FFTs in Graphics and Vision

Invariance of Shape Descriptors

Outline

- Math Overview
  - Translation and Rotation Invariance
  - The 0th Order Frequency Component
- Shape Descriptors
- Invariance

Invariance

Given a unitary representation (ρ, V) of a group G, we can decompose V into sub-representations:

$$V = \bigoplus_{\lambda \in \Lambda} V^\lambda$$

where $\rho_g\left(V^\lambda\right) \subset V^\lambda$ for all $g \in G$ and all $\lambda \in \Lambda$.

Invariance

$$V = \bigoplus_{\lambda \in \Lambda} V^\lambda$$

Let $\pi_\lambda : V \to V^\lambda$ be the projection operator. Then for any $v \in V$ we have:

$$v = \sum_{\lambda \in \Lambda} \pi_\lambda(v)\,.$$

Mashcke's Theorem:

Since (ρ, V^λ) is a sub-representation, we have:

$$\rho_g \circ \pi_\lambda = \pi_\lambda \circ \rho_g$$

Invariance

$$\pi_\lambda : V \to V^\lambda \qquad \rho_g \circ \pi_\lambda = \pi_\lambda \circ \rho_g$$

Consider the map $\Phi: V \to \mathbb{R}^{|\Lambda|}$ with:

$$\Phi(v) = \{|\pi_\lambda(v)|\}_{\lambda\in\Lambda}.$$

Then for any $g \in G$ we have:

$$\begin{aligned}\Phi\left(\rho_g(v)\right) &= \left\{\left|\pi_\lambda\left(\rho_g(v)\right)\right|\right\}_{\lambda\in\Lambda}\\ &= \left\{\left|\rho_g\left(\pi_\lambda(v)\right)\right|\right\}_{\lambda\in\Lambda}\\ &= \{|\pi_\lambda(v)|\}_{\lambda\in\Lambda}\\ &= \Phi(v)\end{aligned}$$

$\Rightarrow$ The map Φ is G-invariant.

Translation Invariance

Given a function f in 2D, we obtain a translation invariant representation of the function by storing the magnitudes of the frequency components:

$$f(x,y) = \sum_{l,m=-\infty}^{\infty} \hat{f}_{lm} \cdot \frac{e^{i(lx+my)}}{2\pi}$$

$$\Downarrow$$

$$\left\{\left\|\hat{f}_{lm}\right\|\right\} \quad l,m \in \mathbb{Z}$$

Rotation Invariance (Circle)

Given a function $f(\theta)$ on a circle, we obtain a rotation invariant representation by storing the magnitudes of the frequency components:

$$f(\theta) = \sum_{l=-\infty}^{\infty} \hat{f}_l \cdot \frac{e^{il\theta}}{\sqrt{2\pi}}$$

$$\Downarrow$$

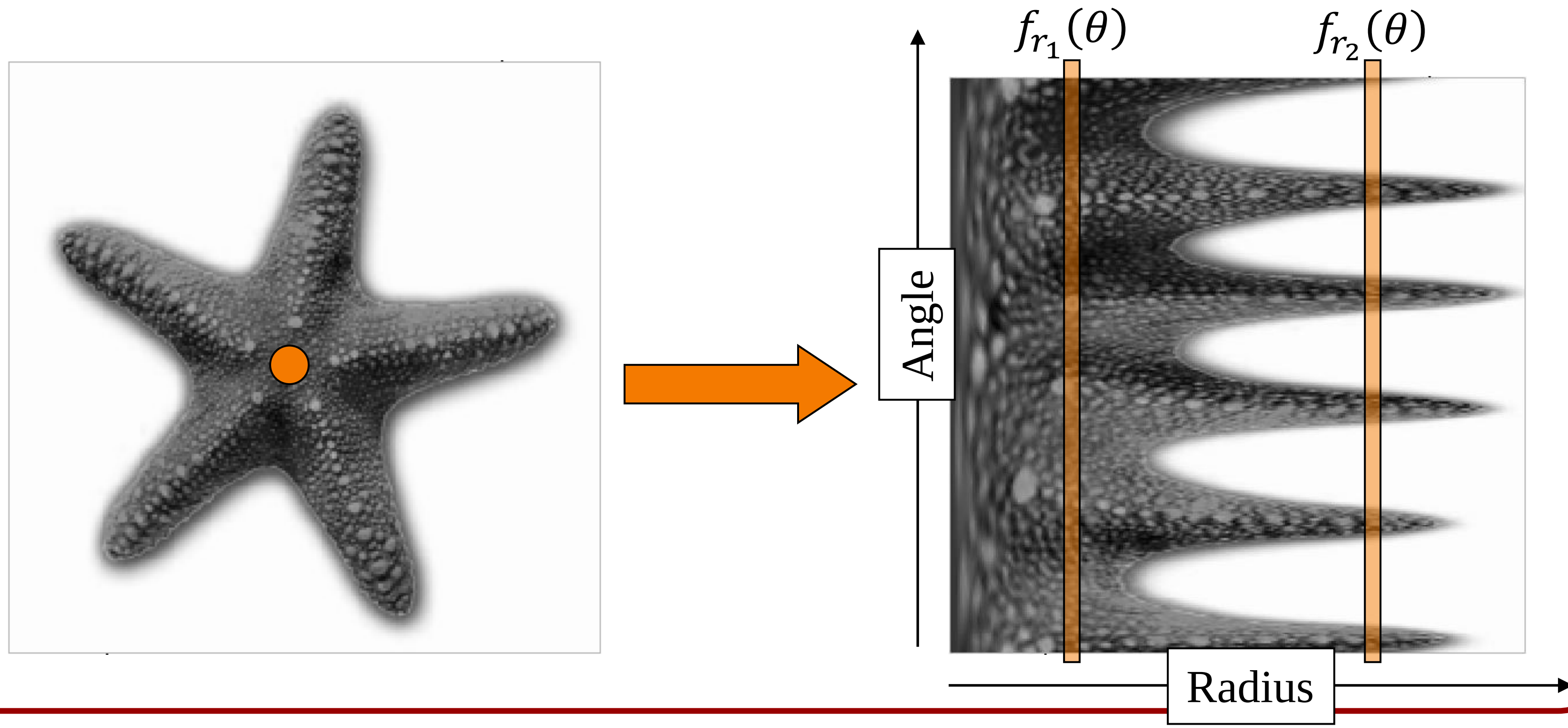
$$\left\{ \left\| \hat{f}_l \right\| \right\} \quad l \in \mathbb{Z}$$

Rotation Invariance (2D)

Given a function $f(x,y)$ in 2D, we obtain a rotation invariant representation of f by:

- Expressing f in polar coordinates:

$$f(f,\theta) = f(r \cdot \cos\theta , r \cdot \sin\theta)$$



Rotation Invariance (2D)

Given a function $f(x,y)$ in 2D, we obtain a rotation invariant representation of f by:

- Expressing f in polar coordinates:
$$f(f,\theta) = f(r\cdot\cos\theta\,, r\cdot\sin\theta)$$
- Expressing each radial restriction in terms of its Fourier decomposition:
$$f(f,\theta) = \sum_{l=-\infty}^{\infty} \hat{f}_l(r)\cdot\frac{e^{il\theta}}{\sqrt{2\pi}}$$
- Storing the magnitude of the frequency components of the different radial restrictions:
$$\left\{\left\|\hat{f}_l(r)\right\|\cdot\sqrt{2\pi r}\right\} \quad l\in\mathbb{Z}, r\in[0,1]$$

Rotation Invariance (Sphere)

Given a function $f(\theta,\phi)$ on a sphere, we obtain a rotation invariant representation by storing the magnitudes of the frequency components:

$$f(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\hat{f}_{lm}\cdot\mathbf{Y}_l^m(\theta,\phi)$$

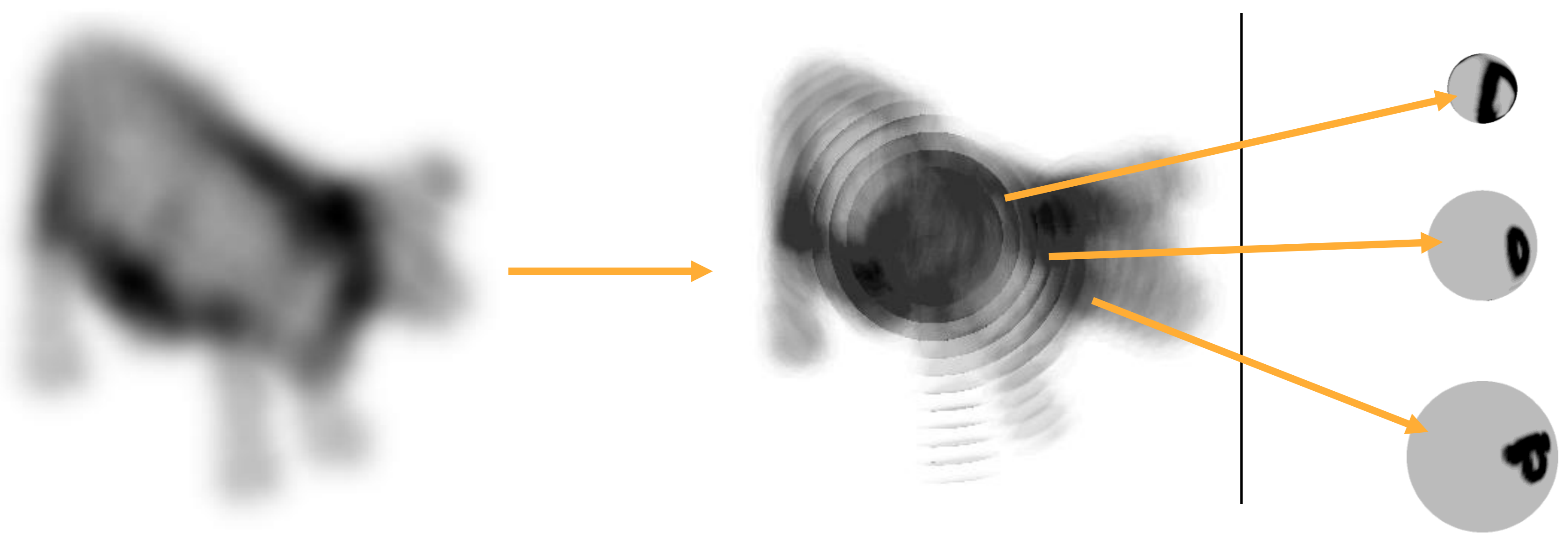
$$\Downarrow$$

$$\left\{\sqrt{\sum_{m=-l}^{l}\left\|\hat{f}_{lm}\right\|^2}\right\}\qquad l\in\mathbb{Z}^{\geq 0}$$

Rotation Invariance (3D)

Given a function $f(x, y, z)$ in 3D, we obtain a rotation invariant representation of f by:

- Expressing f in spherical coordinates:
  $f(r, \theta, \phi) = f(r \cdot \cos\theta \cdot \sin\phi, r \cdot \cos\phi, r \cdot \sin\theta \cdot \sin\phi)$



Rotation Invariance (3D)

Given a function $f(x, y, z)$ in 3D, we obtain a rotation invariant representation of f by:

- Expressing f in spherical coordinates:
  $f(r, \theta, \phi) = f(r \cdot \cos\theta \cdot \sin\phi, r \cdot \cos\phi, r \cdot \sin\theta \cdot \sin\phi)$
- Expressing each radial restriction in terms of its spherical harmonic decomposition:
  $$f(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{f}_{lm}(r) \cdot \mathbf{Y}_l^m(\theta, \phi)$$
- Storing the size of the frequency components coefficients of the different radial restrictions:
  $$\left\{ \sqrt{\sum_{m=-l}^{l} \left\| \hat{f}_{lm}(r) \right\|^2} \cdot \sqrt{4\pi r^2} \right\} \quad l \in \mathbb{Z}^{\geq 0}, r \in [0,1]$$

The 0^{th} Order Frequency Component

Given a function on the circle $f(\theta)$, we can express the function in terms of its Fourier decomposition:

$$f(\theta) = \sum_{l=-\infty}^{\infty} \hat{f}_l \cdot \frac{e^{il\theta}}{\sqrt{2\pi}}$$

What is the meaning of the 0^{th} order frequency component?

