

VC: Cuestionario 2

Gustavo Rivas Gervilla



1. Identificar la/s diferencia/s esencial/es entre el plano afín y el plano proyectivo. ¿Cuáles son sus consecuencias? Justificar la contestación.

Como dice en algunos de los enlaces que se citan a continuación, el plano proyectivo se forma añadiendo al espacio afín los puntos del infinito, de hecho aquí creo que está muy bien explicado el significado de estos puntos del infinito.

En consecuencia lo que ocurre aquí es que perdemos el paralelismo ya que al hacer una proyección añadimos el punto en el infinito creando "la ilusión" de que dos rectas que en la realidad son paralelas se han cortado. Esto se ve muy claro en la tabla de la jerarquía de transformaciones de [Szel. pág 39] donde se ve que lo que se pierde de una transformación afín a una proyectiva es el paralelismo.

<https://loshijosdelagrangre.files.wordpress.com/2013/04/tema-3.pdf>

<http://www.rac.es/ficheros/doc/00894.pdf>

http://isa.umh.es/asignaturas/vc/13-Vision3D_Estereo.pdf

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 ($l = x \wedge x'$). De igual modo el punto de intersección de dos rectas l y l' está dado por $x = l \wedge l'$.

Como podemos leer en [Hart.] y como hemos visto en teoría, en coordenadas homogéneas un punto del plano viene dado por $x = (x_1, x_2, 1)$ pues entonces tenemos lo siguiente:

$x = (x_1, x_2, 1)$ y $x' = (x'_1, x'_2, 1)$ si calculamos ahora el producto vectorial de estos vectores de la forma usual tenemos que $x \wedge x' = (x_2 - x'_2, x'_1 - x_1, x_1x'_2 - x'_1x_2)$. Inicialmente pensé que no podría demostrar lo que me pedía el ejercicio con este cálculo, ya que se espera obtener el vector $(x_1 - x'_1, x_2 - x'_2, 0)$ que es lo usual en la geometría **euclídea**). Pero si recordamos algo de la teoría dada o bien vemos el enunciado de la cuestión 3 nos damos cuenta de que sí que podemos demostrar el resultado con este cálculo.

Lo único que tenemos que notar es que dado el vector que representa una recta, l y un punto x entonces el punto está en la recta sii $l^T x = 0$ pues bien es claro que si la recta la representamos por el vector $x \wedge x'$ ambos puntos están en esa recta ya que el producto vectorial de dos vectores (que es como se representan los puntos y las rectas en homogéneas) es perpendicular a ambos vectores. Con lo que l viene dado por ese vector c.q.d..

Del mismo modo podemos demostrar la segunda parte de esta pregunta ya que ahora el punto $x = l \wedge l'$ es perpendicular a los vectores l y l' con lo que $l^T x = 0 = l'^T x$ con lo que efectivamente x es un punto que está en ambas rectas.

Podemos encontrar también una demostración igual en [Hart. pág 27].

3. Sean x y l un punto y una recta respectivamente en un plano proyectivo P 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 una 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 t.q. $l' = Gl$. Justificar la respuesta.

Vamos a suponer que alguno de los dos puntos tiene norma no nula, esto es así porque de otro modo no podremos hacer el despeje que usaremos a continuación, además, tal y como vemos que se caracteriza la pertenencia de un punto a una recta, si ambos puntos fuesen el $(0,0,0)$ no tendríamos información suficiente pues todas las rectas pasan por este punto. Entonces supuesto esto y observando que los papeles de las parejas de recta y punto son intercambiables procedemos a la demostración:

$$l'^T x' = l'^T Hx = 0 = l^T x$$

Entonces si multiplicamos ahora en ambos extremos de la igualdad, escalarmente, por x^T tenemos lo siguiente:

$$l'^T H \|x\|^2 = l^T \|x\|^2$$

donde $\|\bullet\|$ es la norma euclídea, y por lo que hemos supuesto anteriormente podemos despejarla de ambos miembros, entonces:

$$l'^T H = l^T; \text{ trasponiendo tenemos } H^T l' = l;$$

y ahora ya sólo nos queda por despejar H^T :

$$l' = (H^T)^{-1} l; \text{ con lo que } G = (H^T)^{-1}.$$

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.

