Integration of cues

- Quick review of depth cues
- Cue combination: Minimum variance
- Cue combination: Bayesian
- Nonlinear cue combination: Causal models
- Statistical decision theory

Distance, depth, and 3D shape cues

- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity
- Cue combination

Basic distinctions

- Types of depth cues
  - Monocular vs. binocular
  - Pictorial vs. movement
  - Physiological
- Depth cue information
  - What is the information?
  - How could one compute depth from it?
  - Do we compute depth from it?
  - What is learned: ordinal, relative, absolute depth, depth ambiguities

Definitions

- Distance: Egocentric distance, distance from the observer to the object
- Depth: Relative distance, e.g., distance one object is in front of another or in front of a background
- Surface Orientation: Slant (how much) and tilt (which way)
- Shape: Intrinsic to an object, independent of viewpoint

Distance, depth, and 3D shape cues

- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity

Epstein (1965) familiar size experiment

- How far away is the coin?
Monocular depth cues

Relative size as a cue to depth

Relative size as a cue to depth

Occlusion as a cue to depth

Shading, reflection, and illumination

Shading - prior of light-from-above
Shading (flip the photo upside-down)

Cast Shadows

Dynamic Cast Shadows

Shading and contour

Aerial/Atmospheric Perspective

Geometry of Linear Perspective

Retinal projection depends on size and distance:

Size in the world (e.g., in meters) is proportional to size in the retinal image (in degrees) times the distance to the object
Linear perspective

- Size constancy

Texture

1. Density
2. Foreshortening
3. Size

Distance, depth, and 3D shape cues

- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, [astigmatic blur, chromatic aberration]
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity

Monocular Physiological Cues

- Accommodation – estimate depth based on state of accommodation (lens shape) required to bring object into focus
- Blur – objects that are further or closer than the accommodative distance are increasingly blur
- Astigmatic blur
- Chromatic aberration
Distance, depth, and 3D shape cues

- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, astigmatic blur, chromatic aberration
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity

Motion Parallax

The Kinetic Depth Effect

Dynamic (Kinetic) Occlusion

Distance, depth, and 3D shape cues

- Pictorial depth cues: familiar size, relative size, brightness, occlusion, shading and shadows, aerial/atmospheric perspective, linear perspective, height within image, texture gradient, contour
- Other static, monocular cues: accommodation, blur, astigmatic blur, chromatic aberration
- Motion cues: motion parallax, kinetic depth effect, dynamic occlusion
- Binocular cues: convergence, stereopsis/binocular disparity

Vergence Angle As One Binocular Source
Vergence Angle As One Binocular Source

Vergence Angle As One Binocular Source

Vergence Angle As One Binocular Source

Binocular disparity

Disparity

Uncrossed disparity

Zero retinal disparity

Crossed disparity

Disparity
How to make a random-dot stereogram

Left eye image

Right eye image

Depth Cue Combination: Issues

1. How do you put all of the depth cue information together?

2. What do you do when cues disagree?
   A little ... ?
   A lot ... ?

3. How much weight should each cue get?

When cues disagree ...

Information Fusion Problem

Multiple sources of information, possibly in error, possibly contradictory

How combine the information into a single judgment?
Optimal Cue Combination: Minimum Variance

\[ E(X_1) = \mu_1, \quad E(X_2) = \mu_2 \]

Variances \( \sigma_{1}^2 \leq \sigma_{2}^2 \)

Just use one cue?

Suppose we use a linear cue-combination rule:

\[ X = w_1X_1 + w_2X_2 \]

\[ E[X] = w_1E[X_1] + w_2E[X_2] = (w_1 + w_2)\mu \]

Unbiased?

