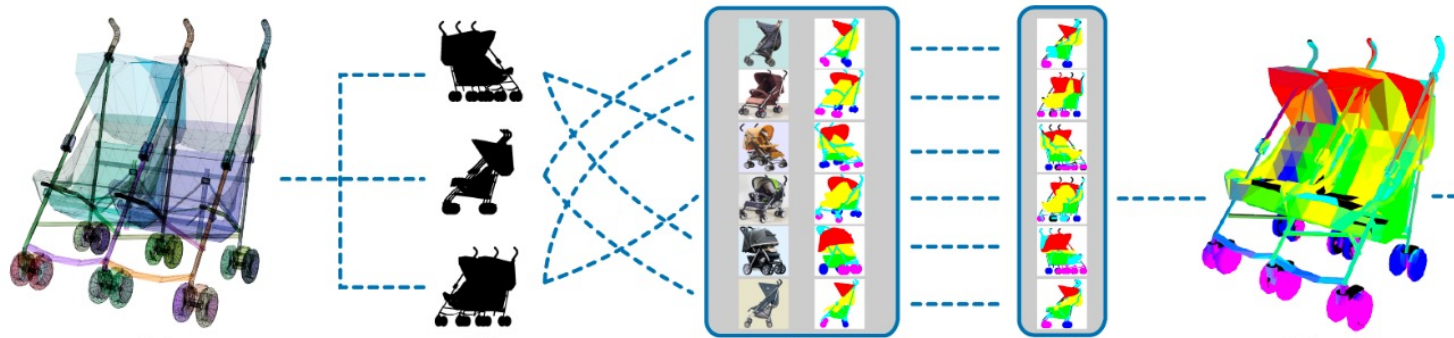
[Wang et al. SIGGRAPH Asia 2013]

Projective Analysis for 3D Shape Segmentation

Yunhai Wang^{**} Minglun Gong^{†*} Tianhua Wang^{‡*} Daniel Cohen-Or^{*} Hao Zhang[§] Baoquan Chen^{*#}

^{*}Shenzhen VisuCA Key Lab/SIAT [†]Memorial University of Newfoundland [‡]Jilin University

^{*}Tel-Aviv University [§]Simon Fraser University [#]Shandong University



- “Back-project” multi-view segmented images into 3D: non-trivial task

Paper: https://www.yunhaiwang.net/public_html/psa.html

Projective shape processing and analysis

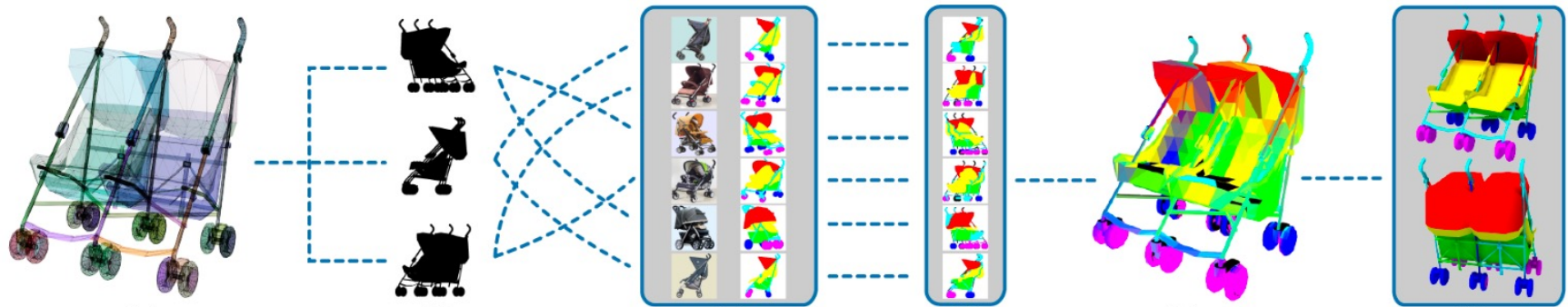
[Wang et al. SIGGRAPH Asia 2013]

Projective Analysis for 3D Shape Segmentation

Yunhai Wang^{**} Minglun Gong^{†*} Tianhua Wang^{‡*} Daniel Cohen-Or^{*} Hao Zhang[§] Baoquan Chen^{*##}

^{*}Shenzhen VisuCA Key Lab/SIAT [†]Memorial University of Newfoundland [‡]Jilin University

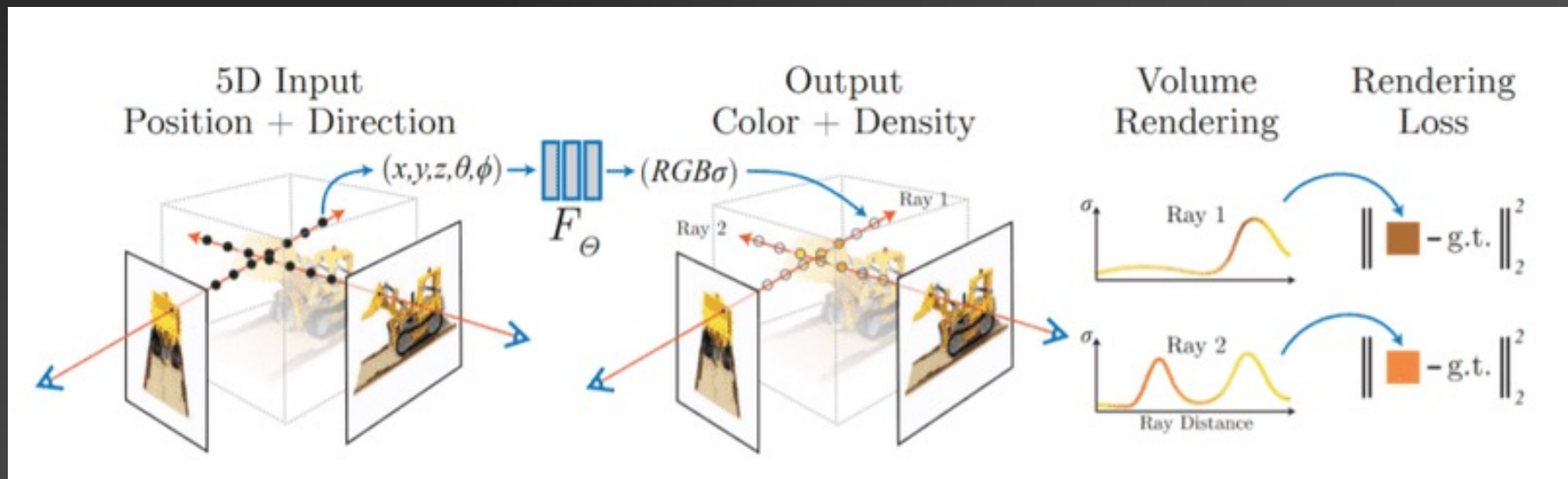
^{*}Tel-Aviv University [§]Simon Fraser University [#]Shandong University



