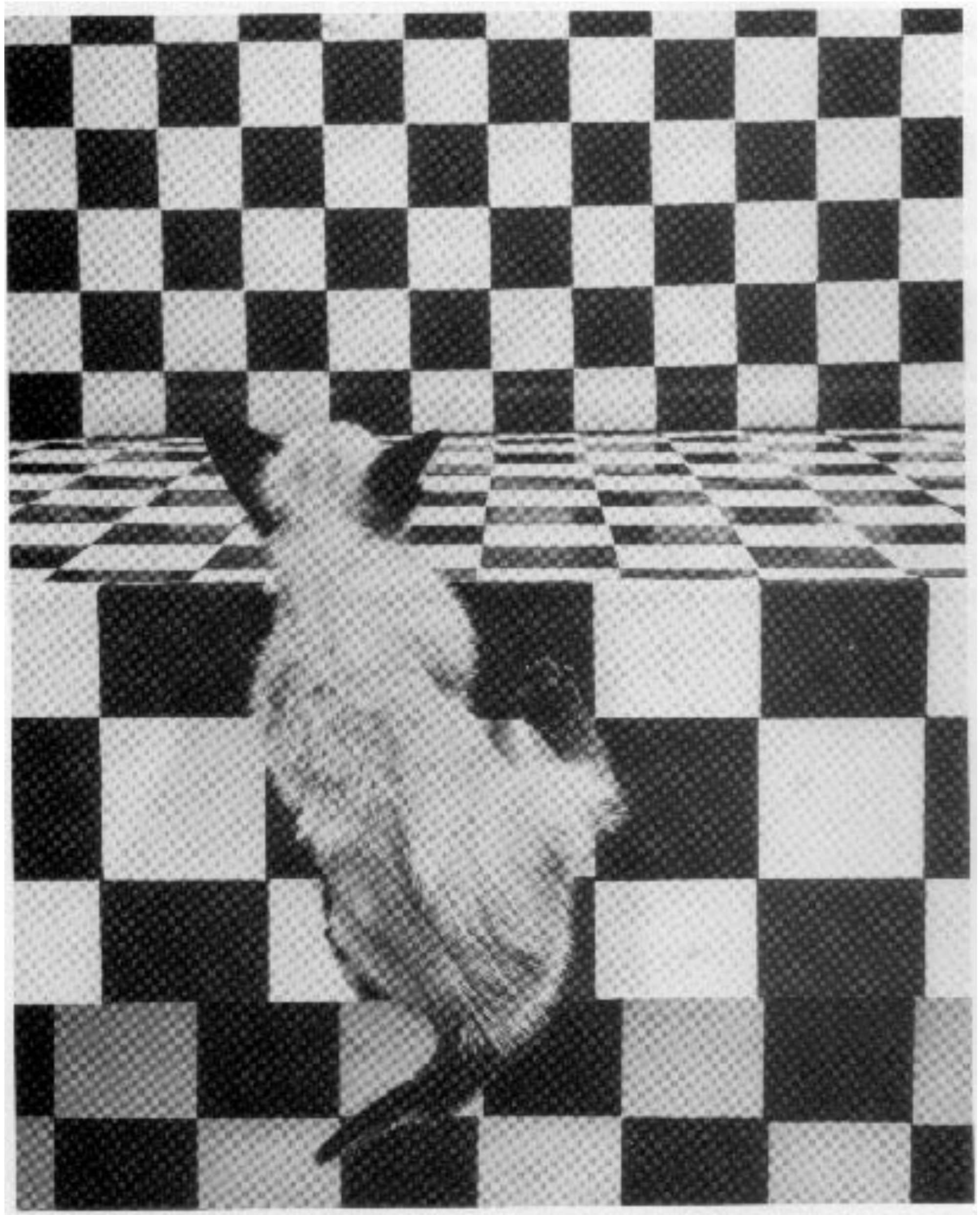
Shape from X

- One image:
  - Shading
  - **Texture**

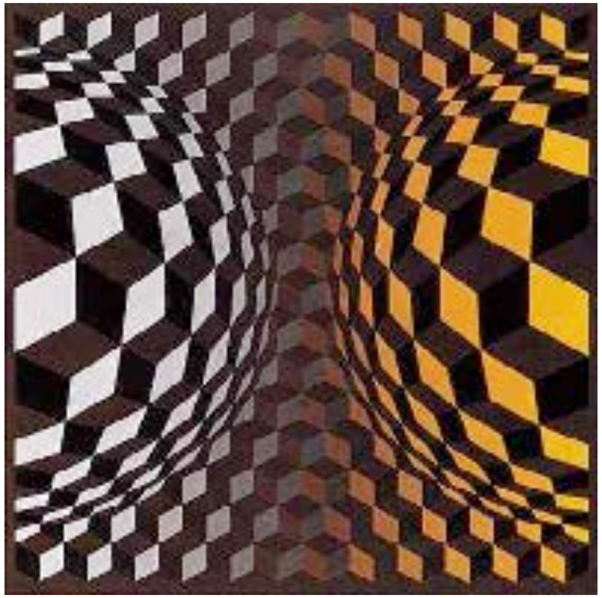
- Two images or more:
  - Stereo
  - Contours
  - Motion

Shape From X

- One image:
  - Shading
  - **Texture**
- Two images or more:
  - Stereo
  - Contours
  - Motion



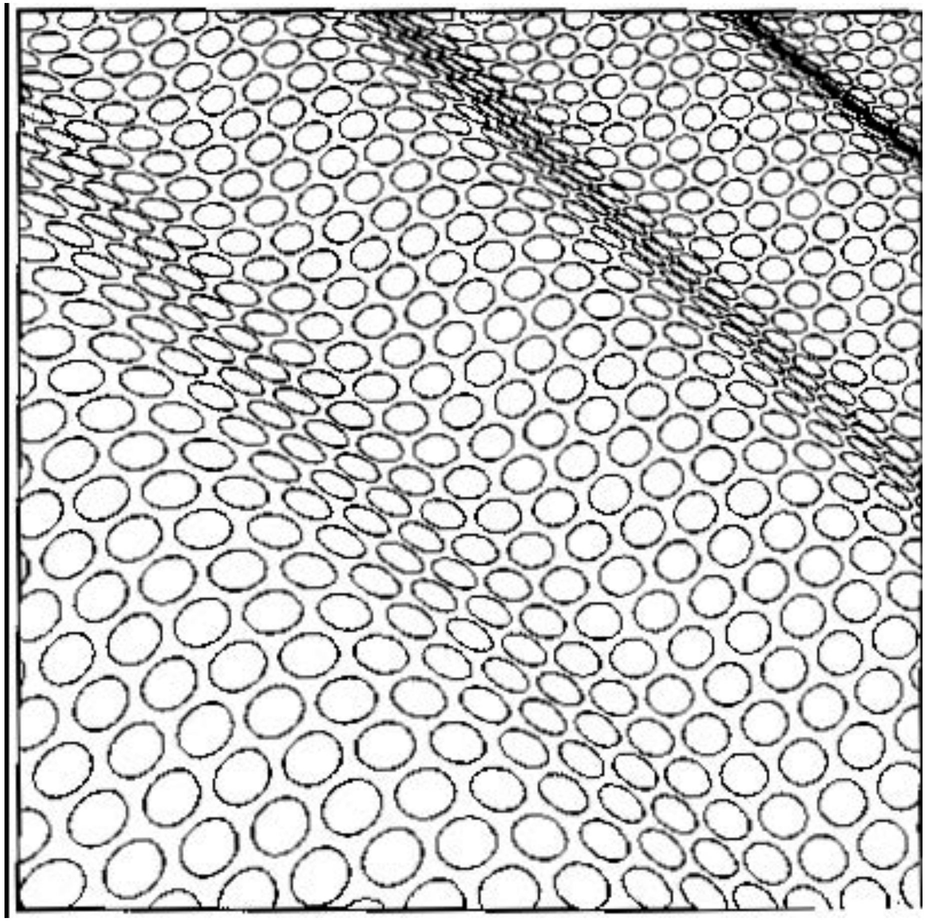
Shape From Texture



Recover surface orientation or surface shape from image texture:

- Assume texture ‘looks the same’ at different points on the surface.
- This means that the deformation of the texture is due to the surface curvature.

Structural Shape Recovery

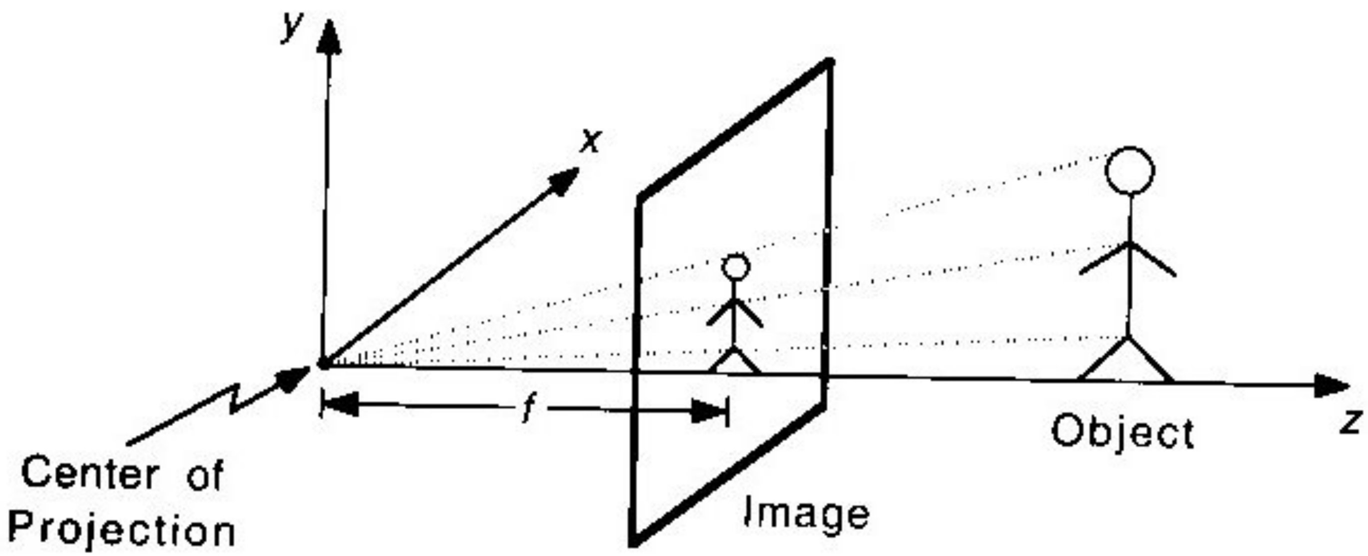


Basic hypothesis: Texture resides on the surface and has no thickness.