Nosotros podemos ver en [Hart. pág49] una idea pero esto lo que hace es encontrar una homografía entre puntos, de modo que cuando calculamos la homografía entre rectas, que como probamos en la cuestión 3, es H^{-T} entonces salga la homografía adecuada, puede comprobarse fácilmente que la homografía H que en el libro se muestra es correcta y solamente habría que hacer algunos cálculos más para salvar digamos las transformaciones que sufre H para aplicarse a rectas.

Ahora bien, como el enunciado simplemente pide una G directamente que cumpla la ecuación que en él se plantea nosotros vamos a hacer los cálculos para calcularla sin tener en cuenta lo que pone en el libro:

Lo primero que observamos es que como la recta por hipótesis no pasa por el punto $(0,0)$ entonces $l_3 \neq 0$ que es algo que vamos a necesitar en los cálculos que vamos a hacer a continuación:

Por comodidad vamos a buscar una homografía H t.q. se cumpla la siguiente ecuación:

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix}$$

Claramente operando tenemos que en principio nos valdría cualquier matriz que fuese de la forma siguiente:

$$\begin{pmatrix} x_{11} & x_{12} & l_1 \\ x_{21} & x_{22} & l_2 \\ x_{31} & x_{32} & l_3 \end{pmatrix}$$

Ahora bien nosotros demostraremos en la cuestión 6 que para que una matriz defina una homografía, esta matriz ha de ser regular, es decir, tiene que tener determinante distinto de cero. Con lo cual vamos a tomar la matriz más sencilla que cumple esto y que tiene la estructura anterior:

$$A = \begin{pmatrix} 1 & 0 & l_1 \\ 0 & 1 & l_2 \\ 0 & 0 & l_3 \end{pmatrix}$$

Claramente $|A| = l_3 \neq 0$, pues bien tenemos que $Al' = l$ (suponemos que l está en formato columna y no fila como en el enunciado para que la notación sea algo más homogénea) y por lo tanto $l' = A^{-1}l$, por tanto $A^{-1} = G$ y ya sólo nos queda calcular tal inversa:

Calculamos la matriz conjugada de A :

$$A^* = \begin{pmatrix} l_3 & 0 & 0 \\ 0 & l_3 & 0 \\ -l_1 & -l_2 & 1 \end{pmatrix}$$

Transponiendo esta matriz y multiplicándola por $\frac{1}{l_3}$ tenemos finalmente que:

$$G = \begin{pmatrix} 1 & 0 & \frac{-l_1}{l_3} \\ 0 & 1 & \frac{-l_2}{l_3} \\ 0 & 0 & \frac{1}{l_3} \end{pmatrix}$$

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

$$H1 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

En esta primera homografía podemos identificar muy fácilmente las transformaciones que la componen ya que es la composición de tres matrices que tienen la estructura de un movimiento reconocible, así leyendo de derecha a izquierda (que es cómo aplicaremos las transformaciones a los puntos afectados) tenemos:

1. Una cizalla que afecta a la x por medio de la y y que deja invariante la y.
2. Un escalado que divide por dos la x y que multiplica por 1/3 aprox. la y.
3. Una traslación por medio del vector (3,5).

$$H2 = \begin{bmatrix} 0 & 1 & -3 \\ -1 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Para ver aquí los movimientos que tenemos vamos a descomponer la matriz un poco más:

$$H2 = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Para hacer esto lo único que tenemos que hacer es buscar los patrones de los movimientos conocidos dentro de las matrices que nos dan y usar que el producto de matrices es asociativo, algunas veces tendremos que reajustar algún valor (no vale con descomponer en dos simplemente los valores dados) para que el producto resulte correcto. Entonces ya hecha esta descomposición lo único que nos queda es decir, como antes, qué movimientos tenemos:

1. Cizalla que deja la x invariante y a la y' la pone como x+y.
2. Escalado que duplica el tamaño.
3. Giro de $\frac{3\pi}{2}$.
4. Traslación por el vector (-3,2).

