Visión por computador. Cuestionario 2.

José Carlos Martínez Velázquez

23 de noviembre de 2016

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

Las transformaciones afines mantienen la relacion de colinealidad de los puntos así como las proporciones y el paralelismo. Esto es, si dos puntos pertenecían a la misma recta, tras una transformación afín lo seguirán estando (colinealidad). En el espacio proyectivo perdemos el paralelismo, provocando ciertas deformaciones. En el espacio proyectivo, por tanto, aunque las rectas cambian sus ángulos, siguen siendo rectas y pasan a cortarse en puntos denominados **puntos de fuga**.

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

Parte 1: verificar que $l = x \times x'$:

Vamos a realizar la primera parte. Verificaremos que, en coordenadas homogéneas, el vector de la recta definida por dos puntos se calcula como el producto vectorial de los dos puntos. Para ello, supongamos dos puntos: $P_1 = (x_1, y_1)$ y $P_2 = (x_2, y_2)$. Pasaremos dichos puntos a coordenadas homogéneas, entonces, pasan a convertirse en los vectores $P_1 = (x_1, y_1, 1)$ y $P_2 = (x_2, y_2, 1)$. Para calcular el producto vectorial, debemos calcular el determinante cuyas filas son: $(\overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$, P_1 y P_2:

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

$$= \overrightarrow{i}(y_1 - y_2) + \overrightarrow{j}(x_2 - x_1) + \overrightarrow{k}(x_1y_2 - x_2y_1) = (y_1 - y_2, x_2 - x_1, x_1y_2 - x_2y_1)$$

Como estamos en coordenadas homogéneas, la última coordenada debería ser 1. El vector director, por tanto, está multiplicado por la constante $w = x_1y_2 - x_2y_1$. Para conseguir 1 en la última coordenada, dividimos por w y obtenemos que el vector director en coordenadas homogéneas es:

$$\overrightarrow{V_l} = \left(\frac{y_1 - y_2}{x_1y_2 - x_2y_1}, \frac{x_2 - x_1}{x_1y_2 - x_2y_1}, 1\right)$$

Las coordenadas del vector director en coordenadas 2D serían: $\overrightarrow{V_l} = (\frac{y_1 - y_2}{x_1y_2 - x_2y_1}, \frac{x_2 - x_1}{x_1y_2 - x_2y_1})$.

Para comprobar que los puntos P_1 y P_2 pertenecen a la recta, se debe cumplir que $l^T P_1 = 0$ y $l^T P_2 = 0$:

$$l^T P_1 = \begin{pmatrix} \frac{y_1 - y_2}{x_1y_2 - x_2y_1} & \frac{x_2 - x_1}{x_1y_2 - x_2y_1} & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}$$

$$= \frac{x_1(y_1 - y_2) + y_1(x_2 - x_1) + x_1y_2 - x_2y_1}{x_1y_2 - x_2y_1}$$

$$= \frac{\cancel{x_1 y_1} - \cancel{x_1 y_2} + \cancel{x_2 y_1} - \cancel{x_1 y_1} + \cancel{x_1 y_2} - \cancel{x_2 y_1}}{x_1 y_2 - x_2 y_1} = 0$$

Del mismo modo, se comprueba que P_2 pertenece a l:

$$l^T P_2 = \begin{pmatrix} \frac{y_1 - y_2}{x_1 y_2 - x_2 y_1} & \frac{x_2 - x_1}{x_1 y_2 - x_2 y_1} & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix}$$
$$= \frac{x_2(y_1 - y_2) + y_2(x_2 - x_1) + x_1 y_2 - x_2 y_1}{x_1 y_2 - x_2 y_1}$$
$$= \frac{\cancel{x_2 y_1} - \cancel{x_2 y_2} + \cancel{x_2 y_2} - \cancel{x_1 y_2} + \cancel{x_1 y_2} - \cancel{x_2 y_1}}{x_1 y_2 - x_2 y_1} = 0$$

Parte 2: verificar que $x = l \times l'$:

Vamos ahora a ver la segunda parte, ver si el punto intersección de dos rectas l y l' puede ser calculado como $x = l \times l'$. Supongamos ahora que lo que acabamos de ver en lugar de ser puntos son rectas, es decir: $l = (l_x, l_y, 1)$ y $l' = (l'_x, l'_y, 1)$, entonces, al calcular el producto vectorial, tenemos que:

$$l \times l' = \begin{pmatrix} \frac{l_y - l'_y}{l_x l'_y - l'_x l_y} & \frac{l'_x - l_x}{l_x l'_y - l'_x l_y} & 1 \end{pmatrix}$$