—> Computation under:

- Perspective projection
- Paraperspective projection
- Orthographic projection

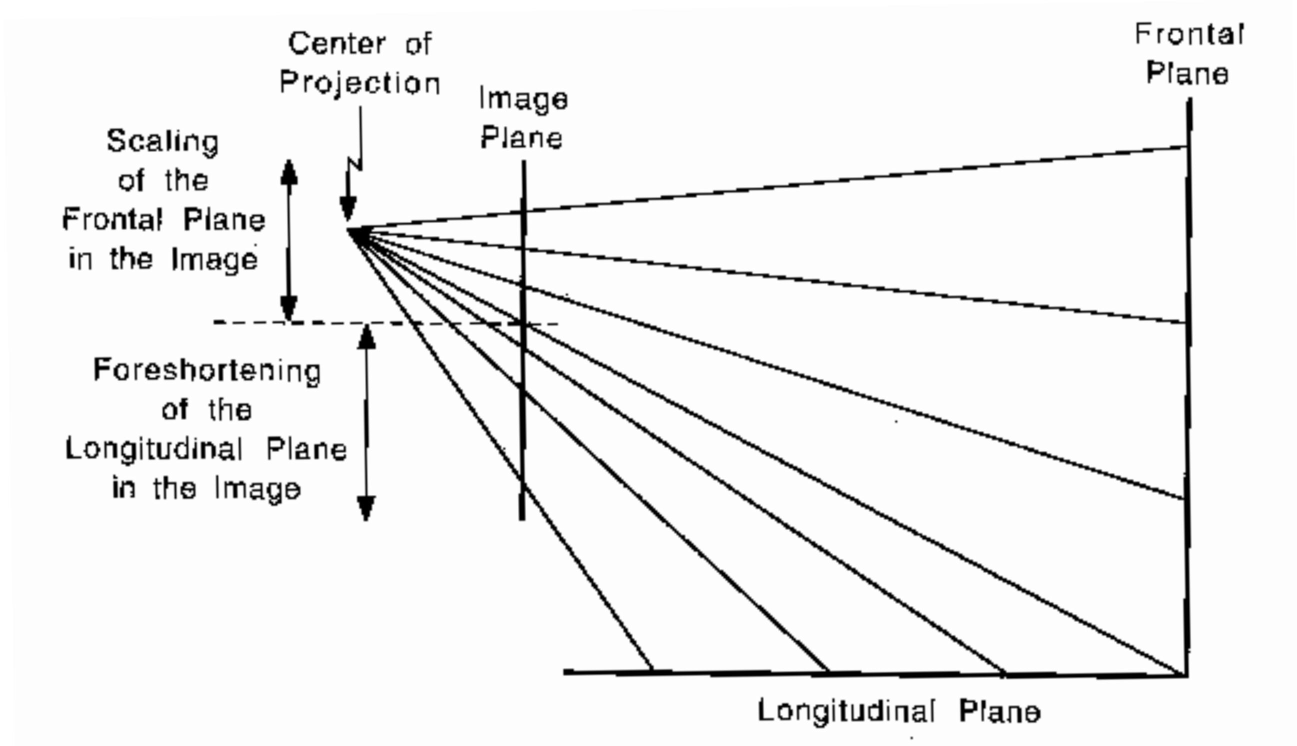
Reminder: Perspective Projection



$$u = f\frac{x}{z}$$

$$v = f\frac{y}{z}$$

Perspective Distortion



The perspective projection distortion of the texture

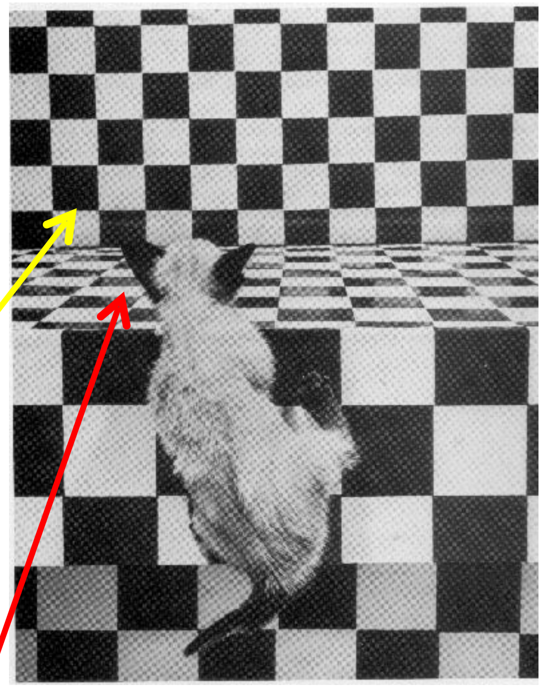
- depends on both depth and surface orientation,
- is anisotropic.

Foreshortening

Depth vs Orientation:

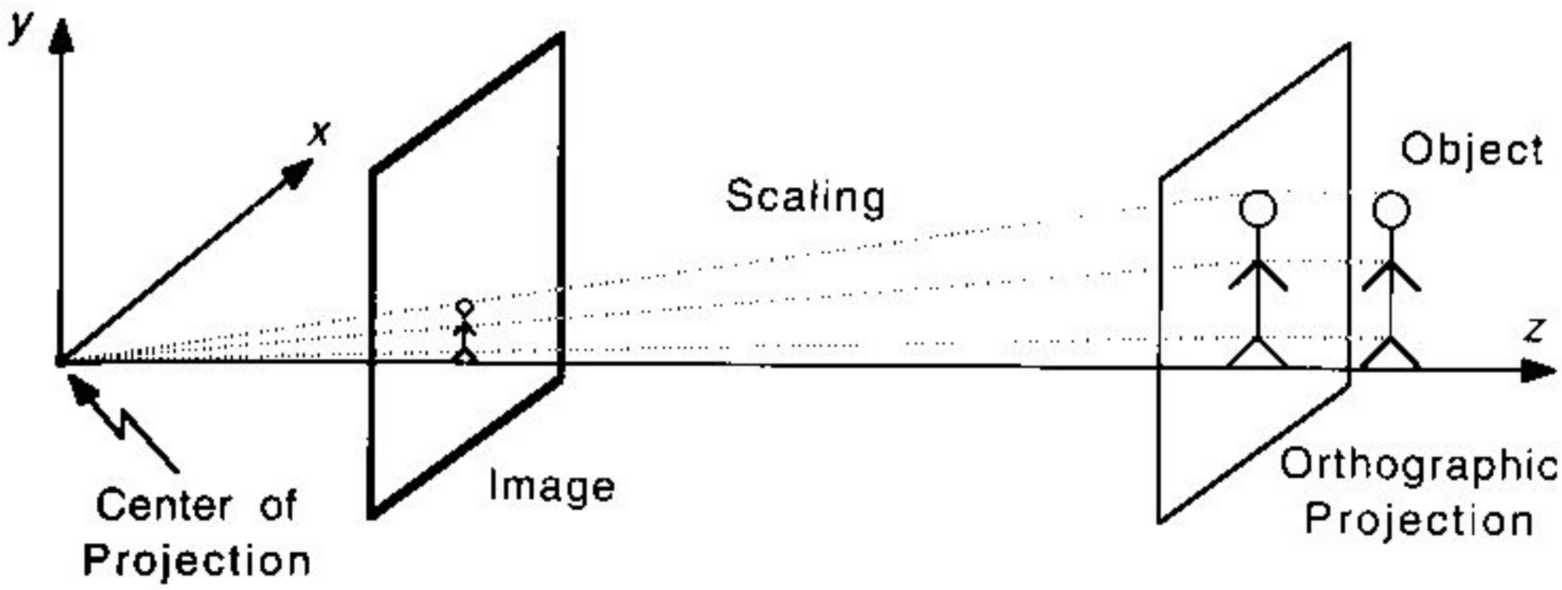
- Infinitesimal vector [Δx,Δy,Δz] at location [x,y,z] image of this vector is

$$\frac{f}{z}[\Delta x - \frac{x}{z}\Delta z, \Delta y - \frac{y}{z}\Delta z]$$



- Two special cases:
  - $\Delta z=0$: The object is scaled
  - $\Delta x=\Delta y=0$: The object is foreshortened

Reminder: Orthographic Projection



$$u = sx$$

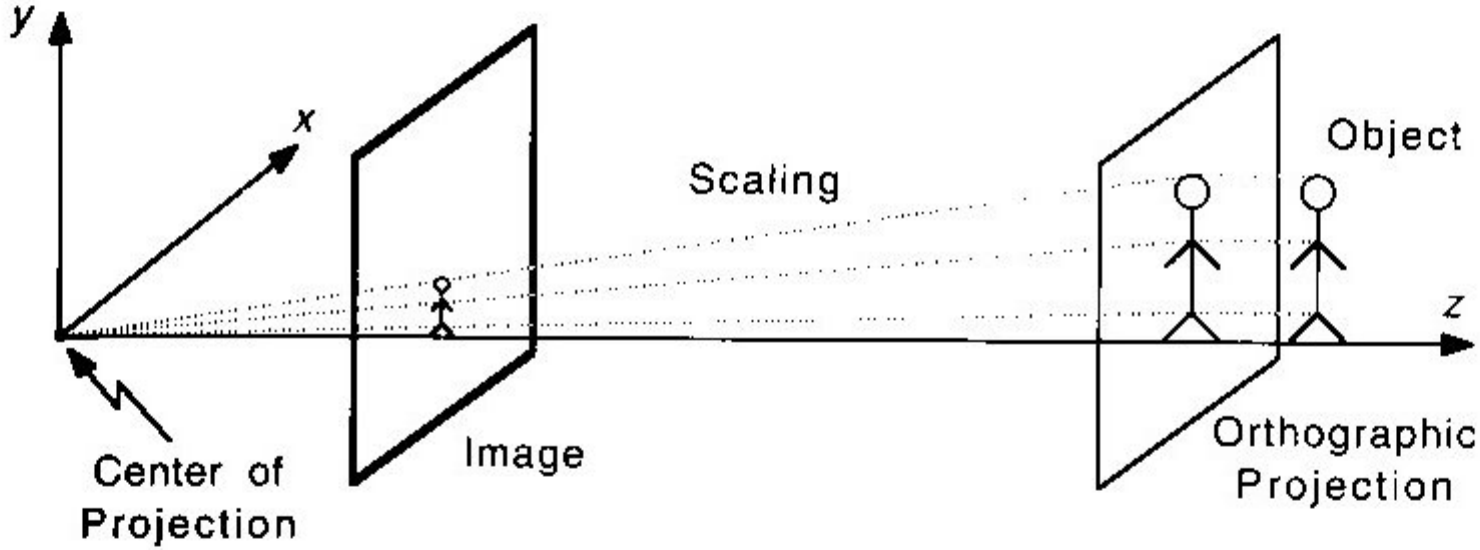
$$v = sy$$

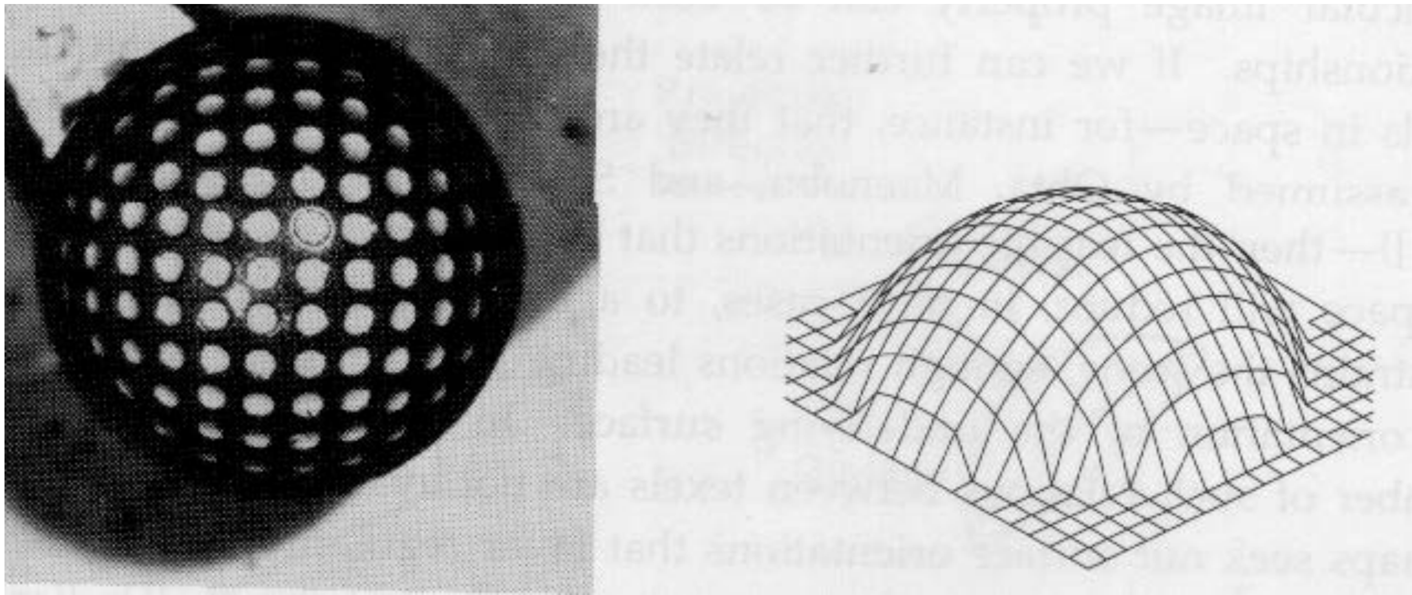
Special case of perspective projection:

