Visión Por Computador (2015-2016)

Grado en Ingeniería Informática

Universidad de Granada

Cuestionario 2

Mª Cristina Heredia Gómez

18 de noviembre de 2015

Índice

Índice de figuras

1. ¿Identificar las diferencias esenciales entre el plano afín y el plano proyectivo? ¿Cuáles son sus consecuencias? Justificar la contestación.

La principal diferencia entre un plano afín y un plano proyectivo, es que en un plano afín tenemos que el paralelismo es una relación de equivalencia entre todas las líneas de tal plano, mientras que en un plano proyectivo, tenemos que cualesquiera dos líneas se cortan en un punto, único. Éste punto puede ser un ”punto en el infinito” , siendo éste donde se cortan las líneas con una línea paralela.
Por esta razón, en un plano proyectivo tendremos que dadas dos rectas diferentes cualesquiera, estas tienen esactamente un punto en común, mientras que en un plano afín, dadas dos rectas diferentes cualesquiera, éstas no tendrán ningún punto en común.

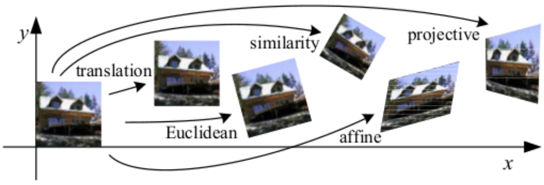

Figura 1.1: figura 3.45 del libro Richard Szeliski

Podemos pasar de un plano proyectivo a uno afín, quitando líneas no paralelas y los puntos que contienen, al igual que podemos pasar de un plano afín a un plano proyectivo, introduciendo una recta en el infinito.
Como podemos ver, la principal consecuencia que tenemos es que ,en un plano afín mantenemos el paralelismo, mientras que en un plano proyectivo mantenemos las líneas rectas y las distancias pero **el paralelismo se pierde**.

2. Verificar que en coordenadas homogéneas el vector de la recta definida por dos puntos puede calcularse como el producto vectorial de los vectores de los puntos ($l = x \cdot x'$). De igual modo el punto intersección de dos rectas l y l' está dado por $x = l \cdot l'$

Sea φ la aplicación que lleva de un plano afín a un plano proyectivo:

$$\varphi : A^2 \rightarrow P^2/(x, y) \rightarrow (x, y, 1)$$

Sean $P_1 = (x_1, y_1)$ y $P_2 = (x_2, y_2)$ ϵ A, dos puntos cualesquiera. Entonces sabemos que un vector de la recta definida por esos dos puntos, es:

$$\vec{v}_r = \overrightarrow{P_2P_1}$$

donde $\vec{v}_{r1} = x_1 - x_2$ y $\vec{v}_{r2} = y_1 - y_2$ entonces, sustituyendo en la ecuación de la forma contigua, tenemos:

$$\frac{x - x_1}{\vec{v}_{r1}} = \frac{y - y_1}{\vec{v}_{r2}} \Rightarrow \frac{x - x_1}{x_1 - x_2} = \frac{y - y_1}{y_1 - y_2}$$

y despejando:

$$(x - x_1)(y_1 - y_2) = (x_1 - x_2)(y - y_1)$$

$$x \cdot y_1 - x \cdot y_2 - x_1 \cdot y_1 + x_1 \cdot y_2 = x_1 \cdot y - x_1 \cdot y_1 - y \cdot x_2 + y_1 \cdot x_2 \Rightarrow$$

$$x(y_1 - y_2) + x_1 y_2 = y(x_1 - x_2) + y_1 x_2 \Rightarrow$$

$$x(y_1 - y_2) - y(x_1 - x_2) + (x_1 y_2 - y_1 x_2) = 0$$

y por tanto el vector de la recta será:

$$\vec{v}_r = (y_1 - y_2, -x_1 + x_2, x_1 y_2 - y_1 x_2)$$

Calculemos ahora el producto vectorial de los dos puntos:

$$\vec{P}1\vec{P}2 = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = \begin{vmatrix} y_1 & 1 \\ y_2 & 1 \end{vmatrix} \vec{i} - \begin{vmatrix} x_1 & 1 \\ x_2 & 1 \end{vmatrix} \vec{j} + \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \vec{k} =$$

$$= (y_1 - y_2)\vec{i} - (x_1 - x_2)\vec{j} + (x_1 y_2 - x_2 y_1)\vec{k}$$

por lo que tenemos que las coordenadas del vector de la recta serían:

$$\vec{v}_r = (y_1 - y_2, -x_1 + x_2, x_1 y_2 - y_1 x_2)$$

obteniendo lo mismo que en el caso anterior. Por lo tanto, tendríamos que el vector de la recta definida por dos puntos puede calcularse como el producto vectorial de tales puntos.