- Perform analysis in image space and then “back-project” into 3D

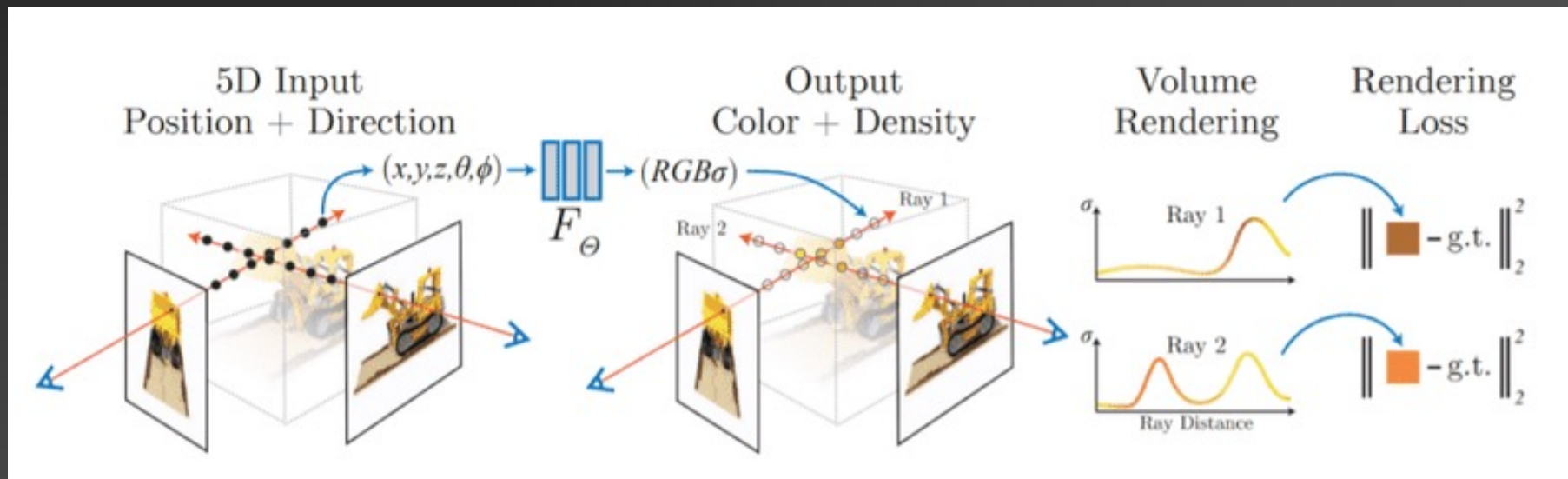
Novel view synthesis via neural networks

- **Neural Radiance Field (NeRF)** [Mildenhall et al. ECCV 2020]



Novel view synthesis via neural networks

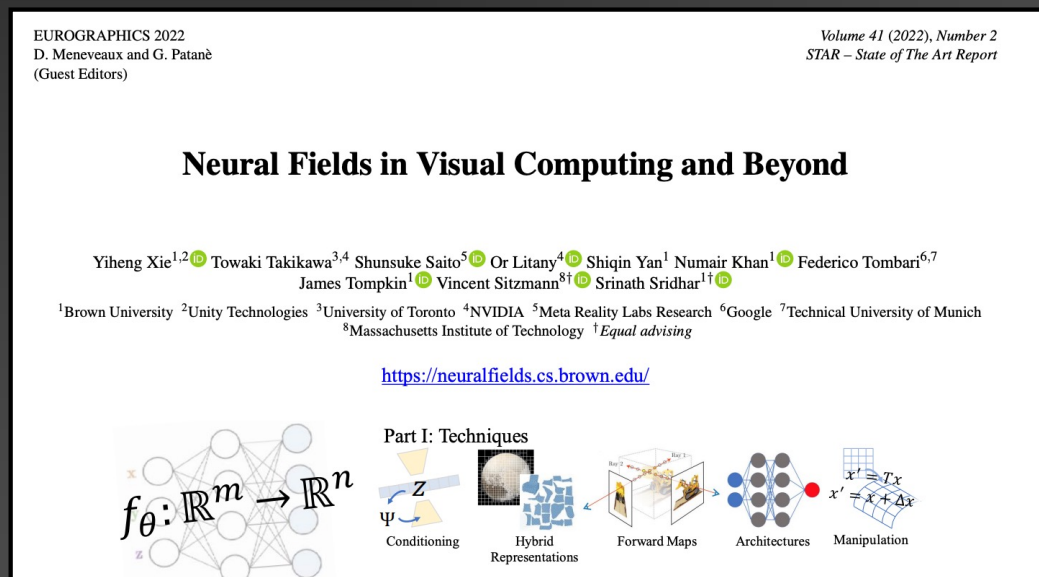
- Neural Radiance Field (NeRF) [Mildenhall et al. ECCV 2020]



- Input: multi-view images (a **dense** set in the initial NeRF paper)
- Output: a neural network trained via **differentiable rendering**, that can be “queried” for novel view synthesis through **volume rendering**

What is a neural field?

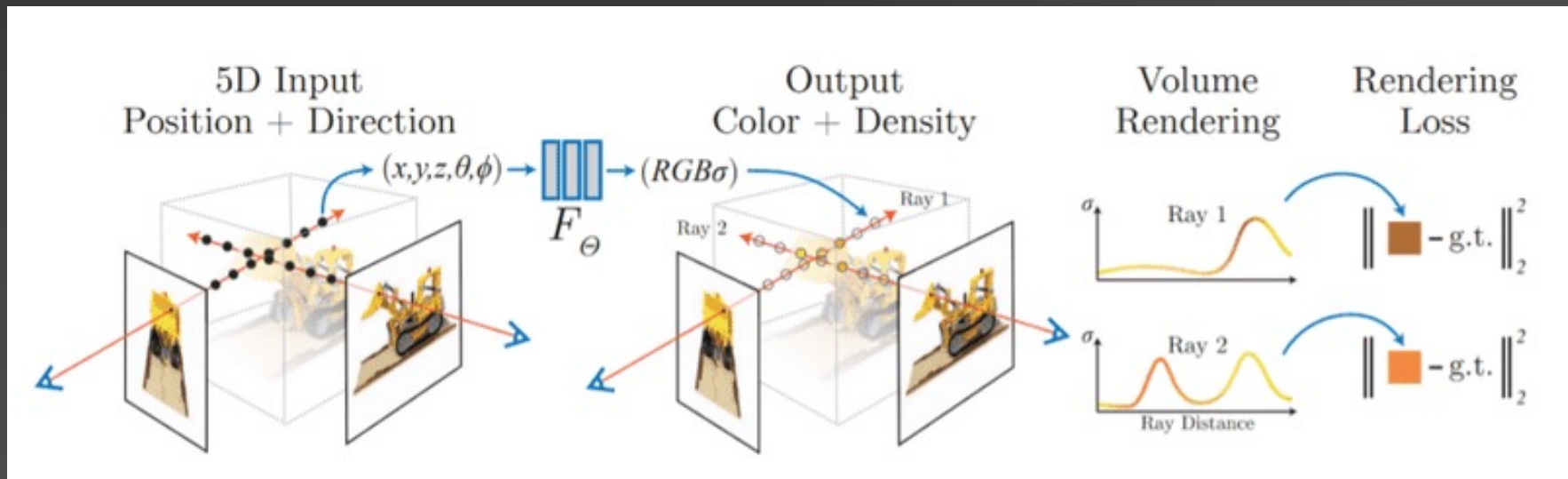
- A field of values (typically in voxel grids for 3D), e.g., occupancy or signed distance function (SDF), predicted using neural networks



Survey: <https://arxiv.org/abs/2111.11426>

What is a neural field?

