

Resumen de tesis de grado (en Español)

Specialized Mappings Architecture, with Applications to Vision-Based Estimation of Articulated Body Pose.

Arquitectura de Mapas Especializados, con Aplicaciones en Estimación de Configuraciones de Cuerpos Articulados Basado en Visión Computacional.

Rómer E. Rosales del Moral

El problema en consideración es el de estimar o inferir configuraciones de objetos articulados en tres y dos dimensiones (3D-2D) basados en descripciones visuales provenientes de sólo una imagen. La estimación de configuraciones en 3D (*e.g.*, estimación *maximum-a-posteriori* o MAP) es considerado generalmente un problema matemáticamente mal propuesto (en el sentido de Hadamard) debido a que no es posible recobrar la configuración del objeto dado que información ha sido perdida luego de su proyección en el plano de la imagen. En esta tesis, el problema es reformulado como un problema de inferencia estadística, donde el objetivo es encontrar una distribución de probabilidad *a-posteriori* dada descripciones de una imagen (esto es llamado inferencia en nuestro contexto) Modelos estadísticos pueden ser obtenidos basados en ejemplos de configuraciones y su imagen respectiva, esto es llamado aprendizaje en el área de inteligencia artificial.

Los modelos probabilísticos generativos ofrecen mecanismos formales de modelar variables aleatorias no observables (aquí esta variable es la configuración del objeto en 3D o 2D) Sin embargo, aunque para este problema es posible definir un modelo generativo con alto grado de exactitud, realizar inferencias no es factible desde un punto de vista computacional (el problema es NP: no es posible encontrar un algoritmo que en tiempo polinomial pueda realizar esta tarea); esto es debido a que el modelo generativo tiene componentes no-lineales. Modelos estadísticos discriminativos podrían ser usados en lugar de modelos generativos, pero aunque inferencia es posible en estos modelos, no es claro como definir estos modelos estadísticos con exactitud para el problema en cuestión (*i.e.*, no es claro como definir una distribución de probabilidad que represente la estructura del problema) Estos dos modelos (generativo y discriminativo) representan puntos de vista complementarios, un método de inferencia ideal debería explotar las ventajas individuales de cada modelo con el fin de hacer inferencia factible y más exacta.

En esta tesis, se proporciona una manera natural y formal de combinar estos modelos. En nuestro método, un modelo discriminativo es obtenido o *aprendido* por medio de data (pares de configuraciones-imágenes) El modelo discriminativo esta basado en una

combinación de distribuciones de probabilidad Gaussianas, cada distribución esta definida por medio de una función o mapa especializado (por ende el título escogido) En lugar de simplemente usar este modelo para definir una distribución *a-posteriori* para la variable aleatoria que representa la configuración en 3D-2D, este modelo es combinado con el modelo generativo, también definido basado en una función o mapa. Como resultado, se hace factible obtener de manera eficiente una aproximación para la verdadera distribución *a-posteriori* definida por el modelo generativo. Esta aproximación es mejor que la proporcionada por el modelo discriminativo. En la tesis se ofrece justificación matemática para el algoritmo de inferencia que resulta de este método y además se detallan dos algoritmos para estimación MAP, uno determinístico y otro probabilístico.

El método es evaluado formalmente por medio de experimentos con imágenes sintéticas (usando computación gráfica) y reales (usando video tomado por una cámara digital standard) en la tarea de estimar la postura de dos cuerpos articulados representados por: (1) 40 parámetros que modelan la posición de articulaciones mayores del cuerpo humano en 2D, (2) 22 parámetros que modelan los ángulos en las articulaciones de la mano, además de la orientación espacial. En estos experimentos, es demostrado que aunque sólo una imagen pareciese insuficiente para estimar las configuraciones, el método propuesto muestra algo grado de exactitud y a la vez el algoritmo diseñado es altamente eficiente (tiempo lineal)