Cuestionario de Teoría-1

Visión por Computador

Daniel Bolaños Martínez

1. Justificar adecuadamente las respuestas.

1.1. Diga en una sola frase cuál cree que es el objetivo principal de la Visión por Computador. Diga también cuál es la principal propiedad de las imágenes de cara a la creación algoritmos que la procesen.

La Visión por Computador tiene como objetivo, escribir programas de ordenador que puedan interpretar imágenes y obtener información relevante y detallada de sus píxeles tal como la procesa el ser humano.

Los algoritmos usados deben ser capaces de operar con representaciones virtuales de las imágenes (matrices de píxeles) y no sólo trabajar con la información de cada píxel de forma aislada, si no también con la de su entorno, ya que aporta información muy útil en el procesamiento.

1.2. Expresar las diferencias y semejanzas entre las operaciones de correlación y convolución. Dar una interpretación de cada una de ellas que en el contexto de uso en visión por computador.

Semejanzas: Ambas son operaciones que aplican una máscara de forma iterada a lo largo de la imagen, realizando operaciones sobre cada píxel que usan información de su entorno local. Ambas utilizan como máscara una matriz H que es la prescripción para los pesos en la combinación lineal. Cuando H es simétrica, ambas operaciones coinciden. Además, la convolución y la correlación cumplen las siguientes propiedades: invariabilidad al cambio y superposición.[1]

Diferencias: Las fórmulas de correlación:

$$H \otimes F = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v] \cdot F[i+u, j+v]$$

y convolución:

$$H \star F = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H[u,v] \cdot F[i-u, j-v]$$

son claramente diferentes cuando H no es simétrica. En la convolución, es necesario, antes de aplicar el algoritmo, invertir H en ambas direcciones. La convolución además de las propiedades mencionadas, cumple las propiedades algebraicas de asociatividad, conmutatividad, bilinealidad y existencia de neutro que no cumple la correlación.

La correlación mide la similitud entre dos señales mientras que la convolución mide el efecto de una señal sobre otra. Por lo que en el contexto de Visión por Computador, la correlación es usada para identificar patrones en imágenes (gracias a que esta operación mide la equivalencia entre señales), mientras que la convolución se utiliza en el procesamiento de imágenes (debido a su eficiencia en la aplicación de filtros).

1.3. ¿Cuál es la diferencia "esencial" entre el filtro de convolución y el de mediana? Justificar la respuesta.

El filtro de convolución es lineal mientras que el de mediana no lo es.

Cuando aplicamos un filtro de convolución, para cada pixel de la imagen, estamos realizando una modificación del valor del pixel por una combinación lineal de los valores de los píxeles de su entorno. Por lo que la función de convolución es lineal.

Por otro lado, la mediana no es una función lineal sobre imágenes. En el filtro de mediana se modifica el valor de cada píxel por la mediana de los valores de su entorno, esto requiere un ordenado previo de los píxeles vecinos y la función ordenar no es lineal, por tanto, tampoco lo será el filtro que la implementa.

1.4. Identifique el "mecanismo concreto" que usa un filtro de máscara para transformar una imagen.

La localidad es el principal medio de operación de un filtro de máscara para transformar una imagen. Funciones basadas en este principio, como la convolución, se basan en la modificación de cada píxel de la imagen, usando la información que aportan los píxeles de su entorno en una matriz centrada en el píxel actual y del mismo tamaño

que la máscara.

Con transformaciones locales podemos implementar diferentes funciones tales como mejorar la calidad de la imagen (suavizado, resize...), detectar patrones o extraer información sobre la textura o bordes.

1.5. ¿De qué depende que una máscara de convolución pueda ser implementada por convoluciones 1D? Justificar la respuesta.

Esto depende de que la matriz de la máscara sea separable, ya que la asociatividad de la convolución nos garantiza el mismo resultado convolucionando por vectores 1D de filas y columnas que si lo hacemos directamente por la matriz completa.

Sea M la matriz de la máscara 2D, h una máscara 1D horizontal y v una vertical, M será separable si se puede representar como producto de ambas máscaras.[2]

$$M = v\ h^T$$

.