The 0th Order Frequency Component

The l^{th} frequency is the dot product of the function with the l^{th} (normalized) complex exponential:

$$\hat{f}_l = \left\langle f(\theta), \frac{e^{il\theta}}{\sqrt{2\pi}} \right\rangle = \int_0^{2\pi} f(\theta) \cdot \frac{e^{-il\theta}}{\sqrt{2\pi}} \, d\theta$$

So the 0th frequency component is:

$$\hat{f}_0 = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(\theta) \, d\theta$$

The 0th Order Frequency Component

Up to a normalization term, the 0th frequency component of a function $f(\theta)$ is the integral of the function over the circle:

$$\hat{f}_0 = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(\theta)\, d\theta$$

The 0th Order Frequency Component

Given a function on the sphere $f(\theta,\phi)$, we can express the function in terms of its spherical harmonic decomposition:

$$f(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\hat{f}_{lm}\cdot\mathbf{Y}_l^m(\theta,\phi)$$

What is the meaning of the 0th order frequency component?

The 0th Order Frequency Component

The $(l,m)^{\text{th}}$ spherical harmonic coefficient is computed by taking the dot product of the function with the $(l,m)^{\text{th}}$ spherical harmonic:

$$\hat{f}_{lm} = \langle f(\theta,\phi), \mathbf{Y}_l^m(\theta,\phi)\rangle$$

So the 0th frequency component is:

$$\hat{f}_{00} = \frac{1}{\sqrt{4\pi}} \int_{|p|=1} f(p)\, dp$$

The 0^{th} Order Frequency Component

Up to a normalization term, the 0^{th} frequency component of a function $f(\theta,\phi)$ is the integral of the function over the sphere:

$$\hat{f}_{00} = \frac{1}{\sqrt{4\pi}} \int_{|p|=1} f(p)\, dp$$

The 0th Order Frequency Component

<u>Note</u>:

When the function f is real, the 0th frequency coefficient will be real.

If, additionally, the function f is positive the 0th frequency coefficient will also be positive:

$$\|\hat{f}_0\| = \hat{f}_0$$
$$\|\hat{f}_{00}\| = \hat{f}_{00}$$

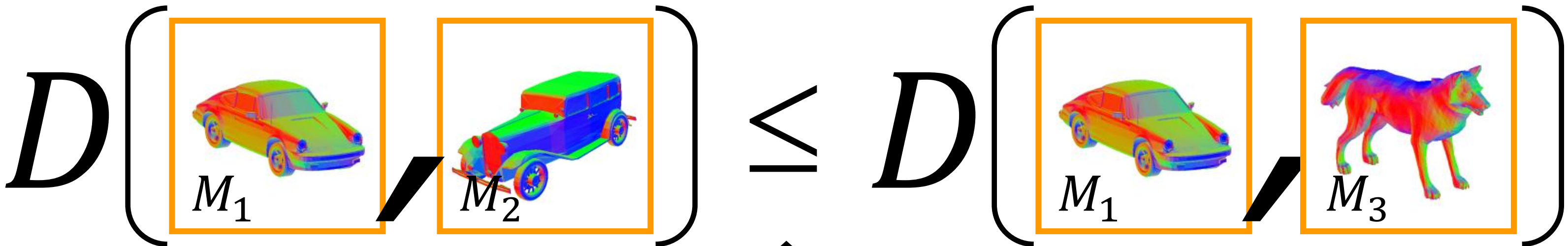
Outline

- Math Overview
- Shape Descriptors
  - Shape Histograms (Ankerst *et al*.)
  - Shape Distributions (Osada *et al*.)
  - Extended Gaussian Images (Horn)
- Invariance

Shape Matching

General Approach

Define a function that takes in two models and returns a measure of their proximity.



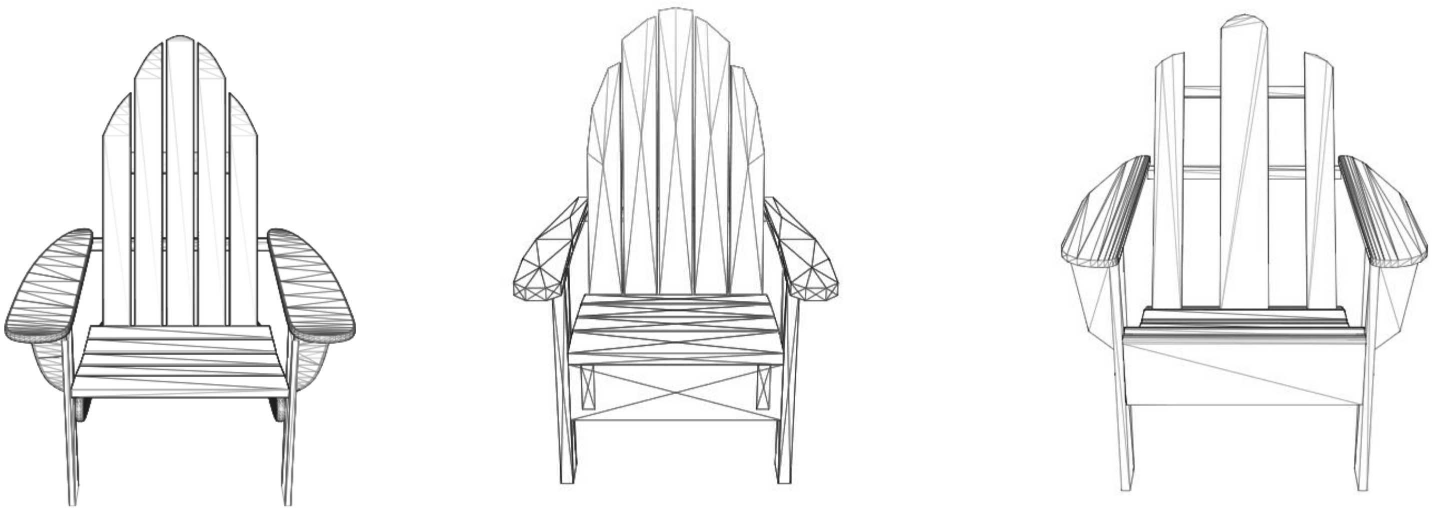
M_1 is closer to M_2 than it is to M_3

Shape Descriptors

<u>Challenge</u>

It is difficult to match shapes directly:

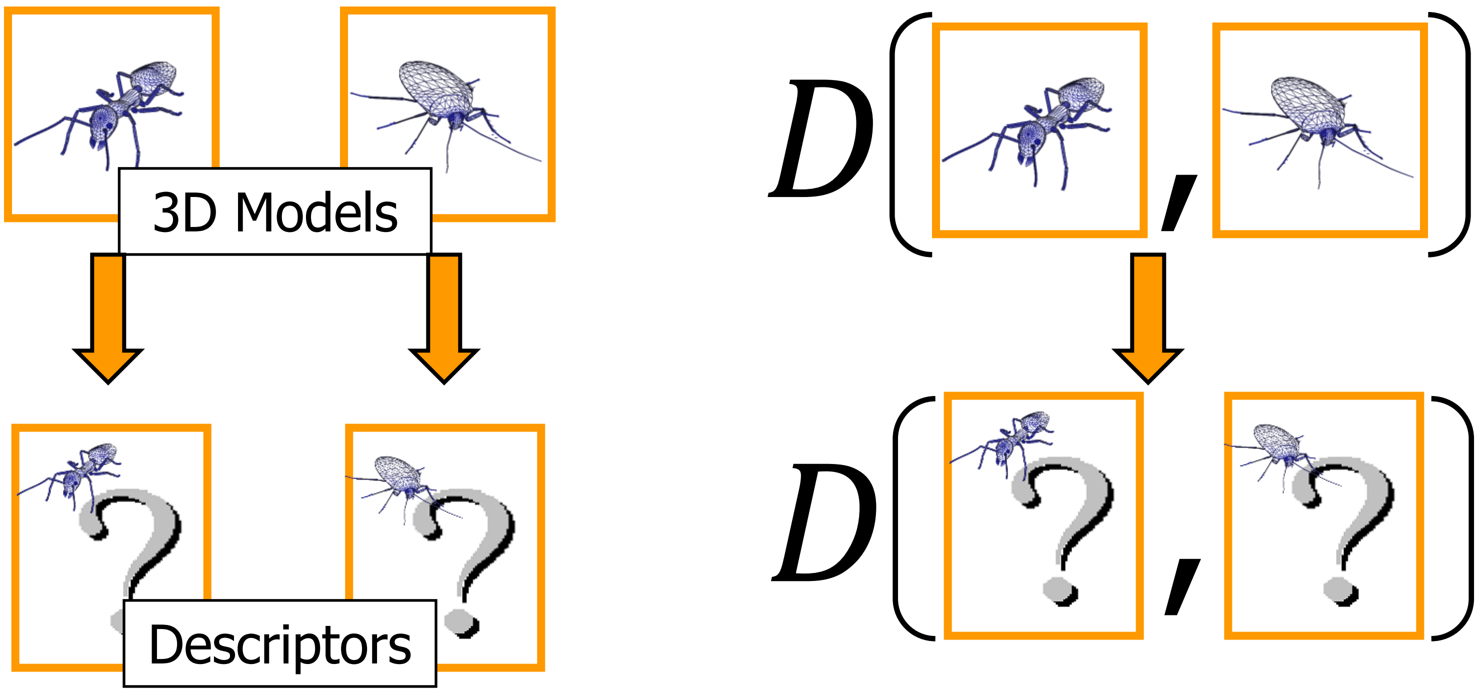
- Similar shapes can have different triangulations
- Similar shapes can have different genera
- Similar shapes may be in different poses
- Etc.



Shape Descriptors

Solution

Represent shapes by a structured abstraction that represents every shape in the same domain.



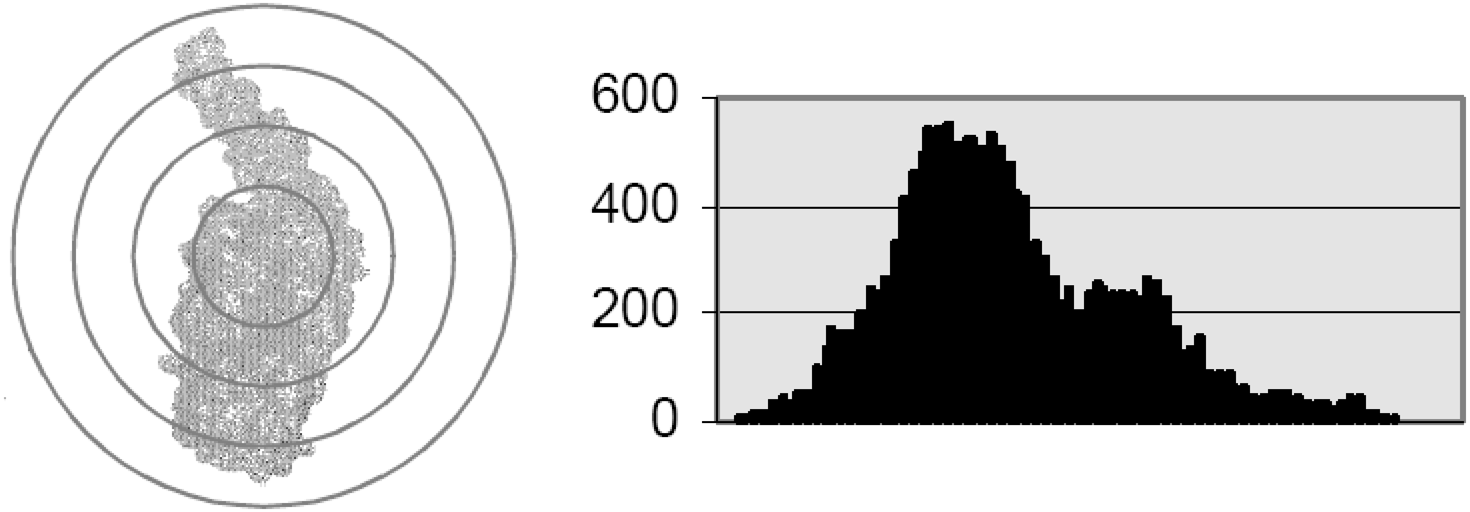
Outline

- Math Overview
- Shape Descriptors
  - Shape Histograms (Ankerst *et al*.)
  - Shape Distributions (Osada *et al*.)
  - Extended Gaussian Images (Horn)
- Invariance