$$H3 = \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Para encontrar la descomposición de la primera matriz por movimiento elementales lo que hacemos es extraer la matriz de escalado que veremos a continuación y resolver unas ecuaciones triviales para saber el valor de los elementos de la otra matriz en la que descompondremos esta primera matriz. La otra matriz dado que la última fila es distinta al (0 0 1) entonces ya sabemos que se trata de un movimiento proyectivo y no podemos extraer de ella ningún otro movimiento elemental excepto si queremos la identidad (el escalado con factor 1 en ambas coordenadas).

$$H3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & 0 \\ 0.25 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Con lo cual con esta homografía aplicamos los siguientes movimientos elementales:

1. Un movimiento proyectivo.
2. Una cizalla tal que $x' = x + 0.5y$ y $y' = 0.25x + y$.
3. Un escalado que deja invariante la x y duplica la y .

$$H4 = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

Lo primero que vamos a hacer es extraer la parte proyectiva, con lo que tendremos la siguiente descomposición en la que aparecerán unos términos incógnita que son fácilmente despejados:

$$H4 = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

Despejando tenemos:

$$H4 = \begin{bmatrix} 2 & -\frac{3}{2} & \frac{3}{2} \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

Y ahora vamos a descomponer la otra matriz que ya tiene la última fila igual a (0 0 1) y lo hacemos poco a poco del mismo modo que lo hemos hecho en la paso anterior, obteniendo finalmente:

$$H4 = \begin{bmatrix} 1 & 0 & \frac{3}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{3}{5} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

Con lo que se aplican los siguientes movimientos:

1. Un proyectivo.
2. Un escalado que duplica la x y multiplica la y por $\frac{5}{2}$.
3. Una cizalla que sólo deja invariante a la y .
4. Una traslación por el vector $\left(\frac{3}{2}, -\frac{1}{2}\right)$

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

Esta presentación ha sido muy útil pues hace un resumen de lo visto en clase y habla de cómo las homografías se representan por medio de una matriz regular con 8 grados de libertad.

Veamos por qué esto, lo primero es que gracias a la introducción de las coordenadas homogéneas podemos hacer que toda transformación proyectiva sea lineal (el ejemplo más claro se ve en cómo se beneficia la translación de esto).

Ahora veamos la definición de homografía que se da aquí lo que tenemos es una aplicación biyectiva entre dos espacios proyectivos que viene inducida por un isomorfismo entre los espacios vectoriales asociados a esos espacios proyectivos.

Entonces ahora que sabemos esto lo único que nos queda es ver cómo ha de ser una matriz para que la aplicación sea biyectiva, y la respuesta es clara (y también se menciona en el enlace anterior) la matriz ha de ser invertible (sería inyectiva pero al estar entre espacios de la misma dimensión también es sobreyectiva).

Además una homografía o proyección conserva la colinealidad, esto es, lleva puntos alineados en puntos alineados. Pero si tenemos la matriz que representa a la aplicación es claro que se preserva mediante ella la colinealidad, una demostración de esto podemos verla en las diapositivas del enlace que adjuntamos al final de la pregunta. Lo único que se hace es algo parecido a lo que hemos hecho en el ejercicio 2:

Dado un punto x que pertenezca a una línea l entonces como sabemos se tiene que $l^T x = 0$ entonces ahora $l^T H^{-1} Hx = 0$ en consecuencia el punto imagen Hx pertenece a la recta imagen de l por lo demostrado en la cuestión 3.

Nos quedaría por ver que toda proyección puede expresarse como una matriz, para ello lo único que tendríamos que ver es que tomados suficientes puntos podemos tomar un sistema de ecuaciones para estimar la matriz de homografía. Señalemos que decimos estimar, ya hemos visto en prácticas que usamos mínimos cuadrados para poder calcular una aproximación a la solución. Por tanto yo diría que podemos calcular cualquier homografía sin más que resolver un sistema en base a imágenes de suficientes puntos pero no puedo probarlo

formalmente.

Con lo que toda matriz regular (3x3 ya que estamos trabajando con planos) define una homografía entre planos y viceversa que es lo que queríamos obtener.

<http://www.umiacs.umd.edu/~ramani/cmssc828d/ProjectiveGeometry.pdf>

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

Serán aquellas que se mantengan por medio de una aplicación proyectiva "pura", podemos ver en [Hart.] que lo que se conserva es la concurrencia, la colinealidad, el orden de contacto (intersección, tangencia e inflexión), las discontinuidades tangenciales y el ratio cruzado de 4 puntos.