- A field of values (typically in voxel grids for 3D), e.g., occupancy or signed distance function (SDF), predicted using neural networks
- NeRF: **color** and **density** per voxel for **volume rendering**



What is a neural field?

- A field of values (typically in voxel grids for 3D), e.g., occupancy or signed distance function (SDF), predicted using neural networks
- NeRF: color and density per voxel for **volume rendering**
- The 3D representation **is the neural network** that is trained on multi-view images: an **“overfit”** network (overfit to a specific scene)

What is a neural field?

- A field of values (typically in voxel grids for 3D), e.g., occupancy or signed distance function (SDF), predicted using neural networks
- NeRF: color and density per voxel for **volume rendering**
- The 3D representation is the **neural network** that is trained on multi-view images: an “**overfit**” network (overfit to a specific scene)
- A **compressed** representation that can be **quickly queried**

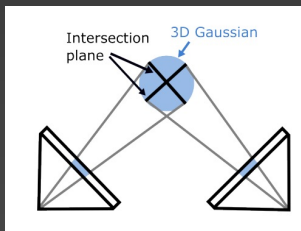
3D Gaussian splatting (3DGS)

- 3DGS [Kerbl et al. SIG 2023]: each **(point) splat** is a **3D Gaussian**
- Also from multi-view images as input and produces a **PBR**
- Surpassing NeRF (2000) in rendering speed, but with higher storage



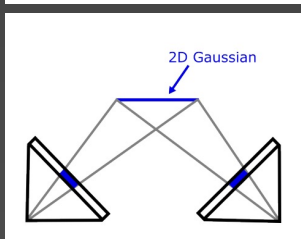
3D Gaussian splatting (3DGS)

- 3DGS [Kerbl et al. SIG 2023]: each (point) splat is a 3D Gaussian
- Also from multi-view images as input and produces a PBR
- Surpassing NeRF (2000) in rendering speed, but with more storage
- No canonical notion of object surface — “What is the **right surface**?”



3D Gaussian Splatting for Real-Time Radiance Field Rendering

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis



2D Gaussian Splatting for Geometrically Accurate Radiance Fields

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, Shenghua Gao

Efficiency: most “3D shapes” are two-dimensional!

Critical issue #1

- NeRF and 3DGS: **rendering, not modeling** primitives

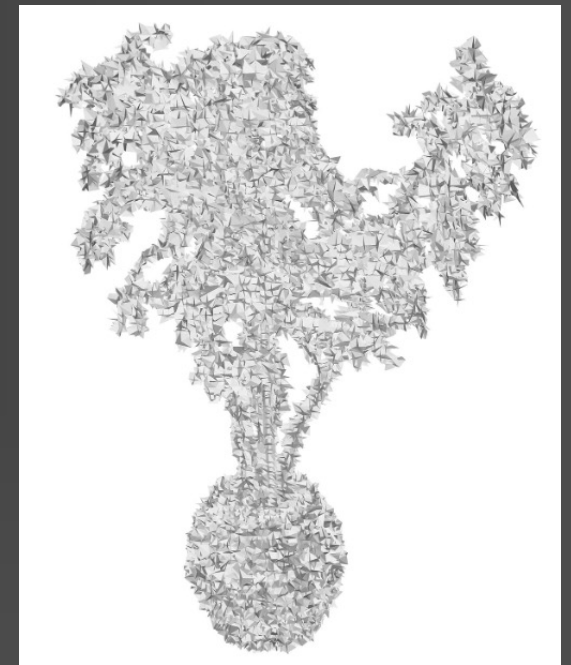
Images from MobileNeRF [Chen et al. CVPR 2023]



Critical issue #1

☹️ NeRF and 3DGS: **rendering, not modeling** primitives

- Unstructured: not how human reasons about 3D
- Not editable/reusable
- Not functional
- Non-compact



Images from MobileNeRF [Chen et al. CVPR 2023]

Critical issue #2

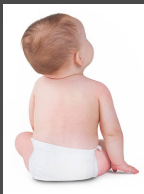
- 🙄 NeRF and 3DGS: rendering vs. modeling primitives
- 🙄 Avoiding 3D supervision (all diff-rendering) is **unnatural**



vs.

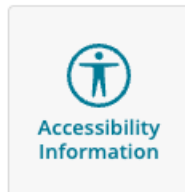
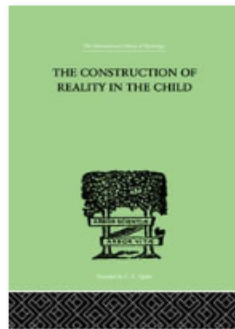


Learning by **interacting** in 3D



How do infants & children “learn 3D?” (aside)

- Well-known theory from Jean Piaget (1954)



Book

The Construction Of Reality In The Child

By Jean Piaget

Edition 1st Edition

First Published 1954

eBook Published 3 July 2013

Pub. Location London

Imprint Routledge

DOI <https://doi.org/10.4324/9781315009650>

Pages 400

eBook ISBN 9781315009650

Subjects Behavioral Sciences



Share



Citation

How do infants & children “learn 3D?” (aside)

- Well-known theory from Jean Piaget (1954)

✦ AI Overview

Piaget's theory of cognitive development proposes that **children progress through four distinct stages (Sensorimotor, Preoperational, Concrete Operational, and Formal Operational) as they construct their understanding of the world**. Key processes like assimilation and accommodation, which involve fitting new information into existing mental schemas or adapting schemas for new information, drive this development. Children move toward mental equilibrium by adjusting their understanding through these processes, leading to more complex and abstract thinking. 



The Four Stages of Cognitive Development

- Sensorimotor Stage (Birth to ~2 years)**
 - Infants learn about the world through their senses and actions. 
 - A key achievement is the development of object permanence, the understanding that objects continue to exist even when out of sight. 

How do infants & children “learn 3D?” (aside)

- Well-known theory from Jean Piaget (1954)

- Preoperational Stage (~2 to ~7 years)**

- Children use words and images to represent things and begin to think symbolically.
 - Egocentrism:** is common, where a child struggles to see things from another person's perspective.

- Concrete Operational Stage (~7 to ~11 years)**

- Children can think logically about concrete events and understand concepts like reversibility, the idea that things can be changed and then returned to their original state.
 - They are limited in their ability to think about abstract or hypothetical situations.

- Formal Operational Stage (~11 years and older)**

- Individuals gain the ability to think abstractly, use hypothetical reasoning, and understand complex ideas like metaphors and analogies.
 - This stage allows for the development of scientific thinking and abstract theories.

How do infants & children “learn 3D?” (aside)

Part of [Stages of Cognitive Development Guide](#) ✓

Support and Criticism of Piaget's Stage Theory

By [Kendra Cherry, MEd](#) | Updated on November 08, 2023

✓ Reviewed by [Amy Morin, LCSW](#)



Credit: Traumrune/Wikimedia Commons/CC BY 3.0

Think of an **object class**, e.g., cars, lamps, chairs, buildings, etc., and what **properties or attributes** you are most sure of about it.