Shape Histograms

Approach

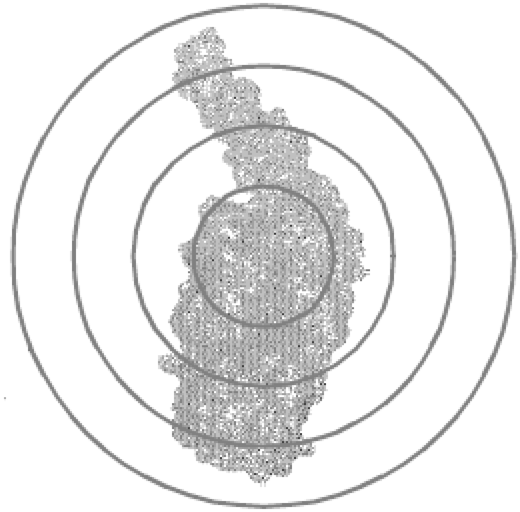
- Decompose space into concentric shells
- Store how much of the shape falls within each shell

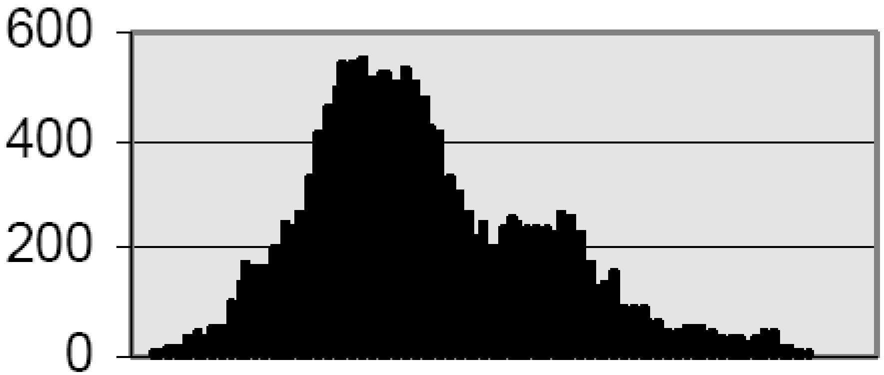


Shape Histograms

Properties

- The shape is represented by 1D array of values.
- The representation is invariant to rotation (about the center)





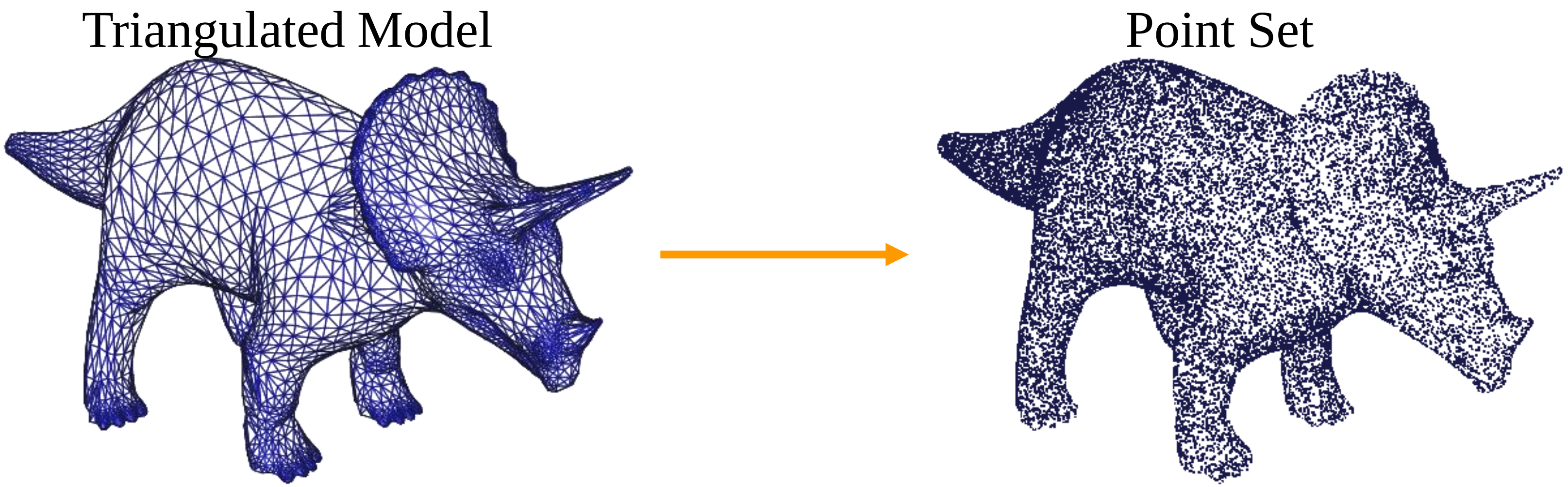
Outline

- Math Overview
- Shape Descriptors
  - Shape Histograms (Ankerst *et al*.)
  - Shape Distributions (Osada *et al*.)
  - Extended Gaussian Images (Horn)
- Invariance

D2 Shape Distributions

Approach

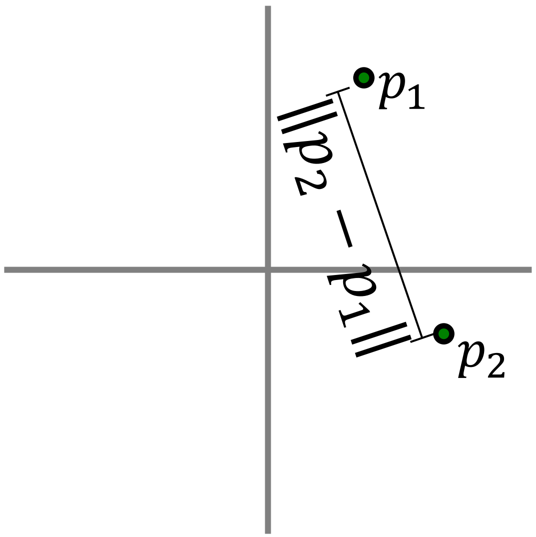
Avoid the whole problem of tesselation, genus, etc. by building the shape descriptor from random samples from the surface of the model:

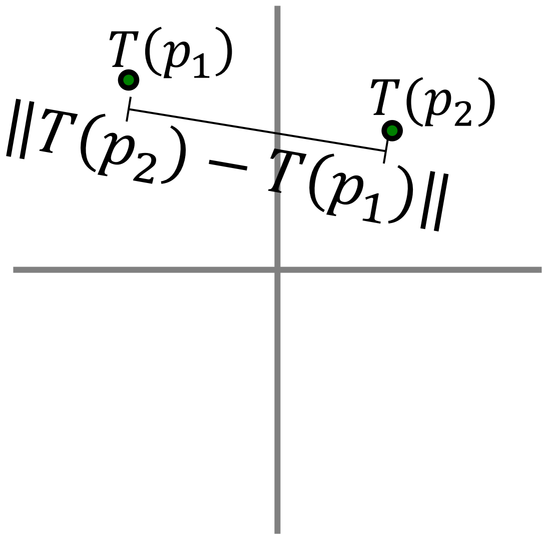


D2 Shape Distributions

Key Idea

Use the fact that the distance between pairs of points on the model does not change if the model is translated and/or rotated.



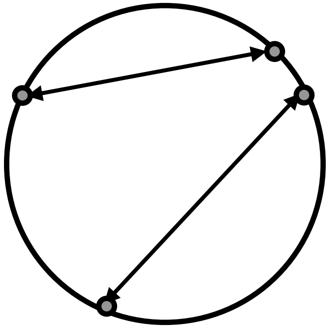


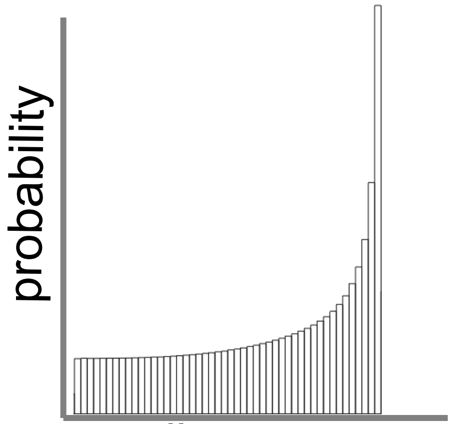
D2 Shape Distributions

Descriptor

Represent shapes by the histogram of distances between pairs of points on the model:

$$D2_P(d) = \frac{|\{p, q \in P \,|\, \|p - q\| = d\}|}{|P|^2}$$

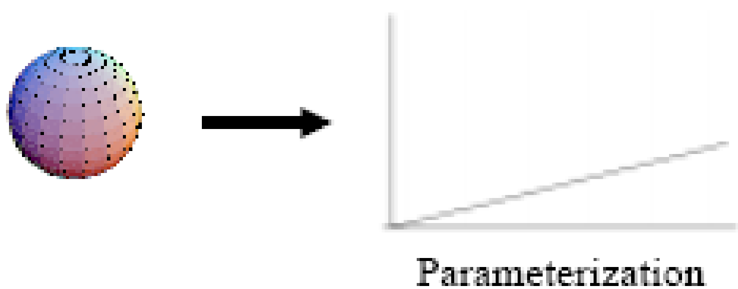


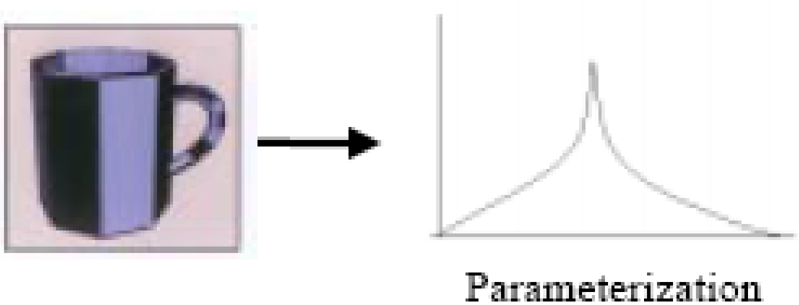


D2 Shape Distributions

Properties

- The shape is represented by a 1D array of values.
- The representation is invariant to translations and rotations





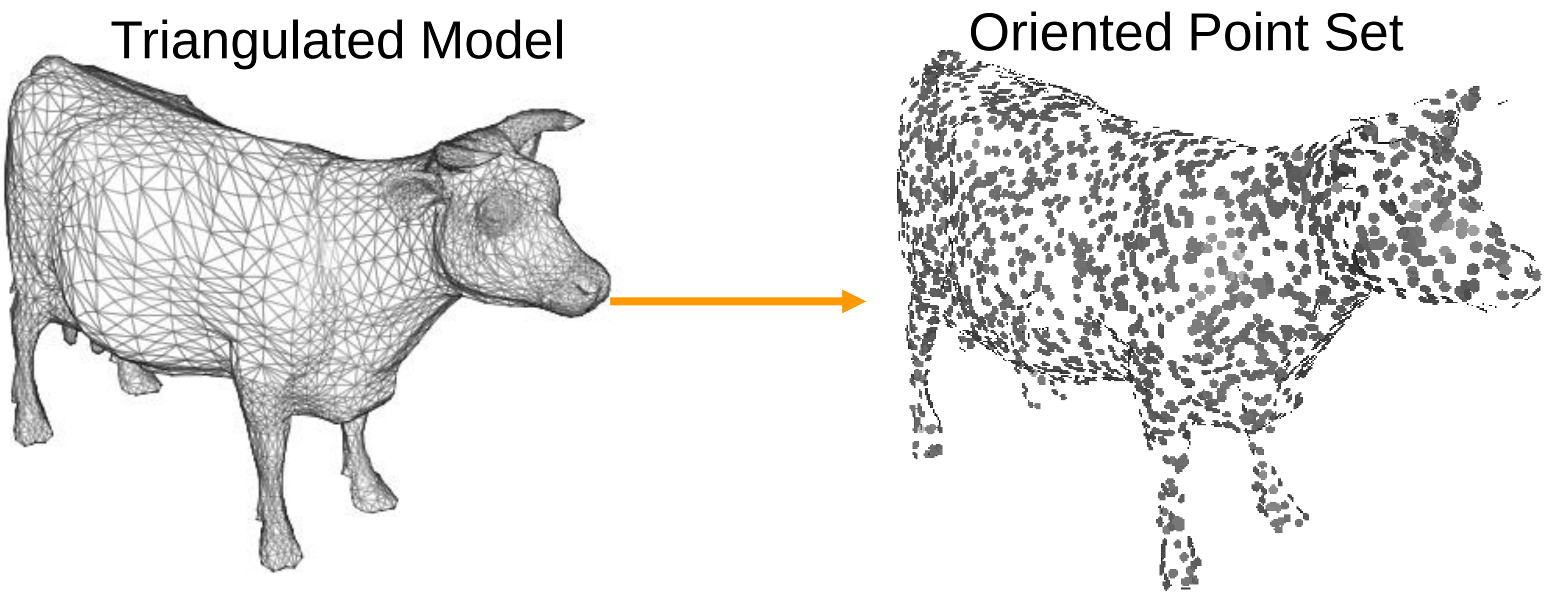
Outline

- Math Overview
- Shape Descriptors
  - Shape Histograms (Ankerst *et al*.)
  - Shape Distributions (Osada *et al*.)
  - Extended Gaussian Images (Horn)
- Invariance

