

# OBJECT PERCEPTION AS BAYESIAN INFERENCE

DANIEL KERSTEN

PASCAL MAMASSIAN

ALAN YUILLE

2004

La visión permite extraer las propiedades de los objetos (forma, tamaño, material, textura, distancia, geometría) para funcionar en el mundo (determinar distancias, madurez de frutos, cuán fáciles los objetos son de agarrar, etc). Las imágenes naturales son complejas y ambiguas, de modo que es difícil determinar cómo la información retinal se traduce en conocimiento útil. Formas tridimensionales similares pueden dar imágenes distintas, y formas distintas pueden dar imágenes similares.

Se tratará a la percepción de objetos como un problema de inferencia estadística. Los problemas de ambigüedad y complejidad pueden manejarse utilizando inferencia Bayesiana.

## INTRODUCTION TO BAYES

### HOW TO RESOLVE AMBIGUITY IN OBJECT PERCEPTION?

La base es el proceso de “inferencia inconsciente” de Helmholtz. Las imágenes retinales son ambiguas y se requiere de conocimiento previo para explicar la percepción. Este conocimiento es combinado de forma automática con la información presente para inferir las propiedades de los objetos.

Los observadores que utilizan el marco Bayesiano para hacer interpretaciones óptimas son llamados observadores ideales. Un tipo de estos observadores que computa la interpretación más probable según la distribución de probabilidades posterior es el observador *máximo a posteriori* (MAP). Esta distribución es producto de la distribución prior (la probabilidad de cada posible estado de la escena antes de observar el estímulo) y la verosimilitud (la probabilidad del estímulo dado cada posible estado de la escena). Las distribuciones prior pueden representar el conocimiento de las regularidades de objetos, formas, iluminaciones, materiales, etc. Un observador ideal no necesariamente llega a la solución correcta en cada estímulo input, pero hace las mejores conjeturas y llega al mejor desempeño promedio a largo plazo. Un observador ideal aún puede ser engañado por ilusiones.

THE GENERATIVE MODEL. El modelo generativo  $S \rightarrow I$  especifica cómo una descripción de imagen  $I$  (e.g., los valores de intensidad de imagen o características extraídas de ellos) son determinados por la descripción de la imagen  $S$  (e.g., un vector con variables que representan la forma de la superficie, reflectividad del material, dirección de la iluminación y punto de vista). La verosimilitud de las características de la imagen,  $p(I|S)$ , y

la probabilidad prior de la descripción de la escena,  $p(S)$ , determinan un modelo generativo externo. Un modelo generativo especifica la relación causal entre variables aleatorias (*e.g.*, objetos, iluminación y punto de vista) que determinan los datos observables (*e.g.*, intensidades de imagen).

**THE TASK SPECIFICATION.** Algunas tareas requieren de mayor precisión para estimar algunos aspectos de las imágenes que para otros, por lo que llegar siempre al estimado más probable puede no ser ideal. Los costos y beneficios asociados con errores de distintos aspectos de la decisión perceptual forman una función de utilidad. Una decisión perceptual óptima será una función de las restricciones de la tarea además de la probabilidad posterior.

A veces se pueden simplificar los requerimientos de la tarea dividiendo  $S$  en componentes ( $S_1, S_2$ ) que especifican cuáles propiedades es importante estimar con precisión y cuáles no.

**THE INFERENCE SOLUTION.** La percepción Bayesiana es una solución inversa  $I \rightarrow S_1$  al modelo generativo, la cual estima las variables  $S_1$  dada la imagen  $I$  y descuenta las variables de confusión. Las decisiones se basan en la distribución posterior  $p(S|I)$ , especificada por Bayes como  $\frac{p(I|S)p(S)}{p(I)}$ . Un observador ideal conjunta esta distribución con una función de utilidad (o una función de pérdida negativa), y elige la interpretación que tiene la mayor utilidad esperada.

### PSYCHOPHYSICS

Los experimentos psicofísicos pueden probar las teorías sobre el conocimiento especificado por  $p(S)$ .

### IDEAL OBSERVERS

Los observadores ideales proveen las mejores pruebas psicofísicas dado que completan modelos de ejecución visual basándose en la distribución posterior y en la tarea. Un observador ideal maximiza el desempeño para una tarea visual, por lo que sirve como un estándar para comparar el desempeño humano. Dado que la visión humana tiene limitaciones en cuanto a la naturaleza de sus cálculos y su hardware físico, se espera bastante desviación de la optimalidad.

### BASIC BAYES: THE TRADE-OFF BETWEEN FEATURE RELIABILITY AND PRIORS

La ambigüedad objetiva en las imágenes aparece si distintos objetos pudieran haber producido la misma descripción de la imagen o características de la imagen. El sistema visual se ve forzado a hacer conjeturas, pero puede hacer conjeturas inteligentes sesgándose hacia objetos o interpretaciones típicos. La fórmula de Bayes indica que estas conjeturas, y por lo tanto la percepción, son una *trade-off* entre confiabilidad en las características de la imagen, como encarna la verosimilitud  $p(I|S)$ , y la probabilidad prior  $p(S)$ . Algunas percepciones serán guiadas más por los priors y otras más por los datos. Cuanto menos confiables las características de la imagen (más ambiguas), mayor la influencia de los priors en la percepción.