Predictability of object attributes from class

Which object attribute is **more predicable** from class label, e.g., chairs?

- ❖ Shape?
- ❖ Topology?



Image taken from [dreamstimes.com](#)

Predictability of object attributes from class

Which object attribute is **more predictable** from class label, e.g., chairs?

- ❖ Color?
- ❖ Texture?
- ❖ Material?



Image taken from pinterest.ca

Is this a chair or not?

Why is this not a chair? Or is it?



What makes a chair a chair?

CVPR
2011

What Makes a Chair a Chair?

Helmut Grabner¹

Juergen Gall¹

Luc Van Gool^{1,2}

¹Computer Vision Laboratory
ETH Zurich

{grabner,gall,vangool}@vision.ee.ethz.ch

²ESAT - PSI / IBBT
K.U. Leuven

luc.vangool@esat.kuleuven.be

Abstract

Many object classes are primarily defined by their functions. However, this fact has been left largely unexploited by visual object categorization or detection systems. We propose a method to learn an affordance detector. It identifies locations in the 3d space which "support" the particular function. Our novel approach "imagines" an actor performing an action typical for the target object class, instead of relying purely on the visual object appearance. So, function is handled as a cue complementary to appearance, rather than being a consideration after appearance-based detection. Experimental results are given for the functional category "sitting". Such affordance is tested on a 3d representation of the scene, as can be realistically obtained through SfM or depth cameras. In contrast to appearance-based object detectors, affordance detection requires only very few training examples and generalizes very well to other sittable objects like benches or sofas when trained on a few chairs.

1. Introduction

"An object is first identified as having important functional relations, [...] perceptual analysis is derived of the functional concept [...]."

Nelson, 1974, [17]

"Affordances relate the utility of things, events, and places to the needs of animals and their actions in fulfilling them [...]. Affordances themselves are perceived and, in fact, are the essence of what we perceive."

Gibson, 1982, [8, p. 60]

"There's little we can find in common to all chairs – except for their intended use."

Minsky, 1986, [16, p. 123]

"[...] objects like coffee cups are artifacts that were created to fulfill a function. The function of an object plays a critical role in processing that object [...] for categorization and naming."

Carlson-Radvansky et al., 1999, [4]



Figure 1. The "chair-challenge" by I. and H. Bülthoff [3] (reprint with the author's permission).

These quotes emphasize that functional properties or affordances¹ are essential for forming concepts and learning object categories. Experiments (e.g. [18, 4]) have demonstrated that both appearance and function are strong cues for learning by infants. Initially they attend only to the form of an object. Later they use form and function and finally (by the age of 18 months) they attend to the relationships between form and function. Furthermore, Booth and Waxman [2] have identified two salient cues that facilitate categorization in infancy, namely (i) object functions and (ii) object names. Moreover, names of objects most often evolve on the basis of function².

Whereas all this is well known for a long time, it has been left mostly unused for object detection in computer vision. Taking a look at the results of the recent Pascal VOC Challenge [5], the performance still strongly depends

¹Affordances: A situation where an object's sensory characteristics intuitively apply its functionality and use. [...] A chair, by its size, its curvature, its balance, and its position, suggests sitting on it". <http://www.usabilityfirst.com/glossary/affordance>, 2010/07/28. Introduced in 1979 by Gibson [9, p. 127] based on the verb *afford*.

²When considering the evolution of a word for an object, most of the time it is based on its function. For example the word "chair": PIE base *sed- (to sit) → Latin *sedentarius* (sitting, remaining in one place) → *sedentary* (meaning "not in the habit of exercise") → *cathedral* → *chair*. <http://www.etymonline.com>, 2010/10/02.

"There's little we can find in common to all chairs – except for their intended use."

Marvin Minsky: "The Society of Mind" [1986]

Marvin Minsky's "The Society of Mind"

12.5 THE FUNCTIONS OF STRUCTURES

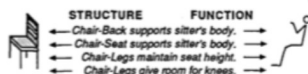
Many things that we regard as physical are actually psychological. To see why this is so, let's try to say what we mean by "chair." At first it seems enough to say:

"A chair is a thing with legs and a back and seat."

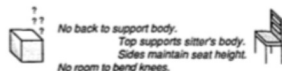
But when we look more carefully at what we recognize as chairs, we find that many of them *do not fit this Association because they don't divide into those categories quite. When all is done, there's little we can find in common to all chairs—except for their intended use.*

"A chair is something you can sit upon."

With many, many, pictures and questions, it is hard to make an attempt to show what we mean by this. The solution is that we need to combine at least two different kinds of descriptions. On one side, we need structural descriptions for recognizing chairs when we see them. On the other side we need functional descriptions in order to know what we can do with chairs. We can capture more of what we mean by interweaving both ideas, but it's not enough merely to propose a vague association, because in order for it to have some use, we need more intimate details about how those chair parts actually help a person to sit. To catch the proper meaning, we need connections between parts of the chair structure and the requirements of the human body that those parts are supposed to serve. Our network needs details like these:



Without such knowledge, we might just crawl under the chair or try to wear it on our head. But with that knowledge we can do amazing things, like applying the concept of a chair to see how we could sit on a box, even though it has no legs or back!



Unframes that include structures like this can be powerful. For example, such knowledge about relations between structure, comfort, and posture could be used to understand when a box could serve as a chair: that is, only when it is of suitable height for a person who does not require a backrest or room to bend the knees. To be sure, such clever reasoning requires special mental skills with which to redescribe or "reformulate" the descriptions of both box and chair so that they "match" despite their differences. Until we learn to make old descriptions fit new circumstances, our old knowledge can be applied only to the circumstances in which it was learned. And that would scarcely ever work, since circumstances never repeat themselves perfectly.

LEARNING MEANING

123

"There's little we can find in common to all chairs — except for their intended use."

"... we need to combine at least two different kinds of descriptions (of objects). On one side, we need **structural descriptions** for recognizing chairs when we see them. On the other side, we need **functional descriptions** in order to know what we can do with chairs."

Predictability of shape structures

- More predictability in the structure of objects from the same class, than other attributes such as color, shape, etc.
- This is due to the common functionality of the objects

Predictability of shape structures

- More predictability in the structure of objects from the same class, than other attributes such as color, shape, etc.
- This is due to the common functionality of the objects
- Functionality is characterized by object parts + relations,