- Large f
- Objects close to the optical axis

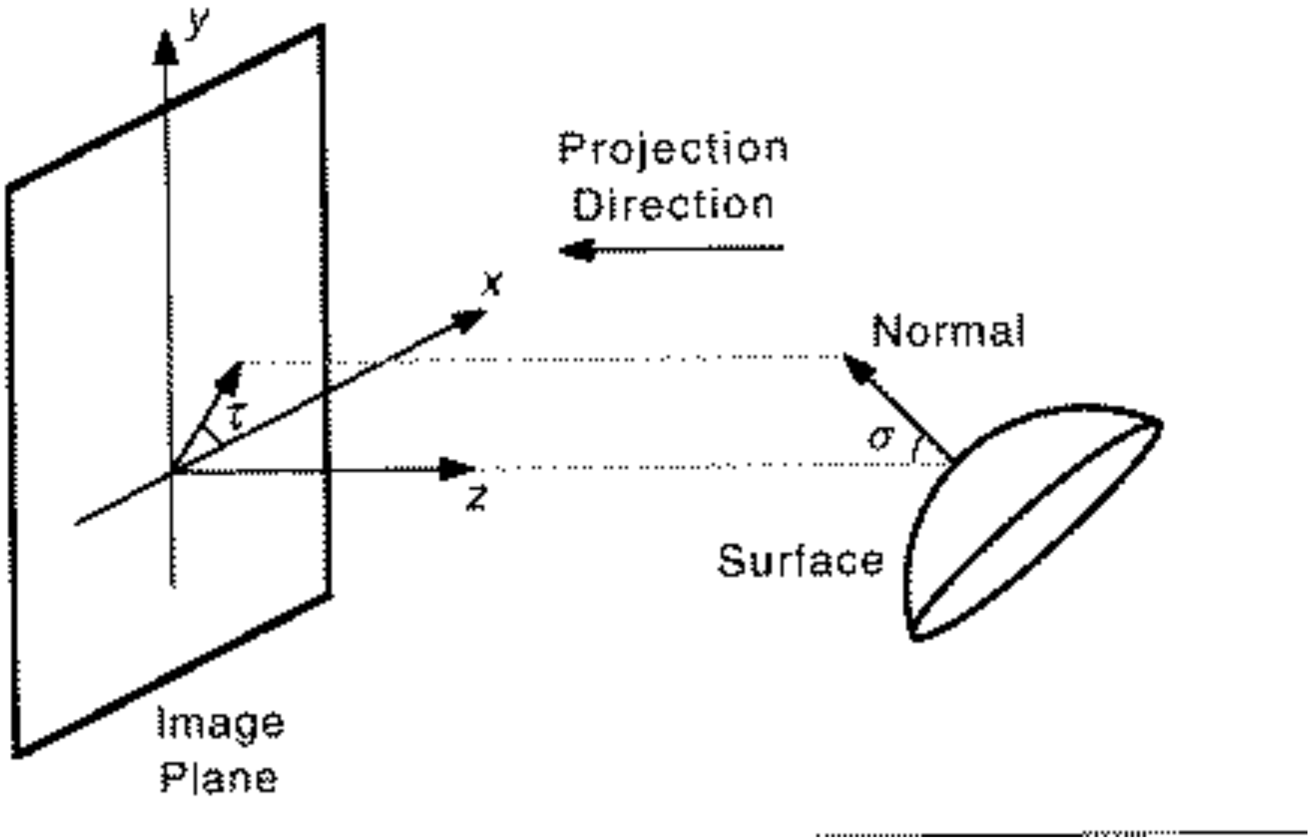
→Parallel lines mapped into parallel lines.

Orthographic Projection

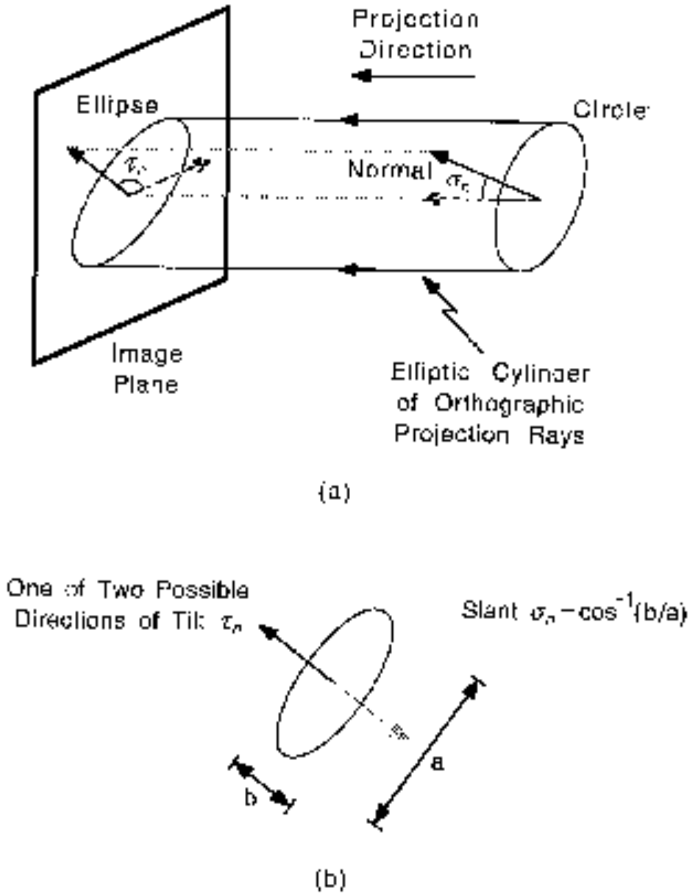




Tilt And Slant

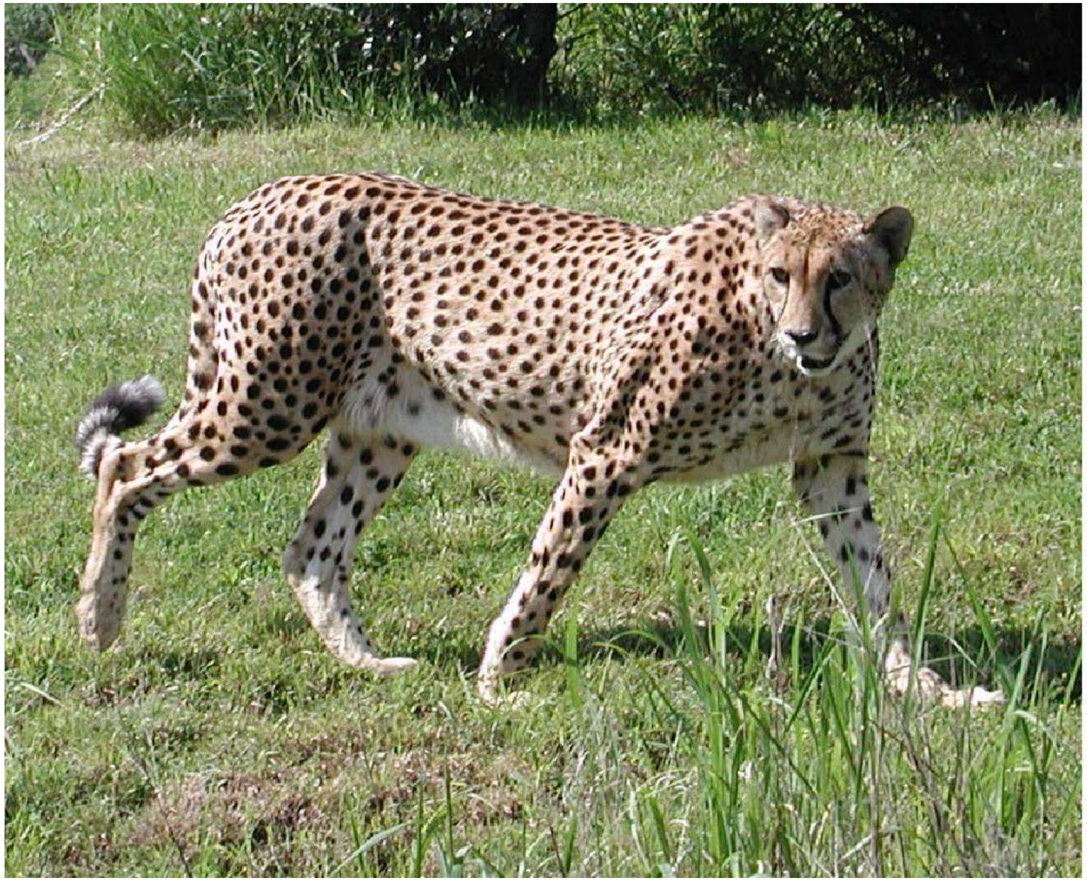


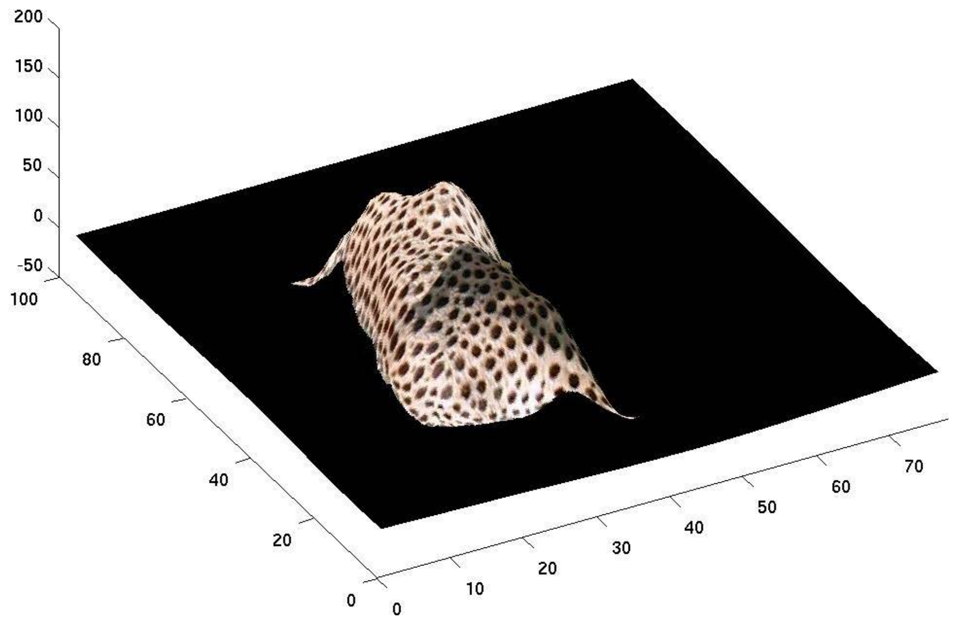
Orthographic Projection



- **Tilt:** Derived from the image direction in which the surface element undergoes maximum compression.
- **Slant:** Derived from the extent of this compression.

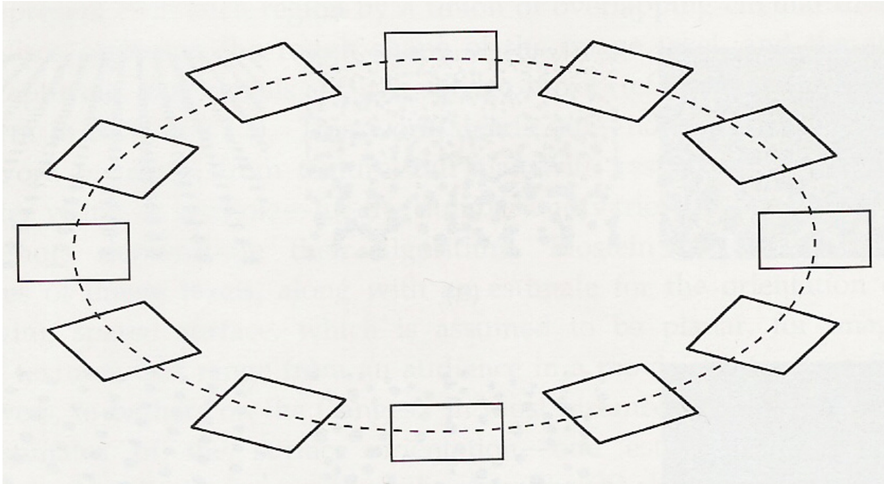
Cheetah





A.M. Low, Phd Thesis, 2006

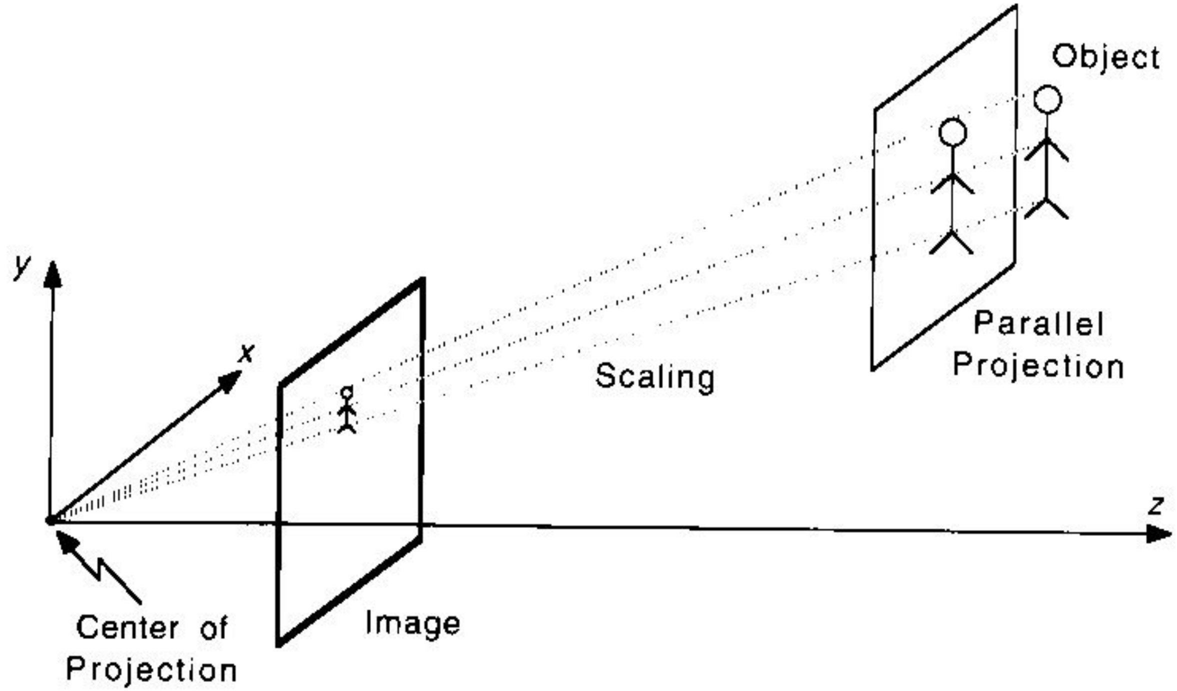
Perpendicular Lines



Orthographic projections of squares that are rotated with respect to each other in a plane inclined at ω=60° to the image plane.

$$\frac{\left\|(\mathbf{p}_1/l_1)\times(\mathbf{p}_2/l_2)\right\|}{\left\|\mathbf{p}_1/l_1\right\|^2+\left\|\mathbf{p}_2/l_2\right\|^2}=\frac{\cos(w)}{1+\cos^2(w)}$$

Parapespective Projection

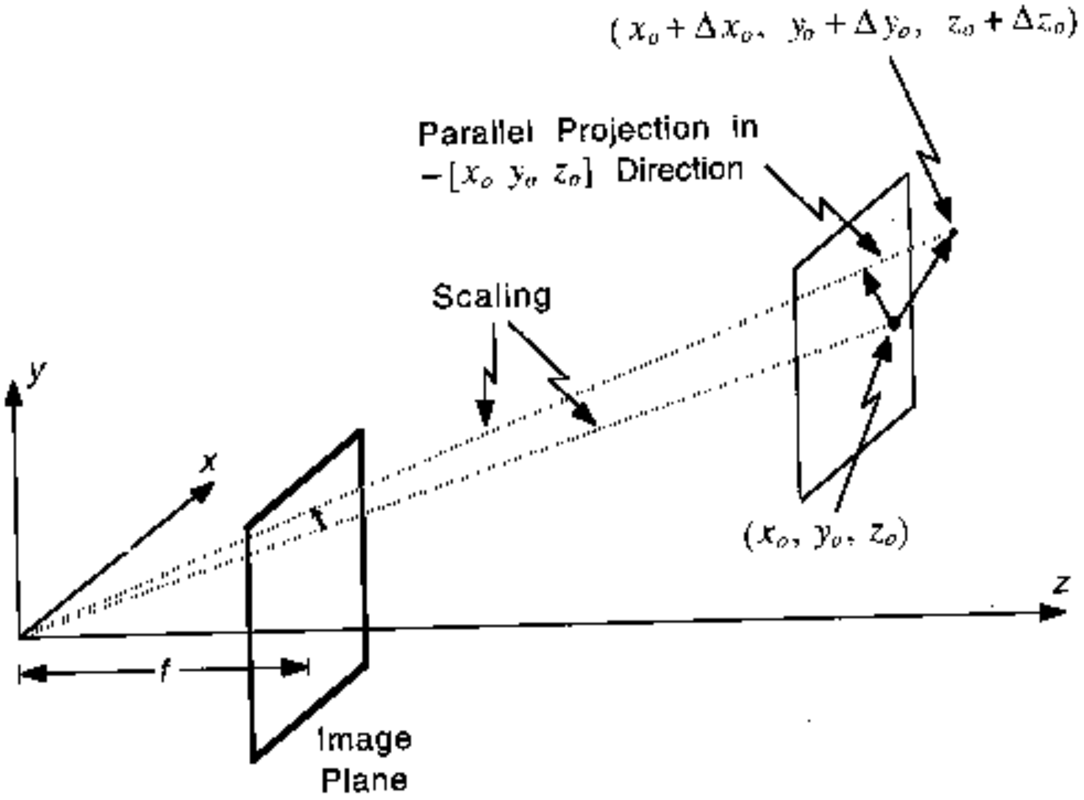


Generalization of the orthographic projection:

- Object dimensions small wrt distance to the center of projection.

→ Parallel projection followed by scaling

Parapespective Projection



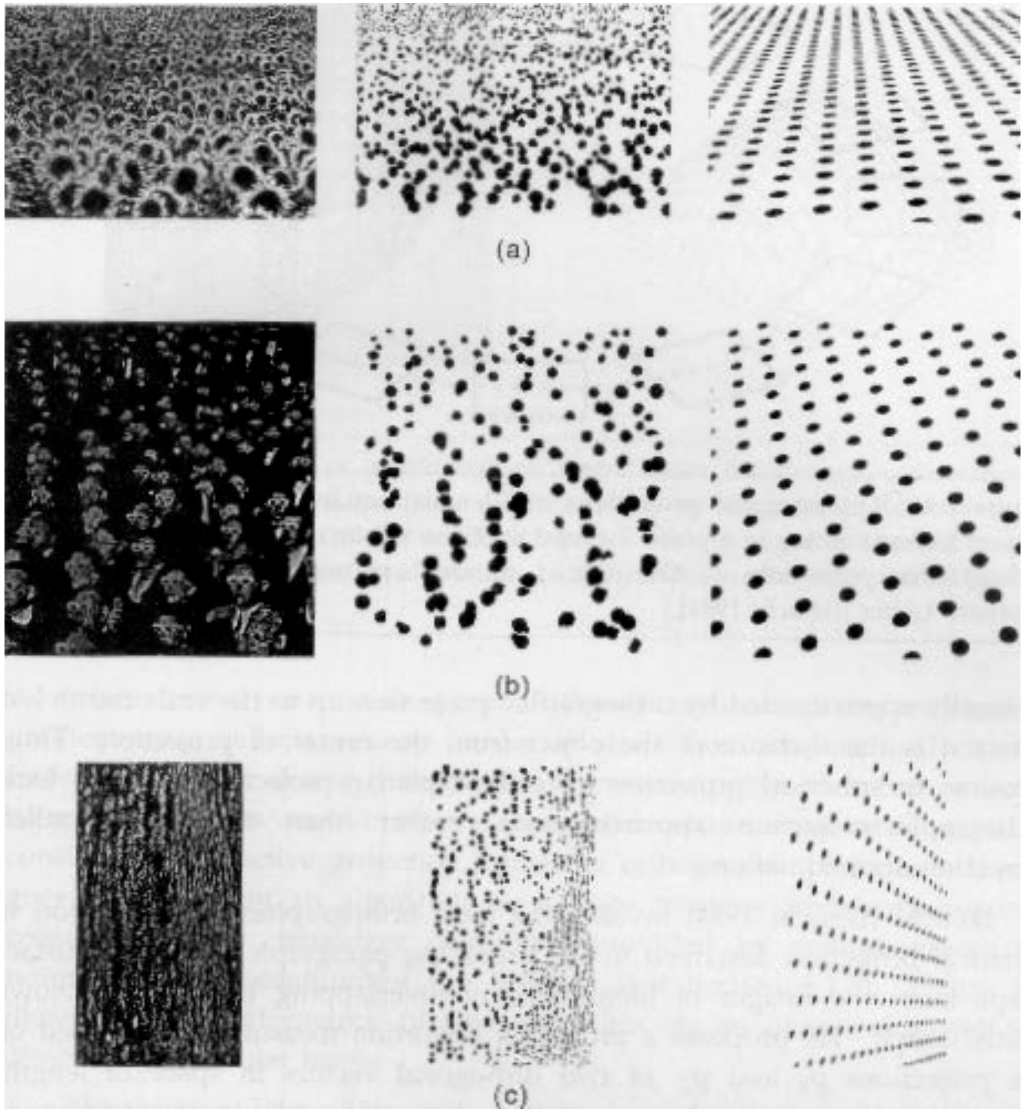
- For planar texels:

Unknown surface normal.