Como sabemos una homografía por definición conserva la colinealidad una demostración de este hecho la podemos encontrar en la cuestión 6 (una vez supuesto que cualquier homografía es una matriz regular). Por lo tanto tenemos que efectivamente esta propiedad se conserva.

La concurrencia es también muy sencilla de probar debido a lo anterior, ya que si dos rectas tienen un punto común entonces por la colinealidad tenemos que el transformado de ese punto estará en las transformadas de ambas rectas. Y el orden de contacto también es sencillo.

Veamos ahora que se conserva el cociente cruzado para ello vamos a resolver un problema que se plantea en [Hart. pág.63]:

El cross ratio de cuatro puntos se define como sigue:

$$Cross(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4) = \frac{|\bar{x}_1\bar{x}_2||\bar{x}_3\bar{x}_4|}{|\bar{x}_1\bar{x}_3||\bar{x}_2\bar{x}_4|}$$

donde:

$$|\bar{x}_i\bar{x}_j| = \det \begin{pmatrix} x_{i1} & x_{j1} \\ x_{i2} & x_{j2} \end{pmatrix}$$

La pista que nos da el ejercicio para abordar el problema es que si tenemos las imágenes de dos puntos expresadas como $\bar{x}'_i = \lambda_i H_{2 \times 2} \bar{x}_i$ y $\bar{x}'_j = \lambda_j H_{2 \times 2} \bar{x}_j$ donde la igualdad no depende de escalas entonces podemos probar que $|\bar{x}'_i\bar{x}'_j| = \lambda_i \lambda_j \det H_{2 \times 2} |\bar{x}_i\bar{x}_j|$. Veámoslo:

Tomamos $H = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ entonces $\lambda_i \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix} = \begin{pmatrix} \lambda_i a x_{i1} + \lambda_i b x_{i2} \\ \lambda_i c x_{i1} + \lambda_i d x_{i2} \end{pmatrix}$. Con lo cual:

$$|\bar{x}'_i\bar{x}'_j| = \begin{vmatrix} \lambda_i a x_{i1} + \lambda_i b x_{i2} & \lambda_j a x_{j1} + \lambda_j b x_{j2} \\ \lambda_i c x_{i1} + \lambda_i d x_{i2} & \lambda_j c x_{j1} + \lambda_j d x_{j2} \end{vmatrix} = \lambda_i \lambda_j [a c x_{i1} x_{j1} + a d x_{i1} x_{j2} + b c x_{i2} x_{j1} + b d x_{i2} x_{j2} - (a c x_{j1} x_{i1} + a d x_{j1} x_{i2} + b c x_{i1} x_{j2} + b d x_{i2} x_{j2})] = \lambda_i \lambda_j [(a d -$$

$$bc)(x_{i1}x_{j2}) + (bc - ad)(x_{i2}x_{j1})] = \lambda_i \lambda_j [(ad - bc)(x_{i1}x_{j2} - x_{i2}x_{j1})] = \lambda_i \lambda_j \det H_{2 \times 2} [\bar{x}_i \bar{x}_j] \text{ c.q.d..}$$

Entonces ahora que hemos probado esto acabamos la demostración de que el cross ratio se conserva mediante homografías:

$$\begin{aligned} Cross(\bar{x}'_1, \bar{x}'_2, \bar{x}'_3, \bar{x}'_4) &= \frac{|\bar{x}'_1 \bar{x}'_2| |\bar{x}'_3 \bar{x}'_4|}{|\bar{x}'_1 \bar{x}'_3| |\bar{x}'_2 \bar{x}'_4|} = \\ \frac{\lambda_1 \lambda_2 \det H_{2 \times 2} [\bar{x}_1 \bar{x}_2] \lambda_3 \lambda_4 \det H_{2 \times 2} [\bar{x}_3 \bar{x}_4]}{\lambda_1 \lambda_3 \det H_{2 \times 2} [\bar{x}_1 \bar{x}_3] \lambda_2 \lambda_4 \det H_{2 \times 2} [\bar{x}_2 \bar{x}_4]} &= \frac{|\bar{x}_1 \bar{x}_2| |\bar{x}_3 \bar{x}_4|}{|\bar{x}_1 \bar{x}_3| |\bar{x}_2 \bar{x}_4|} = Cross(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4) \end{aligned}$$

Con lo que tenemos lo que queríamos.

