**Reporte sobre técnicas de extracción de características**

**Procesamiento multimedial**

**Carlos Andrés Reyes Evangelista 157068**

**Lucas y Kanade**

Lucas y Kanade presentaron una técnica iterativa de registro de imágenes que pretendía disminuir los costos que las ténicas tradicionales representaban. Su técnica utiliza información de intensidad espacial del gradiente para buscar directamente por aquellas localizaciones que arrojaran los mejores resultados. Esto funciona porque en esencia ellos no buscan empates explícitos en las imágenes subsiguientes, sino que analizan la dirección en que un objeto se ha movido (bajo la asumción de que la diferencia temporal entre imágenes es mínima).

En el algoritmo que ellos presentan para el registro de imagen en una dimensión se busca encontrar la disparidad que existe entre dos curvas, esto es efectuado mediante el cálculo de derivadas que pretenden fungir como aproximaciones lineales; al resultado de estas operaciones derivativas se le es designado como *h*. A continuación es requerida una normalización de los distintos valores *h* encontrados –de los cuales se esperan valores pequeños– lo cual ellos deciden lograr con un cálculo de promedio ponderado por valores variables inversamente proporcionales al valor absoluto de la segunda derivada de cada punto *x*, pues es conjeturado que esta aproximación lineal es mejor en puntos *x* donde *F(x)* es casi lineal; gracias a esta operación son mejor conservados los valores donde *|F''(x)|* son pequeños para favorecer las aproximaciones conseguidas. Cuando se haya obtenido el promedio (ponderado) de los *h* candidatos se avanza *F(x)* en esa cantidad y se repite el proceso nuevamente -de ahí que sea llamado iterativo-. Lucas y Kanade plantean que esta secuencia de aproximaciones deberían converger al mejor *h* para cada caso.

Este algoritmo puede generalizarse para más dimensiones y para visiones de tipo estereo.

Las dos técnicas a presentar a continuación giran en torno a la detección de esquinas en una imagen bidimensional, forman parte del grupo de técnicas conocido como Detectores de Características cuyo propósito llano es conseguir encontrar localizaciones en una imagen de las cuales sea posible hallar con cierta facilidad y certeza sus localizaciones correspondientes en otras imágenes relacionadas mediante la comparación de parches en ambas imágenes de los que se espera reflejen gradientes en al menos dos direcciones significativamente diferentes.

**Förstner y Harris**

Förstner y Harris (y Mike Stephens) fueron los primeros en proponer utilizar máximos locales sobre medidas cuantificables obtenidas a partir de los valores propios de la matriz que representa la suma ponderada de gradientes horizontales y verticales, conocida como **matriz de autocorrelación** obtenida de la comparación entre un parche y su correspondiente localización en otra imagen presumiblemente con pequeñas variaciones entre sí. Estos valores propios –designados en ciertas fuentes como *λ1* o *λ2* y en otras como *α* y *β*– son utilizados en una ecuación para asignar una puntuación a cada ventana, con base en estos resultados se reporta la existencia de una esquina (u otra característica clave buscada) en una ventana o no. Esta ecuación mencionada se calcula obteniendo la diferencia del determinante de la matriz de autocorrelación (*λ1 \* λ2*) y el cuadrado del rastro de la misma matriz (*λ1 + λ2*) multiplicado por un escalar arbitrario.

El algoritmo base de esta (y la siguiente técnica) se realiza según Szeliski como sigue:

*«1. Calcular derivadas horizontales y verticales*

*» 2. Computar las tres imágenes correspondientes al producto externo de estos gradientes*

*» 3. Convolucionar estas imágenes contra un Gaussiano más largo que el usado en el paso 1*

*» 4. Computar la medida de interés escalar utilizando la ecuación descrita arriba*

*» 5. Encontrar los máximos locales que sobrepasan cierto umbral y reportarlos como localizaciones donde se halló una característica clave.»*

**Shi y Tomasi**

Shi y Tomasi lidian con el problema de encontrar “buenas características que rastrear” más que detectar características en sí mismas ya que consideran que ese es un problema por demás resuelto, al menos en base. Ellos proponen una métrica para cuantificar y calificar las características de modo que únicamente sean rastreadas las que se consideren tienen una disimilitud baja –se considera disimilitud al residuo de características entre el primer y el actual cuadro–. Para conseguir este resultado consideran, al igual que Harris, los valores propios de la matriz de autocorrelación en un algoritmo exactamente análogo, pero con un enfoque distinto al definir los límites: Shi y Tomasi proponen utilizar esos valores para generar un rango de umbrales que va desde el valor mínimo de esos dos valores propios hasta el mayor de ellos, si las comparaciones sobre ambos ejes se encuentran entre estos rangos se es declarada una esquina o una buena característica a rastrear.

**David Lowe - SIFT**

David Lowe, por su parte, dedica su investigación a la detección de transformaciones de características invariantes de escala (SIFT por sus siglas en inglés), estas características son tan importantes porque brindan la capacidad de realizar empates confiables, robustos y casi asegurados de una imagen a otra incluso en una base de datos que albergue una amplia colección de cuadros relacionados.

El procedimiento propuesto por Lowe consiste en cuatro fases: detección de extremos en escala espacial, localización de puntos clave precisos, asignación de orientación y descriptor de imágenes locales.

El primer paso consiste en generar una escala espacial a partir de la convolución de un kernel Gaussiano de escala variable –decidido con base en los resultados de Koenderik (1984) y Lindenberg (1994)– contra una imagen de entrada. Hay entonces que encontrar localizaciones de puntos de interés sobre esta escala espacial para lo cual es utilizada una función de diferencia Gaussiana que permita comparar entre escalas; una vez conseguido, cada pixel en la imagen es comparado contra sus 8 vecinos en su misma escala y contra los 9 pixeles del vecindario correspondiente en las dos escalas adyacentes más próximas, si el pixel es considerado un máximo o un mínimo extremo local entonces es contemplado como un potencial punto de interés.

Una vez que son conseguidos, cada punto es juzgado junto con su vecindario al tomar en cuenta factores como su localización, escala y radio de sus principales curvaturas, de esta manera es posible rechazar muchos puntos de interés que se vean posiblemente opacados por ruido, falta de contraste o que estén mal posicionados con respecto a un borde.

Según David, vale la pena asignar una orientación constante a cada punto de interés para asegurar que éste no variará independientemente de si la imagen lo hace, esta orientación es obtenida al generar un histograma con 36 entradas que modelen los 360 grados de una orientación estándar, las frecuencias de estas entradas son calculadas con base en los gradientes de un espectro de puntos muestra que sean cercanos a cada punto de interés; se es asignada como orientación aquella entrada que consiga la mayor frecuencia.

Finalmente, el último paso consiste en crear un llamado descriptor para cada punto de interés el cual es formado al considerar un vecindario de 16x16 píxeles alrededor del punto de interés y éste subdividirlo en 16 sub-bloques de 4x4 para los cuales se consideran 8 orientaciones, el descriptor finalmente es formado como un vector que almacena estas direcciones para cada uno de los sub-bloques.

Referencias: