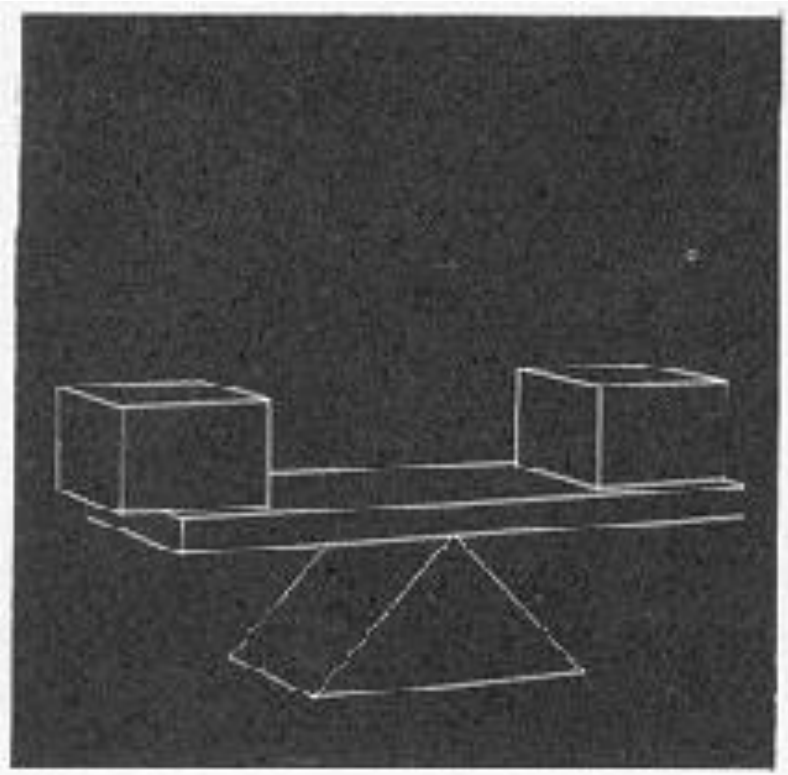
Part-based modelling of compound scenes from images

Anton van den Hengel, Chris Russell, Anthony Dick, John Bastian,
Daniel Pooley, Lachlan Fleming, Lourdes Agapito

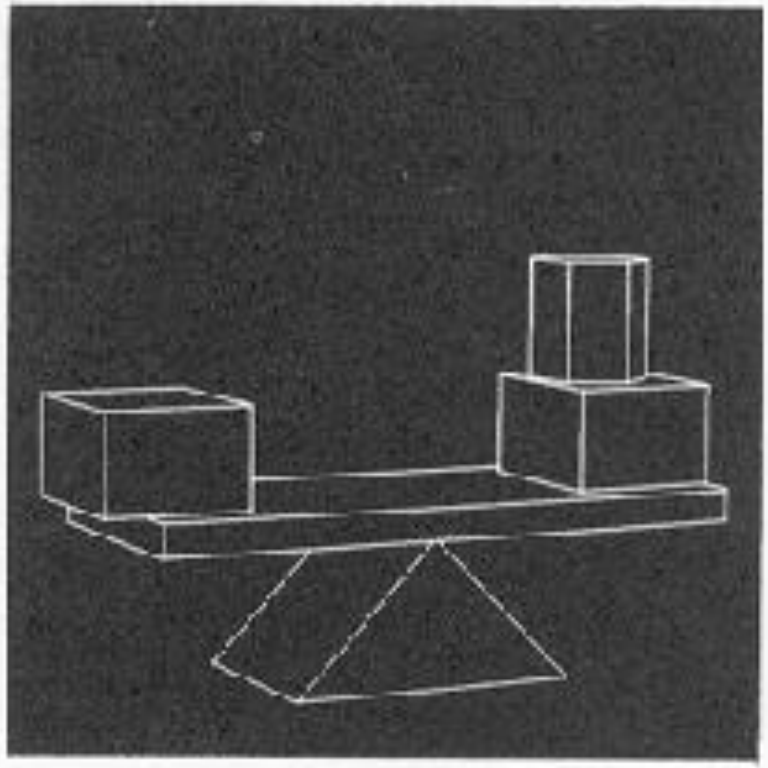
The University of Adelaide,
The Australian Centre for Robotic Vision,
University College London

Australian Centre for Visual Technologies
Innovation and education in visual information processing

Deconstruction has been a theme

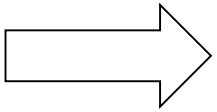


(e) See-saw.