3. **Sean x y l un punto y una recta respectivamente en un plano proyectivo P1 y suponemos que la recta l pasa por el punto x, es decir $l'^T x = 0$. Sean x' y l' un punto y una recta del plano proyectivo P' donde al igual que antes $l'^T x' = 0$. Supongamos que existe un homografía de puntos H entre ambos planos proyectivos, es decir $x' = Hx$. Deducir de las ecuaciones anteriores la expresión para la homografía G que relaciona los vectores de las rectas, es decir G tal que $l' = Gl$. Justificar la respuesta**

4. **Suponga la imagen de un plano en donde el vector l=(l 1 ,l 2 ,l 3) representa la proyección de la recta del infinito del plano en la imagen. Sabemos que si conseguimos aplicar a nuestra imagen una homografía G tal que si $l' = Gl$,siendo $l'T = (0, 0, 1)$ entonces habremos rectificado nuestra imagen llevándola de nuevo al plano afín. Suponiendo que la recta definida por l no pasa por el punto (0,0) del plano imagen. Encontrar la homografía G. Justificar la respuesta**

5. **Identificar los movimientos elementales (traslación, giro, escala, cizalla, proyectivo) representados por las homografías H1, H2, H3 y H4:**

$$H1 = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0{,}5 & 0 & 0 \\ 0 & 0{,}3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.1}$$

$$H2 = \begin{pmatrix} 0 & 1 & -3 \\ -1 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.2}$$

$$H3 = \begin{pmatrix} 1 & 0{,}5 & 0 \\ 0{,}5 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \tag{5.3}$$

$$H4 = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \tag{5.4}$$

5.0.1. H1

tenemos una translación, un escalado y un movimiento afín. La primera matriz

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.5}$$

representa una translación, con $t_x = 3$ y $t_y = 5$, ya que tiene la forma:

$$\begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \tag{5.6}$$

La segunda matriz, representa un escalado, con $S_x = 0{,}5 \neq 0$ y $S_y = 0{,}3 \neq 0$, ya que es de la forma:

$$\begin{pmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.7}$$

Por último, la tercera matriz representa un movimiento afín, con $C_x = 3$ y $C_y = 0$, mediante el cual añade deformación(pierde ángulos), aunque mantiene las rectas. Creo que es un movimiento afín por ser de la forma:

$$\begin{pmatrix} 1 & C_x & 0 \\ C_y & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.8}$$

descartamos la posibilidad de que pudiera ser una rotación, dado que aunque $\begin{vmatrix} 1 & 1 \\ 0 & 3 \end{vmatrix} = 1$, tenemos que $R \cdot R^T \neq id$

5.0.2. H2

La primera matriz:

$$\begin{pmatrix} 0 & 1 & -3 \\ -1 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.9}$$

representa un giro+translación, con $\cos\theta = 0$ y $\sin\theta = 1$ ya que es de la forma:

$$\begin{pmatrix} \cos\theta & \sin\theta & t_x \\ -\sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \tag{5.10}$$

y cumple que $\begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} = 1$ y que $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)^T = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) = id.$

La segunda matriz:

$$\begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.11}$$

representa una transformación afín con $C_x = 0$ y $C_y = 2$.

5.0.3. H3

La primera matriz:

$$\begin{pmatrix} 1 & 0{,}5 & 0 \\ 0{,}5 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.12}$$

representa un movimiento afín, con $C_x = 0{,}5$ y $C_y = 0{,}5$, mientras que la segunda matriz:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \tag{5.13}$$

representa una transformación proyectiva, ya que modifica los puntos " introduciendo perspectiva " .

5.0.4. H4

Por último, la matriz:

$$H4 = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \tag{5.14}$$

que es de la forma:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

(observamos que su última fila es $\neq$ (0,0,1)) aplica una transformación proyectiva sobre los puntos de la imágen, Por tanto, de haber puntos en el infinito etos irían a puntos de fuga.