True Area.

Projected Area. $$A' = -\frac{f^2}{z_0^3}\mathbf{n} \bullet [x_0\, y_0\, z_0]A$$

Parapespective Projection



Texels:

- Image regions being brighter or darker than their surroundings.
- Assumed to have the same area in space.

→ Given enough texels, it becomes possible to estimate the normal.

Texture Gradient

Statistical Shape Recovery

Mesure texture density as opposed to texel area, that is, the number of textural primitives per unit surface.

Assuming the texture to be homogeneous, we have:

$$\psi \mathbf{n} \propto \mathbf{b}$$

$$\psi = \begin{bmatrix} u_1 & v_1 & 1 \\ \dots & \dots & \dots \\ u_n & v_n & 1 \end{bmatrix}^t$$

Unknown surface normal.

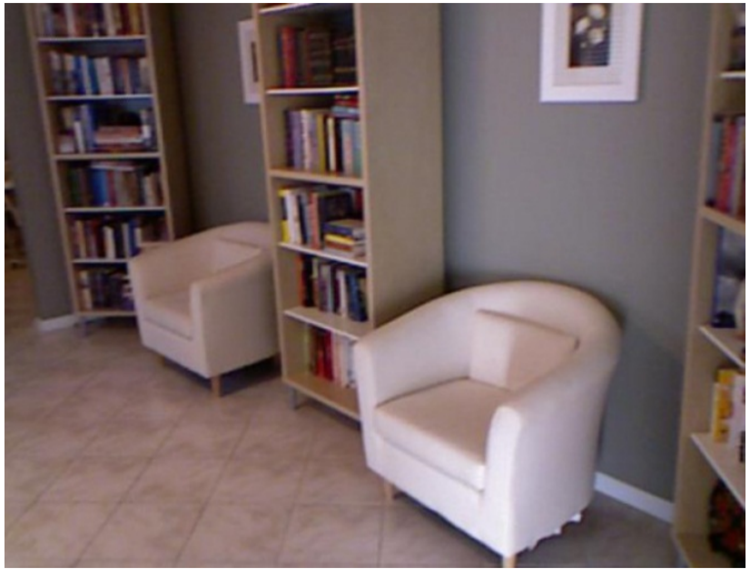
$$\mathbf{b} = [b_1, \dots, b_n]^t$$

Image coordinates.

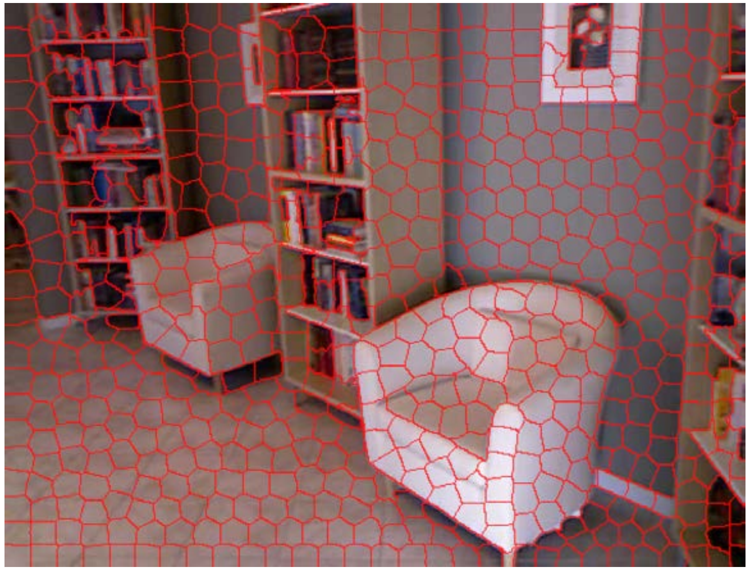
Function of density.

$$\Rightarrow \mathbf{n} = \frac{\psi \mathbf{n}}{\|\psi \mathbf{n}\|}$$

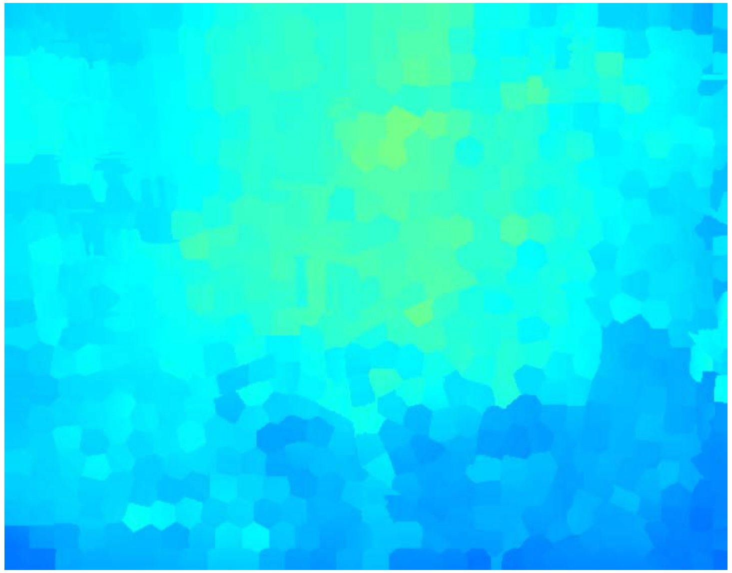
Machine Learning



Input Image



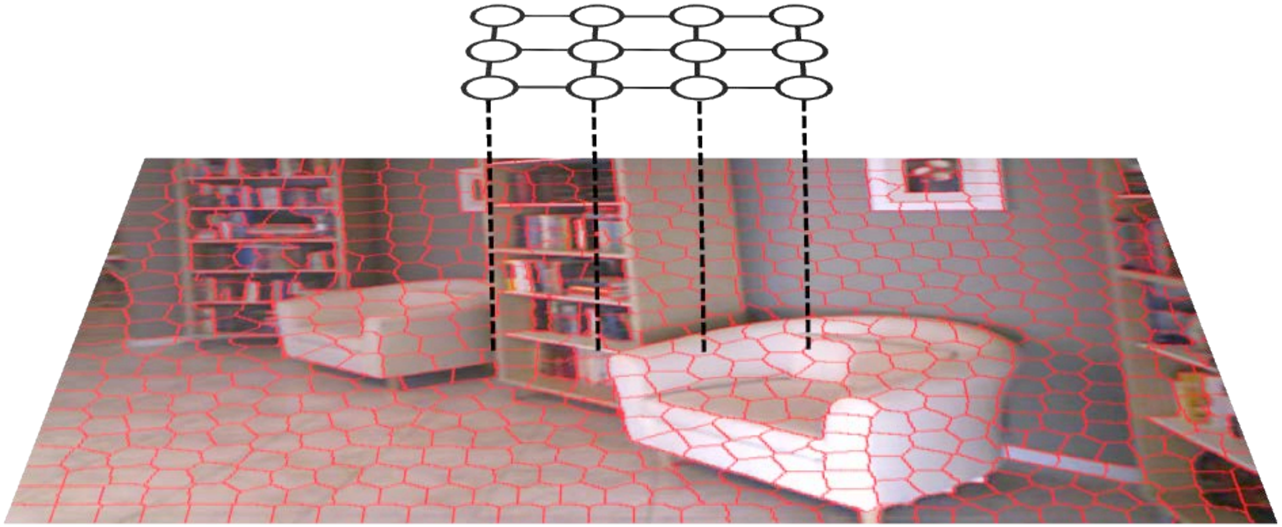
Superpixels



Train a regressor to predict depth —> Noisy predictions

Markov Random Field (MRF)

Graph with vertices and edges



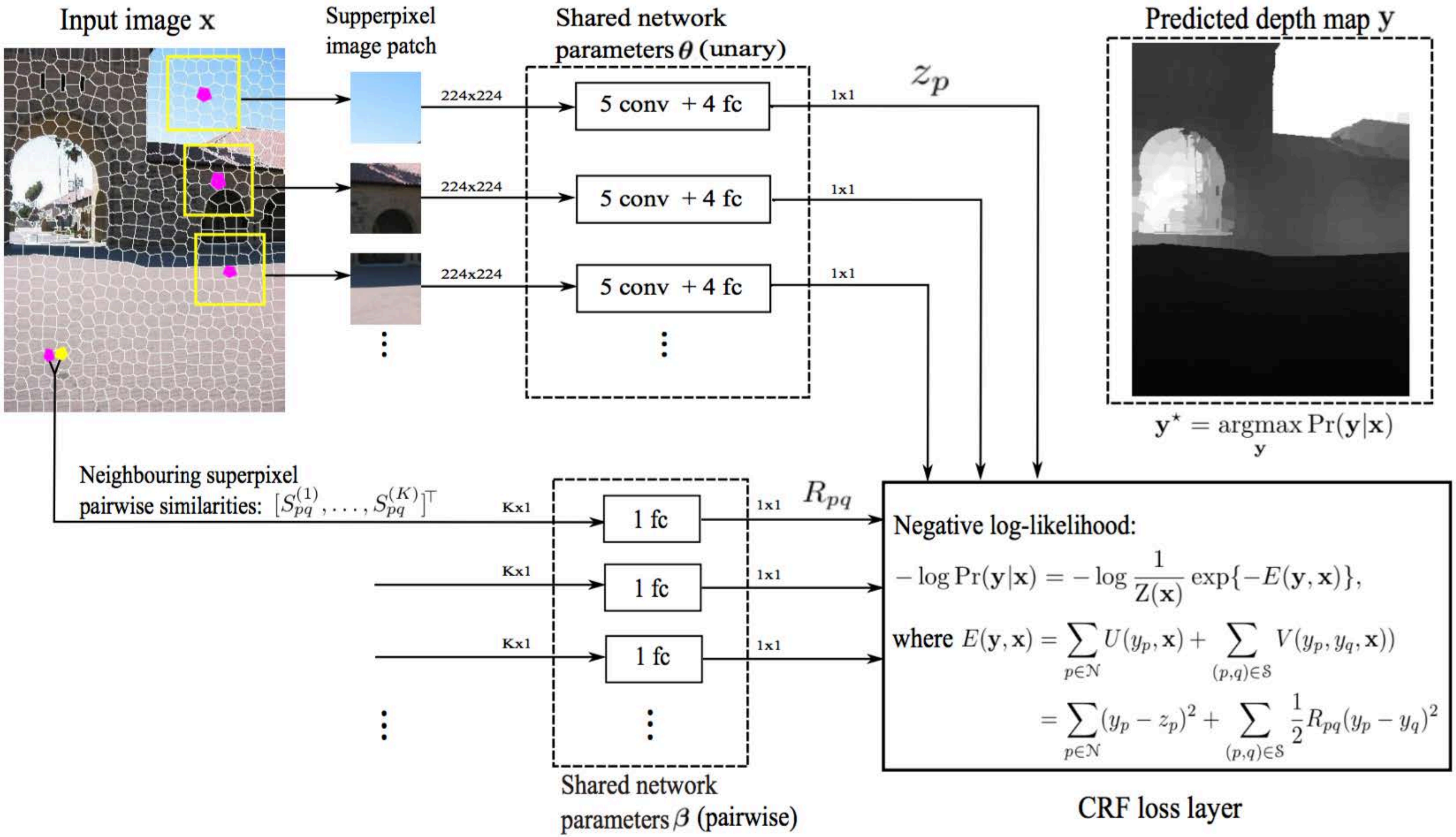
Assign values to the nodes to minimize

$$E(Y) = \underbrace{\sum_i \varphi(y_i)}_{\text{unary}} + \underbrace{\sum_{(i,j)} \psi(y_i, y_j)}_{\text{pairwise}}$$