En cuanto a las discontinuidades tangenciales no sé a qué se refiere.

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.

Como se menciona en [Hart. pág 25] las transformaciones proyectivas modelan la distorsión geométrica que se produce sobre un plano al ser visto desde la perspectiva de la cámara y este es el enfoque que vamos a considerar para constestar a esta pregunta. Es decir lo que nos tenemos que preguntar es cómo afecta una aplicación proyectiva a la geometría de un plano y en consecuencia qué es lo más fuerte que podemos hacerle al plano por medio de una transformación proyectiva.

Para ello vamos a ir subiendo en la jerarquía de transformaciones proyectivas que se muestra en [Hart. pág 38 y sucesivas]:

Lo primero que tenemos son las isometrías que como sabemos son las transformaciones que más propiedades conservan; las que más invariantes tienen.

Luego tenemos las semejanzas, en las cuales se conservan los ángulos. Aquí como se incorpora un coeficiente de escala isotrópico. En cambio aquí ya perdemos la distancia entre puntos ya que estamos aplicando un escalado, pero no perdemos los ratios entre longitudes claro.

Pasamos ahora a las transformaciones afines que es una transformación lineal no singular seguida de una traslación. Estas transformaciones pueden verse como dice en [Hart.] como la composición de una rotación, un escalado no homogéneo, deshacer la rotación anterior y hacer otra rotación. Pues bien cómo la única diferencia con las semejanzas es este escalado no homogéneo lo que ocurre es que perdemos dos invariantes que son los ratios de longitud y los ángulos.

Llegamos finalmente a las transformaciones proyectivas propiamente dichas (que como sabemos engloban a todas las demás, en esta jerarquía cada nivel engloba al anterior) y con ellas llegamos a la transformación geométrica más fuerte que puede sufrir un plano por el punto de vista de la cámara.

Mientras que en las transformaciones afines la orientación de una línea transformada dependía sólo de cómo estuviese orientada en el plano original, en una transformación proyectiva genérica ésta depende también de la posición de la línea en el plano original y lo mismo ocurre con la escala del área, por esto por ejemplo vemos un cuadrado más pequeño cuánto más alejado está.

Pero la diferencia principal entre una aplicación afín y una proyectiva "pura" es el efecto que tiene esta sobre los puntos del infinito, veamos qué sucede al aplicarle una transformación afín y una proyectiva a un punto del infinito. Para ello usaremos la notación por bloques que se usa en [Hart.] para describir las matrices de las transformaciones:

$$\begin{pmatrix} A & t \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} A & t \\ v^T & v \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ v_1 x_1 + v_2 x_2 \end{pmatrix}$$

Con lo que la deformación geométrica más fuerte es la pérdida de paralelismo.



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.

Normalmente el detector de Harris se conoce como un detector de esquinas y esto es así porque usualmente los puntos seleccionados por él son esquinas, son puntos donde hay variaciones de intensidad significable en todas las direcciones.

Pues bien el detector de Harris usa para detectar puntos la intensidad lumínica (o el color) de la imagen, ya que lo que hace es ver al mover la "ventana de análisis" el error cuadrático que se produce, ahora bien, en lugar de calcular los valores propios de la matriz H que se construye (con el gradiente de la función de intensidad I sobre la imagen) hace una cuenta con la traza y el determinante de esta matriz que da un resultado muy similar y es menos costoso.

Yo creo que el detector de Harris detecta patrones fotométricos dado que es sensible a cambios de escala, de modo que para una misma curva dependiendo de su escala puede clasificar a un punto como esquina o como borde y al fin y al cabo con la información con la que trabaja es con cambios en la intensidad.

Además de la bibliografía de la asignatura hemos consultado: https://es.wikipedia.org/wiki/Detector_de_esquinas#El_algoritmo_Harris_.26_Stephens_.2F_Plessey_.2F_el_Shi-Tomasi_detector_de_esquinas

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

Es claro que los datos que han hecho de un punto Harris ser precisamente un punto Harris están recogidos en su región de soporte, es decir, podríamos volver a computar lo que calcula para determinar si un punto es Harris o no.