Minimum-Variance Cue Combination

\[ X = wX_1 + (1 - w)X_2 \]

Unbiased

\[ Var(X) = w^2Var(X_1) + (1 - w)^2Var(X_2) = w^2\sigma_1^2 + (1 - w)^2\sigma_2^2 \]

Minimize

Reparameterization

Define reliability \( r_i = \sigma_i^{-2} \)

\[ X = w_1X_1 + w_2X_2 \]

\[ w = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2} = \frac{r_1}{r_1 + r_2} \]

Weight proportional to reliability

\[ r = r_1 + r_2 \]

Reliabilities add

Perturbation Methodology and Influence Measures

How can we measure the influence of various cues on perceptual judgments in complex scenes?

Goal: Change the stimulus as little as we possibly can.
Perturbation Method

Example: Texture and Motion

Observer adjusts ...

Two Depth Cues

Observer is using only the perturbed cue.

One possibility ...

Another possibility ...

Two Depth Cues

Perturbation Method

Two Depth Cues

Perturbation Method

Two Depth Cues

Perturbation Method

Two Depth Cues

Perturbation Method

Two Depth Cues

Perturbation Method

Two Depth Cues

Perturbation Method

Two Depth Cues

Two Depth Cues

Two Depth Cues

Two Depth Cues
Perturbation Method

Observer is not using the perturbed cue at all.

Another possibility ...

Two Depth Cues

Perturbed Cue ….

Perturbation Method

A final possibility ...

Two Depth Cues

Perturbed Cue ….

We will measure the influence of the cue on the observer’s setting.

Two Depth Cues

Perturbed Cue ….

Influence Measures

\[ I_{\text{cue}} = \frac{\Delta_{\text{setting}}}{\Delta_{\text{cue}}} \]

Change in observer’s setting

Influence of the cue

Perturbation of the cue

Texture and Motion: Data

Optimal Cue Combination: Bayesian

Compute posterior:

\[ p(\text{depth} | x_1, x_2) = \frac{p(x_1, x_2 | \text{depth})p(\text{depth})}{p(x_1, x_2)} \]

Assume conditional independence:

\[ p(\text{depth} | x_1, x_2) \propto p(x_1 | \text{depth})p(x_2 | \text{depth})p(\text{depth}) \]

If likelihoods and prior are Gaussian, so is posterior, and means and reliabilities are as in minimum-variance case. Prior acts like a static cue.
Optimal Cue Combination: Bayesian

\[ p(\text{depth} | x_1, x_2) \propto p(x_1 | \text{depth})p(x_2 | \text{depth})p(\text{depth}) \]

Depending on cost function and priors, choose:

- ML: Maximum-likelihood estimator
- MAP: Maximum a posteriori estimator
  - Mean of the posterior
  - Median of the posterior
  - Etc.

\[ p(d e p t h | x_1, x_2) \propto p(x_1 | d e p t h) p(x_2 | d e p t h) p(d e p t h) \]

Rock & Victor (1964)

View object through distorting lens while exploring object haptically

Visually and haptically specified shapes differed. What shape is perceived?

Visual/Capture?

Why should vision be the “gold standard” all other modalities are compared to?

Visual Capture?

Why should vision be the “gold standard” all other modalities are compared to?
Visual Capture?

Why should vision be the “gold standard” all other modalities are compared to?

\[ S_{VH} = w_V S_V + w_H S_H \]

Weights
\[ w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2} \]

Variance
\[ \frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2} \]
Individual Differences

Nonlinear Cue Combination: Causal models

Visual Auditory combination (Ventriloquist effect)

Modeling: Where do cues come from?

Traditional Bayesian model
What would happen now?

