

Proyecto final: IHC basado en gestos de rasgos faciales

Integrantes del equipo:

Objetivo: Reconocer las expresiones faciales de una persona cuyas imágenes son capturadas por la cámara de una computadora con fines de lograr una interacción humano-computadora, en donde la computadora deberá responder/interactuar de acuerdo al tipo de expresión.

Actividades a realizar:

1) Revisión del estado del arte.

2) Captura de imágenes/video.

Capturar un conjunto de imágenes (video) de una persona realizando diferentes expresiones faciales. La cámara debe estar fija y en la parte central de la pantalla (como la cámara de una laptop preferentemente).

3) Tipos de expresiones.

Las expresiones faciales a reconocer deben representar al menos cuatro de las siguientes emociones: neutral, alegría, tristeza, furia, sorpresa, miedo.

Nota: Se obtendrán puntos extras de acuerdo a la complejidad de la expresión y/o movimientos faciales que la conformen.

4) De la posición y orientación de la cara de la persona

Para facilitar el proceso de reconocimiento se aceptará que la persona esté viendo todo el tiempo hacia la cámara y realizando cambios de expresión de manera lenta y continua, e incluso establecer un intervalo de tiempo específico entre una expresión y otra, si así se desea. Por otro lado, si se permiten movimientos más complejos de la cara (giratorios y/o laterales), así como cambios significativos de iluminación, se obtendrán puntos extras.

3) Conformar el skeleton de cada expresión facial

El skeleton de cada expresión facial se define como un conjunto de puntos (puntos o marcadores fiduciales) de acuerdo al tipo de expresión. En este proyecto el skeleton que representa a cada expresión facial se puede generar de manera manual, ya sea para un sujeto (el cual será el que se examinará) o para un conjunto de sujetos (y calcular un skeleton promedio para cada expresión facial). La cantidad y localización de los puntos fiduciales que conforman al skeleton será definida por ustedes. Recordar que los puntos fiduciales son regiones clave que definen una expresión. Por ejemplo, la sonrisa se puede capturar por los cambios de posición de las comisuras de los labios. El artículo sugerido muestra un ejemplo de skeleton (contornos de características principales en la cara) pero no necesariamente debe ser como ahí se indica.

Nota: Los puntos fiduciales son importantes para estimar el flujo óptico (movimiento 2D) de manera más precisa y por lo tanto facilitará el reconocimiento de la expresión que se está generando de la anterior.

4) Implementación o uso código o librerías existente (en MatLab o OpenCV) de una técnica de flujo óptico

Para detectar el cambio de movimientos de una expresión a otra deben usar una técnica de flujo óptico. Dado que solo interesa seguir los vectores de movimiento (trayectoria) de los puntos fiduciales en el skeleton de la expresión anterior al skeleton de la expresión siguiente (de acuerdo a un intervalo de tiempo establecido), el análisis de cambio se realizarán en esos puntos y sus vecinos y se puede utilizar una técnica de estimación local (ej. Lucas-Kanade).

5) El reconocimiento de la expresión se da en el momento que la posición relativa entre los puntos fiduciales que conforman al skeleton capturado al término del intervalo establecido, se emparejen, con el mínimo error, con los puntos de uno de los skeletons previamente almacenados.

1) *s (frames) y diferencia entre fondo e imagen.* Calcular la diferencia existente (bajo un cierto umbral previamente establecido) entre dos imágenes (frames) consecutivos. Calcular la máscara de diferencia entre frames.

Calcular la diferencia existente (bajo un cierto umbral previamente establecido) entre el frame actual y la imagen de fondo anterior. Calcular la máscara de diferencia de fondo.

2) *Registro del fondo.* Guardar de manera incremental la imagen de fondo que se va obteniendo a partir de la diferencias en los frames anteriores. Requiere de un registro de la máscara indicando si un pixel contiene o no información de fondo.

3) *Detección del objeto.* La máscara del objeto inicial en un frame al tiempo t es generada a partir de las máscaras de diferencias obtenidas al tiempo t tanto de los frames y las máscaras de diferencia de fondo.

4) *Post-procesamiento.* Corrige el ruido existente en algunas regiones del objeto las cuales se obtienen en el paso anterior debido al movimiento irregular del objeto y el ruido inherente en la cámara. Se realiza una eliminación de ruido y un suavizamiento en los contornos.

4) Análisis de resultados y conclusiones.

5) Elaboración de la presentación.

6) Elaboración del reporte final.

Proyecto final: Fotomontaje digital para crear la foto grupal de la G17.

Integrantes del equipo:

Objetivo: Generar una foto grupal de la G17 a partir de las mejores tomas (rostro y posición corporal) de cada alumno. Este proceso se le conoce como fotomontaje digital.