Para que la matriz de la máscara sea separable:

Descomponer en valores singulares la matriz M y presentarla de la siguiente forma:

$$M = \sum_i \sigma_i\ u_i\ v_i^T$$

. Si únicamente el primer valor singular es distinto de cero, la máscara será separable y obtenemos la expresión de las dos máscaras 1D de forma explícita ($\sqrt{\sigma_0}\ u_0$ y $\sqrt{\sigma_0}\ v_0^T$).

1.6. Identificar y justificar con argumentos las diferencias y consecuencias desde el punto de vista teórico y de la implementación entre:
a) Primero alisar la imagen y después calcular las derivadas sobre la imagen alisada.
b) Primero calcular las imágenes derivadas y después alisar dichas imágenes.

Teóricamente ambas operaciones son equivalentes, ya que la convolución cumple la propiedad conmutativa, por lo que el orden en el que se haga el alisamiento y las derivadas sobre la imagen es indiferente.

En la práctica, la opción a) es mejor respecto a la b), ya que como se ha observado en las prácticas, mientras que en el apartado a) realizamos el suavizado sobre la imagen y luego aplicamos las derivadas sobre la imagen suavizada (3 operaciones), en la opción b), primero calculamos las derivadas y luego suavizamos cada una por separado (4 operaciones).

1.7. Identifique las funciones de las que podemos extraer pesos correctos para implementar de forma eficiente la primera derivada de una imagen. Suponer alisamiento Gaussiano.

El operador de Sobel combina suavizado Gaussiano y diferenciación. Primero se realiza un suavizado Gaussiano sobre la imagen y luego se realiza la primera o segunda derivada.

Calcular primero el alisado a partir de la gaussiana y luego derivarla, es equivalente a aplicar la derivada de la función gaussiana:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2}\ e^{-\frac{x^2+y^2}{2\sigma^2}}$$

El filtro será implementable de forma eficiente si conseguimos expresar la primera derivada de la gaussiana como un filtro separable. Con esto reduciremos el número de operaciones.[2]

Descomponemos la función gaussiana y calculamos su primera derivada respecto de cada variable.

$$G(x, y, \sigma) = (\frac{1}{\sqrt{2\pi}\sigma}\ e^{-\frac{x^2}{2\sigma^2}})\ (\frac{1}{\sqrt{2\pi}\sigma}\ e^{-\frac{y^2}{2\sigma^2}})$$

$$\frac{\partial G(x, y, \sigma)}{\partial x} = \frac{\partial G(x, \sigma)}{\partial x}\ G(y, \sigma) = (-\frac{x}{\sigma^2}\ G(x, \sigma))\ G(y, \sigma)$$

$$\frac{\partial G(x, y, \sigma)}{\partial y} = \frac{\partial G(y, \sigma)}{\partial y}\ G(x, \sigma) = (-\frac{y}{\sigma^2}\ G(y, \sigma))\ G(x, \sigma)$$

Podemos descomponer las derivadas de la gaussiana como máscaras de convolución separables por filas y por columnas. Cada máscara define cada componente del operador de Sobel que será empleado para aproximar la derivada a partir del gradiente. Por la simetría de la función Gaussiana, podemos usar el mismo para ambos ejes.

1.8. Identifique las funciones de las que podemos extraer pesos correctos para implementar de forma eficiente la Laplaciana de una imagen. Suponer alisamiento Gaussiano.

Como en el ejercicio anterior, calcular la Laplaciana de la imagen es equivalente a convolucionar la imagen por la Laplaciana de la Gaussiana.

A partir de la fórmula de la Laplaciana:

$$Laplaciana(G) = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2}$$

Podemos conseguir implementar de forma eficiente el filtro tomando la Laplaciana como suma de dos filtros separables.[2]

Aplicando los resultados del apartado anterior, podemos obtener la fórmula de la Laplaciana como:

$$Laplaciana(G) = (-\frac{x}{\sigma^2}\ G(x,\sigma))'\ G(y,\sigma) + (-\frac{y}{\sigma^2}\ G(y,\sigma))'\ G(x,\sigma)$$

$$Laplaciana(G) = (G(x,\sigma)\ (\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}))\ G(y,\sigma) + (G(y,\sigma)\ (\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2}))\ G(x,\sigma)$$

Hemos obtenido la Laplaciana como suma de dos filtros separables. Cada máscara servirá para detectar cambios de frecuencia en cada una de las direcciones.