6. ¿Cuáles son las propiedades necesarias y suficientes para que una matriz defina una homografía entre planos? Justificar la respuesta

Una homografía H, es: $u' = H \cdot u$
tal que si multiplicamos un vector de puntos (u) por H, obtenemos un vector nuevo de puntos (u'), calculados a partir del anterior, aplicándole una serie de transformaciones.
Hablemos de sus propiedades más imporantes. Sabemos que todo movimiento puede ser representado con varios puntos, sin embargo, si vamos a calcular un homografía, necesitaremos que esos puntos estén en correspondencias en las dos imágenes, o la homografía no será correcta.

Otro detalle es que, a pesar de que los puntos pueden ser cualesquiera, no deberá de haber más de dos alineados si queremos obtener un resultado (más o menos) bueno.
Además, una homografía debe tener tamaño 3x3 y debe de tener det $\neq 0$, ya que debe tener inversa, pues igual que es posible pasar de una imágen1 a una imágen2 a través de H, tiene que ser posible pasar de una imágen2 a la imágen1, a través de H^{-1}.

7. ¿Qué propiedades de la geometría de un plano quedan invariantes si se aplica una homografía general sobre él? Justificar la respuesta.

Una homografía general pueden ser varias transformaciones, no es una transformación geométrica concreta. Por lo tanto, hay propiedades como la proyección o el paralelismo que según la homografía, podrían o no conservarse.
Ésto nos lleva a pensar ¿entonces qué se conserva?
Resulta evidente que aplicándole una transformación a una imágen, lo que estaba alineado en la imágen original lo sigue estando en la transformada. Por tanto , ahí tenemos algo que se conserva: **las rectas**.
Si tenemos que unas rectas se intersecaban en la imágen original, entonces también lo harán en la transformada.
Algo parecido ocurre con **las distancias**, que también se conservan, de tal forma que lo que estaba a cierta distancia en la imágen original sigue estando a esa cierta distancia en la transformada.

8. ¿Cuál es la deformación geométrica más fuerte que se puede producir sobre la imagen de un plano por el punto de vista de la cámara? Justificar la respuesta.

desde el punto de vista de la cámara , la deformación **geométrica** más fuerte que se puede producir yo creo que podría ser posicionar la cámara paralela a una esquina, de tal forma que, en función de la inclinación de la cámara se pueda apreciar una recta más o menos gruesa de píxeles del tamaño de la diagonal de la imágen, perdiendo así el paralelismo de la imágen completamente.

9. ¿Qué información de la imagen usa el detector de Harris para seleccionar puntos? ¿El detector de Harris detecta patrones geométricos o fotométricos? Justificar la contestación.

El detector de Harris selecciona puntos basándose en el gradiente.

$$H = \begin{pmatrix} \sum_w (\frac{\partial I}{\partial x})^2 & \sum_w (\frac{\partial I}{\partial x})(\frac{\partial I}{\partial y}) \\ \sum_w (\frac{\partial I}{\partial x})(\frac{\partial I}{\partial y}) & \sum_w (\frac{\partial I}{\partial y})^2 \end{pmatrix}$$

donde las derivadas son invariantes a giros. Se fija en las regiones donde hay cambios fotométricos (más variación luminosa) y cambios en la dirección. Para ello usa dos valores ortogonales que nos indican cuando la variación del gradiente es máxima, permitiendo así detectar las esquinas.
El detector de Harris es monoescala, aunque se pueden hacer múltiples escalas y combinarlas poseriormente.
Además el detector de Harris detecta **patrones fotométricos**, pues como hemos dicho antes, se basa en

la magnitud del gradiente, y sin embargo no detecta patrones geométricos, por lo que no nos da la información que podría obtener de alrededor de los puntos(no hace extracción de información) , a diferencia de otros detectores.

10. ¿Sería adecuado usar como descriptor de un punto Harris los valores de los píxeles de su región de soporte? En caso positivo identificar cuando y justificar la respuesta

Yo creo que sí, ya que los valores de los píxeles al fin y al cabo son coordenadas de puntos que me van a indicar si en una región dada de la imágen los puntos son relevantes o no lo son.
Para ello, habrá que efectuar la suma sobre toda la región (donde tendremos hacia dónde apunta la dirección más importante), eso sí, teniendo en cuenta que no haya variaciones en la dirección.
Como la suma se hace en toda la ventana, será necesario ,aparte de dar las coordenadas de los puntos, también dar la escala.