—> Enforces consistency

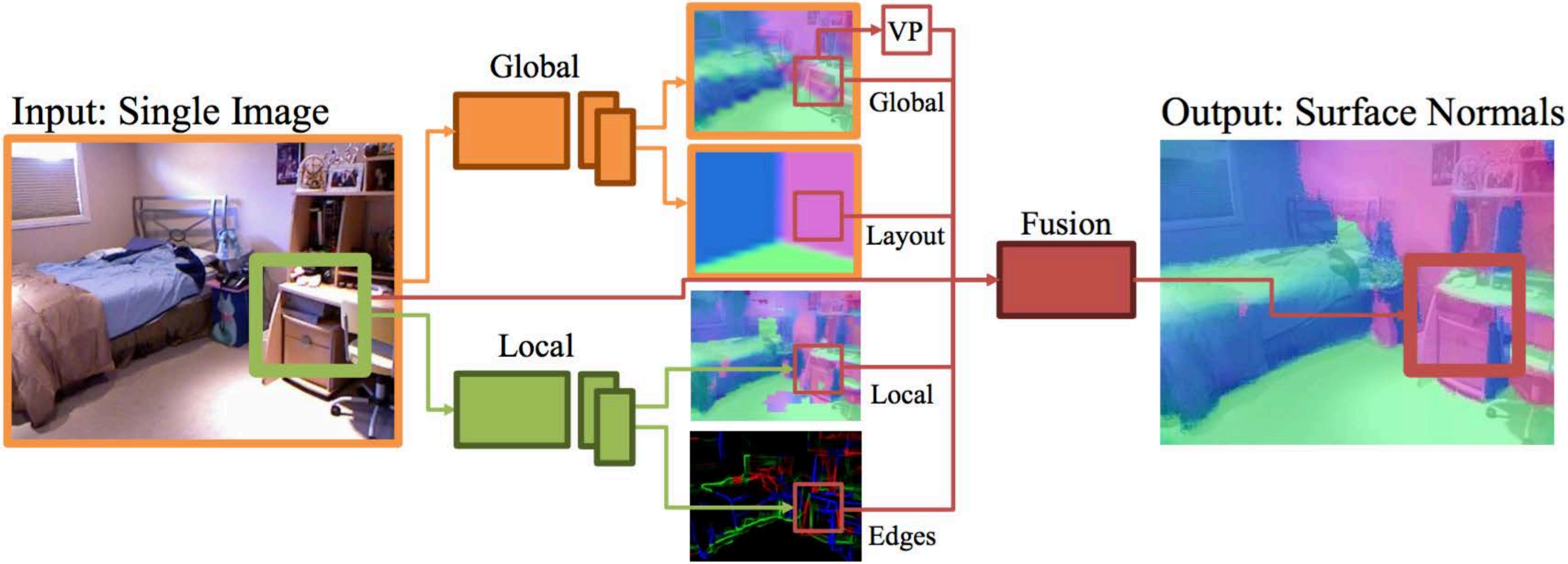
EPFL

Deep Learning with MRF



Liu et al., PAMI 2016

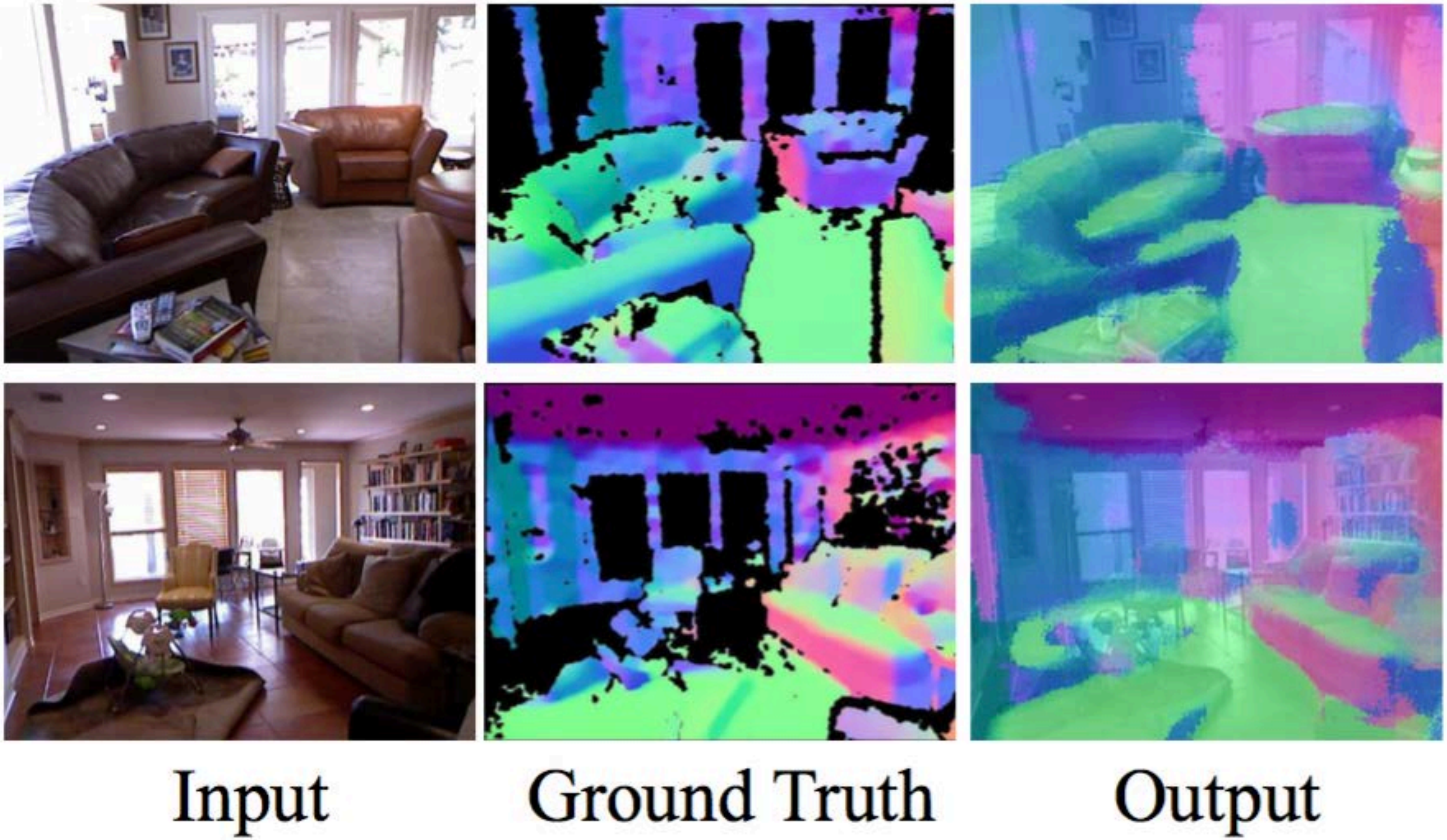
Deep Learning without MRF



- The network can be designed to enforce normal consistency.
- But only for the class of scenes it has been trained for.

