Yuille & Kersten

Vision as Bayesian inference: analysis by synthesis?
• the study of human vision should be aimed at determining how humans perform natural tasks with natural images
• the study of human vision should be aimed at determining how humans perform natural tasks with **natural images**

• generalisation from **artificial stimuli** can lead to faulty generalisations
• The study of human vision should be aimed at determining how humans perform natural tasks with **natural images**

• Generalisation from **artificial stimuli** can lead to faulty generalisations

• Proposal for a theory that can deal with the complexity of natural images:
• the study of human vision should be aimed at determining how humans perform natural tasks with natural images

• generalisation from artificial stimuli can lead to faulty generalisations

• proposal for a theory that can deal with the complexity of natural images:
  
  • perception as Bayesian inference
  • analysis by synthesis
perception as inverse inference
generative process
generative process
inverse inference

generative process

\[
\begin{align*}
\text{H} & \xrightarrow{f} \text{O} \\
\text{O} & \xrightarrow{f^{-1}} \text{H}
\end{align*}
\]
ambiguity
observed image
observed image
need for top down
synthetic images
(a)

Scene ↓

Vocabulary: A B C ...

Image samples: ![Sample Images]...
• ambiguity increases with model complexity
• ambiguity increases with model complexity

• but it can be resolved using Bayesian inference
hypothesis space

image data
hypothesis space

image data

hypotheses that explain data well
hypothesis space

image data

hypotheses that explain data well
(a) Feature extraction → Proposals

(b) Synthesis and verification
generative model for natural images
N

number of nodes 32
\( N \)\hfill number of nodes

\( R(L) \)\hfill region

32
number of nodes

region

type of model

text/face/background

32
type of model

region

number of nodes

parameters
text/face/background

32

N → R(L) → θ → 0
(a)

Scene

\[ W \]
(a)

Scene

Text: $(\zeta_1, L_1, \Theta_1)$

Face: $(\zeta_2, L_2, \Theta_2)$

Background: $(\zeta_3, L_3, \Theta_3)$

$W$
(a) Scene
  - Text: \((\zeta_1, L_1, \Theta_1)\)
  - Face: \((\zeta_2, L_2, \Theta_2)\)
  - Background: \((\zeta_3, L_3, \Theta_3)\)

\[ W \]

\[ I \]
image generation $P(I|W)$
$W$ \quad \text{image generation} \quad P(I|W) \quad \text{argmax } P(W|I)

$I$ \quad \text{parsing}

$W^*$
inference
Data driven Markov Chain Monte Carlo (DDMCMC)
\[ W_1 \]
this is a Markov chain.
$K_i(W_1, W_2) = q_i(W_1, W_2) \ a_i(W_1, W_2)$

Metropolis-Hastings kernel
\[ K_i(W_1, W_2) = q_i(W_1, W_2) \, a_i(W_1, W_2) \]
\[ K_i(W_1, W_2) = q_i(W_1, W_2) a_i(W_1, W_2) \]

“bottom-up proposals”

proposal distribution

acceptance probability

“high-level model”
original image  bottom-up proposals  note the tree-face!
original image  bottom-up proposals

note the tree-face!

detection
original image  bottom-up proposals

segmentation  detection

note the tree-face!
original image  bottom-up proposals

segmentation  detection  synthesised image

note the tree-face!
conclusions

• major goal in vision should be to determine how visual system works under natural conditions and natural tasks
conclusions

• major goal in vision should be to determine how visual system works under natural conditions and natural tasks

• probability distributions over structured representations offer the promise to model natural images
conclusions

• major goal in vision should be to determine how visual system works under natural conditions and natural tasks

• probability distributions over structured representations offer the promise to model natural images

• DDMCMC inference algorithm follows “analysis by synthesis” strategy, which may correspond to forward and backward pathways in the cortex