

Regiones de interés (keypoints)

Alan Gabriel Vargas Ponce

Ingeniería de Software

4-D

Introducción:

Los keypoints, también conocidos como puntos característicos o de interés, son pequeñas áreas de una imagen que se consideran relevantes para la identificación y descripción de esta. Estos puntos se caracterizan por tener una alta probabilidad de ser parte de un objeto o elemento importante en la escena.

Encontrar estos puntos clave es una tarea fundamental en el procesamiento de imágenes, ya que abre la puerta a un sinfín de posibilidades como, por ejemplo:

- actúan como faros que guían a los algoritmos de detección de objetos hacia las regiones de interés. Al identificar estos puntos clave, podemos determinar la presencia de objetos específicos en una imagen, incluso si están parcialmente ocultos o bajo condiciones de iluminación difíciles.
- Los keypoints faciales, como los ojos, la nariz y la boca, son esenciales para el reconocimiento facial. Al identificar y comparar estos puntos clave en diferentes imágenes, podemos determinar si se trata de la misma persona, incluso con cambios de expresión o de ángulo.
- Permiten seguir el movimiento de objetos a lo largo de una secuencia de imágenes. Al rastrear estos puntos clave, podemos obtener información sobre la trayectoria del objeto, su velocidad y su dirección.
- Ayudan a unir diferentes imágenes para crear un panorama completo. Al identificar puntos clave coincidentes en las imágenes adyacentes, podemos combinarlas sin problemas para obtener una vista más amplia de la escena.
- Permiten encontrar imágenes similares, incluso si hay pequeñas diferencias en la perspectiva, la iluminación o el tamaño. Al comparar los puntos clave de dos imágenes, podemos determinar si se trata de la misma escena capturada desde diferentes ángulos o con diferentes condiciones.

Los keypoints son mucho más que simples puntos en una imagen. Son la puerta de entrada a una comprensión más profunda del contenido visual, una herramienta fundamental para el desarrollo de aplicaciones de visión artificial que transforman la forma en que interactuamos con el mundo que nos rodea.

Desarrollo:

Scale Invariant Feature Transform (SIFT): SIFT es un método que permite detectar puntos característicos en una imagen y luego describirlos mediante un histograma orientado de gradientes. Y además, lo hace de forma que la localización y la descripción presenta una gran invarianza a la orientación, la posición y la escala. Cada punto característico queda, por lo tanto, definido mediante su vector de características de 128 elementos, y se obtiene la información de su posición en coordenadas de la imagen, la escala a la que se encontró la orientación dominante de la región alrededor de dicho punto.

Para obtener un conjunto de descriptores SIFT de una imagen es necesario por un lado obtener los puntos característicos y después, para cada punto de interés, calcular su vector descriptor a partir de la información de los píxeles que lo rodean. SIFT fue propuesto para imágenes en escala de grises, por lo que el vector de características de 128 elementos que define cada píxel, contiene información sobre cómo se distribuyen los niveles de intensidad alrededor de cada punto de interés previamente obtenido. Por lo tanto, el algoritmo consta de dos partes claramente diferenciadas:

- a) Obtención de los puntos característicos
- b) Descripción de la región alrededor de cada punto de interés

Ventajas:

- Invariante a la escala, la rotación y la iluminación
- Robusta a cambios de perspectiva y oclusión
- Alta precisión y repetibilidad
- Eficiente en términos de tiempo de procesamiento
- Desventajas:

Desventajas:

- Sensible al ruido
- No es tan efectivo para imágenes con texturas repetitivas
- Puede ser computacionalmente costoso para grandes conjuntos de datos

Si bien Lowe presenta por primera vez SIFT en el CVPR'99 (Computer Vision and Pattern Recognition Conference), (Lowe, 1999), es el artículo publicado cinco años después, en la International Journal of Computer Vision (Lowe, 2004) donde el método es explicado con un mayor nivel de detalle. En esta última publicación se dice que las etapas para el cálculo del método son las

siguientes:

1. Detección de extremos en el espacio-escala

Se buscan puntos de interés en toda la imagen y en todas las escalas consideradas utilizando una diferencia de Gaussianas.

2. Localización precisa de puntos característicos

Para cada uno de los puntos de interés anteriores se ajusta un modelo que permite determinar su localización y escala. Además, se seleccionan los puntos característicos, keypoints, eliminando los que están próximos a los bordes o tienen bajo contraste.

3. Asignación de la orientación

A cada punto característico se le asigna una o varias orientaciones en función de las direcciones del gradiente local. Esta orientación conjuntamente con la ubicación y la escala calculadas anteriormente permiten que el descriptor sea invariante a estas tres situaciones.

4. Descripción del punto característico

Alrededor de cada punto característico se miden los gradientes locales de la imagen y se utiliza su histograma para obtener una representación de esa región que es robusta a cambios significativos en la iluminación y a pequeñas distorsiones en la forma.