Extended Gaussian Images

Approach

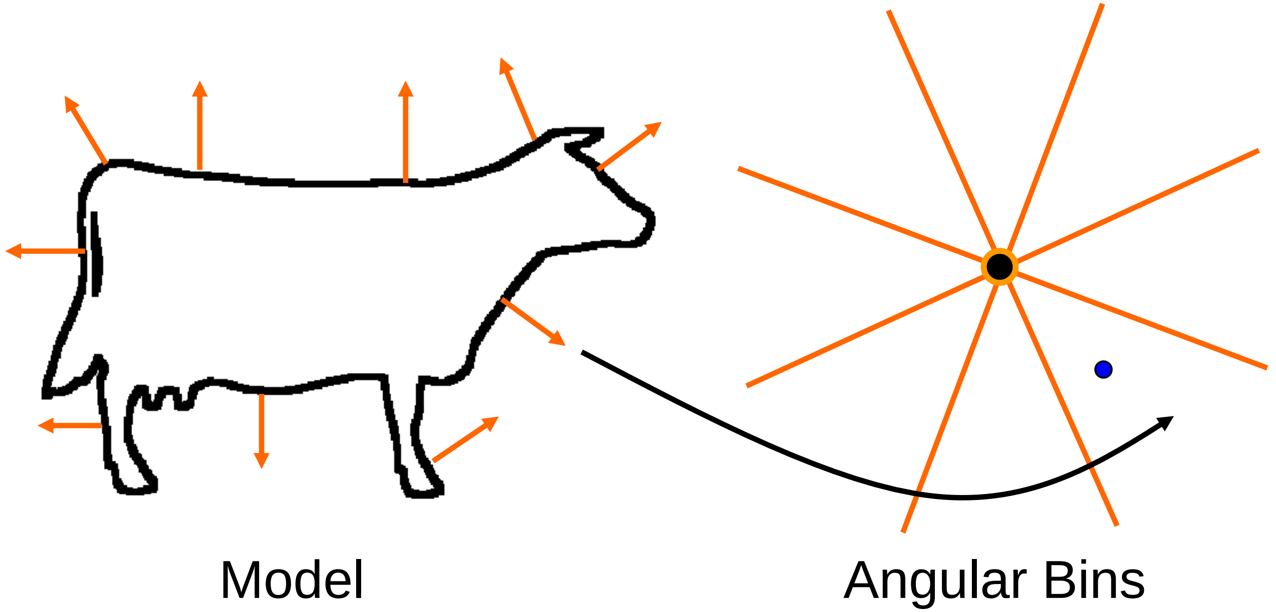
Use the fact that every point on the surface has a position and a normal.



Extended Gaussian Images

Descriptor

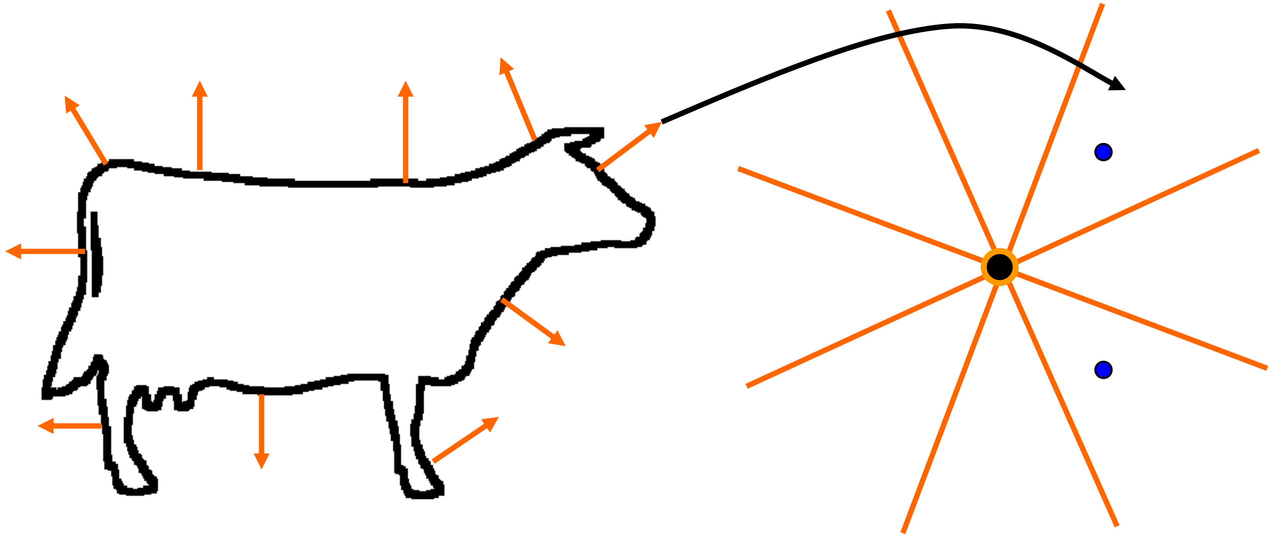
Represent a model by binning points based on the associated surface normal



Extended Gaussian Images

<u>Descriptor</u>

Represent a model by binning points based on the associated surface normal



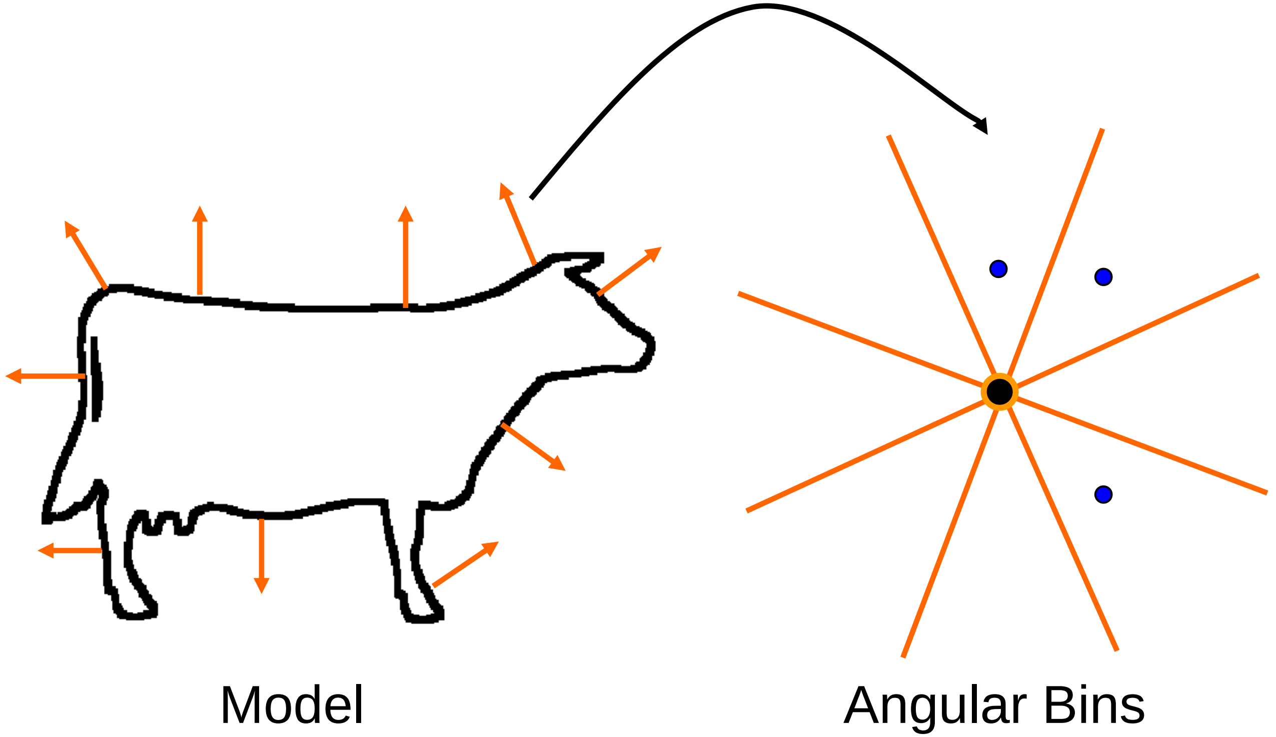
Model

Angular Bins

Extended Gaussian Images

<u>Descriptor</u>

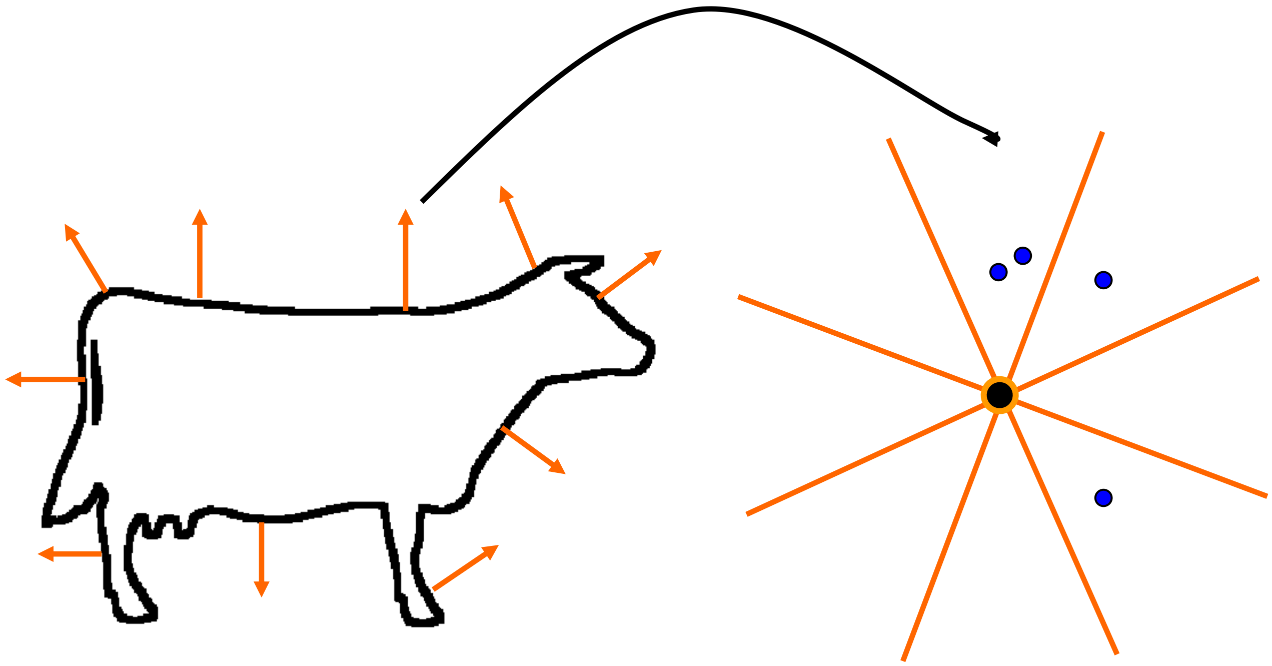
Represent a model by binning points based on the associated surface normal