Ahora bien lo que se pretende con un descriptor es al fin y al cabo establecer correspondencias, la principal motivación de la extracción de características es poder comparar distintas regiones y ver cuáles se parecen, al menos esta es la utilidad que nosotros le hemos dado en la asignatura.

Entonces convendría recordar lo que se busca cuando se construye un descriptor, es decir, cuándo es verdaderamente un descriptor bueno. En las transparencias podemos ver que las características más importantes de un descriptor son la invarianza frente a transformaciones y el nivel de discriminación que se puede conseguir por medio del descriptor.

Entonces lo que vamos a hacer es ver cómo sería esto para un descriptor basado en la región de soporte de un punto Harris (realmente lo que estamos diciendo es que primero aplicamos el filtro Harris y después a los puntos señalados por él les asociamos como descriptor los valores de su región de soporte).

La primera cosa a señalar como buena de este descriptor es que es un vector de $n \times n$ componentes siendo $n \times n$ la dimensión de la región de soporte con la que aplicamos Harris (suponemos que será cuadrada como lo es usualmente). Por lo tanto al ser un vector podemos calcular la distancia euclídea a la hora de hacer comparaciones algo que es positivo.

En cuanto a la invarianza sabemos por las transparencias que los puntos Harris son invariantes a rotaciones con lo cual esto no plantea ningún problema. También tenemos que son parcialmente invariantes para cambios en la intensidad, lo que nos lleva al primer caso donde este descriptor no funcionaría y es cuando el cambio en la intensidad de la imagen (debido por ejemplo a cambios en la luminosidad de la escena) sean muy bruscos entonces ya este descriptor no encontraría matches que sí que son buenos.

Donde más sufren los puntos Harris es en los cambios de escala por tanto cuando tenemos una imagen a la que se le ha aplicado un zoom es muy probable que el descriptor no nos sirva para establecer matches con otra imagen de la misma

escena con otro zoom. Lo que se hace para evitar este problema es ir cambiando la escala (esto se puede hacer por medio de una pirámide Gaussiana) pero entonces ya tendríamos que modificar nuestro descriptor para añadir la escala (o el nivel) en el que se ha localizado dicho punto Harris.

Y en cuanto a la discriminación que establece eso dependerá de la imagen, si no hay demasiadas regiones con la misma composición de color (que es al fin y al cabo lo que estamos almacenando en nuestro descriptor), porque nosotros distinguimos "esquinas" pero si hay varias "esquinas similares" pues no será un buen descriptor.

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

Lo que hacemos es tomar una región de 16x16 píxeles al rededor del punto en cuestión, dividimos esta región en subregiones de 4x4 y calculamos en cada una un histograma de direcciones del siguiente modo: dividimos las posibles direcciones en octantes, entonces en cada píxel calculamos la dirección y magnitud del gradiente y sumamos la magnitud en el octante correspondiente, obteniendo así 16 histogramas. Teniendo una información muy buena de cómo varía la intensidad al rededor del punto de interés que estemos estudiando en ese momento.

Esto para dar mayor robustez al descriptor se hace con un poco más de procesamiento sobre lo calculados anteriormente: lo primero es que para evitar que los gradientes de puntos alejados del centro tenga el mismo peso lo que hacemos es darles pesos a sus magnitudes en base a una Gaussiana y así evitamos el problema que estos presentan y es que son más vulnerables a errores en los datos. Además para evitar otros errores sobre la dirección dominante lo que hacemos es usar interpolación trilineal para sumar las magnitudes de los píxeles.

Además para reducir los efectos de contraste lo que hacemos es normalizar el vector que conforma el descriptor. Y para hacer este descriptor aún más robusto ante otras variaciones fotométricas recortamos los valores a 0.2 y volvemos a normalizar el vector de 128 dimensiones.

Hemos

consultado [Szel. pág. 223] y el siguiente enlace: https://en.wikipedia.org/wiki/Scale-invariant_feature_transform#Keypoint_descriptor

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.