Speeded Up Robust Features (SURF): Speeded-Up Robust Features (SURF) es un algoritmo de visión por computador, capaz de obtener una representación visual de una imagen y extraer una información detallada y específica del contenido. Esta información es tratada para realizar operaciones como por ejemplo la localización y reconocimiento de determinados objetos, personas o caras, realización de escenas 3D, seguimiento de objetos y extracción de puntos de interés. Este algoritmo forma parte de la mencionada inteligencia artificial, capaz de entrenar un sistema para que interprete imágenes y determine el contenido. El Algoritmo SURF se presentó por primera vez por Herbert Bay en ECCV 9.^a conferencia internacional de visión por computador celebrada en Austria en mayo de 2006.

SURF, es un detector y un descriptor de alto rendimiento de los puntos de interés de una imagen, donde se transforma la imagen en coordenadas, utilizando una técnica llamada multi-resolución. Consiste en hacer una réplica de la imagen original de forma Piramidal Gaussiana o Piramidal Laplaciana, y obtener imágenes del mismo tamaño pero con el ancho de banda reducido. De esta manera se consigue un efecto de borrosidad sobre la imagen original, llamado Scale-Space. Esta técnica asegura que los puntos de interés son invariantes en el escalado. El algoritmo SURF está basado en el predecesor SIFT.

Los pasos o métodos que sigue este algoritmo es el siguiente:

Detección

El algoritmo de SURF está basado en los mismos principios y pasos que el SIFT, pero utiliza un esquema diferente y esto debería proveer mejores resultados: más rapidez. Con el fin de detectar puntos característicos en una escala de manera invariable SIFT utiliza filtros de aproximación en cascada. Donde la Diferencia de Gaussianos, DOG, se calcula sobre imágenes re-escaladas progresivamente.

Imagen Integral

Semejante al SDoG.

En vez de utilizar gaussianas para promediar la imagen, se utilizan cuadrados (aproximaciones). Hacer la convolución de la imagen con un cuadrado es mucho más rápido si se utiliza la imagen integral.

La imagen integral se define como:

$$I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(x, y)$$

donde $\mathbf{x} = (x, y)$.

La suma de la imagen original dentro de un rectángulo D de la imagen se puede evaluar rápidamente utilizando esta imagen integral. I (x, y) sumada sobre el área seleccionada requiere 4 evaluaciones de S (x, y) (A, B, C, D)

Puntos de interés en la matriz Hessiana

SURF utiliza un detector de BLOB (Binary Large Object) basado en el Hessiano para encontrar puntos de interés. El determinante de la matriz Hessiana expresa la extensión de la respuesta y es una expresión de un cambio local alrededor del área.

El detector se basa en la matriz Hessiana, debido a su buen desempeño en la precisión. Más precisamente, se detectan estructuras BLOB en lugares donde el factor determinante es el máximo. En contraste con el detector de Hess - Laplace para Mikolajczyk y Schmid, se basa en el determinante de la Hessiana también para la selección de escala, como se hace por Lindeberg. Dado un punto en una imagen I, la matriz H Hessiana (\mathbf{x}, σ) en \mathbf{x} a escala σ se define de la siguiente manera:

$$H(\mathbf{x}, \sigma) = \begin{pmatrix} Lxx(\mathbf{x}, \sigma) & Lxy(\mathbf{x}, \sigma) \\ Lxy(\mathbf{x}, \sigma) & Lyy(\mathbf{x}, \sigma) \end{pmatrix}$$

Donde $Lxx = (x, \sigma)$ es la convolución del segundo orden derivativo $\partial x / \partial x^2 g(\sigma)$ con la imagen en el punto x, y de manera similar para $Lxy = (x, \sigma)$ y $Ly y = (x, \sigma)$.

Representación espacio escala y localización de los puntos de interés

Los puntos de interés deben ser encontrados en diferentes escalas, entre otras cosas porque la búsqueda de correspondencias a menudo requiere su comparación en las imágenes donde se les ve a diferentes escalas. Los espacios escala se aplican en general como una pirámide de imagen. Las imágenes se suavizan repetidamente con un filtro gaussiano y luego, se submuestra a fin de conseguir un nivel superior de la pirámide. Por lo tanto, se calculan varios pisos o escaleras del "det H" con varias medidas de las máscaras:

$$\sigma_{approx} = Currentfiltersize * \left(\frac{(BaseFilterscale)}{BaseFilterSize} \right)$$

Después se buscan los máximos 3D en (x, y, n) utilizando el cubo de vecindad 3x3x3. A partir de ahí se procede a hacer la interpolación del máximo. Lowe restó estas capas de la pirámide para conseguir que el DOG (Diferencia de gaussianas) encuentre en las imágenes los contornos y las manchas.