(f) With hexagonal prism.

Structure and semantics interact

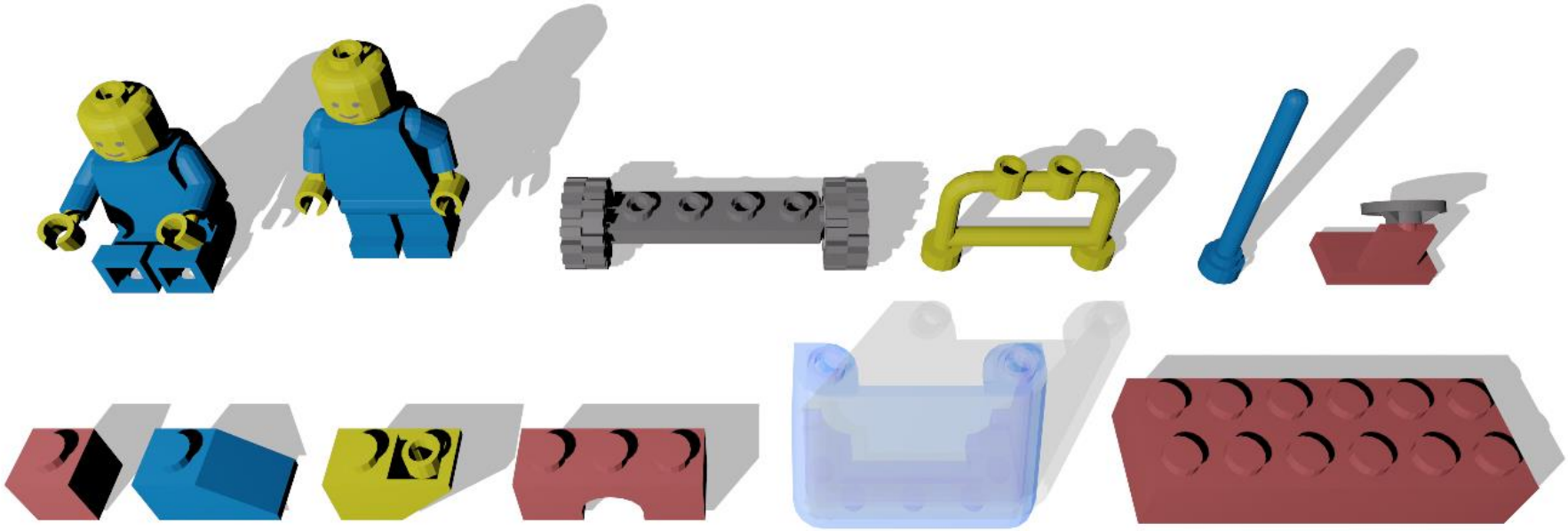


Structure and semantics interact

Deconstruction

- We want to be able to take the world apart
  - And put it back together
- We want to reason about the solution in 3D
  - Not least so we can ensure the solution makes sense
- We need a set of parts to build our models from

Lego!

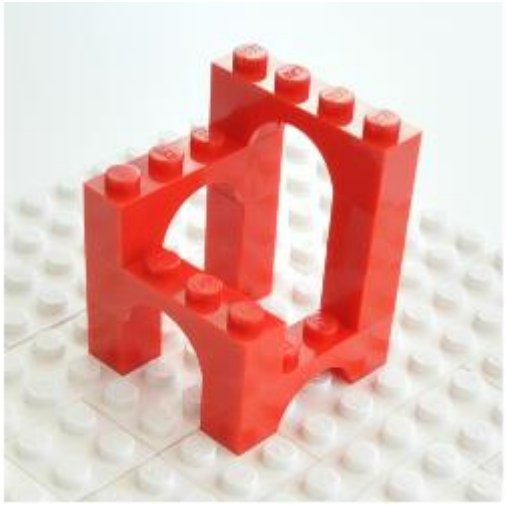


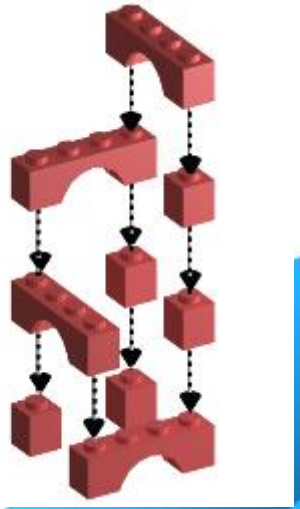
Silhouette-based composition

- We're only interested in shape

- Silhouettes can be composed linearly

Deconstruction



- Render all possible building blocks in every possible position, and record the silhouette of each
- Then reconstruct object silhouettes as combinations of these block silhouettes