Actividades a realizar:

1) Revisión del estado del arte.

2) Utilizar fotos de la web para definir la postura corporal de cada alumno. Cada uno de sus compañeros deberá enviarles varias fotos de su rostro y cuerpo en la postura corporal indicada (al menos dos opciones de postura). Dado que ellos conocen el objetivo de su proyecto se les debe solicitar que de manera intencional generen una mala toma (cerrando los ojos, mirando hacia otro lado, serios, etc.) así como una buena toma (mirando hacia la cámara, con ojos abiertos y sonriendo).

3) Las etapas a considerar son:

a. *Selección manual de las zonas que se van a “fotomontar”.* En el artículo esta selección se realiza marcando con el mouse con una línea semigruesa sobre la región o regiones que se tomarán en cada una de las fotos de entrada (imágenes fuente). Ustedes pueden realizarlo así, o de manera exacta seleccionar pedazos completos en cada imagen fuente.

b. *Implementar el algoritmo de Optimización de Corte de Grafos (graph-cut).* Pueden utilizar código existente.

4) Análisis de resultados y conclusiones.

5) Elaboración de la presentación y del reporte final.

Artículo de referencia:

A. Awargala *et al.*, Interactive Digital Photomontage, SIGGRAPH04

Proyecto final: Correspondencia de características en video

Integrantes del equipo:

Objetivo: Encontrar la correspondencia de una serie de puntos de interés en una secuencia de imágenes (video).

Actividades a realizar:

- 1) Revisión del estado del arte.
- 2) Utilizar videos en la web y/o generar sus propios videos para evaluar el desempeño de sus algoritmos.

Nota: se tomará en cuenta el grado de complejidad de los videos a utilizar.

- 3) Las etapas a considerar son:

a. *Detección de características invariantes y su representación mediante descriptores en espacio-temporales de imágenes.* Métodos posibles: puede ser cualquiera de los mencionados en el artículo de referencia, pero también pueden utilizar un descriptor espacial de los más conocidos y recientes. No necesitan implementarlo de cero, pueden usar código existente, en las librerías o en la Web, del lenguaje que utilizarán.

b. *Estimación del flujo óptico.* En caso de haber seleccionado un descriptor espacial en el punto a. Entonces deberán estimar el flujo óptico sobre los puntos descriptores (vectores).

c. *Obtener la correspondencia.* Ver sección 4 del artículo de referencia. En resumen, deben encontrar un conjunto de puntos de interés espacio-temporales dentro de los descriptores resultantes de las etapas previas.

- 4) Análisis de resultados y conclusiones.
- 5) Elaboración de la presentación y del reporte final.

Artículo de referencia:

Laptev, I. Lindeberg, T., Local Descriptors for Spatio-temporal Recognition, *Lecture Notes of Computer Science*, 2006.

Proyecto final: Detección y reconocimiento de gestos de mano (*hand gesture detection and recognition*)

Integrantes del equipo:

Objetivo: Detectar y reconocer gestos de mano en 2D a partir de su forma comparándolos con gestos existentes en una base de datos. Implementar un algoritmo para la detección y reconocimiento de gestos basado en los componentes principales.

Actividades a realizar:

- 1) Revisión del estado del arte.
- 2) Utilizar y crear una base de datos de gestos de mano

Utilizar bases de datos disponible en la web (o en su caso, crear una). Las imágenes deben contener gestos realizados con una mano en 2D centrados en la imagen con fondo estático.

Generar una base pequeña con gestos para las pruebas experimentales.

Nota: Puntos extras se darán según la variedad/complejidad en los gestos de mano a ser detectados y reconocidos.

- 3) Implementar el algoritmo descrito en el artículo de Turk y Pentland pero enfocarlo para gestos de mano ó implementar otro algoritmo de su selección.

M. Turk, A. Pentland, Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86. (Disponible en la página del curso)

En general, el método utiliza la transformación de Karhunen-Loeve. Cada cara almacenada en la base de datos es un vector de dimensión N . El análisis de los componentes principales (PCA) se utiliza para encontrar un subespacio de dimensión M cuyos vectores de base corresponden a las direcciones de máxima varianza en el espacio original de la imagen. Este nuevo subespacio es normalmente de dimensión más baja ($M < N$).

- 4) Análisis de resultados y conclusiones.
- 5) Elaboración del reporte final.

Proyecto final: Corrección del color (*color correction*)

Integrantes del equipo:

Objetivo: Corregir o restaurar el color de una imagen en escalas de grises basándose en métodos de aprendizaje estadístico de imágenes.