1.9. Suponga que le piden implementar de forma eficiente un algoritmo para el cálculo de la derivada de primer orden sobre una imagen usando alisamiento Gaussiano. Enumere y explique los pasos necesarios para llevarlo a cabo.

Supongamos que nos dan el filtro como matriz o vectores. Para ello implementaremos una máscara Sobel para obtener el filtro de derivación. Seguiremos los siguientes pasos:

1. Aplicamos el filtro gaussiano mediante una convolución para alisar la imagen y eliminar las altas frecuencias y posible ruido.

2. Aplicamos una convolución sobre los filtros de las derivadas.

3. Debemos normalizar el filtro (una vez convolucionado) multiplicando por σ. Obteniendo así la máscara de Sobel que aproxima el gradiente.

4. Aplicamos el filtro por filas y columnas con una convolución sobre la imagen. Obteniendo así la primera derivada.

1.10. Identifique semejanzas y diferencias entre la pirámide gaussiana y el espacio de escalas de una imagen, ¿cuándo usar una u otra? Justificar los argumentos.

Semejanzas:

- En ambas obtenemos la aplicación sucesiva de un filtro de Gaussiana sobre una imagen a la que se le realiza por cada nivel un subsampling.
- La pirámide Gaussiana es un caso concreto de espacio de escalas gaussiano donde cada escala tiene solo un nivel.

Diferencias:

- En el espacio de escalas tenemos por cada escala un número de niveles a partir de los que se calculan las escalas con la diferencia de las diferentes matrices a las que se les aplicado el filtro gaussiano.

1.11. ¿Bajo qué condiciones podemos garantizar una perfecta reconstrucción de una imagen a partir de su pirámide Laplaciana? Dar argumentos y discutir las opciones que considere necesario.

La condición necesaria y suficiente sería poder operar las matrices de números flotantes con la suficiente precisión como para no perder información en las sucesivas operaciones sobre las imágenes.

Si verificamos esto, partiendo del último nivel de la pirámide Laplaciana, podremos realizar el proceso inverso a su construcción para obtener la imagen original. Cada nivel de la pirámide Laplaciana, viene definido como $L_k = G_k - f(G_{k+1})$, siendo G_k el k-ésimo nivel de la pirámide Gaussiana y f la función upsampling, despejando G_k y aplicando f^{-1} (subsampling), podemos reconstruir la pirámide Gaussiana y obtener la imagen original como primer nivel de la pirámide Gaussiana.

1.12. ¿Cuáles son las contribuciones más relevantes del algoritmo de Canny al cálculo de los contornos sobre una imagen? ¿Existe alguna conexión entre las máscaras de Sobel y el algoritmo de Canny? Justificar la respuesta

Las contribuciones más importantes del algoritmo de Canny al cálculo de los contornos de una imagen son:[5]

- Reducción del ruido: aplica un filtro de Gaussiana para suavizar la imagen.
- Cálculo de la intensidad del gradiente de la imagen: calcula el valor del gradiente en cada punto de la imagen usando máscaras de derivadas parciales.
- Supresión de no máximos: elimina los puntos de los bordes en los que en la dirección del gradiente, no sean máximo local.
- Histéresis: usa un umbral alto y otro bajo en el filtrado de frecuencias. El bajo se utilizará para completar los bordes que el umbral alto haya detectado en algún punto.

El filtro de Sobel es un paso del algoritmo de Canny, lo utilizamos para calcular el gradiente a partir de la aproximación de derivadas.

1.13. Identificar pros y contras de k-medias como mecanismo para crear un vocabulario visual a partir del cual poder caracterizar patrones. ¿Qué ganamos y que perdemos? Justificar los argumentos

El mecanismo k-medias es un algoritmo de clustering utilizado en el modelos "Bolsa de palabras" para crear un vocabulario visual minimizando la distancia de cada muestra al representante de cada cluster al que le adjudica una etiqueta.[2]

Pros:

- Simplicidad del algoritmo en su implementación
- Dentro de los algoritmos de partición que se pueden usar, es de los más eficientes.
- Asegura, al menos, la convergencia al mínimo local.