Template shapes

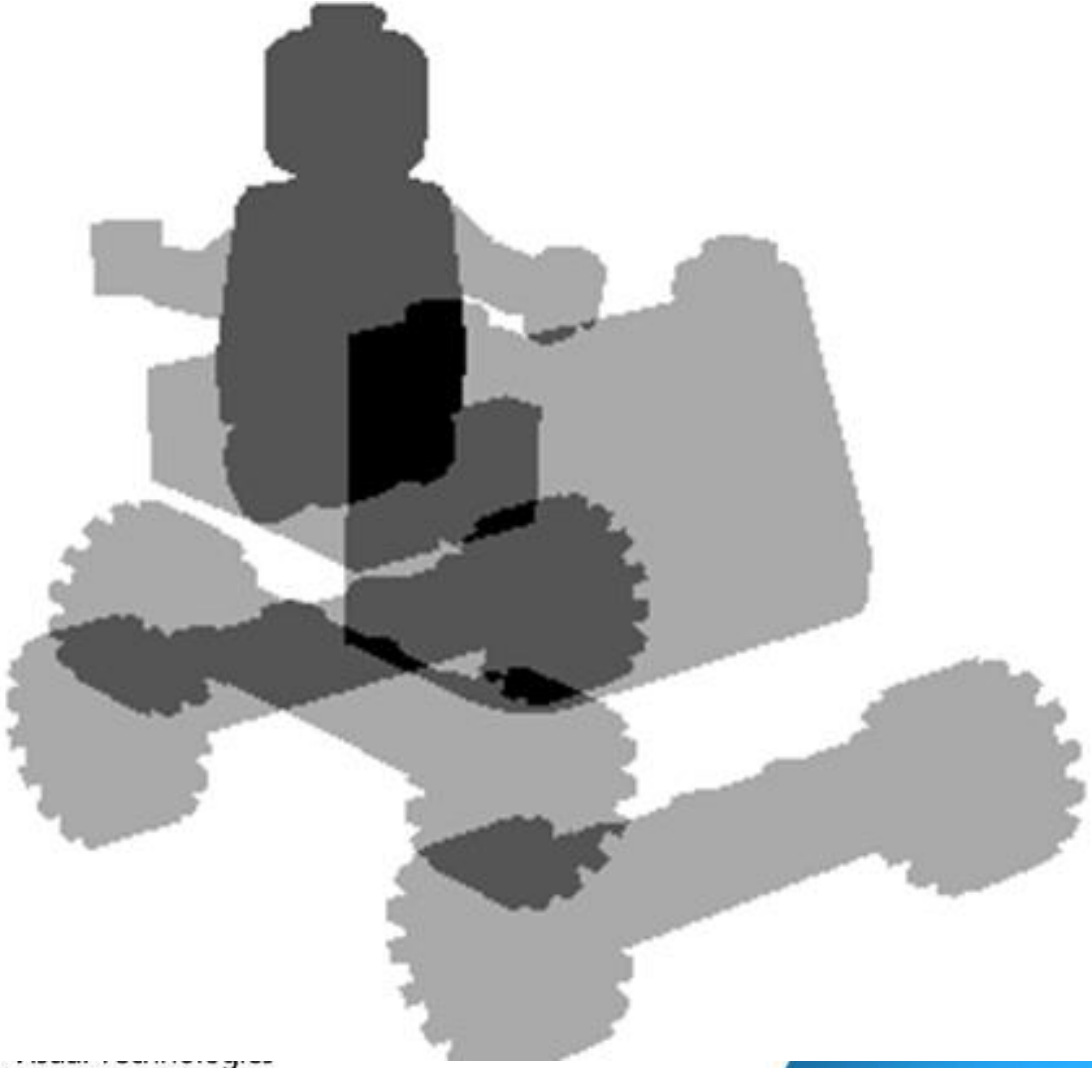
- nTemplates = nShapes x nPositions x nRotations
  - So there are lots of them
  - But they are sparsely used