Dictionary

Definitions from [Oxford Languages](#) · [Learn more](#)



struc·ture

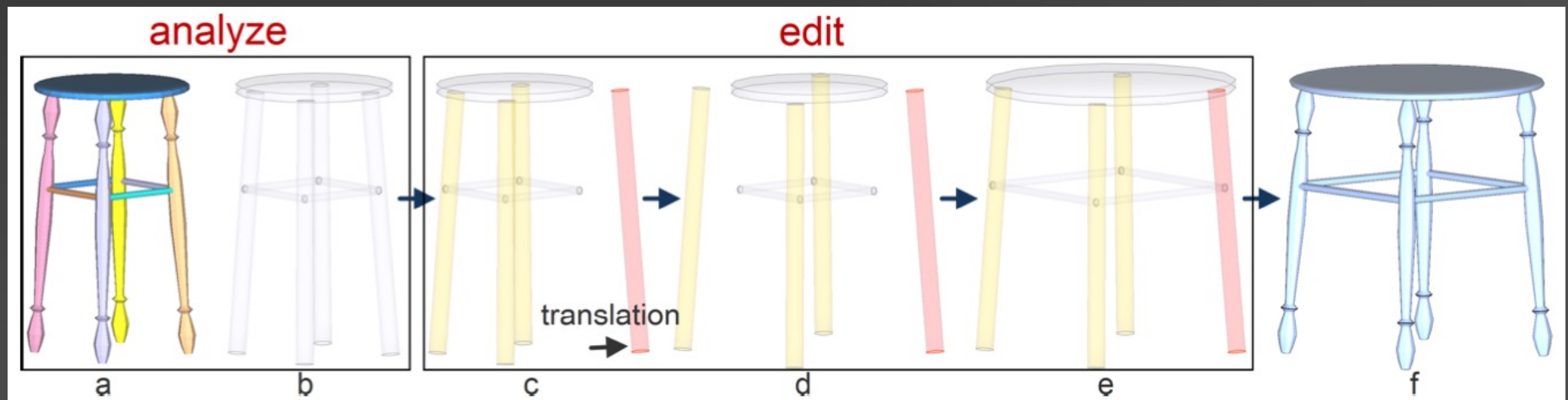
/ˈstræk(t)SHər/

noun

the arrangement of and relations between the parts or elements of something complex.

Structure relations and application

- Part symmetry, repetitions, proximity, angles, relative positioning (of parts) such as co-planarity, etc.
- Application: structure-aware editing

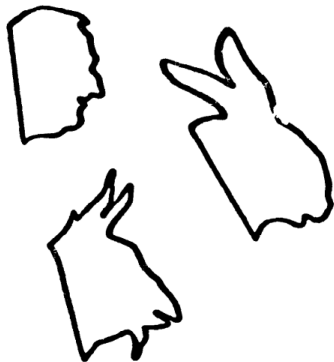


[Zheng et al. 2010]

Mimicking human perception

Cognition, 18 (1984) 65–96

2



Parts of recognition*

D.D. HOFFMAN
University of California, Irvine

W.A. RICHARDS
Massachusetts Institute of Technology

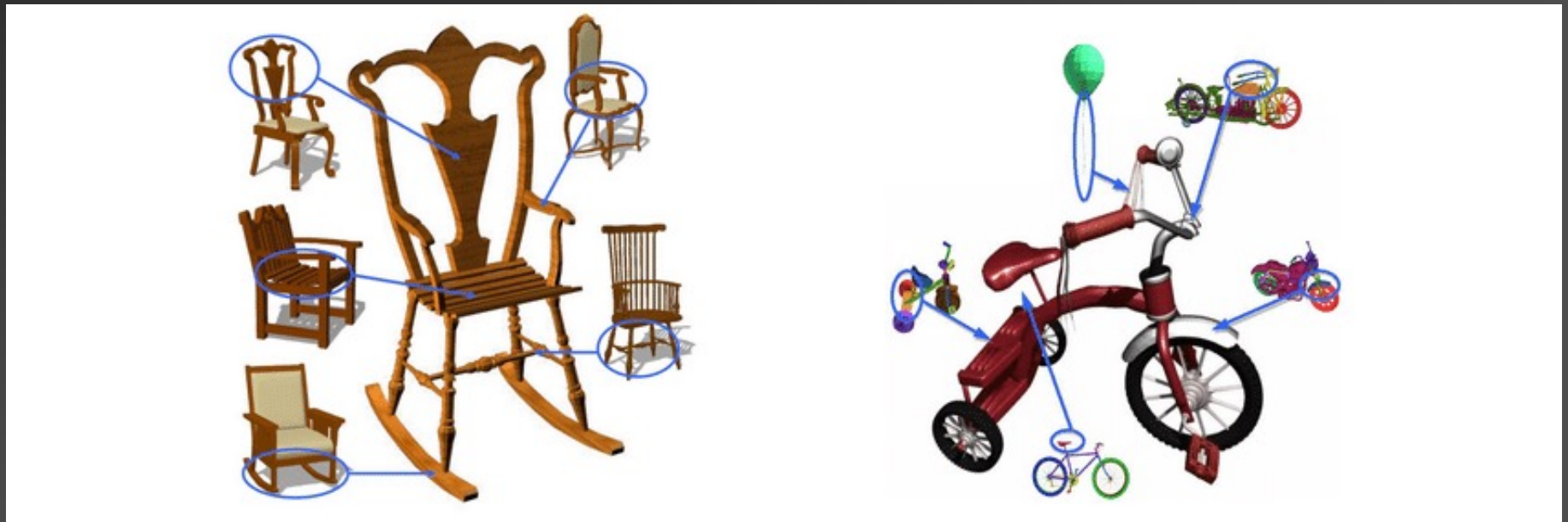
Abstract

We propose that, for the task of object recognition, the visual system decomposes shapes into parts, that it does so using a rule defining part boundaries rather than part shapes, that the rule exploits a uniformity of nature—transversality, and that parts with their descriptions and spatial relations provide a first index into a memory of shapes. This rule allows an explanation of several visual illusions. We stress the role inductive inference in our theory and conclude with a précis of unsolved problems.

“For the task of object recognition, the visual system decomposes shapes into **parts**, . . . , parts with their **descriptions and spatial relations** provide a first index into a memory of shapes ...

Another benefit of structured models

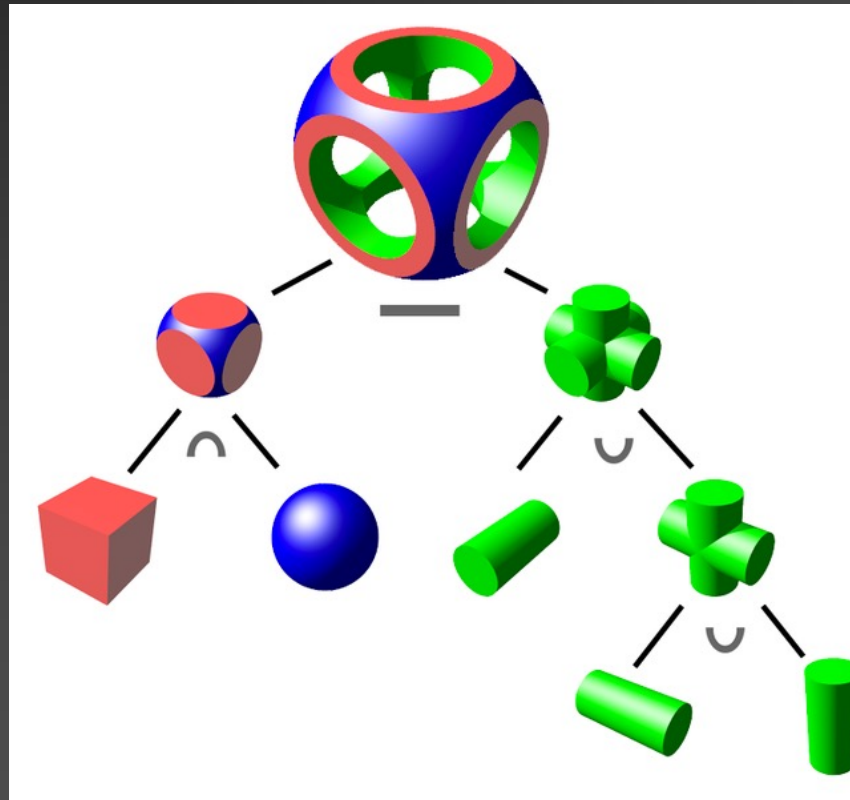
- **Reusability**, e.g., part re-assembly for 3D model creation