Features from Accelerated Segment Test (FAST): FAST (por sus siglas en inglés Features from Accelerated Segment Test) es un algoritmo de visión artificial utilizado para la detección de esquinas en imágenes.

Las esquinas son puntos de interés que pueden ser útiles para diversas tareas de visión artificial, pueden ayudar en el rastreo de objetos, las esquinas son relativamente estables ante cambios de iluminación y perspectiva, lo que las hace ideales para rastrear objetos en movimiento a través de diferentes cuadros de video, reconocimiento de objetos, las esquinas pueden ayudar a identificar la forma y la orientación de un objeto en una imagen, reconstrucción 3D, las esquinas correspondientes en imágenes tomadas desde diferentes puntos de vista se pueden utilizar para reconstruir la estructura 3D de una escena.

Fast realiza un conjunto de pasos para llegar a su objetivo, los pasos que realiza minuciosamente son:

Prueba de segmento, se analiza un círculo pequeño alrededor del píxel en cuestión. Si se encuentran suficientes píxeles consecutivos dentro del círculo que sean significativamente más brillantes u oscuros que el píxel central, entonces se considera como un candidato a esquina.

Prueba de alta velocidad, para descartar rápidamente a los no-candidatos, FAST examina solo un subconjunto de píxeles específicos dentro del círculo. Si un candidato no cumple con este criterio, se descarta de inmediato.

Una vez que se identifican los candidatos a esquinas, se pueden aplicar técnicas de refinamiento para mejorar la precisión de la ubicación de la esquina.

Ventajas de FAST:

Velocidad: Es uno de los algoritmos de detección de esquinas más rápidos disponibles.

Simplicidad: El algoritmo es relativamente simple de implementar.

Eficiencia computacional: Requiere poca potencia de procesamiento.

Desventajas de FAST:

Menos preciso: Puede ser menos preciso que otros algoritmos de detección de esquinas más complejos.

Binary Robust Independent Elementary Features (BRIEF):

El método Binary Robust Independent Elementary Features (BRIEF) es un algoritmo de detección y descripción de características utilizado en visión por computadora y reconocimiento de imágenes. Fue propuesto por Michael Calonder, Vincent Lepetit, Christoph Strecha, y Pascal Fua en 2010. BRIEF está diseñado para extraer puntos clave (o características) de una imagen y describirlos de una manera eficiente y robusta para su posterior uso en tareas como el emparejamiento de imágenes, la detección de objetos y la reconstrucción tridimensional.

A diferencia de otros algoritmos de características, como SIFT (Scale-Invariant Feature Transform) o SURF (Speeded-Up Robust Features), que generan descripciones de características utilizando información de gradientes y otros cálculos complejos, BRIEF se basa en un enfoque más simple y eficiente.

El proceso de funcionamiento de BRIEF se puede resumir de la siguiente manera:

Selección de puntos clave, en primer lugar, se seleccionan los puntos clave en la imagen de entrada utilizando un detector de características, como el detector de esquinas Harris o el detector de puntos de interés FAST (Features from Accelerated Segment Test). Estos puntos representan áreas de interés en la imagen donde se pueden encontrar características distintivas.

Generación de descriptores binarios, una vez que se han identificado los puntos clave, BRIEF genera descriptores binarios para cada punto clave. Para hacer esto, BRIEF utiliza un conjunto predefinido de pares de píxeles relativos, donde cada par especifica una comparación de intensidad entre dos píxeles en una ventana alrededor del punto clave. La comparación de intensidad entre estos pares de píxeles se utiliza para determinar el valor binario del descriptor. Por ejemplo, si el

valor de intensidad del primer píxel es mayor que el del segundo, se asigna un valor binario de 1; de lo contrario, se asigna un valor de 0.

Selección y codificación de descriptores, dado que los descriptores binarios generados por BRIEF son sensibles al ruido y a las pequeñas variaciones en la intensidad, es necesario aplicar una estrategia de selección y codificación para mejorar la robustez y la discriminación de los descriptores. Esto puede implicar el uso de técnicas como la selección de pares de píxeles más informativos o la codificación de los descriptores mediante técnicas de hashing para reducir la dimensionalidad y mejorar la eficiencia computacional.

Correspondencia de características, una vez que se han generado los descriptores binarios para los puntos clave en múltiples imágenes, se pueden emparejar estos descriptores para identificar correspondencias entre las características de diferentes imágenes. Esto puede lograrse mediante técnicas de búsqueda de vecinos más cercanos (nearest neighbor) o mediante algoritmos de emparejamiento robustos que tienen en cuenta la geometría y la coherencia de las correspondencias.