Contras:

- Los resultados obtenidos dependen en gran medida del valor de K usado que debe ser conocido de antemano y a veces se debe probar con diferentes valores para ver cual obtiene mejores resultados.
- El resultado puede converger a mínimos locales debido a que opera con la distancia entre el elemento actual y su representante más cercano.
- Los resultados dependen de la semilla utilizada ya que al principio del algoritmo se inicializan los representantes de cada cluster de forma aleatoria. Dependiendo del valor de la semilla obtendremos mayor o menor tasa de convergencia.

1.14. Identifique pros y contras del modelo de "Bolsa de Palabras" como mecanismo para caracterizar el contenido de una imagen. ¿Qué ganamos y que perdemos? Justificar los argumentos.

El modelo Bolsa de Palabras se utiliza para clasificar imágenes a partir de la extracción de características tratadas como palabras a partir de las cuales se genera un vocabulario visual. A partir del vocabulario, se calcula una distribución de las imágenes representadas como un histograma según la frencuencia de sus palabras visuales.[2][3]

Pros:

- Fácil representación de la categoría del objeto. Solo es necesario detectar las palabras en la imagen y almacenarlas como un vector numérico, lo que permite compartir representaciones entre múltiples clases.
- Pueden aplicarse directamente algoritmos basados en Aprendizaje Automático. Utiliza k-medias para la construcción del vocabulario y SVM y Naive Bayes para la clasificación.
- Presenta robustez frente a oclusiones. Al ignorar la información de ubicación, podemos obtener la representación de características consistentes incluso si la característica local aparece en una ubicación diferente mediante cambios de situación.

Contras:

- La localidad en las imágenes es problemática debido a que el algoritmo no almacena información sonre la distribución espacial de las palabras por lo que todas tienen la misma probabilidad para los métodos de bolsa de palabras. Esto dificulta la clasificación de imágenes.

1.15. Suponga que dispone de un conjunto de imágenes de dos tipos de clases bien diferenciadas. Suponga que conoce como implementar de forma eficiente el cálculo de las derivadas hasta el orden N de la imagen. Describa como crear un algoritmo que permita diferenciar, con garantías, imágenes de ambas clases. Justificar cada uno de los pasos que proponga.

Las gráficas proporcionadas por las derivadas de una imagen se caracterizan por mostrar cambios morfológicos que incluyen pendientes, curvatura y amplitud, todos los

cuales dependen del punto de observación relativo. Por lo tanto, la primera y segunda derivada pueden proporcionar, directa e indirectamente, información sobre las tres variaciones morfológicas.[4]

Representaremos las gráficas de su primera y segunda derivada aplicadas a la imagen. Y obtendremos los descriptores a partir de los valores más representativos, por ejemplo los máximos y mínimos absolutos de las derivadas.

Generamos un vocabulario visual y aplicaremos el método de Bolsa de Palabras para crear un histograma de la clase a partir de las frecuencias de los descriptores obtenidos para proceder a la clasificación de las imágenes.

Finalmente, tomaremos un conjunto de entrenamiento que suponga por ejemplo el 80 % del conjunto de las imágenes. Usamos el algoritmo de k-medias para su clasificación, aplicado para 2 clústers, ya que el enunciado nos dice que contamos con dos tipos de imágenes en el dominio del conjunto. Deberíamos ser capaces de garantizar una buena tasa de clasificación para el conjunto test.

Quizás podríamos aumentar la tasa de acierto añadiendo nuevas características como por ejemplo la información ofrecida hasta la N-ésima derivada fijando un N que nos proporcione buenos resultados sin abusar de un tiempo de cálculo demasiado alto.

2. Bibliografía.

[1]. Diapositivas de clase.

[2]. Richard Szeliski. Computer Vision: Algorithms and Applications.

[3]. Bernt Schiele and Mario Fritz. Bag of Words Model and Part-Based Models for Object Class Recognition.

[4]. S.Paraskevopouloua, D.Barsakcioglua, M. Saberia, A. Eftekhara and T. Constandinoua. FSDE for Real-time and Hardware-Efficient Spike Sorting.

[5]. Canny, J. A Computational Approach To Edge Detection.