“Modeling by Example” [Funkhouser et al. SIGGRAPH 2004]

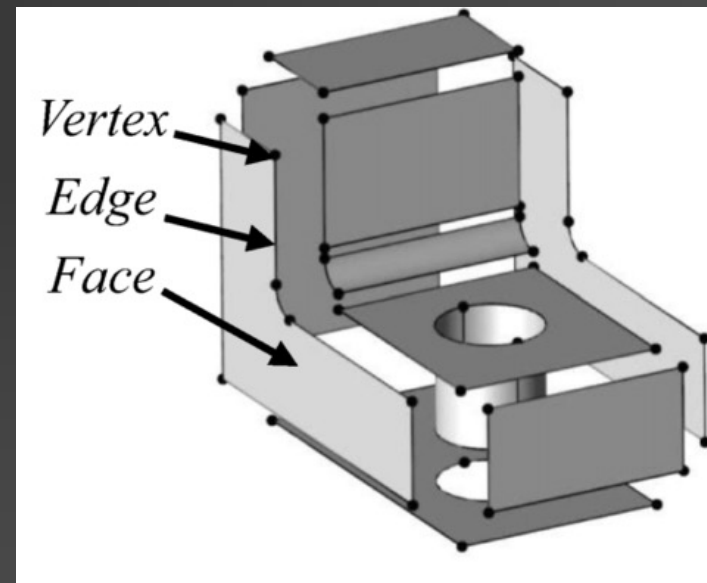
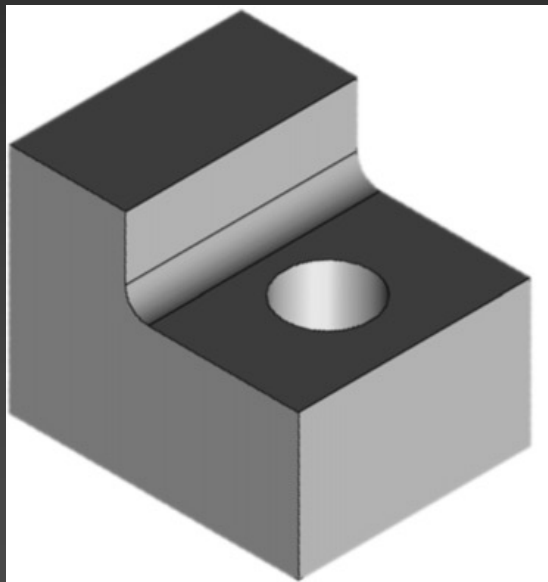
Structured representation: CSG

- Constructed solid geometry (CSG) for CAD models



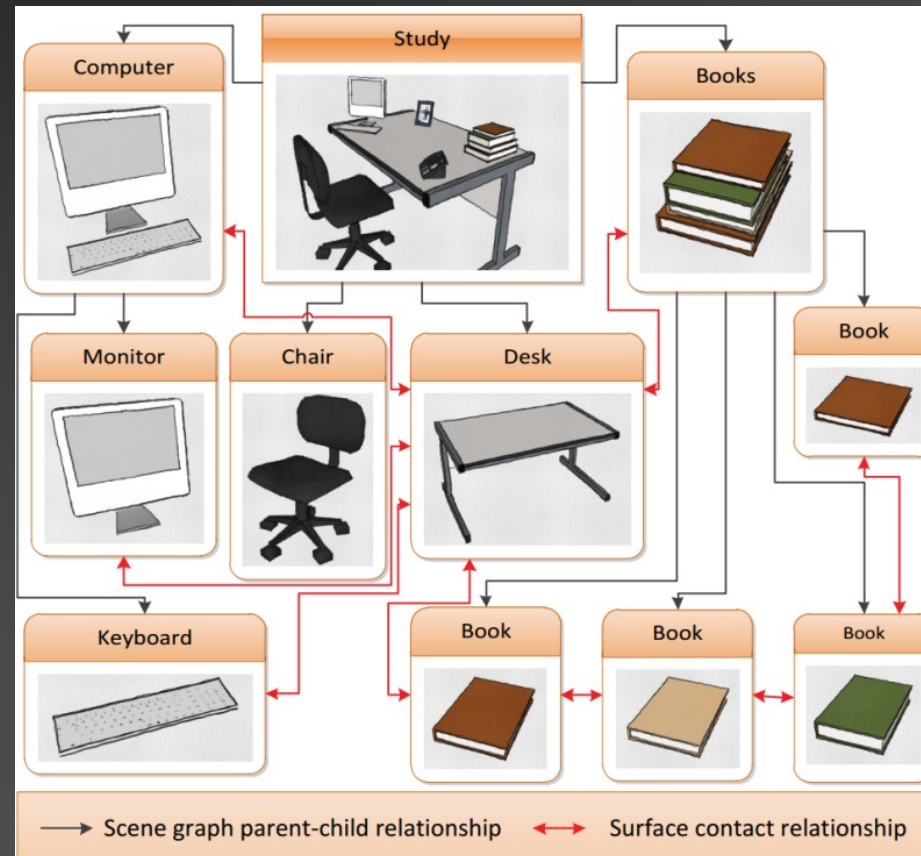
Structured representation: B-Rep

- Boundary representation (B-Rep) for CAD models



Structured representation: scene graphs

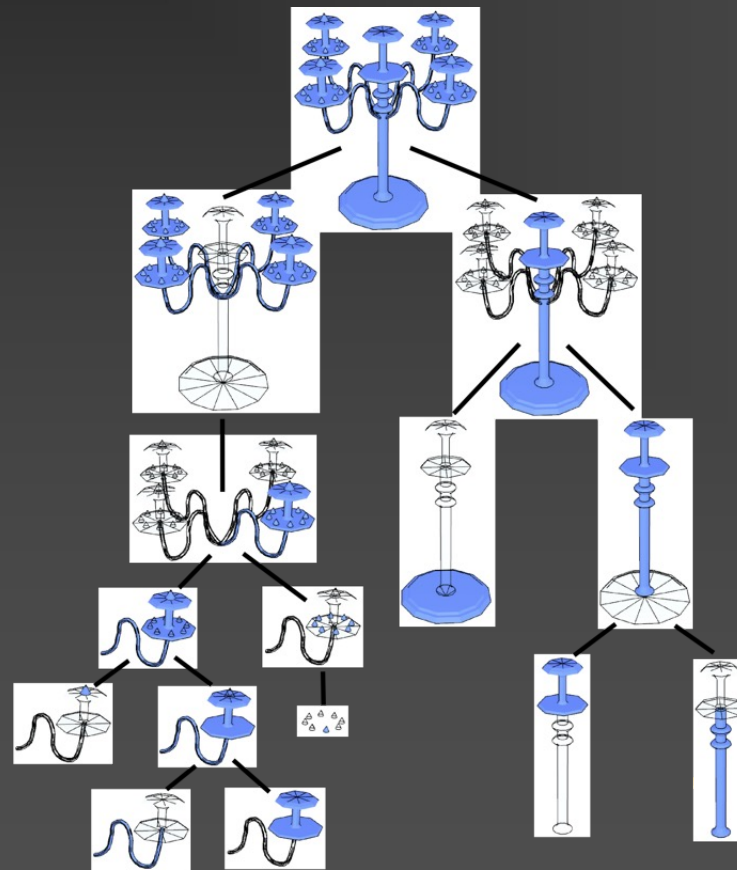
- For indoor scenes and object layouts/arrangements



Structured representation: symmetry hierarchy

Symmetry hierarchy: symmetry guides **grouping and** assembly of shape parts to form a meaningful **hierarchical** part organization.