Extended Gaussian Images

<u>Descriptor</u>

Represent a model by binning points based on the associated surface normal



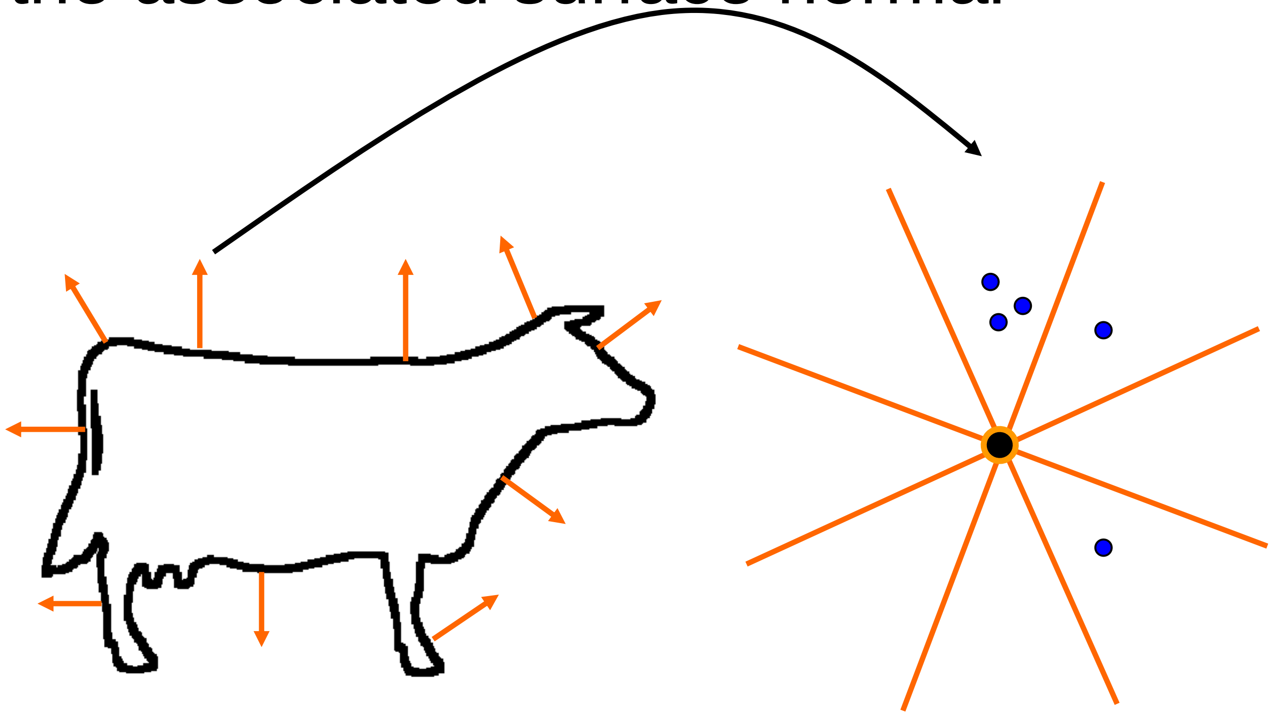
Model

Angular Bins

Extended Gaussian Images

<u>Descriptor</u>

Represent a model by binning points based on the associated surface normal



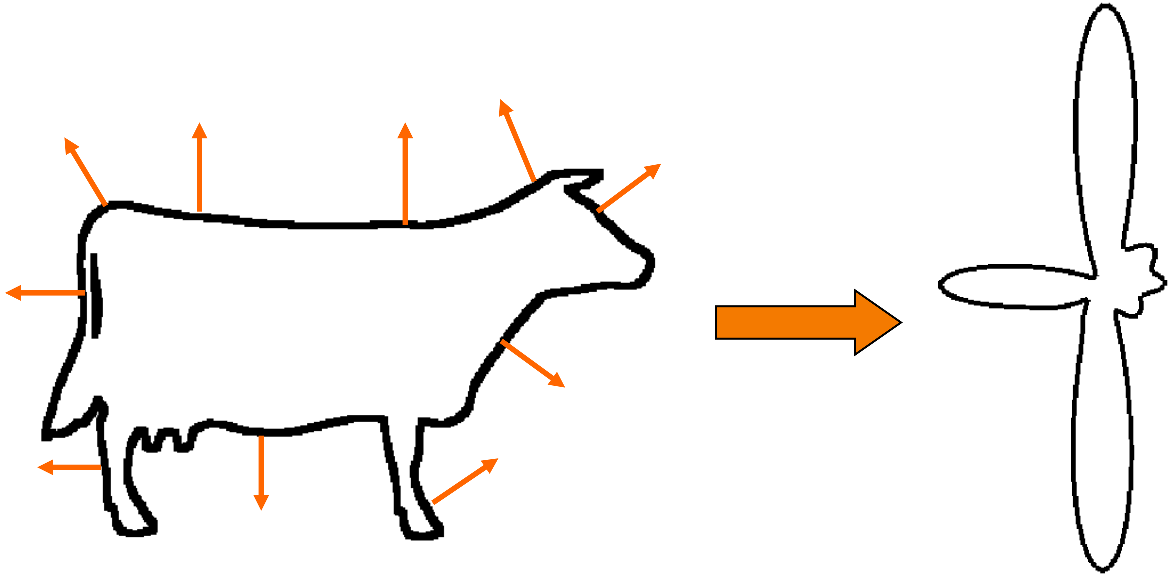
Model

Angular Bins

Extended Gaussian Images

Descriptor

Represent a model by binning points based on the associated surface normal



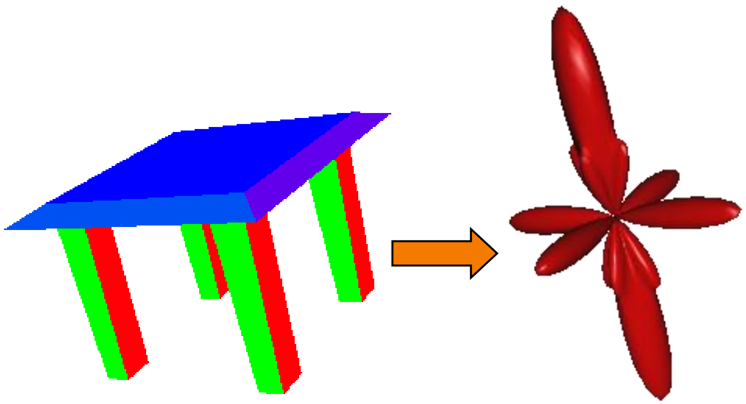
Model

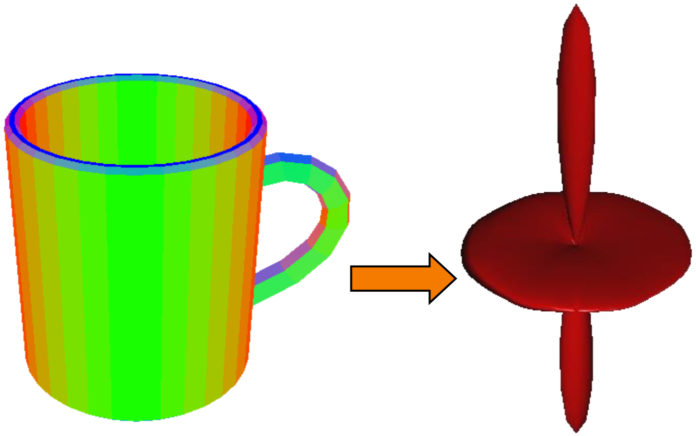
EGI

Extended Gaussian Images

Properties

- A 2D curve / 3D surface is represented by a histogram over a circle / sphere.
- The representation is invariant to translations.





Outline

- Math Overview
- Shape Descriptors
- Invariance

Normalization vs. Invariance

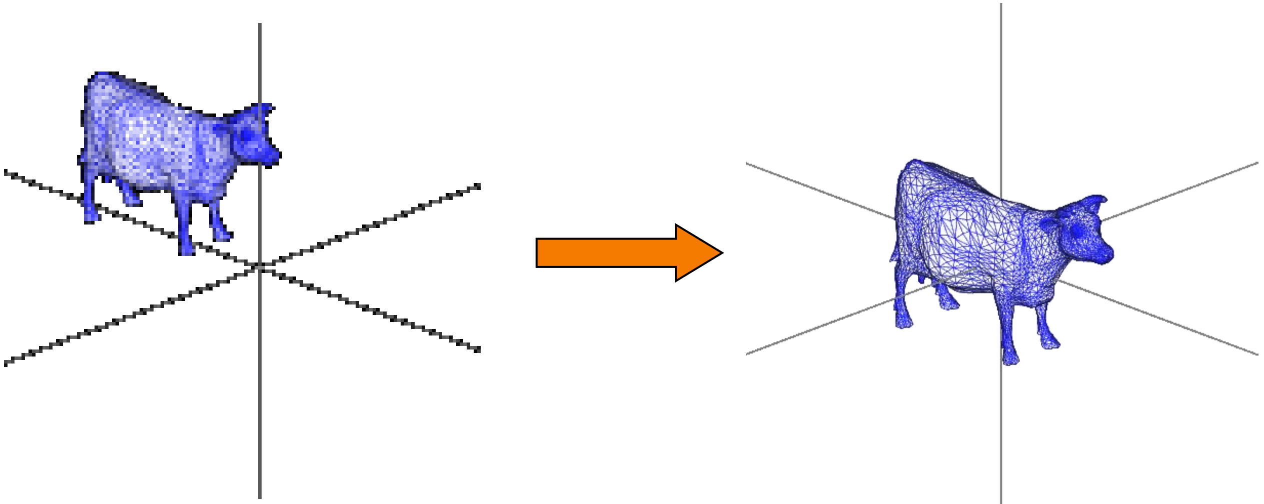
We say that a shape representation is normalized with respect to translation / rotation if the shape is placed into a canonical pose.

Normalization vs. Invariance

We say that a shape representation is <u>normalized</u> with respect to translation / rotation if the shape is placed into a canonical pose.

<u>Example</u>:

We can normalize for translation by moving the surface so that the center of mass is at the origin.



Normalization vs. Invariance

We say that a shape representation is <u>invariant</u> with respect to translation / rotation if the representation discards (all) information that depends on translation / rotation.

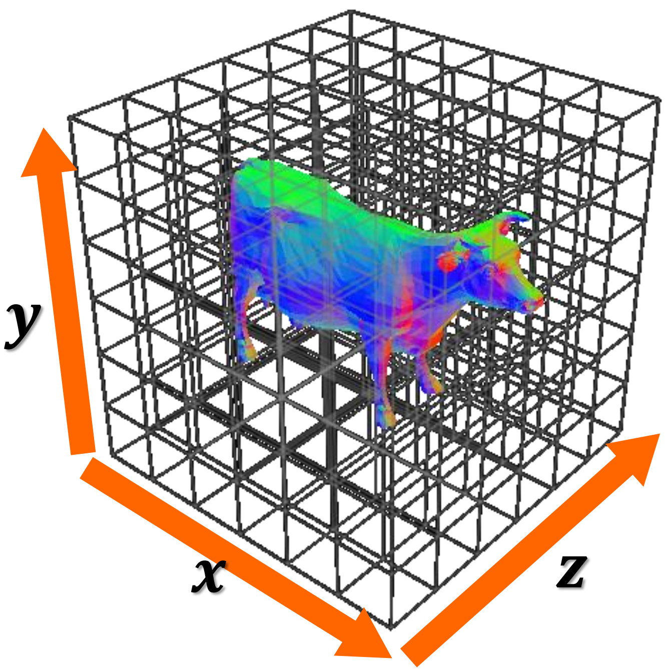
Invariance

We have seen a general method for making functions invariant to translation and rotation.

