Cuestionario 3

Luis Antonio Ortega Andrés

26 de diciembre de 2019

Pregunta 1. *¿Cúal es la transformación más fuerte de la geometría de una escena que puede introducirse al tomar una foto de ella?. Dar algún ejemplo.*

La transformación más fuerte es la pérdida de paralelismo en las rectas, resultado de una homografía no afín. Un ejemplo de esto es una fotografía de las vías de tren, que pese a saber que son paralelas, parecen intersecarse en el horizonte.

Figura 1: Fotografía de unas vías de tren.

Pregunta 2. *¿Por qué es necesario usar el plano proyectivo para estudiar las transformaciones en las imágenes de fotos de escenas?. Dar algún ejemplo.*

Las transformaciones geométricas afínes preservan el paralelismo y como hemos visto, esto no es una propiedad de las fotografías. Por ello, necesitamos un modelo que permita transformaciones más generales, como es el plano proyectivo.

Por ejemplo, si buscamos componer un mosaico con dos imágenes distintas de un mismo lugar, las transformaciones afines podrían no ser suficiente para encajar las imágenes mientras que utilizando transformaciones propias del plano proyectivo, como las homografías, si podemos lograrlo.

Pregunta 3. *Sabemos que en el plano proyectivo un punto no existe en el sentido del plano afín, sino que se define por una clase de equivalencia de vectores definida por* $\{k(x,y,1)\ k\neq 0\}$*. Razone usando las coordenadas proyectivas de los puntos afines de una recta que pase por el* $(0,0)$ *del plano afín que los puntos de la recta del infinito del plano proyectivo son necesariamente vectores del tipo* $(\star,\star,0)$ *con* $\star$ *cualquier número.*

Si escribimos la ecuación paramétrica de una recta vectorial $ax+by=0$ con a o b no nulo, obtenemos $x=bt$ e $y=-at$, de forma que los puntos de la recta son de la forma $(bt,-at)$, si introducimos una coordenada $z=1/t$, los puntos de la recta (salvo el origen) son $(b/z,-a/z)$, cuyas coordenadas homogéneas son $(b,-a,z)$.
Si en esta situación nos alejamos del origen por la recta a la recta del infinito, es decir, tomamos límite cuando $t\to+\infty$, el punto del plano proyectivo resultante es $(b,-a,0)$. Por esto, los puntos de la recta del infinito tienen tercera coordenada nula.

Pregunta 4. *¿Qué propiedades de la geometría de un plano quedan invariantes cuando se toma una foto de él? Justificar la respuesta.*

Sabemos que esta situación se corresponde con la aplicación de una homografía general, de las que también sabemos que sólo se preserva la colinealidad. Es decir, lleva rectas en rectas y puntos no alineados en puntos no alineados.

En general sabemos que las homografías no respetan longitudes o ángulos, pero esto tampoco se mantiene mediante transformaciones afines, lo novedoso es que tampoco se mantiene el paralelismo.

Pregunta 5. *En coordenadas homogeneas los puntos y rectas del plano se representan por vectores de tres coordenadas (notados x y l respectivamente), de manera que si una recta contiene a un punto se verifica la ecuación* $x^Tl=0$*. Considere una homografía* H *que transforma vectores de puntos,* $x'=Hx$*. Dado que una homografía transforma vectores de tres coordenadas también existen homografías* G *para transformar vectores de rectas* $l'=Gl$*. Suponga una recta* l *y un punto* x *que verifican* $x^Tl=0$ *en el plano proyectivo y supongo que conoce una homografía* H *que transforma vectores de puntos. En estas condiciones ¿cuál es la homografía* G *que transforma los vectores de las rectas?. Deducirla matemáticamente.*

Sabemos que $x\in l \iff x^Tl=0$, luego si $x'=Hx$ y $l'=Gl$, se da $x'^Tl'=0 \iff (Hx)^TGl=x^TH^TGl=0$. Para conseguir esto, basta imponer $H^TG=I$, de forma que $x^TH^TGl=x^Tl=0 \iff x'\in l'$.

Como H es regular, también lo es su traspuesta, lo cual implica que $G=(H^T)^{-1}$.