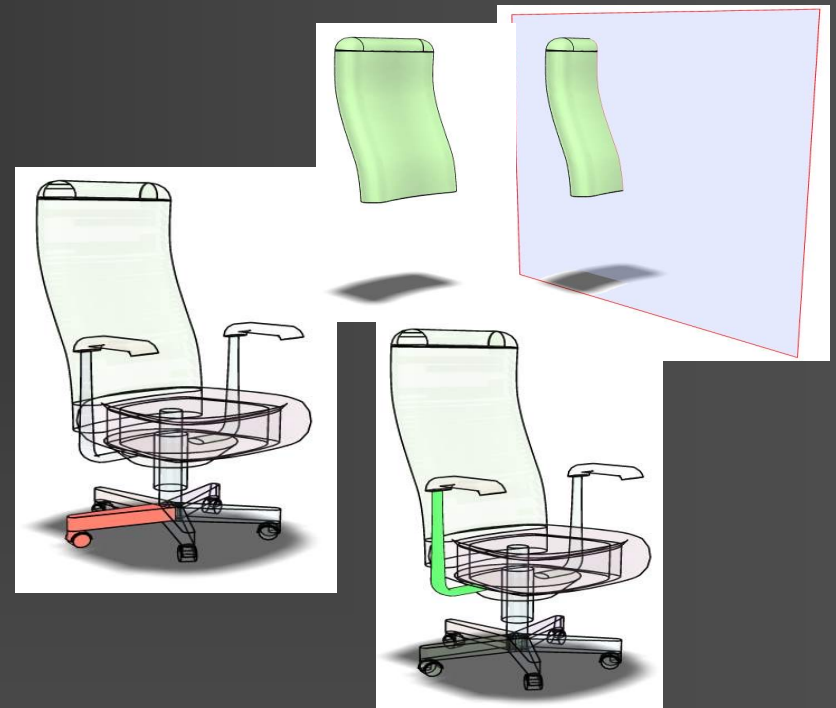
[Wang et al. 2011]