Invariance

Translation:

Compute the Fourier decomposition and store just the magnitudes of the Fourier coefficients.



Cartesian Coordinates

$$f(x,y,z) = \sum_{l,m,n} \hat{f}_{lmn} \cdot \frac{e^{i(lx+my+zn)}}{(2\pi)^{1.5}}$$

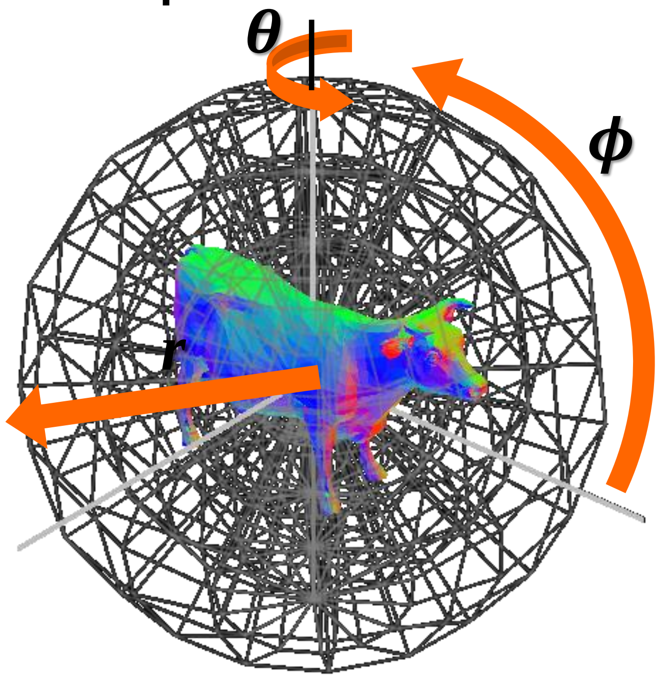
$$\left\{\left\|\hat{f}_{lmn}\right\|\right\}_{l,m,n}$$

Translation Invariant Representation

Invariance

Rotation:

Compute the spherical harmonic decomposition and store just the sizes of the different frequency components of the different radial restrictions.



Spherical Coordinates

$$f(r,\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m} \hat{f}_{lm}(r) \cdot \mathbf{Y}_l^m(\theta,\phi)$$

$$\left\{ \sqrt{\sum_{m=-l}^{l} \left\| \hat{f}_{lm}(r) \right\|^2 \cdot \sqrt{4\pi r^2}} \right\}_{l=0}^{\infty}$$

Rotation Invariant Representation

Overblown Claim

All methods that represent 3D shapes in either a translation-invariant or rotation-invariant method implicitly use these invariance approaches.

Goal

Given the three shape descriptors:

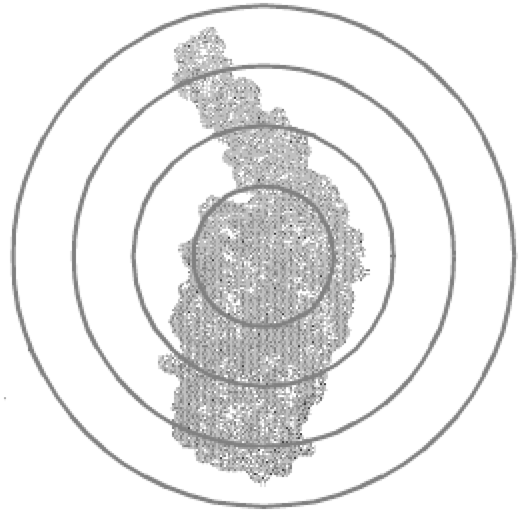
- Shape Histograms
- Shape Distributions
- Extended Gaussian Images

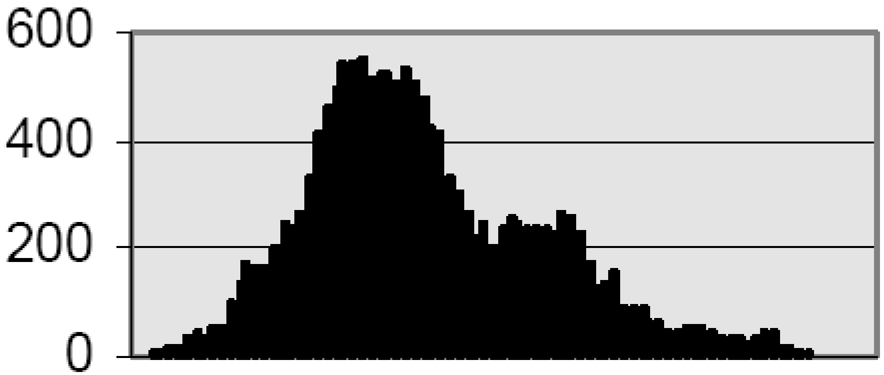
- How does the descriptor obtain its invariance?
- How can the descriptiveness of the descriptor be improved while maintaining invariance?

Shape Histograms

The shape descriptor represents a 3D shape by binning points by their distance from the center.

It is rotation invariant.

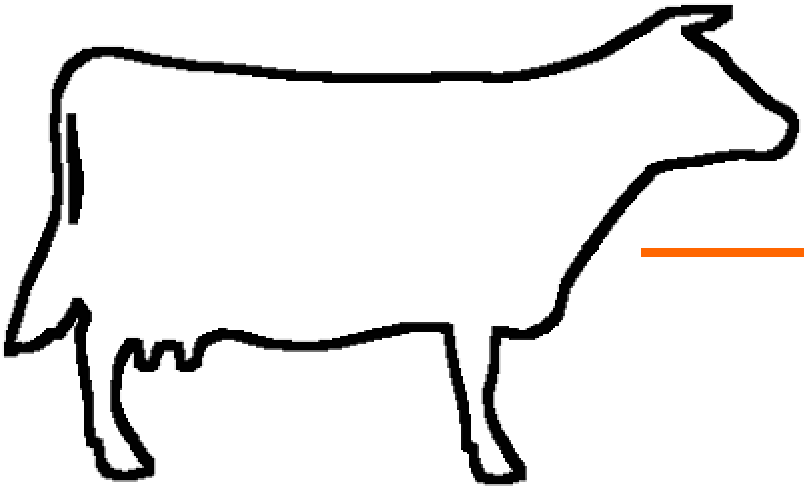




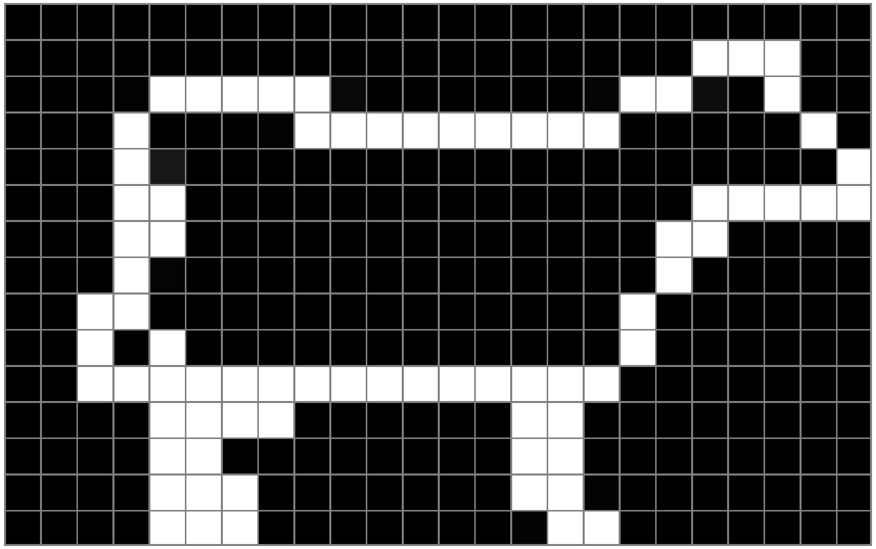
Shape Histograms

The shape histogram starts by representing the surface by a 3D function, obtained by rasterizing the boundary into a voxel grid:

- A voxel has value 1 if intersects the boundary
- A voxel has value 0 otherwise.



Model



Rasterization

Shape Histograms

The shape histogram can be obtained by setting the value for the bin corresponding to radius r to the "size" of the rasterization restricted to the sphere of radius r:

$$\text{ShapeHistogram}(r) = \int_{|p|=r} \text{Raster}(p)\, dp$$

Shape Histograms

We can express the rasterization in spherical coordinates:

$$R(r,\theta,\phi) = \mathrm{Raster}(r\cdot\cos\theta\cdot\sin\phi\,, r\cdot\cos\phi\,, r\cdot\sin\theta\cdot\sin\phi)$$

Then, fixing the radius, we can express the function in terms of spherical harmonics:

$$R(r,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\hat{R}_{lm}(r)\cdot\mathbf{Y}_l^m(\theta,\phi)$$

Shape Histograms

In this formulation, the value of the shape histogram at a radius of r is the value of the 0th spherical harmonic coefficient:*

$$\text{ShapeHistogram}(r) = \hat{R}_{00}(r) \cdot \sqrt{4\pi} \cdot r^2$$

*The scale factor of $\sqrt{4\pi} \cdot r^2$ accounts for the fact that the area of the sphere scales quadratically with radius.

Shape Histograms

So the shape histogram obtains its rotation invariance by storing the (size of the) 0^{th} order frequency component:

$$\text{ShapeHistogram}(r) = \hat{R}_{00}(r) \cdot \sqrt{4\pi} \cdot r^2$$

Extension:

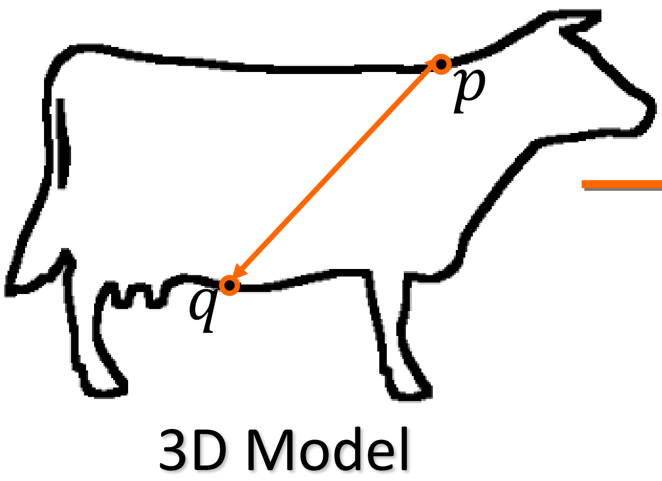
We can obtain a more descriptive representation, without giving up rotation invariance, by storing the size of every frequency component:

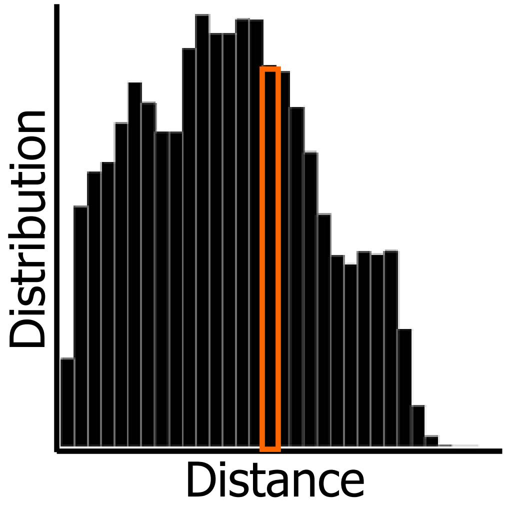
$$\text{EShapeHistogram}(r,l) = \sqrt{\sum_{m=-l}^{l} \left\| \hat{R}_{lm}(r) \right\|^2} \cdot \sqrt{4\pi} \cdot r^2$$

$D2$ Shape Distribution

The shape descriptor represents a 3D shape by binning point-pairs by their distance.

It is both translation and rotation invariant.