A linear model of composition

$$\arg\min_{\boldsymbol{\alpha}\in\Omega} \|\mathbf{y} - \boldsymbol{\Pi}\boldsymbol{\alpha}\|_1 + \lambda\|\boldsymbol{\alpha}\|_1$$

- $\mathbf{y}$ the silhouette of the shape to be recovered
- $\boldsymbol{\Pi}\boldsymbol{\alpha}$ the reconstructed silhouette
  - $\boldsymbol{\alpha}$ a binary vector of block coefficients
  - $\boldsymbol{\Pi}$ a matrix with one block silhouette per column
- Ω the set of plausible models
- λ a scale factor controlling parsimony

Composition problems



A better model of composition

- Separate foreground from background (rows)
  - $\mathbf{y} \rightarrow \mathbf{y}_0, \mathbf{y}_1$ and $\mathbf{\Pi} \rightarrow \mathbf{\Pi}_0, \mathbf{\Pi}_1$
- Add slack variables for foreground

$$\arg\min_{\boldsymbol{\alpha}\in\Omega,\boldsymbol{\xi}} \|\mathbf{\Pi}_0\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\xi}\|_1 + \lambda\|\boldsymbol{\alpha}\|_1$$
$$\text{s.t. } \mathbf{y}_1 - \mathbf{\Pi}_1\boldsymbol{\alpha} \leq \boldsymbol{\xi},\ 0 \leq \boldsymbol{\xi} \leq 1$$

Constraints - Intersection

- Form matrix $\mathbf{\Gamma}$ where every row represents a constraint
  - If templates i and j intersect then insert a row (k) in $\mathbf{\Gamma}$ such that

$$\mathbf{\Gamma}_{ki} = 1, \mathbf{\Gamma}_{kj} = 1$$

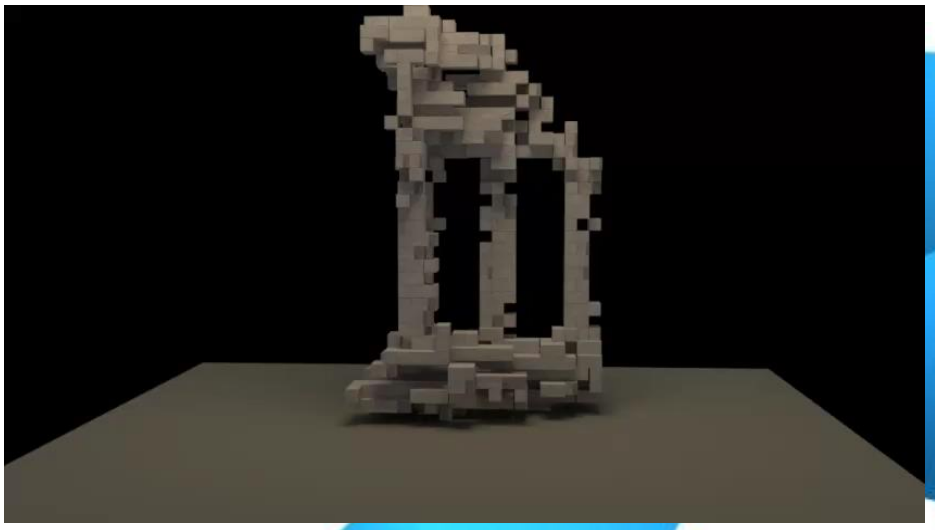
  - Insert constraint into optimisation problem that

$$\mathbf{\Gamma}\boldsymbol{\alpha} \leq \mathbf{1}$$

Constraints - Gravity

- Form Υ where every row represents a constraint
  - If template i needs support t set $\Upsilon_{ii} = t$
  - If template j provides s support to i set $\Upsilon_{ij} = -s$
- And require that

$$\Upsilon \alpha \leq 0$$



Integer programming

$$\arg\min_{\boldsymbol{\alpha}\in\mathbb{Z}_2^T,\boldsymbol{\xi}} \|\boldsymbol{\Pi}_0\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\xi}\|_1 + \lambda\|\boldsymbol{\alpha}\|_1$$
$$\text{s.t. } \mathbf{y}_1 - \boldsymbol{\Pi}_1\boldsymbol{\alpha} \leq \boldsymbol{\xi},\ \boldsymbol{\Gamma}\boldsymbol{\alpha} \leq \mathbf{1},\ \boldsymbol{\Upsilon}\boldsymbol{\alpha} \leq \mathbf{0}$$

