Cuestionario de teoría 1

Antonio Álvarez Caballero
analca3@correo.ugr.es

21 de octubre de 2015

1. Cuestiones

Cuestión 1. ¿Cuáles son los objetivos principales de las técnicas de visión por computador? Poner algún ejemplo si lo necesita.

Solución. La visión por computador busca extraer información acerca de una imagen,a la que llamamos *significado*. Esta información puede ser de tipo semántica, como identificar objetos en la imagen o reconocer caras, o de tipo geométrica, como contornos, formas, distancias o perspectiva. También se puede usar para almacenar o extraer mensajes, que es lo que llamamos *esteganografía*.

Cuestión 2. ¿Una máscara de convolución para imágenes debe ser siempre una matriz 2D? ¿Tiene sentido considerar máscaras definidas a partir de matrices de varios canales como p.e. el tipo de OpenCV CV_8UC3? Discutir y justificar la respuesta.

Solución. No, ya que hemos visto que muchas máscaras son separables por filas y columnas. Así conseguimos una complejidad computacional menor, y por tanto deberíamos buscar siempre este tipo de máscaras. La Gaussiana, por ejemplo, es separable.

Sí tiene sentido máscaras multicanal, ya que podríamos aplicar una máscara a cada canal de color, modificando cada uno de una manera distinta o combinando varios canales. Aunque generalmente se suelen separar los canales, aplicar filtros, y por último reconstruir la imagen.

Cuestión 3. Expresar y justificar las diferencias y semejanzas entre correlación y convolución. Justificar la respuesta.

Solución. Ambas operaciones consisten en pasar la máscara por todos los píxeles de la imagen, realizando una combinación lineal sobre los valores de luminosidad de los píxeles vecinos.

La diferencia entre ambas operaciones es que en la correlación (formalmente es correlación cruzada) primero refleja la máscara sobre ambos ejes antes de aplicar la convolución.

Cuestión 4. ¿Los filtros de convolución definen funciones lineales sobre las imágenes? ¿y los de mediana? Justificar la respuesta.

Solución. Los filtros de convolución definen funciones lineales, lo demostramos tomando la definición de convolución y sean $\alpha \in \mathbb{R}$ una constante y F_1, F_2 dos imágenes. Entonces:

$$
\begin{aligned}
H \star (\alpha(F_1 + F_2)) = \sum_{u=-k}^{k} \sum_{v=-k}^{k} H(u,v)\alpha\Big(F_1(i-u,j-v) + F_2(i-u,j-v)\Big) = \\
= \alpha\Big(\sum_{u=-k}^{k} \sum_{v=-k}^{k} H(u,v)(F_1(i-u,j-v) + F_2(i-u,j-v))\Big) = \\
= \alpha\Big(\sum_{u=-k}^{k} \sum_{v=-k}^{k} H(u,v)F_1(i-u,j-v) + \\
+ \sum_{u=-k}^{k} \sum_{v=-k}^{k} H(u,v)F_2(i-u,j-v)\Big) = \\
= \alpha(H \star F_1 + H \star F_2)
\end{aligned}
$$

Los filtros de mediana no son lineales, ya que la mediana de un conjunto de números no lo es. Tomamos F_1 y F_2 las dos imágenes siguientes:

$$
F_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 6 \\ 8 & 7 & 10 \end{pmatrix}; F_2 = \begin{pmatrix} 0 & 0 & 0 \\ 5 & 6 & 0 \\ 10 & 7 & 6 \end{pmatrix}; F_1 + F_2 = \begin{pmatrix} 0 & 0 & 0 \\ 5 & 9 & 6 \\ 18 & 14 & 16 \end{pmatrix}
$$

Es claro que no es una operación lineal, ya que en F_1 vale 3, en F_2 vale 5 y en $F_1 + F_2$ vale $6 \neq 3 + 5$.

Cuestión 5. ¿La aplicación de un filtro de alisamiento debe ser una operación local o global sobre la imagen? Justificar la respuesta.

Solución. Debe ser una operación local. Esto se debe a una de las reglas básicas de la visión por computador: El valor asociado un píxel en la matriz asociada a la imagen está relacionado con los de su alrededor. No tendría sentido tomar todos los píxeles de una imagen, ya que entonces todos los píxeles se alisarían de la misma manera, no obteniendo un buen resultado. Además, así no cumplimos la misión del alisamiento, que es reducir la diferencia lumínica entre píxeles vecinos.