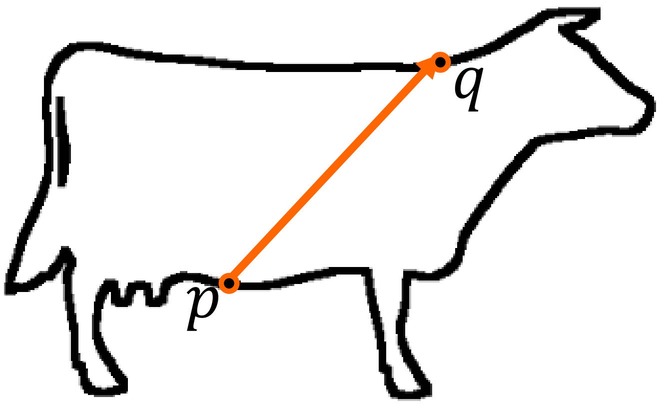
$D2$ Distribution

$D2$ Shape Distribution

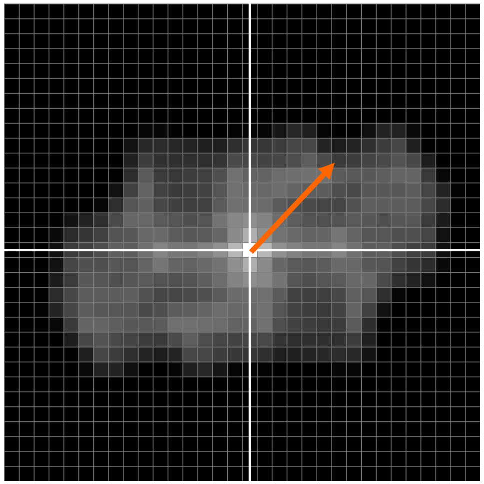
Let's consider rotation invariance first.

$D2$ Shape Distribution

One way to think of the $D2$ shape descriptor is by binning the difference vector between pairs of points on the surface.



3D Model

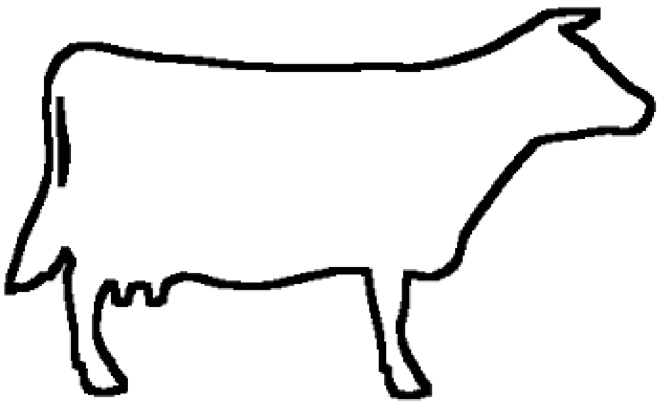


Binned Difference Vectors

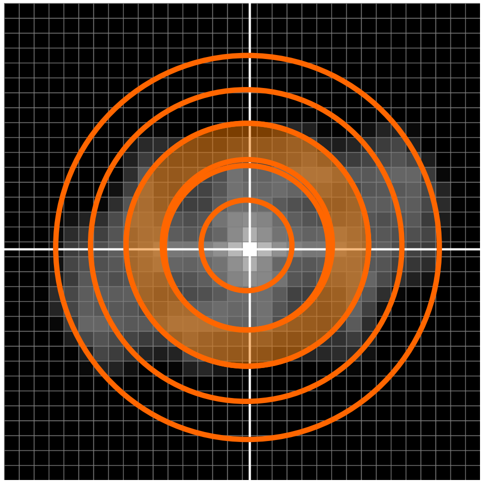
$D2$ Shape Distribution

One way to think of the $D2$ shape descriptor is by binning the difference vector between pairs of points on the surface.

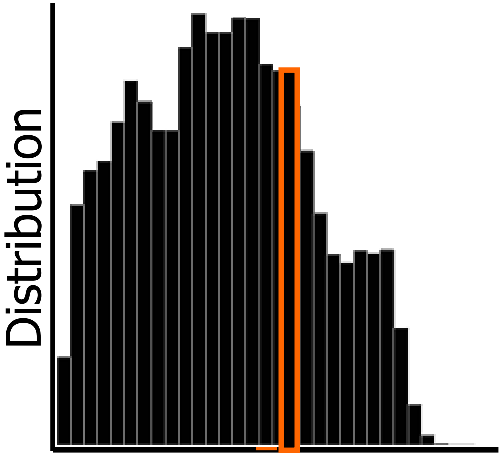
Then the shape distribution can be obtained by computing the Shape Histogram of the binning:



3D Model



Binned Difference Vectors



$D2$ Shape Distribution

As with the Shape Histogram, the $D2$ Shape Distribution can be realized by storing 0^{th} order frequency components of the spherical harmonic decomposition.

Extension:

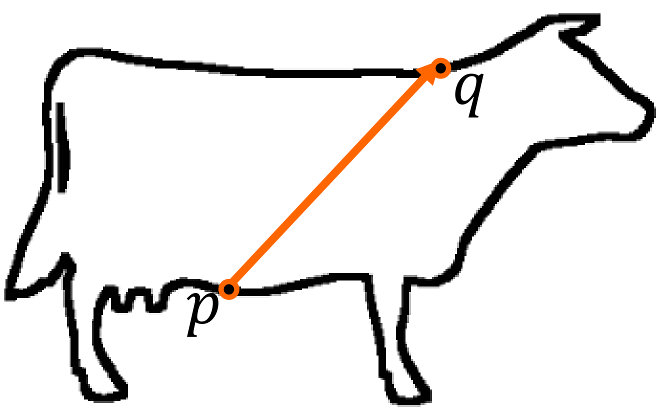
As with the Shape Histogram the representation can be made more descriptive, without sacrificing rotation invariance, by storing the size of every frequency component.

$D2$ Shape Distribution

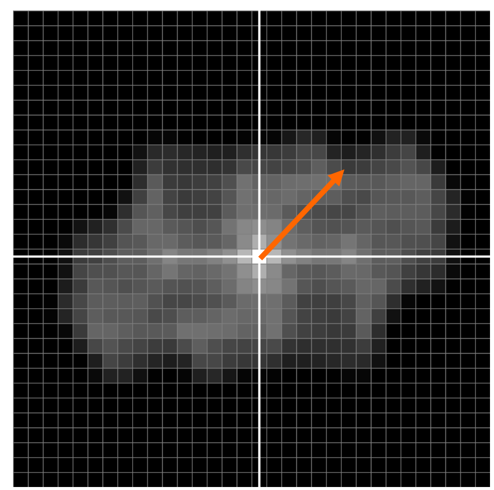
What about the translation invariance?

$D2$ Shape Distribution

The Shape Distribution is computed from the binning of point-pair differences. How is this function computed?



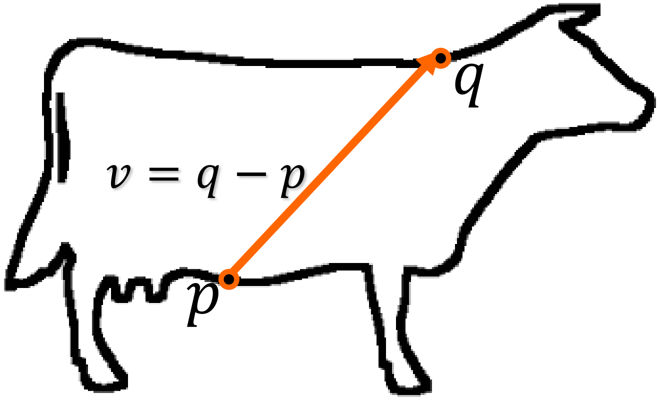
3D Model



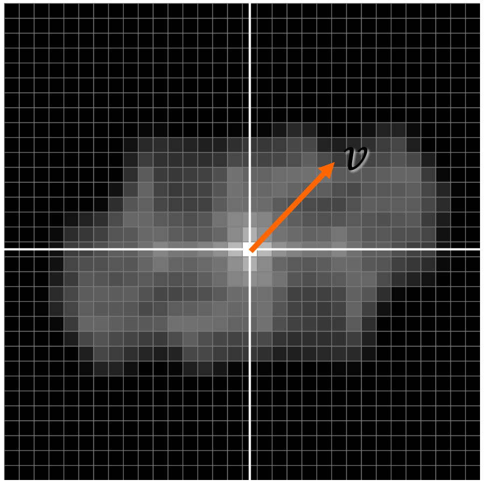
Binned Difference Vectors

$D2$ Shape Distribution

A point q on the surface will contribute to bin v if the point $q - v$ is also on the surface.



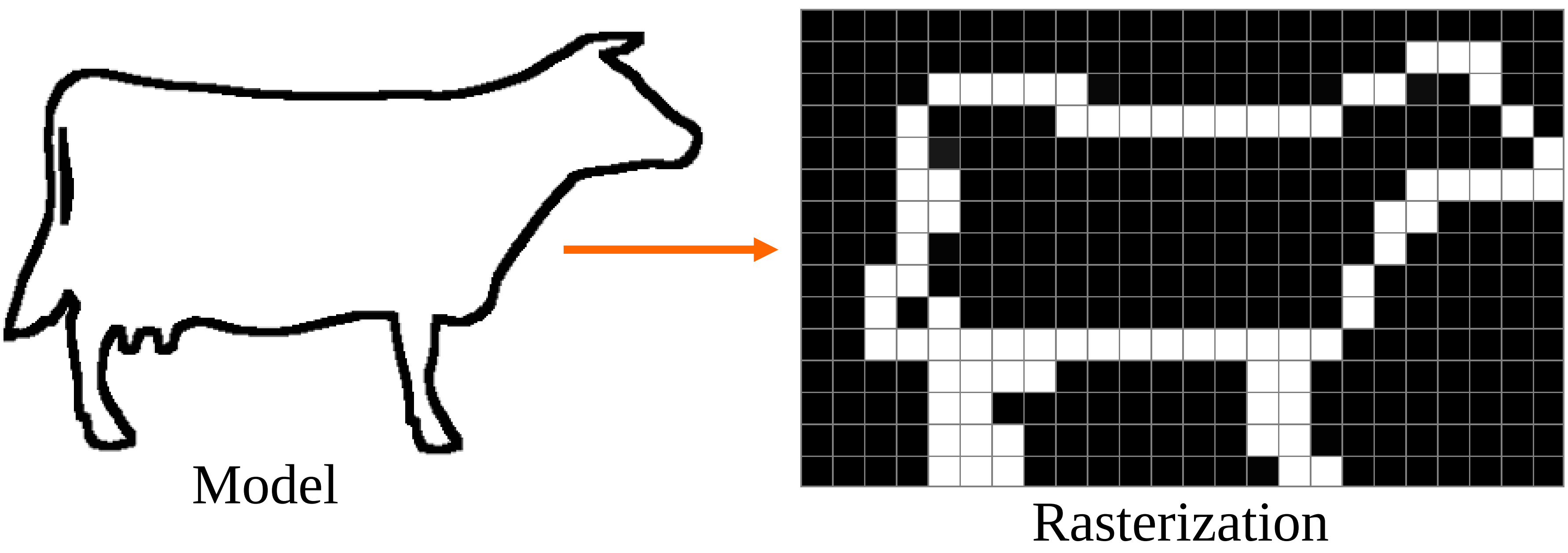
3D Model



Binned Difference Vectors

$D2$ Shape Distribution

Consider the rasterization of the surface into a regular voxel grid.