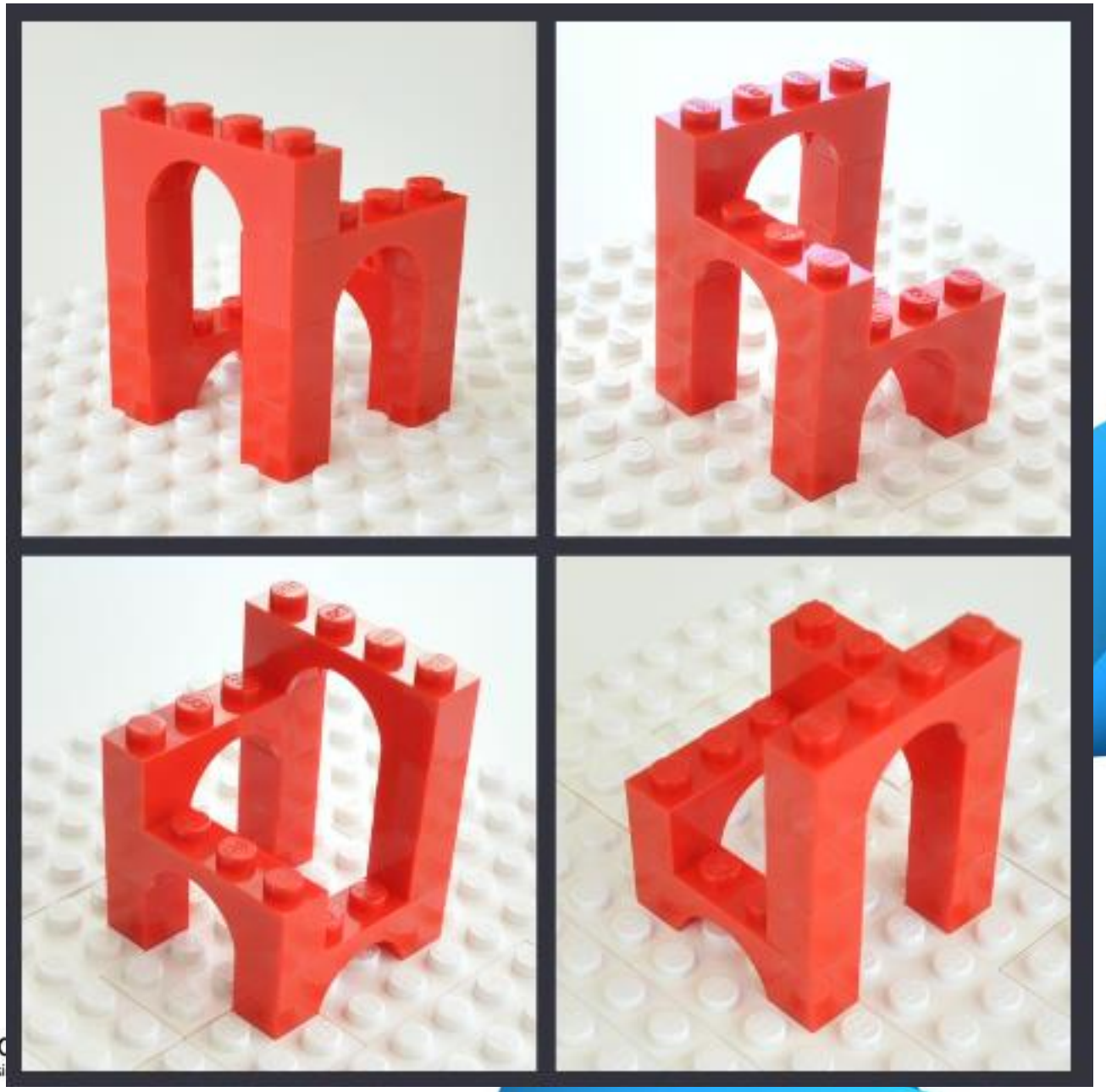
- But there are too many pixels in $\mathbf{y}_1$