Oriented FAST and Rotated BRIEF (ORB): El método Oriented FAST and Rotated BRIEF (ORB) es un algoritmo de detección y descripción de características utilizado en visión por computadora y reconocimiento de imágenes. Fue desarrollado por Ethan Rublee, Vincent Rabaud, Kurt Konolige y Gary Bradski en 2011 como una mejora de los métodos FAST (Features from Accelerated Segment Test) y BRIEF (Binary Robust Independent Elementary Features). ORB combina la rapidez y robustez del detector FAST con la eficiencia y robustez de los descriptores BRIEF, añadiendo además una orientación estimada para cada característica, lo que lo hace adecuado para aplicaciones que requieren invarianza a la rotación.

El funcionamiento de ORB se puede describir de la siguiente manera:

Detección de puntos clave, al igual que FAST, ORB utiliza un detector de características para identificar puntos clave en una imagen. Este detector es rápido y eficiente, identificando áreas de interés que pueden representar características distintivas, como esquinas o puntos de interés. A diferencia de FAST, ORB también calcula una orientación inicial para cada punto clave.

Estimación de la orientación, después de la detección de puntos clave, ORB calcula una orientación para cada punto clave. Esto se logra utilizando un método como el algoritmo de momentos de Hu o el gradiente de la imagen para determinar la dirección principal de la característica.

Generación de descriptores BRIEF rotados, una vez que se han detectado los puntos clave y se ha estimado su orientación, ORB utiliza una versión rotada de los descriptores BRIEF para describir las características en función de su orientación. Esto significa que los descriptores BRIEF se calculan con respecto a la orientación estimada de cada punto clave, lo que mejora la invarianza a la rotación de las características.

Selección y codificación de descriptores, al igual que en el caso de BRIEF, ORB puede aplicar técnicas de selección y codificación de descriptores para mejorar la robustez y discriminación de los descriptores generados. Esto puede incluir estrategias como la selección de pares de píxeles más informativos o la codificación de los descriptores mediante técnicas de hashing.

Correspondencia de características, una vez que se han generado los descriptores BRIEF rotados para los puntos clave en múltiples imágenes, se pueden emparejar estos descriptores para identificar correspondencias entre las características de diferentes imágenes. Esto se puede hacer utilizando técnicas de búsqueda de vecinos más cercanos o algoritmos de emparejamiento robustos que tienen en cuenta la geometría y la coherencia de las correspondencias.

Conclusión:

Los keypoints en el procesamiento de imágenes desempeñan un papel fundamental en una gran variedad de aplicaciones de ingeniería de software, desde la detección de objetos hasta la realidad aumentada. Cada técnica, como SIFT, SURF, FAST, BRIEF y ORB, ofrece enfoques únicos para la detección y descripción de características, con diferentes niveles de robustez, eficiencia y capacidad de adaptación a diversas condiciones.

La importancia de estas técnicas se debe a su capacidad para identificar puntos distintivos en las imágenes, lo que les permite realizar tareas críticas como el emparejamiento de imágenes, la detección de objetos y la reconstrucción tridimensional. Además, al ser invariantes a ciertas transformaciones como la escala, la rotación o la iluminación, estas técnicas son fundamentales para aplicaciones donde se requiere una alta precisión y fiabilidad.

En el campo de la ingeniería de software, estas técnicas son fundamentales para el desarrollo de sistemas de visión por computadora y aplicaciones relacionadas. Se otorgan la capacidad de crear sistemas robustos y eficientes que pueden procesar y comprender imágenes de manera rápida y precisa, lo que abre la entrada a una amplia variedad de aplicaciones innovadoras en ámbitos como la robótica, la medicina, la seguridad y el entretenimiento.

En resumen, las técnicas de detección y descripción de keypoints son elementos fundamentales en el desarrollo de sistemas de visión por computadora y tienen un impacto significativo en la ingeniería de software al facilitar la creación de aplicaciones avanzadas y de alto rendimiento en diversos campos de aplicación.

Fuentes bibliográficas:

- Alegre, E., & Fernández-Robles, L. (s/f). SIFT (SCALE INVARIANT FEATURE). Unileon.es. Recuperado el 17 de marzo de 2024, de <https://buleria.unileon.es/bitstream/handle/10612/11065/cap%208%20ConceptosyMetodosenVxC.pdf;jsessionid=50CD31E1620828E8E593FB7E2D0A07F4?sequence=1>
- Wikipedia contributors. (s/f-c). SURF. Wikipedia, The Free Encyclopedia. <https://es.wikipedia.org/w/index.php?title=SURF&oldid=155954358>
- Wikipedia contributors. (2023, agosto 26). Features from accelerated segment test. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Features_from_accelerated_segment_test&oldid=1172289243
- Wikipedia contributors. (2023b, noviembre 12). Oriented FAST and rotated BRIEF. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Oriented_FAST_and_rotated_BRIEF&oldid=1184781309