Actividades a realizar:

- 1) Revisión del estado del arte.
- 2) Generar la base de datos de entrenamiento

Recopilar imágenes a color que contengan personas. El proyecto se centrará en recuperar el color particularmente de la piel y cabello de una persona. Para generar la base de datos de entrenamiento tomar las imágenes a color y convertirlas a escalas de grises y almacenar los pares.

- 3) Implementar el algoritmo de aprendizaje estadístico basado en Campos Aleatorios de Markov No-Paramétrico

Artículos de base:

Alexei A. Efros and Thomas K. Leung, "Texture Synthesis by Non-parametric Sampling" IEEE International Conference on Computer Vision (ICCV'99), Corfu, Greece, 1999.

L. Abril Torres-Méndez and Gregory Dudek. "Color Correction of Underwater Images for Aquatic Robot Inspection". *Lecture Notes in Computer Science* 3757, Springer A. Rangarajan, B.C. Vemuri, A.L. Yuille (Eds.), pp. 60-73, 2005.

Se utiliza un modelo de Campos Aleatorios de Markov, y la distribución condicional de un pixel dado sus vecinos sintetizados al momento es estimado al encontrar la vecindad de mayor similitud de la imagen de muestra, a través de la construcción de un histograma. Para el caso de restaurar el color de las imágenes en niveles de gris, se puede aplicar el mismo algoritmo utilizando en lugar de una imagen de textura, pequeños pares de pequeñas regiones provenientes de las imágenes a color y no color de la base de datos de entrenamiento y a partir de ellas realizar el histograma de donde se seleccionará la vecindad con mayor similitud a la vecindad del pixel a sintetizar su color.

- 4) Análisis de resultados y conclusiones.
- 5) Elaboración del reporte final.

Proyecto final: Síntesis de texturas (*texture synthesis*)

Integrantes del equipo:

Objetivo: Sintetizar texturas artificiales a partir de una pequeña (semilla).

Actividades a realizar:

- 1) Revisión del estado del arte.
- 2) Generar la base de datos de entrenamiento

Recopilar imágenes con diferentes texturas para generar la base de datos de entrenamiento y de experimentación.

- 3) Implementar el algoritmo de aprendizaje estadístico basado en Campos Aleatorios de Markov No-Paramétrico

Artículo base:

Alexei A. Efros and Thomas K. Leung, "Texture Synthesis by Non-parametric Sampling". *IEEE International Conference on Computer Vision (ICCV'99)*, Corfu, Greece, 1999.

El proceso de síntesis de textura inicia creciendo una nueva imagen a partir de una semilla inicial, un pixel a la vez. Se utiliza un modelo de Campos Aleatorios de Markov, y la distribución condicional de un pixel dado sus vecinos sintetizados al momento es estimado al encontrar la vecindad de mayor similitud de la imagen de muestra, a través de la construcción de un histograma.

- 4) Análisis de resultados y conclusiones.
- 5) Elaboración del reporte final.

Proyecto final: Reemplazo de textura en vestimentas en imágenes a color

Integrantes del equipo:

Objetivo: Reemplazar la textura de la vestimenta de una persona en una imagen a color basándose en métodos de aprendizaje estadístico de imágenes.

El proyecto se centrará en reemplazar la textura de una vestimenta de una persona, tomando en cuenta las distorsiones naturales de la tela debido al contorno o silueta de la persona y las sombras generadas deberán permanecer igual.

Actividades a realizar:

1) Revisión del estado del arte.

2) Generar la base de datos de entrenamiento:

- Capturar/generar imágenes a color de 5 o más tipos de texturas diferentes, pero todas ellas deberán tener en común que el tamaño del patrón que se repite, con diferentes colores inclusive, es más o menos similar sobre un fondo de color liso.

- Realizar y capturar sobre las texturas distorsiones de diferente tipo (un máximo de 10), ya sea manualmente sobre tela real o de manera simulada sobre la imagen generada/tomada de la textura, se espera que dichas distorsiones generen sombras y cambios en la forma del patrón respectivo.

- Para cada uno de los casos obtener la textura base (es decir en realidad el color de fondo, sin textura) conteniendo las sombras generadas por las distorsiones del paso anterior.

3) Implementar el algoritmo de aprendizaje estadístico basado en los campos aleatorios de Markov no-paramétrico explicado en:

Artículos de base:

Efros and Leung.

Se utiliza un modelo de Campos Aleatorios de Markov, y la distribución condicional de un pixel dado sus vecinos sintetizados al momento es estimado al encontrar la vecindad de mayor similitud de la imagen de muestra, a través de la construcción de un histograma. Para el caso de reemplazar la textura de la vestimenta segmentada manualmente, se puede aplicar el mismo algoritmo utilizando en lugar de una sola imagen de textura, pares de pequeñas regiones provenientes de las imágenes de diferentes texturas y la imagen de base (sin textura) de la base de datos de entrenamiento y a partir de ellas realizar el histograma de donde se seleccionará la vecindad con mayor similitud a la vecindad del pixel a reemplazar la textura nueva.