Pregunta 6. *¿Cúal es el mínimo de escalares necesarios para fijar una homografía general? ¿Y si la homografía es afín? Justificar la respuesta.*

Sabemos que una homografía general es una matrix 3×3 regular, esto implica que necesitamos 9 escalares. sin embargo, como los puntos proyectivos no son afectados por reescalados, solo nos harán falta 8 escalares. Veamos esto, sea H una homografía y (x,y,z) un punto proyectivo.

$$H\begin{pmatrix}x & y & z\end{pmatrix}^T=\begin{pmatrix}a & b & c\\ d & e & f\\ g & h & i\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}1 & b/a & c/a\\ d/a & e/a & f/a\\ g/a & h/a & i/a\end{pmatrix}a\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}1 & b/a & c/a\\ d/a & e/a & f/a\\ g/a & h/a & i/a\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}$$

Donde hemos supuesto sin pérdida de generalidad que $a\neq 0$, de no serlo, elegiriamos otro escalar no nulo (debe existir por ser regular). Además utiliamos que $a(x,y,z)=(x,y,z)$, ya que representan el mismo punto.

Supongamos ahora H_A una homografía afín, sabemos que deja fija la recta del infinito, es decir, $H_A(x,y,0)^T = (x',y',0)^T$. Esto quiere decir que H_A es una matriz regular de la forma

$$H_A = \begin{pmatrix} a & b & c \\ d & e & f \\ 0 & 0 & i \end{pmatrix}$$

Siguiendo la misma idea que antes vemos que necesitamos 6 escalares únicamente.

Pregunta 7. *Defina una homografía entre planos proyectivos que haga que el punto* $(3,0,2)$ *del plano proyectivo 1 se transforme en un punto de la recta del infinito del plano proyectivo 2.*

Como queremos que $(3,0,2)$ vaya a un punto de la forma $(x,y,0)$, debe cumplirse que si (a,b,c) es la última fila de la homografía, entonces

$$\begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} = 0$$

Luego $3a + 2c = 0$, podemos elegir $a = -2, c = 3$. El resto de escalares los podemos elegir libremente (siempre que la matriz resultante sea regular) pues ya tenemos garantizado que la imagen del punto caiga sobre la recta infinito del plano proyectivo 2. Por ejemplo

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 3 \end{pmatrix}$$

cuyo determinante es 3.

Pregunta 8. *Una homografía general* H *admite una descomposición única en movimientos elementales de la forma* $H = H_S H_A H_P$*, donde* H_S *representa una homografía de una similaridad (escala, giro y traslación),* H_A *la homografía de un movimiento afín puto y* H_P *una transformación proyectiva pura. Describir un algoritmo que permita encontrar las matrices de la descomposición de una matriz* H *dada. Aplicarlo para encontrar la descomposición de*

$$H = \begin{pmatrix} 1{,}707 & 0{,}586 & 1{,}0 \\ 2{,}707 & 8{,}242 & 2{,}0 \\ 1{,}0 & 2{,}0 & 1{,}0 \end{pmatrix}$$

Pregunta 9. *¿Cuáles son las propiedades necesarias y suficientes para que una matriz defina un movimiento geométrico no degenerado entre planos?. Justificar la respuesta.*

Para que una matrix 3×3 defina un movimiento geométrico no degenerado entre planos proyectivos solo necesitamos que su determinante no sea nulo. De ser nulo podría llevar un punto al $(0,0,0)$, que no corresponden con las coordenadas homogéneas de ningún punto.

Si buscamos que una matriz defina un movimiento geométrico no degenerado entre planos afines, debemos fijar la recta del infinito y el plano afín. Ya vimos en el ejercicio 6, que la matríz debe ser de la forma

$$H = \begin{pmatrix} a & b & c \\ c & d & e \\ 0 & 0 & i \end{pmatrix}, i \neq 0$$

Veamos ahora que es suficiente, si (x, y, z) es un punto afín ($z \neq 0$), entonces $H(x, y, z)^T = (\star, \star, iz)$ donde $iz \neq 0$, luego el punto es afín.

Pregunta 10. *¿Qué información de la imagen usa el detector de Harris para seleccionar los puntos?. ¿El detector de Harris detecta patrones geométricos o fotométricos?. Justificar la respuesta.*

El detector de Harris utiliza información local del gradiente de la imagen para seleccionar esquinas. Para hacerlo realiza los siguientes pasos:

- Calcula el gradiente en cada punto de la imagen.
- Crea una matriz H con los valores del gradiente en una región de la iagen.
- Calcula los valores propios y aplica el operador de Harris.
- Se queda con aquellos valores que superen un cierto umbral y sean máximos locales.

Detecta patrones geométricos (esquinas) y además es invariante a transformaciones constantes de intensidad de las imágenes, pues utiliza las derivadas de la misma para obtener información.

Pregunta 11. *¿Sería adecuado usar como descriptor de un punto Harris los valores de los píxeles de su región de soporte?. Identifique ventajas, inconvenientes y mecanismos de superación de estos últimos.*

Por lo general no sería adecuado, ya que si solo tomamos los valores de los píxeles en la región de soporte no tenemos garantizada la invarianza frente a reescalados, rotaciones u otras transformaciones afines.