$D2$ Shape Distribution

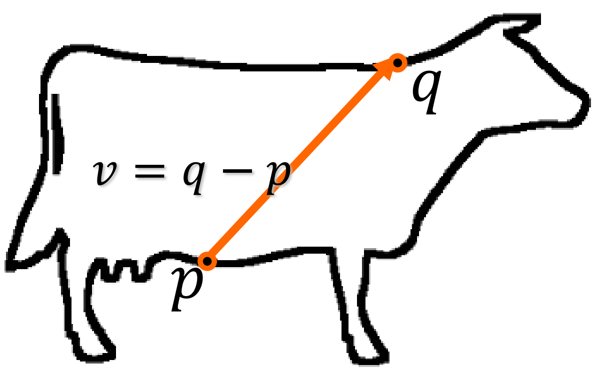
A point q on the surface will contribute to bin v if the point $q - v$ is also on the surface.

$$\Downarrow$$

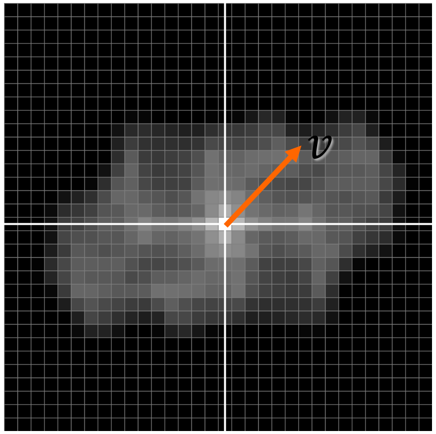
$$\text{Raster}(q - v) = 1$$

$$\Downarrow$$

$$\text{DBin}(v) = \int_{q \in \text{Surface}} \text{Raster}(q - v)\, dq$$



3D Model



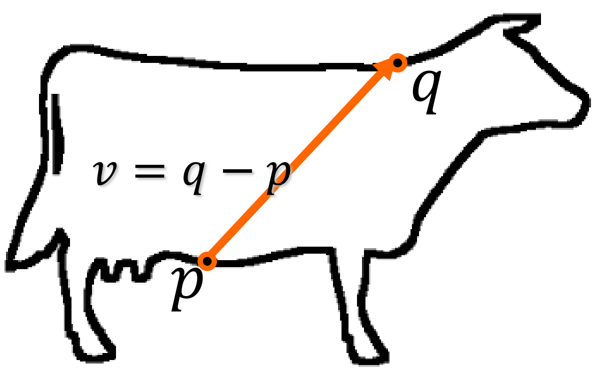
Binned Difference Vectors

D2 Shape Distribution

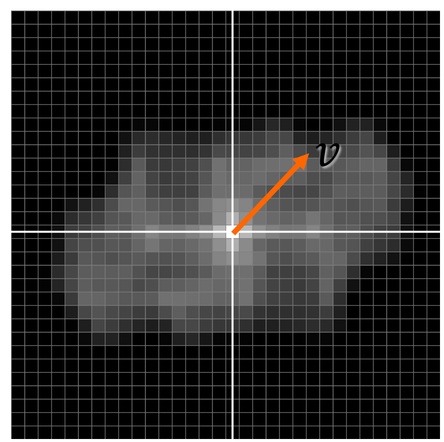
For a point $q \in \mathbb{R}^3$, the point will only contribute to bin v if q and $q - v$ are both on the surface.

That, is q contribute to bin v if and only if:

$$\text{Raster}(q) \cdot \text{Raster}(q - v) = 1$$
$$\Downarrow$$
$$\text{DBin}(v) = \int_{q \in \mathbb{R}^3} \text{Raster}(q) \cdot \text{Raster}(q - v)\, dq$$



3D Model



Binned Difference Vectors

D2 Shape Distribution

Thus, the binning function is the correlation of the rasterization with itself:

$$\begin{aligned}\mathrm{DBin}(v) &= \int_{q\in\mathbb{R}^3} \mathrm{Raster}(q)\cdot\mathrm{Raster}(q-v)\,dq \\ &= (\mathrm{Raster}\star\mathrm{Raster})(v)\end{aligned}$$

D2 Shape Distribution

Recall:

To compute the correlation of f with g we multiply the Fourier coefficients of f by the conjugates of the Fourier coefficients of g:

$$(f \star g)(\theta) = \sum_{l=-\infty}^{\infty} \sqrt{2\pi} \cdot \left(\hat{f}_l \cdot \bar{\hat{g}}_l\right) \cdot e^{il\theta}$$

When $f = g$, this gives:

$$(f \star f)(\theta) = \sum_{l=-\infty}^{\infty} \sqrt{2\pi} \cdot \left\|\hat{f}_l\right\|^2 \cdot e^{il\theta}$$

D2 Shape Distribution

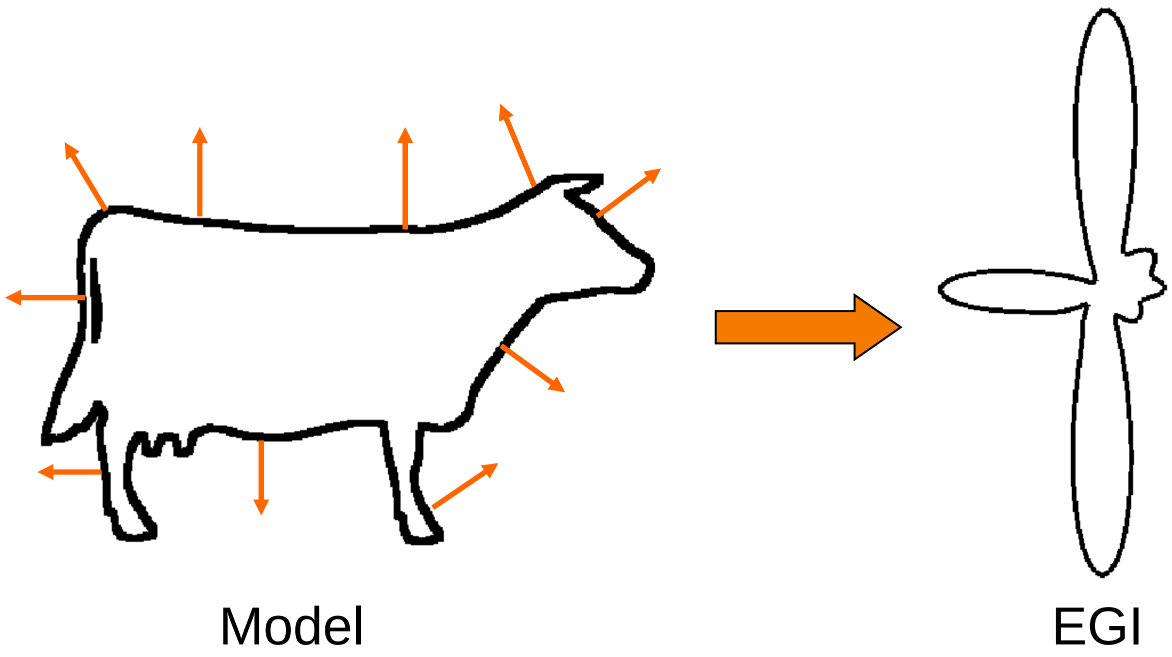
⇒ The binning function implicitly converts the rasterization function into a function whose Fourier coefficients are the square norms of the Fourier coefficients of the rasterization.

Which is what we do to make a function translation invariant.

Extended Gaussian Image

This spherical shape descriptor represents a 3D shape by a histogram on the sphere.

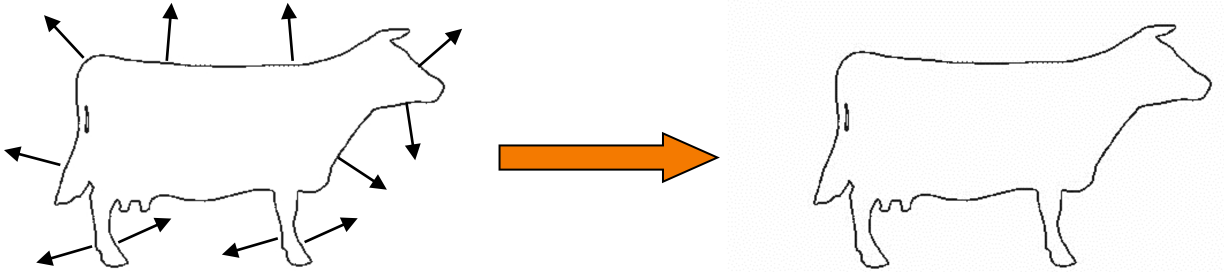
It is obtained by binning points by their normal direction, and is translation invariant.



Extended Gaussian Image

To obtain the EGI representation, we can think of points on the model as living in a 5D space:

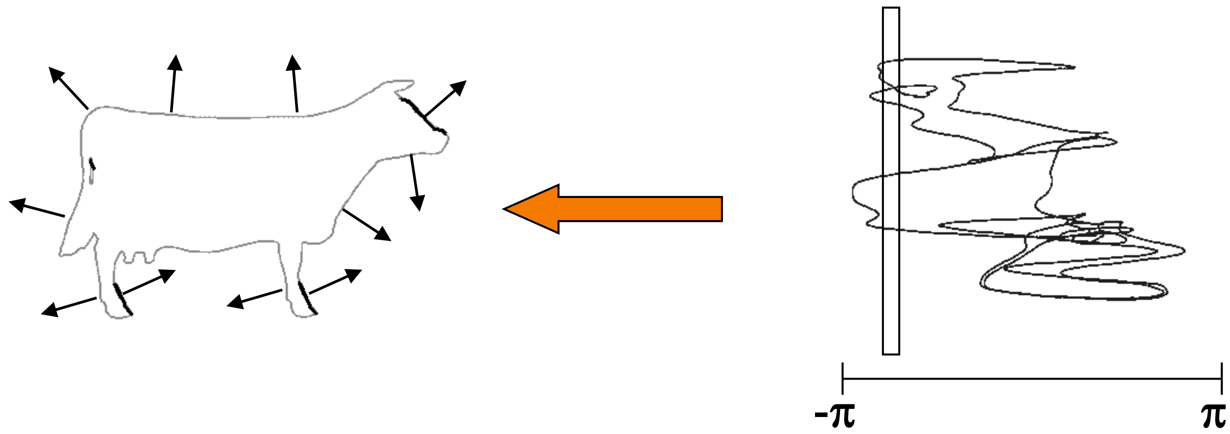
- The first 3 dimensions are indexed by the position.
- The last 2 are indexed by the normal direction.



Extended Gaussian Image

To obtain the EGI representation, we can think of points on the model as living in a 5D space.

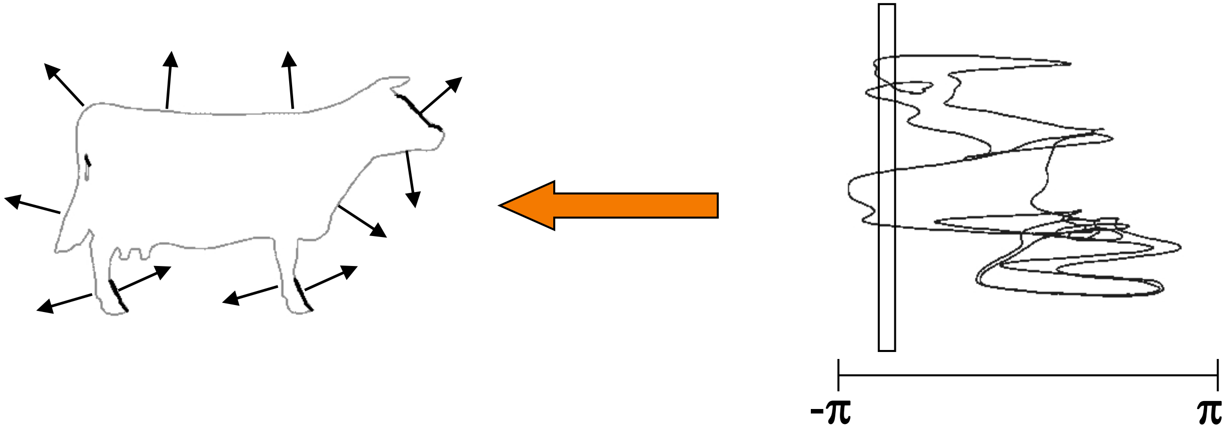
If we fix the normal angle, we get a 3D slice of the 5D space, corresponding to all the points on the surface with the same normal:



Extended Gaussian Image

For each normal n, the EGI stores the "size" of the points in the normal slice corresponding to n.

This is the 0th order frequency component of the rasterization of the points on the model with normal n.



Extended Gaussian Image

For each normal n, the EGI stores the "size" of the points in the normal slice corresponding to n.

This is the 0^{th} order frequency component of the rasterization of the points on the model with normal n.

<u>Extension</u>:

We can get a more discriminating descriptor, without giving up translation invariance, by storing the size of <u>every</u> frequency component.