Cuestión 6. Para implementar una función que calcule la imagen gradiente de una imagen dada pueden plantearse dos alternativas:

- Primero alisar la imagen y después calcular las derivadas sobre la imagen alisada.
- Primero calcular las imágenes derivadas y después alisar dichas imágenes.

Discutir y decir que estrategia es la más adecuada, si alguna lo es. Justificar la decisión.

Solución. Primero debemos alisar la imagen y luego calcular sus derivadas, puesto que la imagen original es muy probable que tenga mucho ruido. Esto hace que las derivadas no sean estables, pudiendo obtener datos muy falseados u erróneos. Alisando primero la imagen, obtenemos unos datos más uniformes y útiles.

El *teorema de la derivación de la convolución* nos dice que si primero derivamos el filtro y luego convolucionamos, es lo mismo. Así que otra estrategia podría ser derivar primero el filtro, lo cual es menos costoso que hacerlo en toda la imagen.

Cuestión 7. Verificar matemáticamente que las primeras derivadas (respecto de x e y) de la Gaussiana 2D se puede expresar como núcleos de convolución separables por filas y columnas. Interpretar el papel de dichos núcleos en el proceso de convolución.

Solución. Usando iPython Notebook resolveremos este ejercicio fácilmente.

```
In [1]: import sympy
        from mpmath import *
        from sympy import separatevars
        from sympy.abc import sigma
        sympy.init_session()
        s = symbols('s')
```

```
IPython console for SymPy 0.7.7.dev (Python 3.5.0-64-bit) (ground types: gmpy)

These commands were executed:
>>> from __future__ import division
>>> from sympy import *
>>> x, y, z, t = symbols('x y z t')
>>> k, m, n = symbols('k m n', integer=True)
>>> f, g, h = symbols('f g h', cls=Function)
>>> init_printing()

Documentation can be found at http://docs.sympy.org/dev
```

```
In [2]: f = 1/(2*pi*sigma**2)*exp(-(x**2+y**2)/(2*sigma**2))
```

```
In [3]: f
```

Out[3]:

$$\frac{e^{\frac{-x^2-y^2}{2\sigma^2}}}{2\pi\sigma^2}$$

```
In [4]: separatevars(diff(f,x),force=True)
```

Out[4]:

$$-\frac{xe^{-\frac{x^2}{2\sigma^2}}}{2\pi\sigma^4}e^{-\frac{y^2}{2\sigma^2}}$$

```
In [5]: separatevars(diff(f,y),force=True)
```

Out[5]:

$$-\frac{ye^{-\frac{x^2}{2\sigma^2}}}{2\pi\sigma^4}e^{-\frac{y^2}{2\sigma^2}}$$

Podemos afirmar entonces que son separables ambas derivadas. Para la derivada en x nos queda, separando también el denominador,

$$\frac{\partial G}{\partial x} = -\frac{xe^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma^2}\frac{e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma^2}$$

Y para la derivada en y,

$$\frac{\partial G}{\partial y} = -\frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma^2}y\frac{e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma^2}$$

Los papeles de ambas en la convolución son calcular una imagen que resalte los cambios de intensidad en cada uno de los ejes: En el eje X la parcial con respecto a x y respectivamente en el eye Y la parcial sobre y.

Cuestión 8. Verificar matemáticamente que la Laplaciana de la Gaussiana se puede implementar a partir de núcleos de convolución separables por filas y columnas. Interpretar el papel de dichos núcleos en el proceso de convolución.

Solución. Ayudándonos de iPython Notebook (pero no colocando ahora la salida), conseguimos la laplaciana de la gaussiana:

$$\bigtriangledown^2 h_\sigma(u,v) = \frac{1}{2\pi\sigma^2}e^{-\frac{u^2+v^2}{2\sigma^2}}\left(\frac{u^2+v^2-2\sigma^2}{\sigma^4}\right) =$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{v}{2\sigma^2}}\right)e^{-\frac{u}{2\sigma^2}}\left(\frac{u^2-\sigma^2}{\sqrt{2\pi}\sigma^4}\right)+$$

$$+\left(\frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{u}{2\sigma^2}}\right)e^{-\frac{v}{2\sigma^2}}\left(\frac{v^2-\sigma^2}{\sqrt{2\pi}\sigma^4}\right)$$

Se observa que es suma de funciones separables de u y v.

Cuestión 9. ¿Cuáles son las operaciones básicas en la reducción del tamaño de una imagen? Justificar el papel de cada una de ellas.

Solución. Las operaciones básicas son alisamiento y submuestreo.