11. ¿Qué información de la imagen se codifica en el descriptor de SIFT? Justificar la contestación.

Sift es una técnica de detección, pero también de extraccción de información, a diferencia de los puntos Harris. Se trata de un descriptor también basado en gradientes, que no se basa ni en niveles de gris ni en intensidad luminosa.
Lo que Sift hace es : teniendo los vectores directores de los gradientes para cada píxel, toma las direcciones en las que el gradiente es mayor y con eso, construye un hibstograma. Para construir el hibstograma, vota a los números vecinos , siguiendo algún criterio, como el vecino más cercano, por ejemplo.
Teniendo en cuenta que hay 16 hibstogramas en total y que cada una tiene 8 números, entonces tendremos que el descriptor Sift viene dado por 128 números, que son la información que se extrae para describir la región.

12. Describa un par de criterios que sirvan para establecer correspondencias (matching) entre descriptores de regiones extraídos de dos imágenes. Justificar la idoneidad de los mismos

En prácticas hemos usado dos: **fuerza bruta+cruce**, que lo que hace es comparar todos con todos; comparando todos los descriptores de una imágen con el descriptor de un punto de la otra (para cada punto), y quedarnos con los mejores. La opción de cruce además lo que añade es que se compare izquierda con derecha y derecha con izquierda.
Otro matcher que hemos usado en pŕacticas ha sido **Flann**, que son las siglas de librería rápida para aproximar vecinos cercanos.
Ésta librería contiene varios algoritmos optimizados, que se pueden seleccionar a la hora de hacer el matching. Pero lo importante, es que aproxima vecinos cercanos, es decir, que en lugar de comparar todos

con todos como el anterior, utiliza los descriptores que están más cerca. Para establecer qué es estar cerca y qué no, en clase se dijo que se podía establecer un umbral ϵ tal que si $d(x^{'}, Hx) < \epsilon$ se acepta y sino se descarta.

13. Cual es el objetivo principal en el uso de la técnica RANSAC. Justificar la respuesta

El objetivo principal en el uso de RANSAC es determinar qué puntos, de entre todos los que ya tenemos, son los revelantes/correctos para nuestro problema. Es decir, hacer la discriminación es función de lo que le indiquemos que queremos estimar.
Para lograrlo, lo que hace es muestrear y luego propone varias homografías y les pregunta al resto de puntos, que votarán la que crean que es mejor. Finalmente, RANSAC devuelve la homografía que ha recibido más votaciones.
Eso sí, hay que garantiar de que en toda la trayectoria hay, al menos, una H sin error.
Por otra parte, tenemos que al discriminar RANSAC elimina puntos que no considera que son buenos pero que quizás si lo son. Para intentar arreglar esto, se podría volver a lanzar todos los puntos descartados por la homografía, en lugar de por el matcher usado antes(por ej, fuerza bruta) , pudiendo así recuperar algunos puntos qe poder añadir a la lista de válidos.

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 mínimo de puntos en correspondencias necesarios para montar un mosaico? Justificar la respuesta

Creo que en ese caso, el número mínimo de puntos en correspondencias, aunque las imágenes estén solapadas, debe de ser 4 , por cada pareja de imágenes. Ésto es porque, como bien se sabe, 4 es el mínimo de puntos que deben darse **en correspondencias** para poder calcular una homología válida. Por tanto, en este caso, si hay 4 imágenes necesitaríamos 12 puntos en correspondencias , 4 para cada homología, suponiendo que la imágen 4 no engancha con la imágen 1.

15. En la confección de un mosaico con proyección rectangular es esperable que aparezcan deformaciones de la realidad. ¿Cuáles y porqué?.¿Bajo qué condiciones esas deformaciones podrían desaparecer? Justificar la respuesta

Teniendo en cuenta que en la confección de un mosaico se realiza la fusión geométrica y no fotométrica (al menos nosotros la fotométrica no la hacemos), tendremos que una posible deformidad de la realidad será que las imágenes tendrán distintas tonalidades, pues no suele ser que se tomen todas las fotos el mismo día y bajo las mismas condiciones de luminosidad.

Otra posible deformación de la realidad que se me ocurre es que se pierde el paralelismo real que hay en la imágen, pues al adaptarlas para que encajen entre sí estamos cambiando la proyectividad de estas, por lo que se pueden ver figuras más alargadas o rectangulares etc que en las imágenes originales.