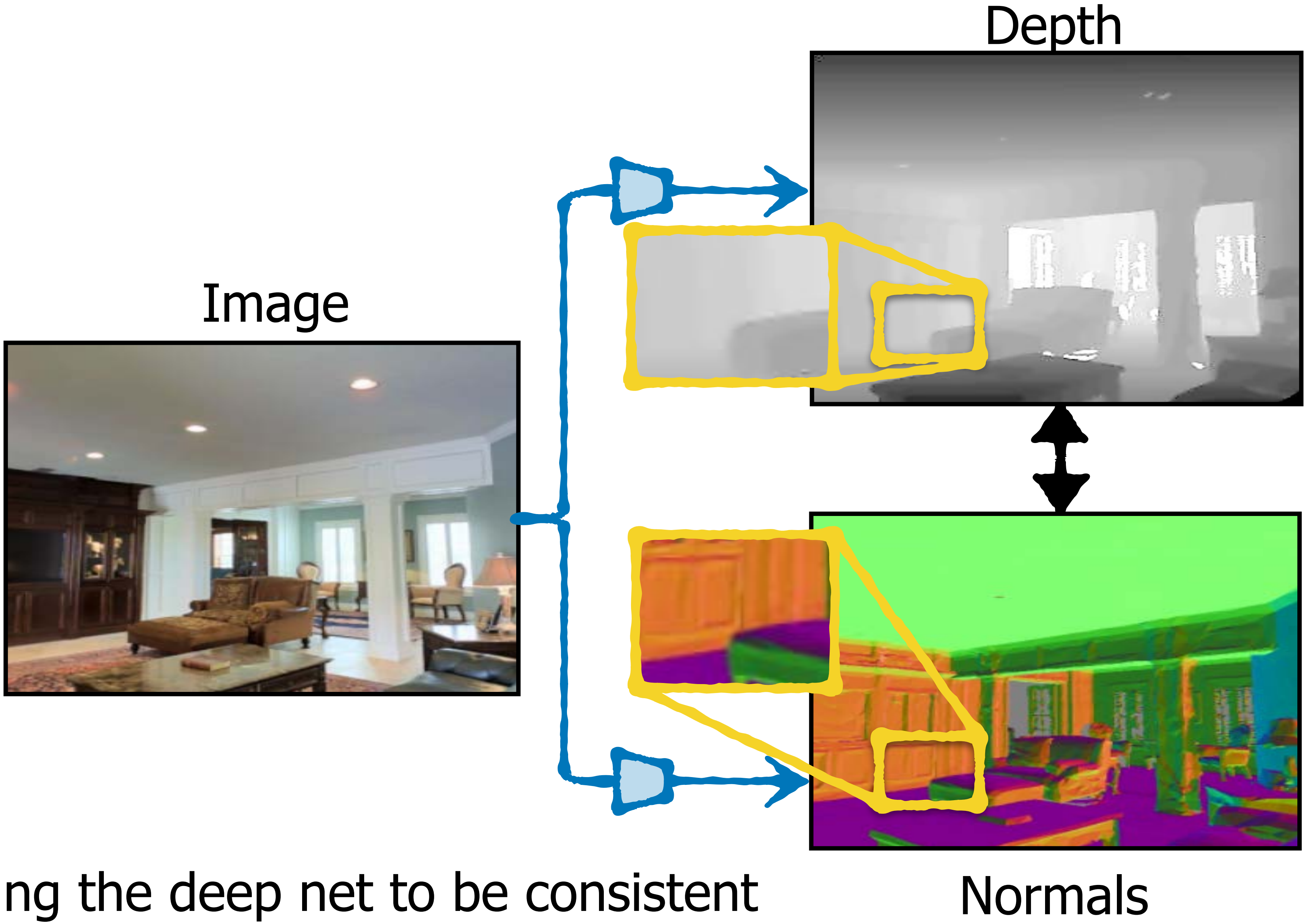
Wang et al., CVPR 2015

Normals from a Single Image



Wang et al., CVPR 2015

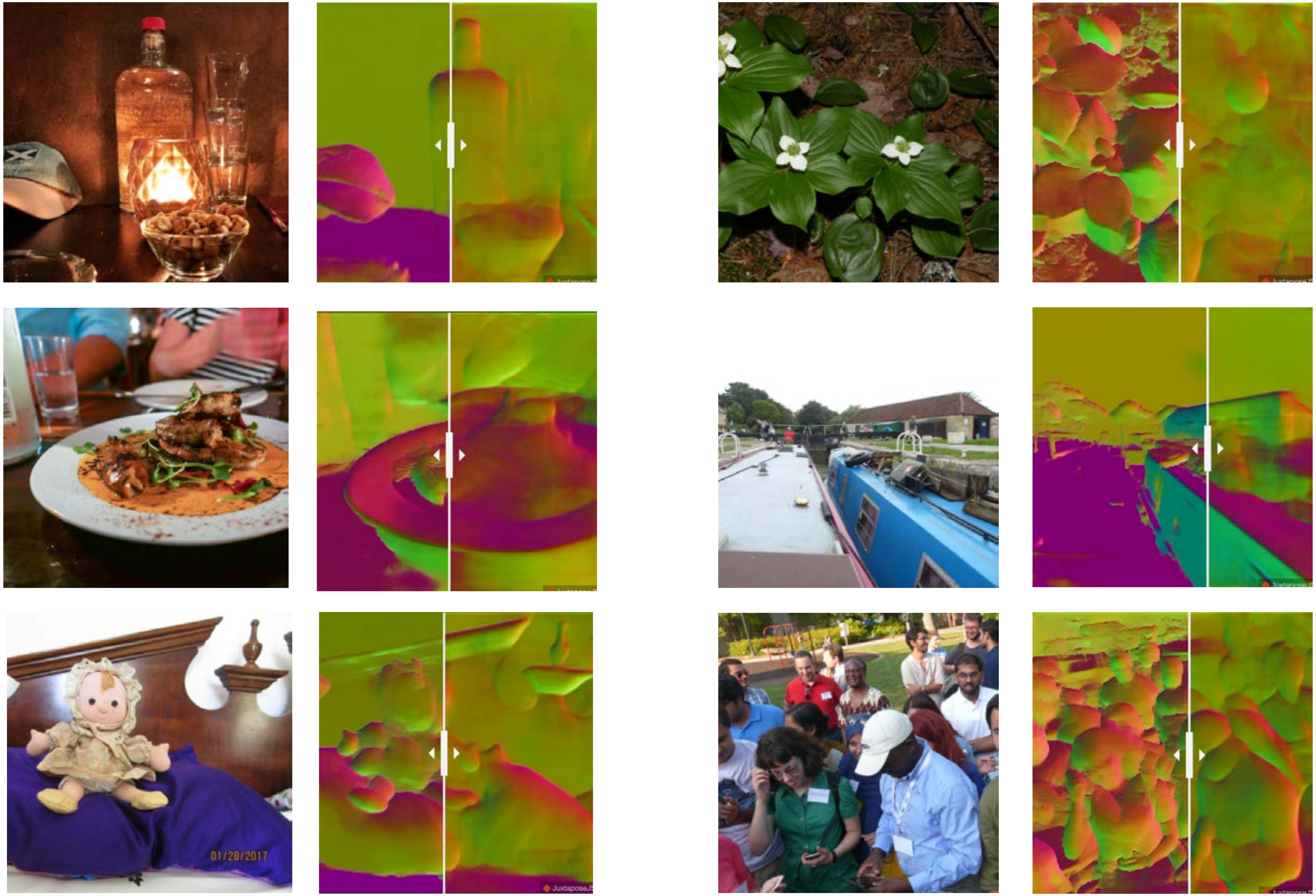
Enforcing Task Consistency



Forcing the deep net to be consistent across tasks increases robustness.

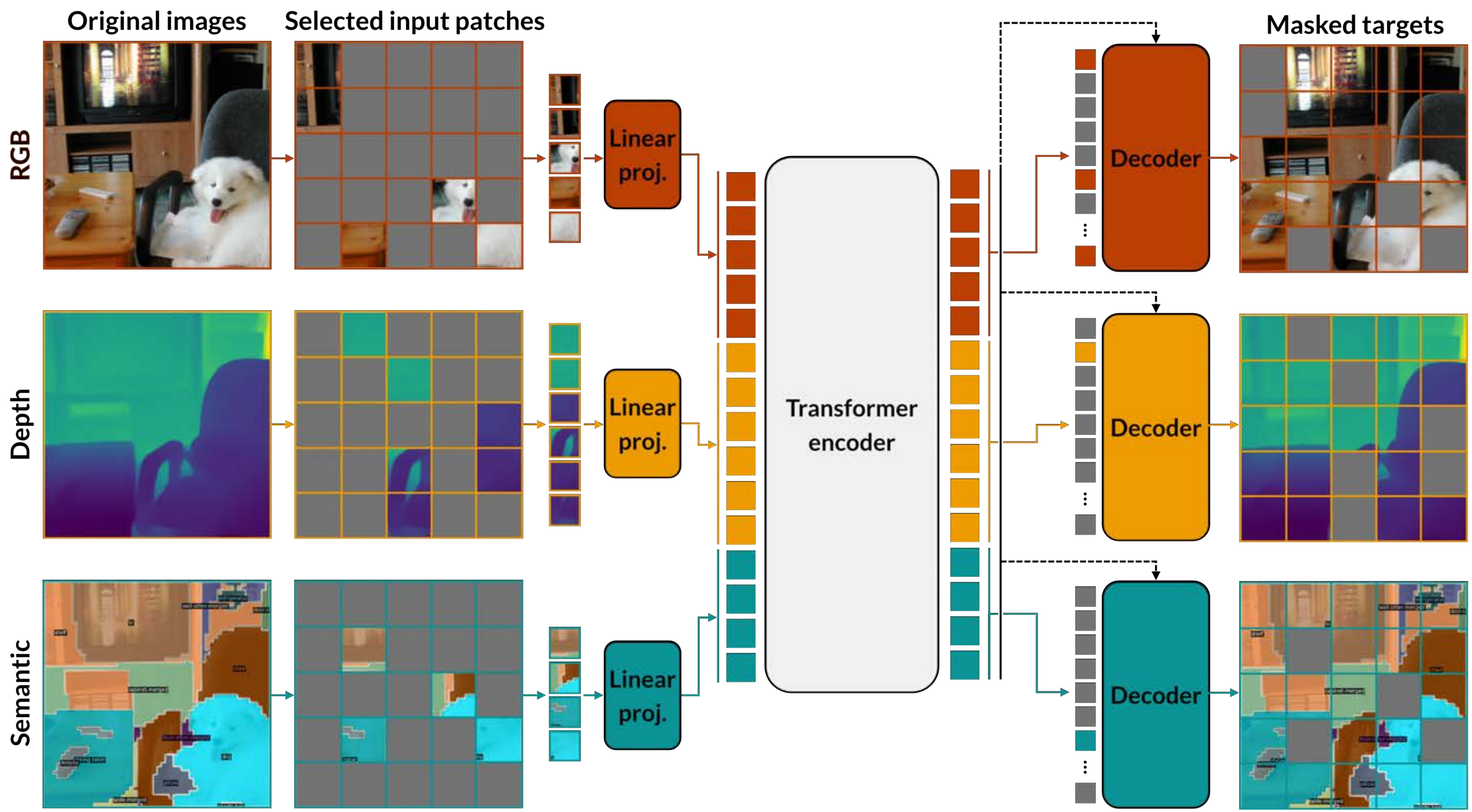
Zamir et al. , CVPR'20

A Very Diverse Training Database Helps



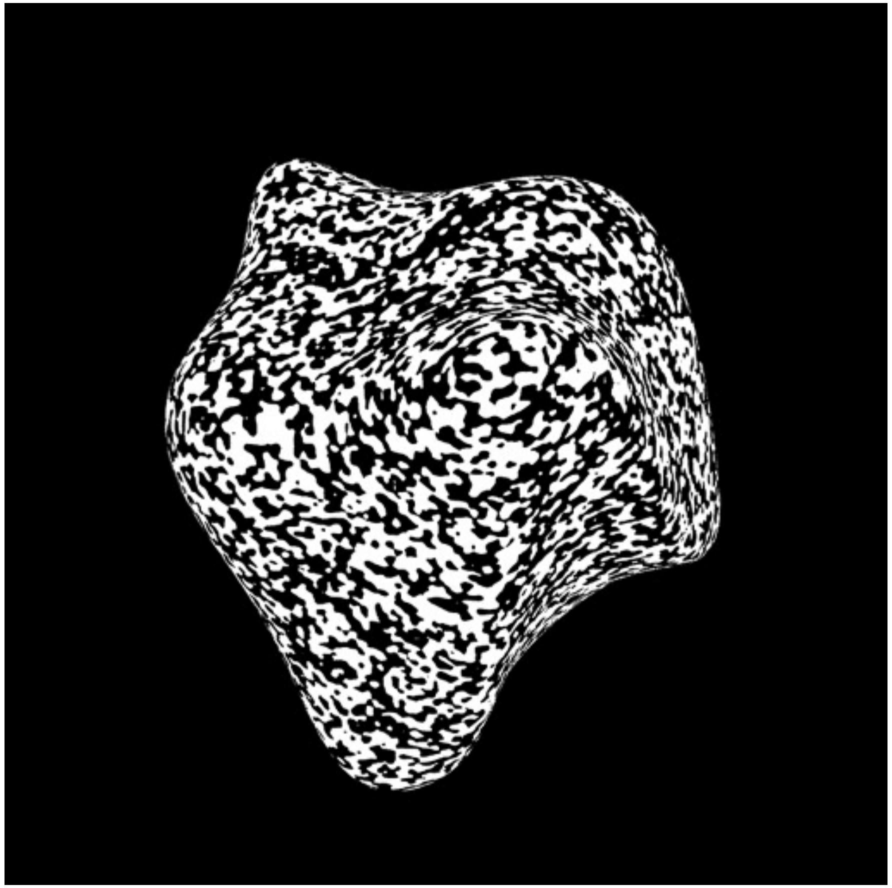
Eftekhar et al. , ICCV'21 vs Chen et al. , CVPR'20