Obviously there may be more than one source.

\[ p(\text{causal model}) \]

- Using Bayes rule:

\[
p(\text{common}|x_{\text{vis}}, x_{\text{aud}}) = \frac{p(\text{common})p(x_{\text{vis}}, x_{\text{aud}}|\text{common})}{p(\text{common})p(x_{\text{vis}}, x_{\text{aud}}|\text{common}) + (1-p(\text{common}))p(x_{\text{vis}}, x_{\text{aud}}|\neg\text{common})}
\]

Common cause: where is the auditory stimulus

Mean squared error estimate

\[
X = P(C|x_{\text{vis}}, x_{\text{aud}})\bar{x} + P(\neg C|x_{\text{vis}}, x_{\text{aud}})\bar{x}_c
\]
**Experimental test**

Wallace et al. 2005; Hairston et al. 2004

**Measured gain**

How can the gain be negative?

Predicting the variance

Worse prediction if we assume model selection

**Statistical Decision Theory**

The Three Elements of SDT

\[
W = \{w_1, w_2, \ldots, w_m\} \quad \text{possible states of the world}
\]

\[
A = \{a_1, a_2, \ldots, a_p\} \quad \text{possible actions}
\]

\[
X = \{x_1, x_2, \ldots, x_n\} \quad \text{possible sensory events}
\]

\[
X \sim f(x; \theta)
\]
**The Three Functions of SDT**

**Goal:** select

\[ d : X \rightarrow A \]
Partial ordering: dominance

Maximin Criterion

Partial ordering of rules
\(d: X \rightarrow A\)
Maximin Criterion

Bayesian Decision Theory

We add a prior probability distribution on the state of the world and define the Bayes Risk of each rule.

\[ EBR(d) = \sum_{i=1}^{n} EG(d, w_i)\pi(w_i) \]

The rule \( d^* \) that maximizes Bayes Risk is the Bayes Rule (it need not be unique).

Bayesian Decision Theory (Continuous Form)

Maximize expected Bayes gain

\[ EBG(D) = \int G(d(x), w) L(w | x)\pi(w) dx dw \]

by choice of a decision rule

\[ d : X \rightarrow A \]

Bayesian Decision Theory

Geisler (1989)

\[ \int G(d(x), w) L(w | x)\pi(w) dx dw \]
Bayesian Decision Theory

\[ \int G(d(x), w)L(w|x)\pi(w)dx dw \]

Geisler (1989)

Knill & Richards (1996)

Perception as Bayesian Inference

A choice

Normative

Descriptive (process)

How to test?

Optimal Cue Combination

Individual Differences

What is the gain/loss function?

Quadratic loss (least squares)

\[ S = \omega_H S_H + \omega_V S_V \]

A weak test of BDT …

Process model is weighted linear combination minimizing quadratic loss
Bayesian Decision Theory

\[ \int G(d(x), w) L(w | x) \pi(w) dx dw \]

Knill & Richards (1996)

All that really matters...

Can the visuo-motor system, presented with arbitrary gain functions, select decision rules that maximize expected gain?

Direct manipulation of gain/loss function

Strong test of BDT

Motor Timing Experiment:

Practice phase:

400 450 500 550 600 650 700

Target

Time bar
Practice phase:

Produce a movement to target
Of duration 650 msec

Practice phase:

Practice phase:

Try again!

Practice phase:

Better!

Calibration SD's

Experiment: Main Task

Make money ....
**Experiment: Main Task**

*How to maximize expected gain?*

\[ EG(t_0) = p_G(t_0)G + p_L(t_0)L \]

**Configurations**

1. 
   \[ -15 \quad +5 \quad 0 \]

2. 
   \[ 0 \quad +5 \quad -15 \]

3. 
   \[ -15 \quad 0 \quad +5 \quad 0 \]

4. 
   \[ 0 \quad +5 \quad 0 \quad -15 \]
Summary

*Subjects chose movements whose mean time came close to maximizing expected gain.*

*No patterned deviations.*

Bayesian Decision Theory

\[
\iint G(d(x), w)L(w \mid x)\pi(w)\,dxdw
\]
Conclusions

Gain/loss functions are problems posed by the environment to the organism.

They are an useful as independent variables in exploring visuo-motor function.

Their manipulation allows us to test performance against ideal in a wide range of economic games.