Lo primero que se asume en [Szel. pág. 226] es que los descriptores extraídos de los puntos de interés están preparados para poder aplicar sobre ellos la distancia euclídea a modo de cálculo para poder comparar y encontrar los matches.

En esta misma página se describe la estrategia de matching más simple que es la que se conoce como *fixed threshold* que consiste en simplemente fijar un umbral y

todos aquellas comparaciones (entre parejas de descriptores de ambas imágenes) que queden por debajo (la distancia euclídea se entiende) los tomamos como matches correctos. Evidentemente si ponemos el umbral muy alto daremos lugar a muchos falsos positivos, y si es muy bajo aparecerán muchos falsos negativos con lo que fijar el umbral adecuado es muy importante y puede resultar muy complicado. Este mismo problema se pone de manifiesto en las transparencias de clase, incluso una mejor suya que es usar ratios para ver cuál de las diferencias (en distancia euclídea) es mejor puede dar lugar a falsos positivos. Otra cosa a señalar es que un mismo descriptor puede tener varios matches lo que complica la elección.

Entonces vamos a ver otra estrategia que es la de considerar sólo los dos matches más cercanos (NNDR) y hacer el cociente entre estos dos de modos que si éste queda por debajo de un valor entonces tomamos esta pareja. Este criterio mejora mucho el anterior puesto que en primer lugar sólo da un match por descriptor y además al seleccionar sólo las parejas con una distancia por debajo de un umbral entonces disminuye el número de falsos positivos. Además es más estable a variaciones por transformaciones de la imagen. Y mejora la precisión pues penaliza a aquellos descriptores que tienen varios candidatos muy similares en distancia. El problema que tiene este criterio es cuando el conjunto de descriptores con el que comparar es muy grande.

Aquí tenemos un artículo que aunque se centra en la comparación de distintos descriptores también da algunas ideas sobre cómo funcionan distintas técnicas de matching (y es el que hemos usado para contestar esta pregunta): https://www.robots.ox.ac.uk/~vgg/research/affine/det_eval_files/mikolajczyk_pami2004.pdf. En él podemos ver, al igual que en [Szel. pág 231], una gráfica mostrando cómo funcionan estos criterios con varios descriptores.

13. ¿Cuál es el objetivo principal en el uso de la técnica RANSAC? Justificar la respuesta.

Como sabemos el algoritmo básico para poner dos imágenes en relación y, por ejemplo, formar un mosaico es: extraer características de ambas imágenes (los puntos de interés), emparejar dichas características entre ambas imágenes y después calcular la homografía con dichos emparejamientos.

¿Cuál es el problema? Pues que nuestros algoritmos por muy buenos que sean pueden producir fallos, no son exactos y en consecuencia pueden darse emparejamientos que no sean ciertos, lo que se llama en inglés *outliers*.

Entonces supongamos que ahora queremos calcular la homografía pero que tenemos estos emparejamiento falsos de por medio, entonces si usamos mínimos cuadrados la recta calculada intentará aproximar lo mejor posible a **todos** los puntos que le demos, produciendo una homografía que no es del todo correcta, puesto que la recta obtenida por mínimos cuadrados no se es correcta.

Para esto aparece el algoritmo de RANSAC un algoritmo que trata de ajustar una recta pero de modo que sólo se tengan en cuenta emparejamientos buenos,

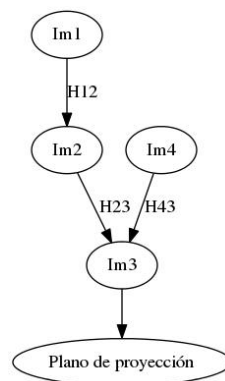
los que llamamos *inliers*.

Hemos consultado el [Szel.] y el siguiente enlace: <https://es.wikipedia.org/wiki/RANSAC>. Además de las transparencias.

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

Vamos a suponer que queremos formar un mosaico medianamente bueno, es decir, que vamos a tener en cuenta el solapamiento de las imágenes en lugar de decir nosotros "a ojo" dónde queremos calcular cada una de las imágenes en el mosaico.