4) Análisis de resultados y conclusiones.

5) Elaboración del reporte final.

Proyecto final: Super-resolución de imágenes (*image super-resolution*)

Integrantes del equipo:

Objetivo: A partir de una imagen de baja resolución obtener una de alta resolución.

Actividades a realizar:

- 1) Revisión del estado del arte.
- 2) Utilizar/crear la base de datos de entrenamiento

Recopilar imágenes con pares de imágenes de baja y alta resolución para generar la base de datos de entrenamiento y de experimentación.

- 3) Implementar el algoritmo Freeman *et al.* basado en ejemplos.

Artículo base:

W.T. Freeman, T.R. Jones, E.C. Pasztor Example-Based Super- Resolution. *Image-Based Modelling, Rendering, and Lighting*. IEEE. March/April, 2002.

La imagen de alta resolución resultante debe presentar un mayor detalle en su contenido comparada con la de baja resolución. Esto implica una restauración del contenido de alta frecuencia, lo cual puede requerir un incremento en la densidad de cada pixel. El método propuesto por Freeman *et al* se basa en utilizar pares de regiones (patches) correspondientes a vecindades de baja resolución (usualmente 7x7 pixeles) y su correspondiente vecindad de alta resolución (usualmente 5x5) como base de entrenamiento para poder obtener información de frecuencias medias y restaurar la información de altas frecuencias.

- 4) Análisis de resultados y conclusiones.
- 5) Elaboración del reporte final.

Proyecto final: Super-resolución utilizando epítomes de imágenes de alta resolución

Integrantes del equipo:

Objetivo: Estimar imágenes con alta resolución a partir de imágenes de baja resolución basado en epítomes de imágenes de alta resolución similares a la escena de baja resolución.

El proyecto se centrará en implementar un algoritmo de super-resolución que dada una imagen de baja resolución de un escenario determinado, realice una búsqueda dentro de una base de datos de imágenes de alta resolución de importancia y genere epítomes y dé como salida una imagen de alta resolución.

Actividades a realizar:

- 1) Revisión del estado del arte.
- 2) Generar las bases de datos con imágenes de alta resolución de diferentes escenarios (paisajes, personas, edificios, etc.)
- 3) Implementar el algoritmo propuesto en [1] para generar los epítomes.
- 4) Implementar el algoritmo de super-resolución propuesto en [2] utilizando como base de entrenamiento sólo las imágenes contenidas en los epítomes.
- 5) Análisis de resultados y conclusiones.
- 6) Elaboración del reporte final.

Artículos de base:

- [1] Nebosa Jojic, Brendan J. Frey and Anitha Kannan. Epitomic analysis of appearance and shape. Ninth IEEE International Conference on Computer Vision (ICCV) Vol. 1 Nice, France. October 13-16, 2003.
- [2] William T. Freeman, Thouis R. Jones and Egon. C. Pasztor. Example-based super-resolution. Technical Report. Mitsubishi Electric Research Laboratories. 2001.

Proyecto final: Predicción de vejez en personas

Integrantes del equipo:

Objetivo: Utilizar un modelo de Campos Aleatorios de Markov no paramétrico más una combinación de los componentes de los espacios de color de personas de diferentes edades para predecir la apariencia en la edad adulta hasta senil de personas jóvenes.

Actividades a realizar:

- 1) Revisión del estado del arte.
- 2) Recopilar imágenes de personas de diferentes rangos de edades, es vital que estas imágenes se encuentren registradas una con la otra. Se pueden utilizar las imágenes que aparecen en el artículo y buscar más en la web.
- 3) Implementación del algoritmo de Campos Aleatorios de Markov no paramétrico (NP-MRF por sus siglas en inglés).
- 4) Utilizar el espacio de color l -alpha-beta. La imagen de entrada en el espacio de color sRGB es convertida al formato l -alpha-beta, cada imagen en el conjunto de entrenamiento, el cual está conformado por pequeños patches de las características faciales que cambian con la edad (ej. ojos y piel que los rodea, frente, labios y piel que los rodea), también es convertida en sus componentes l -alpha-beta. El algoritmo de NP-MRF utiliza solo los canales l de la imagen de entrada y las de entrenamiento para sintetizar la nueva imagen l_{edad} . Por último, l_{edad} se combina con los canales alpha y beta de la imagen de entrada para generar la imagen a color, la cual se convierte nuevamente al formato sRGB para su visualización.
- 5) Análisis de resultados y conclusiones.
- 6) Elaboración del reporte final.