Sparse recovery – Projection

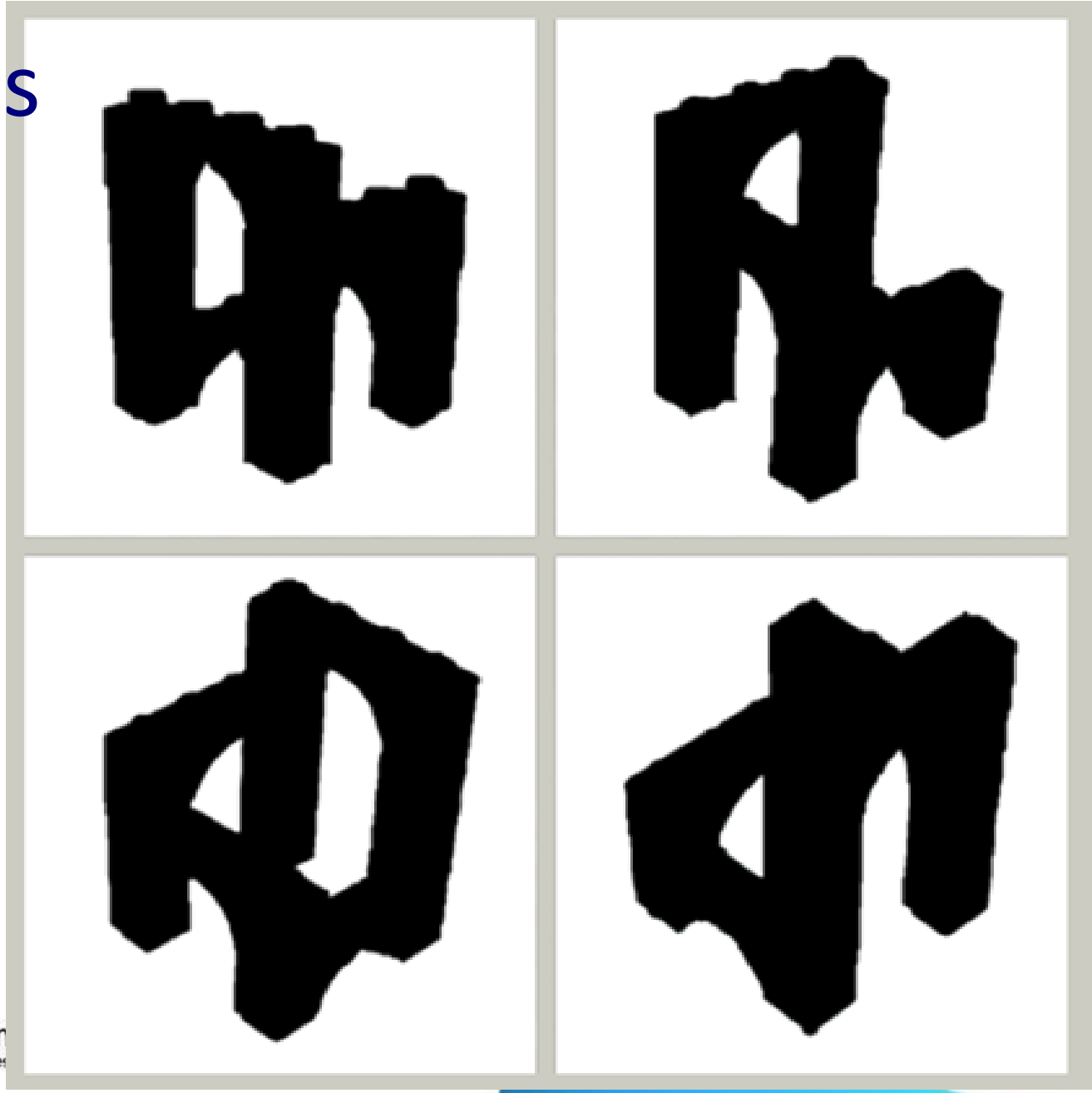
$$\arg\min_{\boldsymbol{\alpha}\in\mathbb{Z}_2^T,\boldsymbol{\xi}} \|\boldsymbol{\Pi}_0\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\xi}\|_1 + \lambda\|\boldsymbol{\alpha}\|_1$$
$$\text{s.t. } \boldsymbol{\Phi}\mathbf{y}_1 - \boldsymbol{\Phi}\boldsymbol{\Pi}_1\boldsymbol{\alpha} \leq \boldsymbol{\xi},\ \boldsymbol{\Gamma}\boldsymbol{\alpha} \leq \mathbf{1},\ \boldsymbol{\Upsilon}\boldsymbol{\alpha} \leq \mathbf{0}$$