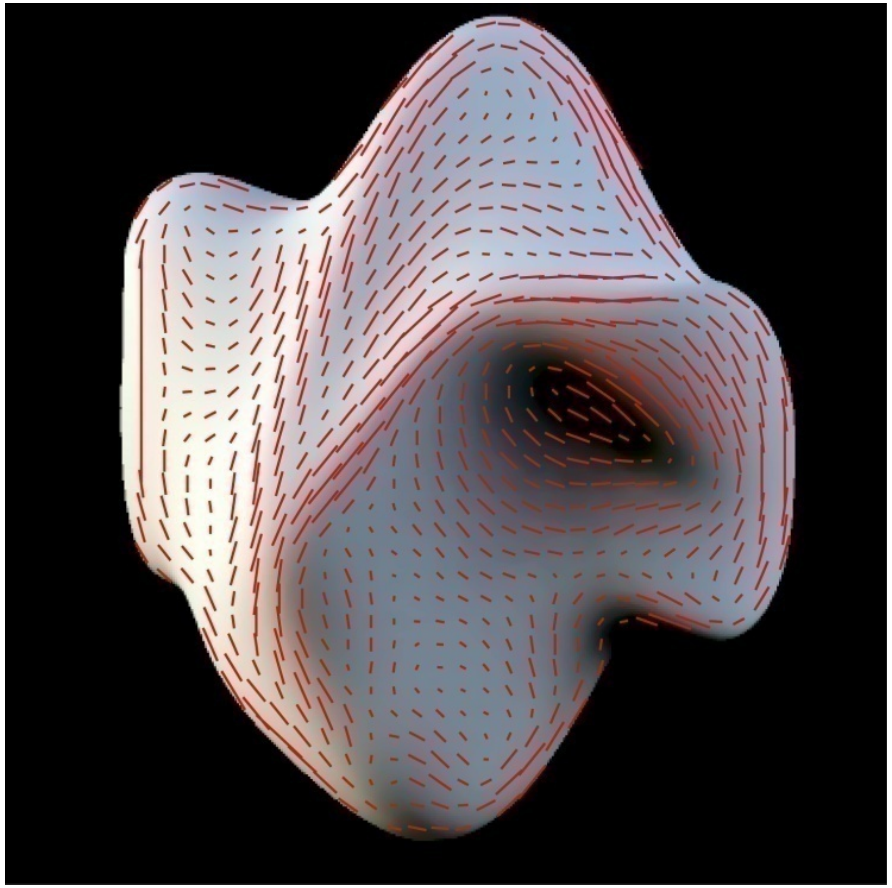
.. and so does a Transformer Architecture



Ranftl et al, ICCV'21

Optional: Illusory Shape Distorsion





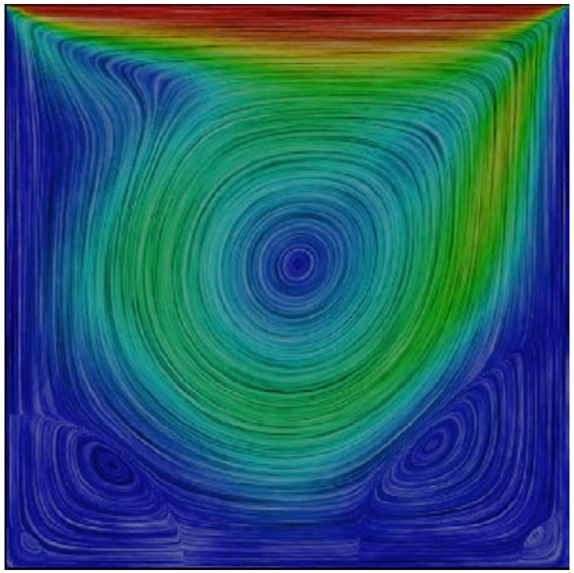
People seem to be sensitive to orientation fields in the cases of both texture and shading.

Flemming et al. PNAS'10

Optional: Shape from Smear

Hypothesis: If orientation and scale fields are the key source of information for 3D shape perception, it should be possible to induce a vivid sense of 3D shape by creating 2D patterns with appropriate scale and orientation fields.

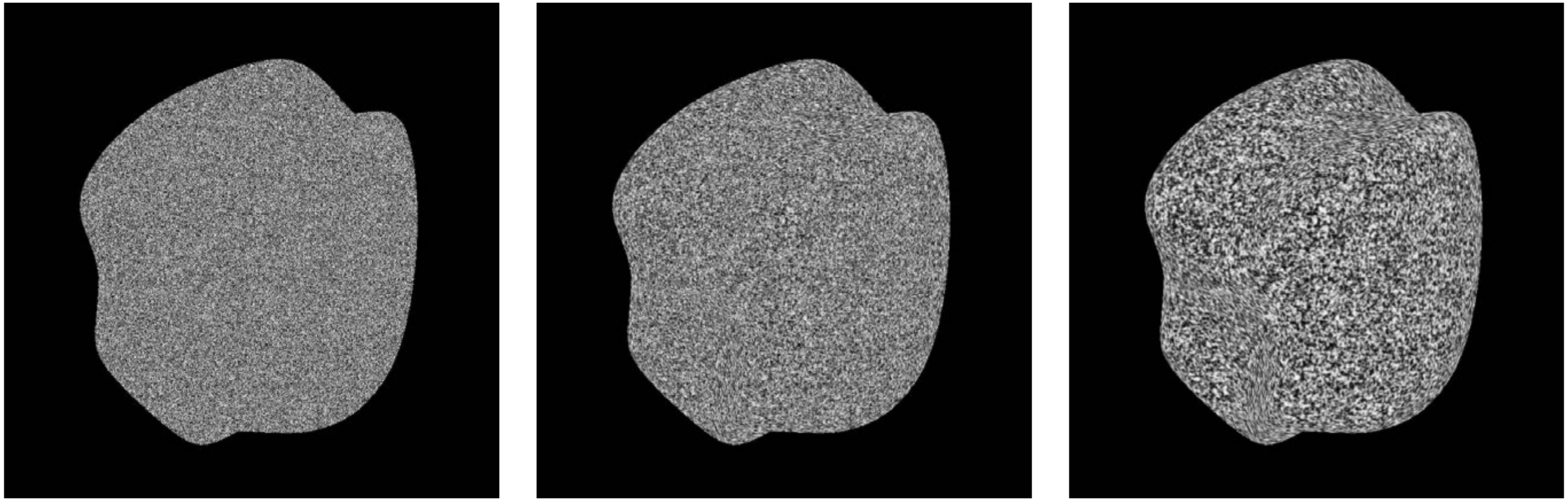
Test: Use a technique known as Line Integral Convolution to smear the texture along specific orientations and scale appropriately.



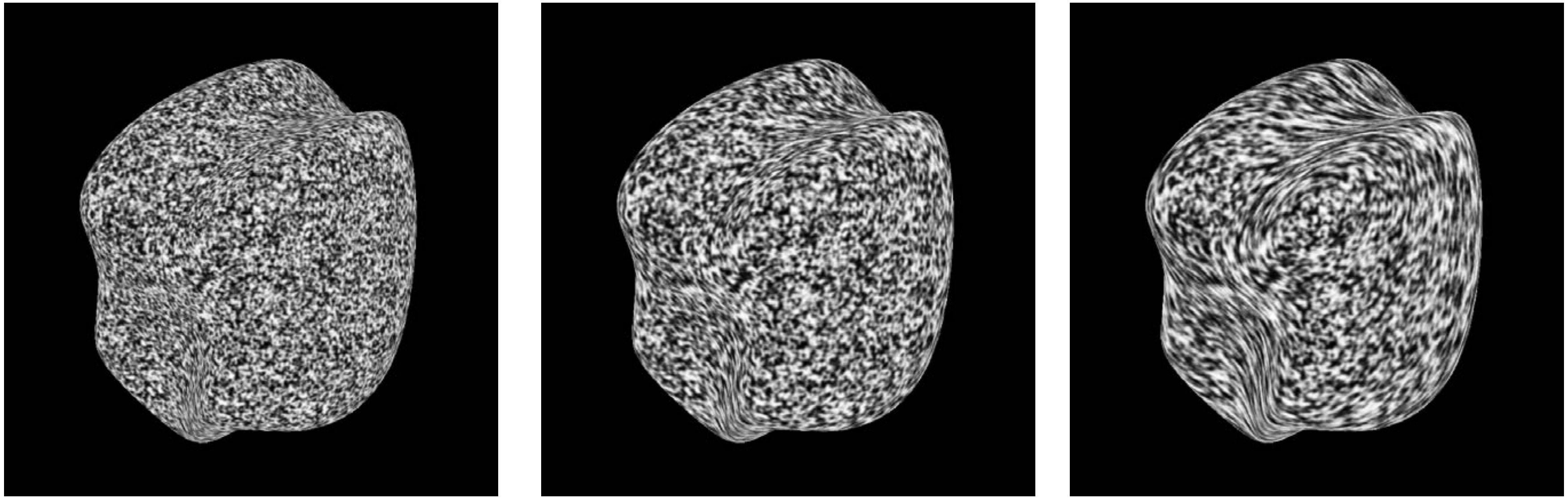
Flemming et al. PNAS'10

Optional: Scaling and Smearing

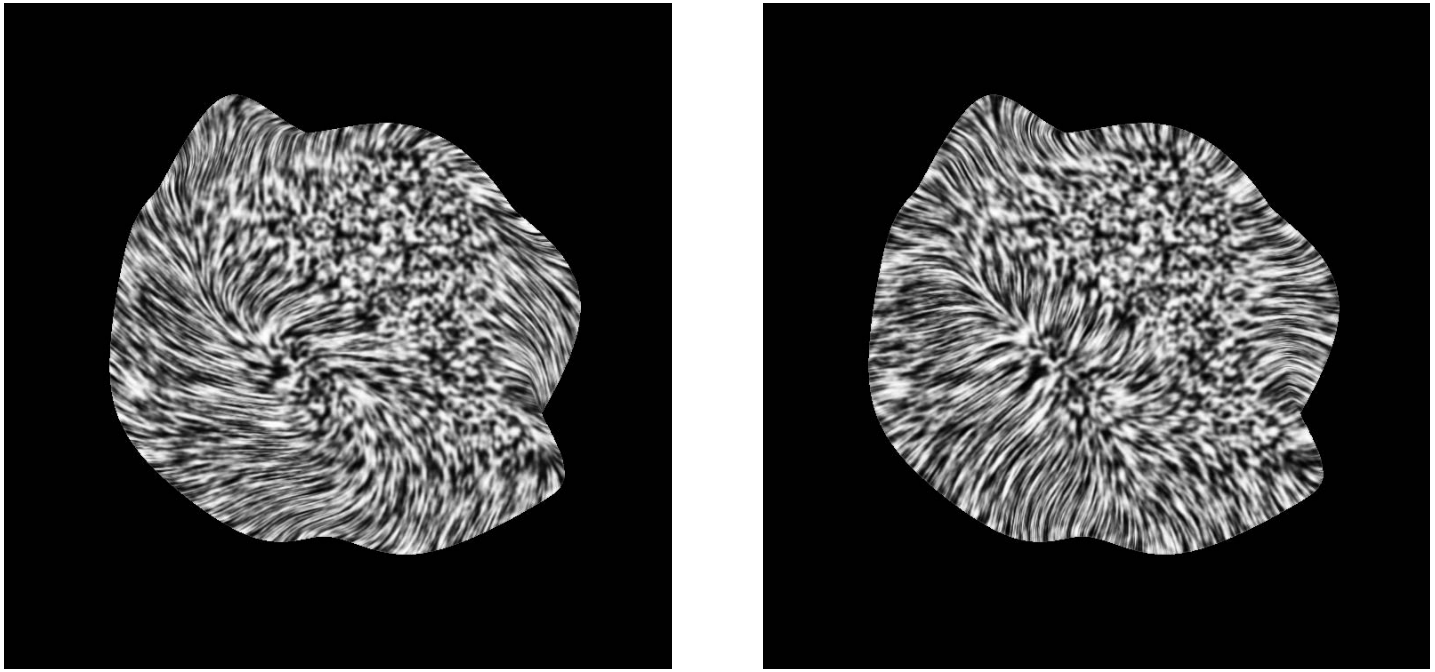
Scaling:



Smearing:



Optional: Inconsistent Stimulus



The orientation field cannot be integrated

- ➢ No depth perception.
- ➢ Do we integrate in our heads?
- ➢ Is this what the deep nets learn to do?

Strengths and Limitations

Strengths:
- Emulates an important human ability.

Limitations:
- Requires regular texture.
- Involves very strong assumptions.
- Deep learning can be used to weaken them.