La ventaja que presenta este descriptor es su sencillez y facil implementación. Podría ser útil en situaciones donde sepamos que los valores de los píxeles no se van a alterar (salvo traslaciones).

La falta de invarianza frente a transformaciones afines se podría corregir extendiendo a un descriptor multiescala o combinar la detección de puntos Harris con otro descriptor como SIFT.

Pregunta 12. *Describa un par de criterios que sirvan para seleccionar parejas de puntos en correspondencias («matching») a partir de descriptores de regiones extraídos de dos imágenes. ¿Por qué no es posible garantizar que todas las parejas son correctas?.*

Supongamos que tenemos dos descriptores D_1 y D_2 extraidos de dos imágenes, buscamos seleccionar elementos de $D_1 \times D_2$ tal que los puntos asociados estén en correspondencia. Consideramos dos criterios, en ambos casos necesitamos una función que mida distancias entre dos descriptores de D_1 y D_2:

- Fuerza bruta con crosscheck. Para cada elemento d_1 de D_1, elegimos el más cercano d_2 en D_2. Si además d_1 es el más cercano para d_2, elegimos la correspondencia.
- Lowe-Average-2NN. Para cada d_1 en D_1, tomamos los dos más cercanos en D_2, d_2, d_2'. Calculamos

  $$r = \frac{|d_1 - d_2|}{|d_1 - d_2'|}$$

  Si r es mayor que un umbral descartamos la correspondencia, pues la distancia entre la mejor correspondencia y la segunda mejor es demasiado pequeña y hay ambigüedad. Si es menor que el umbral, añadimos la correspondencia con el más cercano de los dos.

Con ninguno de los dos métodos se puede garantizar que todas las parejas sean correctas, podría ocurrir que en una de las dos imágenes hubiera dos regiones iguales y en la segunda imagen solo apareciera una vez, de forma que se puedan crear correspondencias incorrectas.

Pregunta 13. *¿Cuál es el objetivo principal del uso de la técnica RANSAC en el cálculo de una homografía? Justificar la respuesta.*

El objetivo de la técnica RANSAC es estimar de forma robusta una homogafía a partir de un conjunto de correspondencias de puntos entre dos imágenes con posibles *outliers*. Se diferencia de la técnica de mínimos cuadrados en que esta es sensible a *outliers*, mientras que son RANSAC pretendemos eliminar esta sensibilidad.

Para lograrlo toma conjuntos de 4 correspondencias (el mínimo para calcular una homografía) en cada iteración, estima una homografía H y computa el error total entre dicha homografía y las correspondencias que tenemos. Luego devuelve aquella que produce menos error.

Pregunta 14. *Si tengo 4 imágenes de una escena de manera que se solapan la 1-2, 2-3 y 3-4. ¿Cuál es el número minimo de parejas de puntos en correspondencias necesarios para montar un mosaico? Justificar la respuesta.*

Ya sabemos que para estimar una homografía necesitamos mínimo 4 parejas de puntos en correspondencias. Con esas 4 parejas tenemos un sistema de 8 ecuaciones con 9 incógnitas, determinando 8 de los 9 escalares que definen la homografía. Pero ya vimos en la pregunta 6, que era suficiente para definirla en su totalidad.

Si buscamos crear un mosaico con 4 imágenes, podemos trasladar la segunda al mosaico mediante la homografía (traslación) H_0, después tendríamos que calcular 3 homografías, $H_{1,2}$, $H_{3,2}$ y $H_{4,2} = H_{4,3}H_{3,2}$. Donde $H_{i,j}$ es la homografía que lleva la imagen i en la imagen j. Si componemos cada una de ellas con H_0, llevariamos todas las imágenes al mosaico.

Esto quiere decir que necesitamos estimar 3 homografías $H_{2,1}$, $H_{3,2}$ y $H_{4,3}$ mediante las correspondencias que tenemos. Para ello necesitamos un total de $4 \cdot 3 = 12$ parejas de puntos como mínimo.

Pregunta 15. *En la confección de un mosaico con proyección rectangular, ¿Es esperable que aparezcan deformaciones geométricas de la escena real? ¿Cuáles y por qué? ¿Bajo qué condiciones esas deformaciones podrían no estar presentes?. Justificar la respuesta.*

Podemos esperar la aparición de deformaciones geométricas si cambiamos el punto de vista de la cámara, pues estamos proyectando en un mismo plano. Para evitarlo, podemos tomar todas las fotos en el mismo eje, es decir, trasladando la camara a lo lardo del mismo eje en todas las fotos. Otra opción es cambiar la proyección a una esferica o cilíndrica.

También podriamos observar deformaciones debido a acumulación de errores en el cálculo de homografías.