Symmetry hierarchy construction

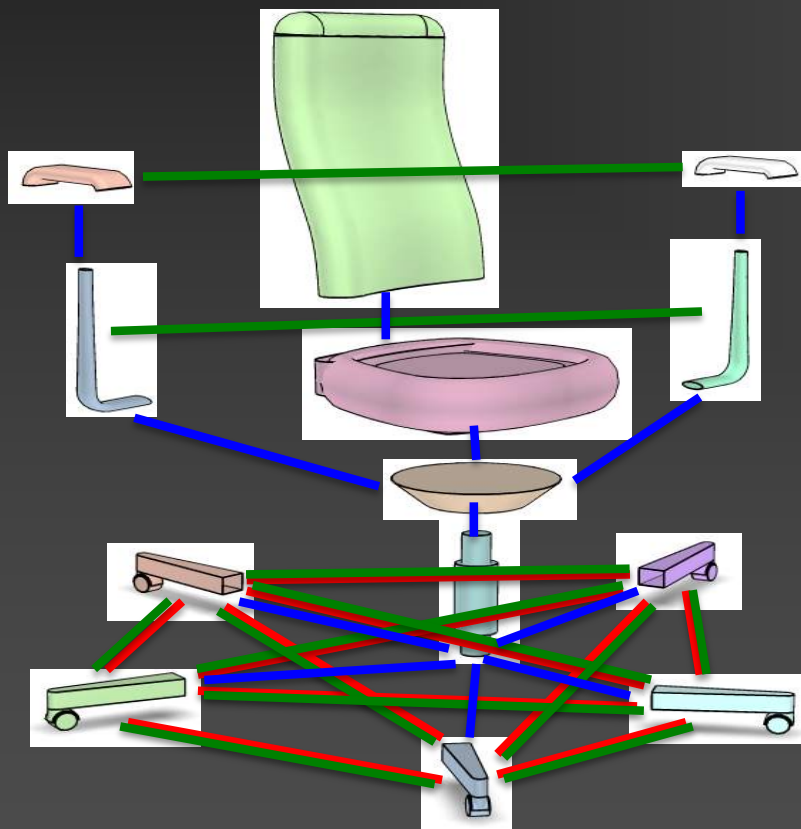


Part segmentation



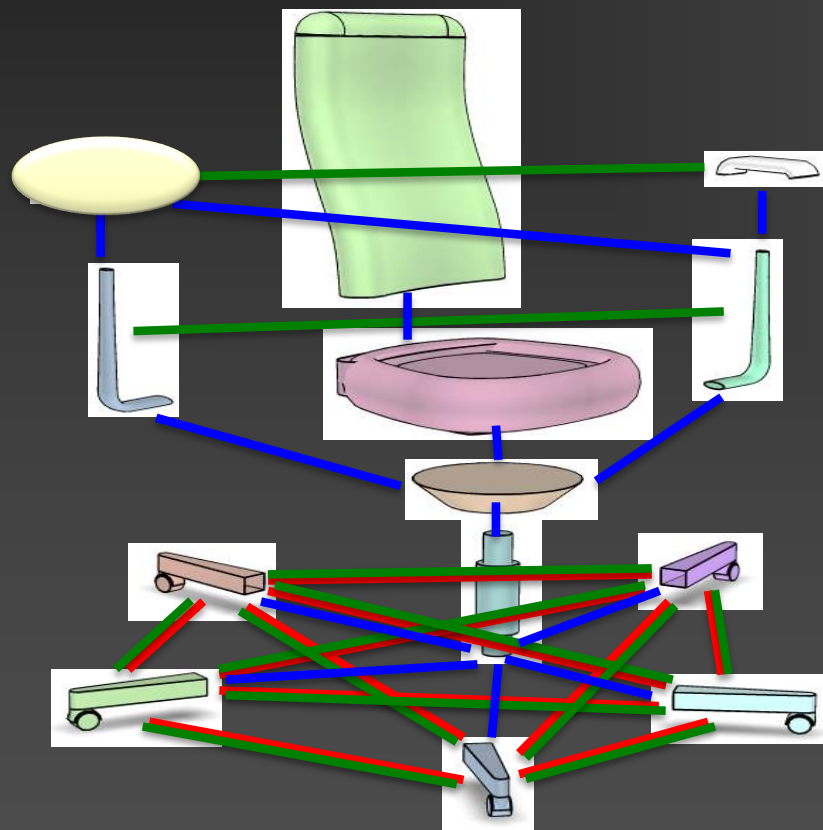
Symmetry detection

Initial graph



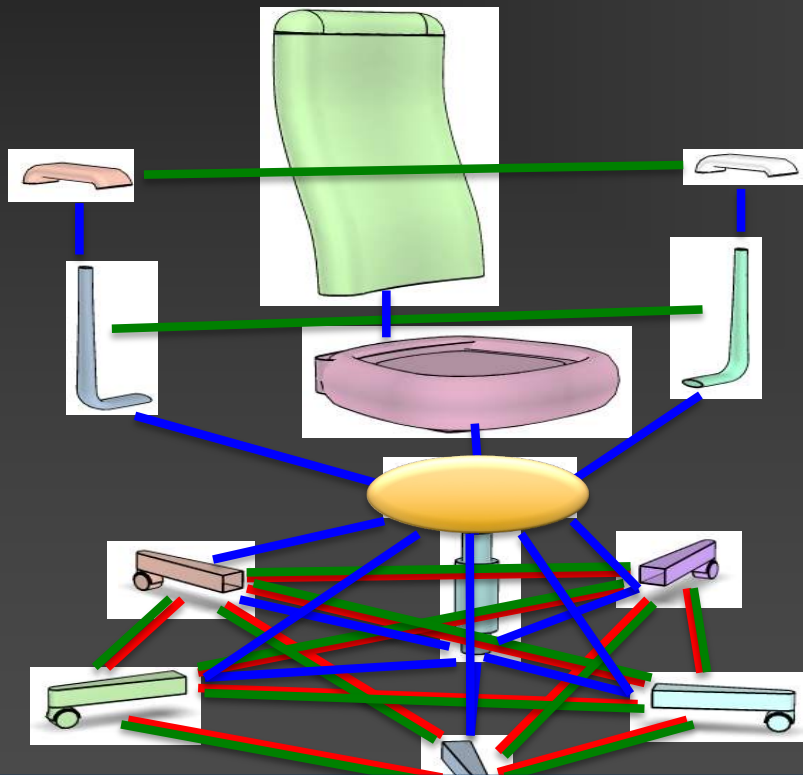
- Rotational symmetry
- Reflection symmetry
- Connectivity

Bottom-up graph construction



Two operations:
Grouping by symmetry

Bottom-up graph construction



Two operations:

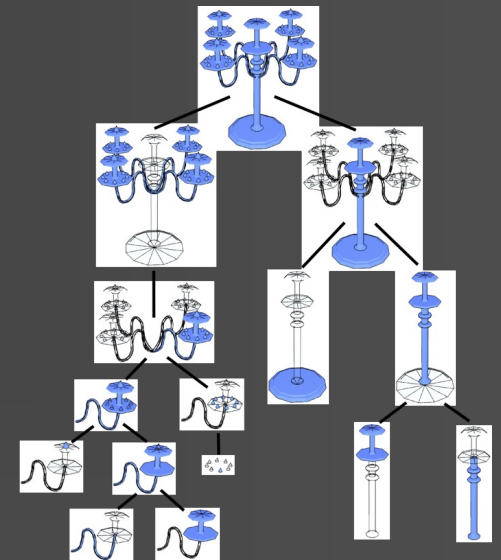
Grouping by symmetry

Assembly by proximity

Key question: in which order should we group the parts?

Symmetry hierarchy properties

- **Structure-aware:** hierarchical part organization
- **Functionality-aware** (a bit): symmetric parts tend to perform the same function
- Decouple structure and fine part geometry
 - Tree organization reveals part structure
 - Leaf nodes reveal part geometries



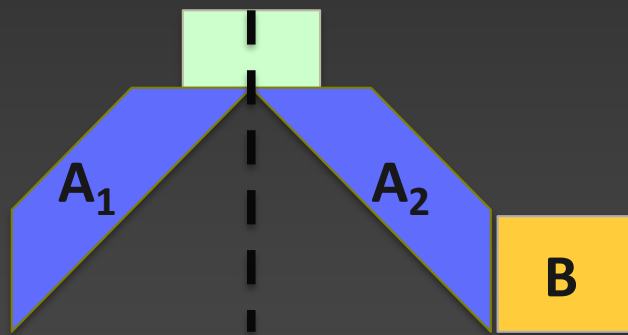
How to order graph contractions

- First attempt (2011): a **model- or rule-driven** approach
- Two guiding principles
 1. Perceptual grouping based on symmetry
 2. **Compactness** of representation = **fewest grouping operations**
- Results in a set of **handcrafted** “precedence rules”

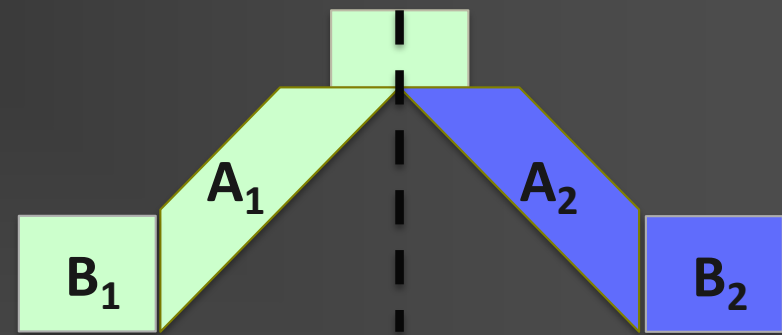
[Wang et al. EG 2011]

Example of handcrafted precedence rules

- Grouping-assembly mixing rules, e.g.,
 - **M1**: symmetry grouping takes precedence over assembly



unless

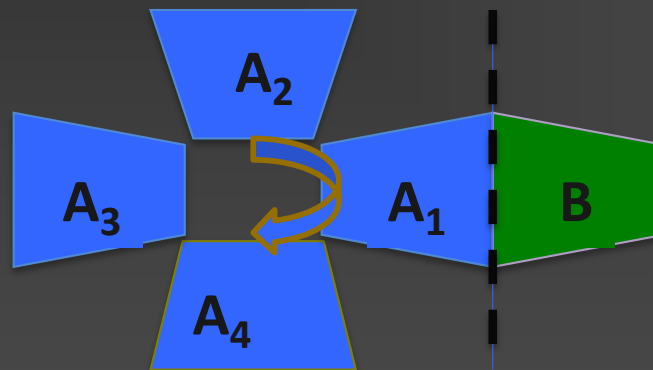


M2: Assembly before grouping: 2 ops

Grouping before assembly: 3 ops

Example of handcrafted precedence rules

- Symmetry grouping rules, e.g.,
 - **G1**: rot-symmetry takes precedence over ref-symmetry



- “Neuralization” in 2017: GRASS = Generative Recursive Autoencoder

GRASS (later, aside for now)

GRASS: Generative Recursive Autoencoders for Shape Structures

Jun Li¹ Kai Xu^{1,2,3*} Siddhartha Chaudhuri⁴ Ersin Yumer⁵ Hao Zhang⁶ Leonidas Guibas⁷

¹National University of Defense Technology ²Shenzhen University ³Shandong University ⁴IIT Bombay

⁵Adobe Research ⁶Simon Fraser University ⁷Stanford University