- Projection by $\boldsymbol{\Phi} \in \mathbb{R}^{D\times S}$ where D << S
- Solve the reconstruction problem in a (much) lower dimensional space

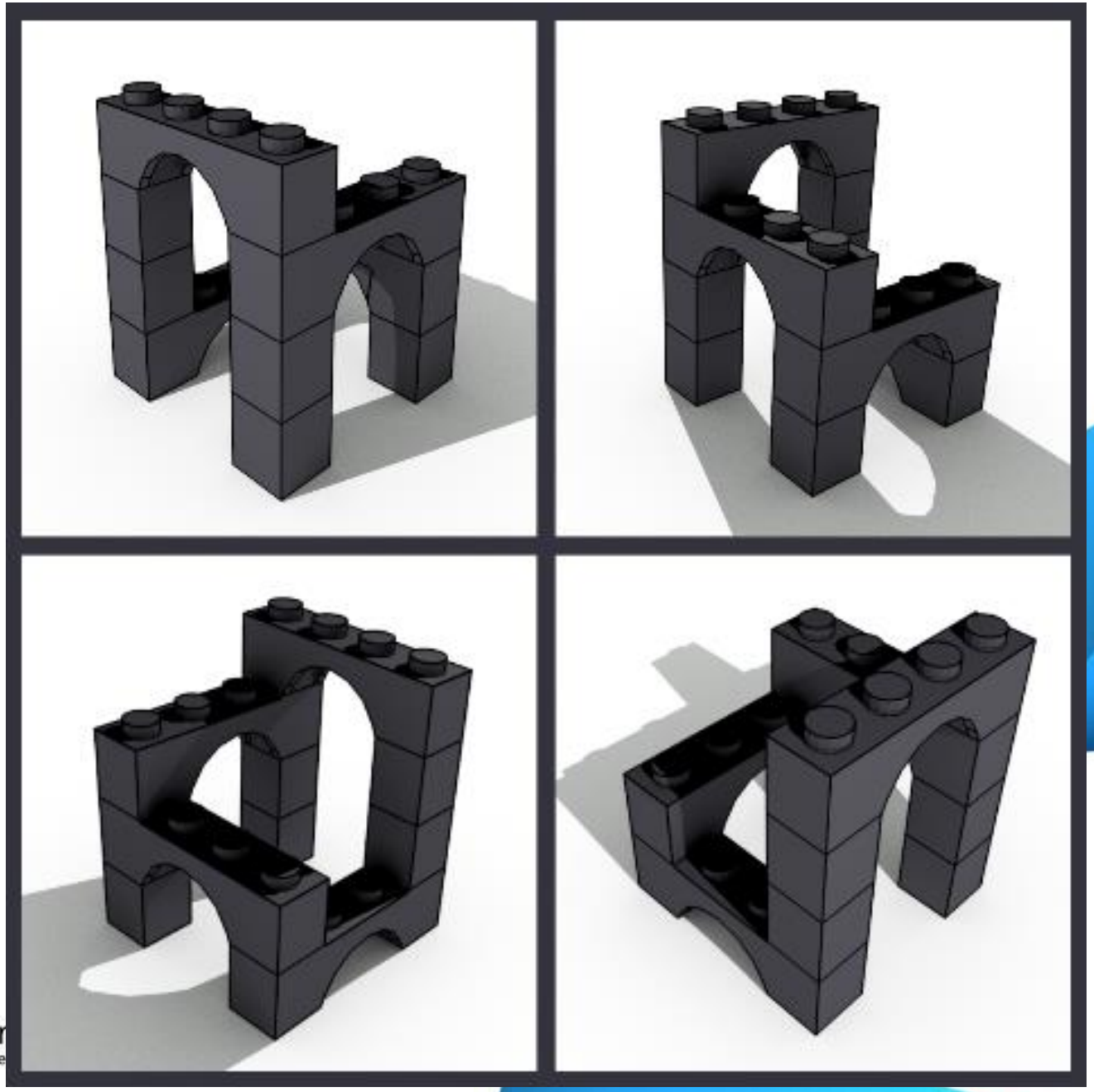
Images



Silhouettes



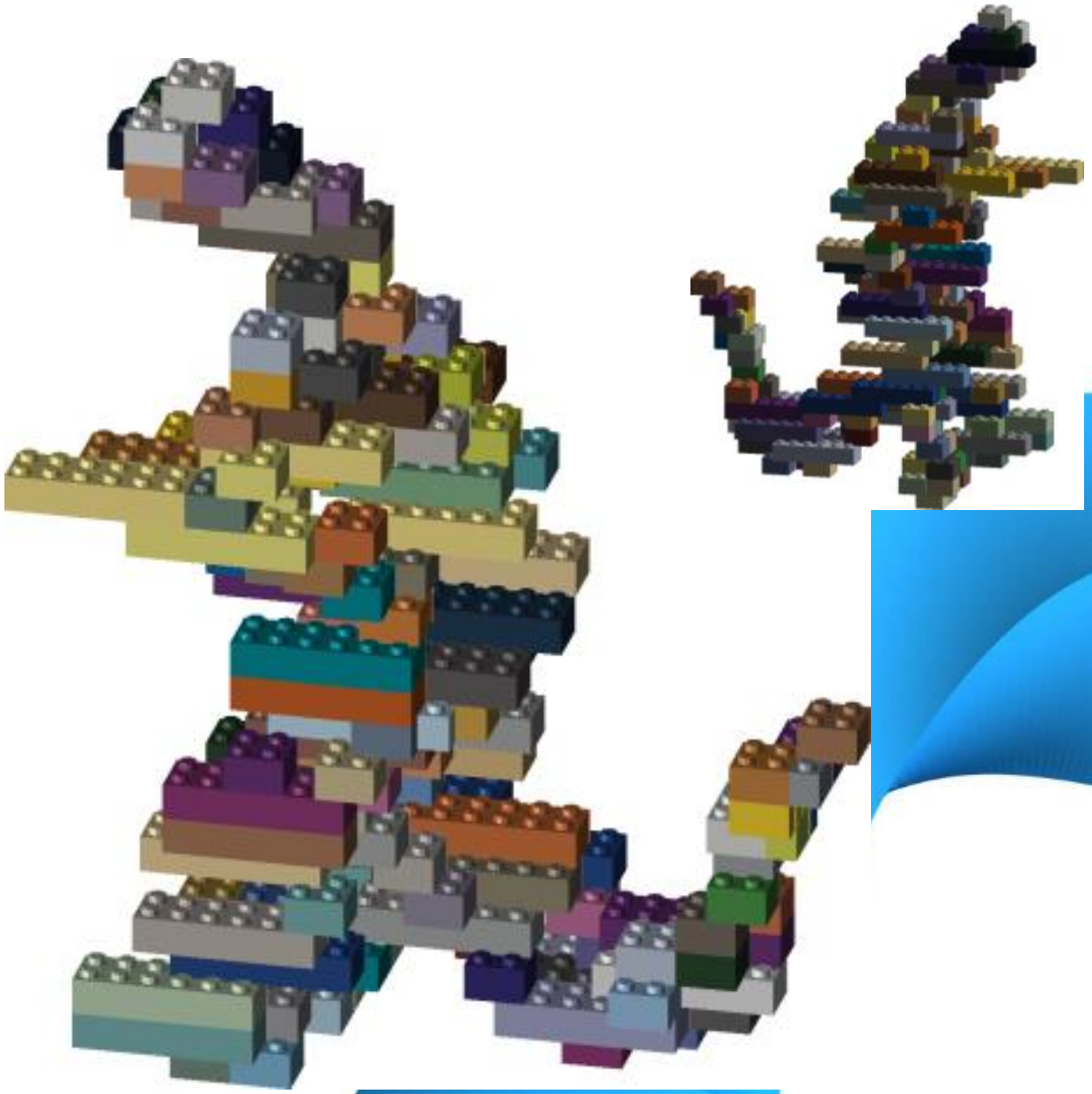
Structure



Structure

Results

Deconstructing general objects



Australian Centre for Visual Technologies
Innovation and education in visual information processing

isual Technologies
formation processing

Deconstruction

- Part-based model
- Estimated using Integer Programming
- Incorporates physical constraints
- Reasoning in 3D
- Projected to a lower dimension without losing accuracy