Entonces tendremos que calcular una homografía que lleve una imagen en aquella con la que se solape para poder tener en cuenta esta relación entre las imágenes que van a formar el mosaico, entonces tendremos un esquema como el siguiente:



donde cada H es la homografía que lleva una imagen en su solapada de acuerdo a la dirección de las aristas del grafo, observemos que no hemos puesto una homografía de $Im3$ al plano de proyección. Esto no significa que no exista tal homografía, simplemente es que no la calculamos como las otras, es decir, esta sí que la podemos calcular nosotros "a ojo" simplemente indicando dónde queremos colocar la $Im3$, normalmente en centro del mosaico. Entonces aquí no necesitamos tener puntos en correspondencias, sólomente introducir los valores que nosotros deseemos. Un ejemplo más claro es si queremos colocar la primera imagen en la esquina superior izquierda del mosaico, con lo que la homografía que lleva esta imagen al plano de proyección (lo que en clase llamamos H_0) es simplemente la identidad, con lo que no necesitamos ningún cálculo.

Con lo cual lo que necesitamos son los puntos en correspondencias para calcular las homografías que aparecen en el grafo anterior. Sabemos por teoría que para poder estimar una homografía necesitamos un mínimo de 4 parejas de puntos

en correspondencias, como lo que en total necesitaremos $12 \cdot 2 = 24$ puntos.

Estamos suponiendo una situación cualquiera si ya se dan otras condiciones más especiales como que por ejemplo sepamos que tipo de homografía relaciona cada imagen del mosaico con la anterior y esta homografía tenga menos de 8 grados de libertad (como tienen las homografías generales) entonces podremos tomar menos puntos o si podemos reutilizar puntos para establecer correspondencias pues también reducimos el número de puntos necesarios. Nosotros hemos hecho el cálculo para la situación más general posible.

Por supuesto la bondad del resultado obtenido dependerá de cómo de buenos sean las 4 parejas de puntos (que no es una cantidad suficientemente grande en la mayoría de caso para obtener un buen resultado) elegidos, pero el mínimo es ese.

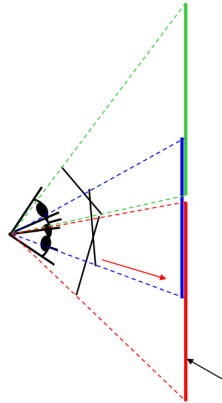
15. En la confección de un mosaico con proyección rectangular es esperable que aparezcan deformaciones de la realidad. ¿Cuáles y por qué? ¿Bajo qué condiciones esas deformaciones podrían desaparecer? Justificar la respuesta.

Aquí lo que vamos a usar es la metáfora que dimos en clase sobre la esfera de rayos. Las deformaciones que suelen producirse son normalmente escalados y cizallas, esto se debe a que normalmente nuestros panoramas no son planos con lo cual lo normal sería proyectar cada plano en uno distinto pero claro cuando estamos con proyección rectangular lo que hacemos es proyectar todas las imágenes en el mismo plano. También se pueden producir pérdidas de paralelismo y todas las que se producen cuando hacemos una transformación proyectiva general. Todo dependerá de qué transformación tengamos que hacer para llevar las imágenes al plano de proyección.

Entonces al situar las imágenes en este plano es cuando se producen estos "estiramientos". Lo que hacemos normalmente para minimizar este efecto es situar la imagen más central (con respecto a la realidad) en el centro.

Y ahora veamos cuándo se puede evitar esto: imaginemos que la cámara es una esfera a la que le llegan rayos entonces si los rayos de incidencia no cambian las deformaciones no suceden. Veamos esto más concretamente: lo que estoy diciendo es que si tomamos un panorama de modo que la cámara sólo se mueva en los ejes X e Y entonces estas deformaciones no se producen ya que digamos que todos los planos pueden ser proyectados de forma paralela en el plano de proyección (siempre y cuando el canvas escogido sea lo suficientemente grande como para no tener que redimensionar las imágenes).

Las deformaciones de las que hablamos se ven expresadas en la siguiente imagen de las transparencias:



Bibliografía

1. Transparencias.
2. Apuntes de clase.
3. Szel. = Computer Vision: Algorithms and Applications. Richard Szeliski.
4. Hart. = Multiple View Geometry in computer vision. Richard Hartley and Andrew Zisserman.
5. Distintos enlaces que se indican en la cuestión correspondiente.