Al realizar submuestreo aparece el *aliasing*, ya que no tomamos una señal lo suficientemente grande y se pierden las frecuencias más altas, perdiendo mucho detalle. Por eso es importante primero alisar y luego submuestrear, ya que así la imagen queda suavizada ünuniformizandol̈as frecuencias, evitando así que se pierdan las más altas.

Cuestión 10. ¿Qué información de la imagen original se conserva cuando vamos subiendo niveles en una pirámide Gausssiana? Justificar la respuesta.

Solución. Conservamos las bajas frecuencias, ya que al realizar los pasos de alisamiento y submuestreo la imagen se va haciendo cada vez más uniforme, quedando las zonas donde los cambios de iluminosidad son más suaves. Se pierden las altas frecuencias, y con ello las zonas de gran detalle de la imagen.

Cuestión 11. ¿Cuál es la diferencia entre una pirámide Gaussiana y una Piramide Lapalaciana? ¿Qué nos aporta cada uan de ellas? Justificar la respuesta. (Mirar en el artículo de Burt-Adelson)

Solución. Si tomamos G_i y G_{i+1} dos niveles sucesivos de una pirámide gaussiana, y contruimos G'_{i+1} como G_{i+1} pero con el tamaño de G_i, entonces la pirámide laplaciana la podemos contruir como

$$L_i = G_i - G'_{i+1}$$

Es claro que si la gaussiana tiene N niveles, $L_N = G_N$, ya que no existe G_{N+1}. Entonces la diferencia es clara: la laplaciana conserva las frecuencias medias y altas, mientras que en la gaussiana las bajas.

La pirámide laplaciana es muy usada en codificación eficiente de imágenes, ya que la compresión es mayor en imágenes con frecuencias más altas, que es justo lo que caracteriza a la laplaciana. Esto se consigue debido a que las imágenes de la pirámide laplaciana tienen entropía y varianza pequeñas, haciendo la codificación más agresiva, pero sin distorsionar lo suficiente como para que los humanos lo notemos.

Cuestión 12. Cual es la aportación del filtro de Canny al cálculo de fronteras frente a filtros como Sobel o Robert. Justificar detalladamente la respuesta.

Solución. Mientras que los filtros de Sobel y Robert aplican dos máscaras de convolución y detectan bordes horizontales y verticales, el filtro de Canny filtra usando la derivada de la Gaussiana. Además, añade la *supresión de los no-máximos* (compara la intensidad de un píxel con sus vecinos en la dirección del gradiente, conservando dichos píxeles si su intensidad es máxima en comparación con los demás) y la *histéresis y enlazado* (se definen umbrales alto y bajo, los píxeles que superan el umbral alto forman un borde, y los más bajos que el umbral bajo se desechan).

Esto consigue una imagen con bordes de anchura un píxel, consiguiendo mucha precisión en la situación del borde. Además, el filtro de Canny devuelve una imagen binaria y no en escala de grises.

Cuestión 13. Buscar e identificar una aplicación real en la que el filtro de Canny garantice unas fronteras que sean interpretables y por tanto sirvan para solucionar un problema de visión por computador. Justificar con todo detalle la bondad de la elección.

Solución. En el ámbito de la seguridad informática existe la *esteganografía*, una rama que se encarga de la ocultación de información en imágenes. Como comentamos en la primera cuestión, la visión por computador trata tanto el cifrado como descifrado de información en este ámbito, así que comentamos un caso: investigadores de la Universidad Nacional de la Matanza, Buenos Aires[1] han utilizado el filtro de Canny para *categorizar* imágenes según su nivel de ruido, ya que a más ruido, más difícil es de sacar información oculta de la imagen, ya que es más difícil seguir los cambios de luminosidad de los píxeles.

La elección sobre este filtro se realizó mayormente por ensayo y error, ya que hicieron pruebas con los filtros de *Sobel*, *Prewitt*, *Roberts*, *Laplaciano*, *Cruce por cero* y *Canny.* Pero es de esperar que el filtro de Canny obtenga mejores resultados, sobre todo por su capacidad de *unir bordes* y su gran precisión.

BONUS. Usando la descomposición SVD (Singular Value Decomposition) de una matriz, deducir la complejidad computacional que es posible alcanzar en la implementación de la convolución 2D de una imagen con una máscara 2D de valores y tamaño cualesquiera (suponer la máscara de tamaño inferior a la imagen).

[1] http://sedici.unlp.edu.ar/bitstream/handle/10915/43224/Documento_completo.pdf?sequence=1

Solución. contenidos...

Para escribir este documento he recibido la ayuda de Alejandro García Montoro.