Supongamos que es cierto que la operación $l \times l' = x$, entonces se debe cumplir que: $l^T \cdot x = l^T \cdot (l \times l') = 0$, es decir que l pasa por el punto de corte x. Veamoslo:

$$l^T \cdot \underbrace{(l \times l')}_{x} = 0 \Rightarrow$$
$$\Rightarrow \begin{pmatrix} l_x & l_y & 1 \end{pmatrix} \begin{pmatrix} \frac{l_y - l'_y}{l_x l'_y - l'_x l_y} \\ \frac{l'_x - l_x}{l_x l'_y - l'_x l_y} \\ 1 \end{pmatrix} =$$
$$= \frac{l_x(l_y - l'_y) + l_y(l'_x - l_x) + l_x l'_y - l'_x l_y}{l_x l'_y - l'_x l_y}$$
$$= \frac{\cancel{l_x l_y} - \cancel{l_x l'_y} + \cancel{l'_x l_y} - \cancel{l_x l_y} + \cancel{l_x l'_y} - \cancel{l'_x l_y}}{l_x l'_y - l'_x l_y} = 0$$

Del mismo modo, para terminar la demostración, vamos a demostrar que $l'^T \cdot (l \times l') = 0$, es decir, que l' pasa por x:

$$l'^T \cdot \underbrace{(l \times l')}_{x} = 0 \Rightarrow$$
$$\Rightarrow \begin{pmatrix} l'_x & l'_y & 1 \end{pmatrix} \begin{pmatrix} \frac{l_y - l'_y}{l_x l'_y - l'_x l_y} \\ \frac{l'_x - l_x}{l_x l'_y - l'_x l_y} \\ 1 \end{pmatrix} =$$
$$= \frac{l'_x(l_y - l'_y) + l'_y(l'_x - l_x) + l_x l'_y - l'_x l_y}{l_x l'_y - l'_x l_y}$$
$$= \frac{\cancel{l'_x l_y} - \cancel{l'_x l'_y} + \cancel{l'_x l'_y} - \cancel{l_x l'_y} + \cancel{l_x l'_y} - \cancel{l'_x l_y}}{l_x l'_y - l'_x l_y} = 0$$

Vemos cómo se puede sustituir $l \times l' = x$ en ambos casos y, por tanto, x pertenece a l y a l' simultáneamente, es decir, x es el punto de corte entre l y l'.

3. Sean x y l un punto y una recta respectivamente en un plano proyectivo P_1 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 tal que $l' = Gl$. Justificar la respuesta.

La expresión $x' = Hx$ significa que tenemos una homografía H que convierte coordenadas de los puntos relativas al plano P_1 a coordenadas relativas al plano P'. Si suponemos que existe H, entonces estas expresiones son equivalentes:

$$l'^T x' = 0 \Rightarrow l'^T \underbrace{Hx}_{x'} = 0$$

Como sabemos que l pasa por x, es decir $l^T x = 0$, estas expresiones son equivalentes:

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

Si eliminamos los términos comunes, obtenemos que

$$l'^T H \cancel{x} = l^T \cancel{x} \Rightarrow l'^T H = l^T$$

Ahora, en la parte derecha podemos obtener l trasponiendo la parte derecha, pero también hay que hacerlo en la parte izquierda, entonces:

$$l'^T H = l^T \Rightarrow (l'^T H)^T = l$$

Recordemos que $(AB)^T = B^T A^T$, aplicamos esta regla:

$$(l'^T H)^T = l \Rightarrow H^T l' = l$$

Sólo falta despejar l', para ello, multiplicamos por la inversa de H^T por la izquierda en ambos lados de la ecuación:

$$\cancel{H^{-T} H^T} l' = H^{-T} l$$

Si renombramos $G = H^{-T}$, entonces hemos encontrado la expresión que buscábamos:

$$l' = H^{-T} l \Rightarrow l' = Gl$$

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

$$H_1 = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad H_2 = \begin{pmatrix} 0 & 1 & -3 \\ -1 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$H_3 = \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad H_4 = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix}$$

Para realizar más fácilmente este ejercicio, vamos a identificar la forma que tienen las matrices de cada una de las cuatro primeras transformaciones afines y la proyección:

- Translación: Sea un punto $P = (x, y)$, tras la translación queremos que quede $P' = (x + t, y + t)$. Como esta transformación afín no puede representarse como $Matriz \cdot punto = punto$, surge la necesidad de usar coordenadas homogéneas. Entonces, transformaremos un punto $P = (x, y, 1)$ en otro punto $P' = (x + t, y + t, 1)$. Entonces, la matriz de translación tiene la forma:

$$M_t = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix}$$

  Si multiplicamos la matriz por un punto, conseguimos lo que queremos:

$$\begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x + t_x \\ y + t_y \\ 1 \end{pmatrix}$$

- Escalado: Ya hemos visto en el caso de la translación por qué surge la necesidad de usar coordenadas homogéneas. Para todas las transformaciones ya deben usarse. El escalado consiste en, dado un punto $P = (x, y, 1)$, convertirlo en $P' = (x \cdot e_x, y \cdot e_y, 1)$. Lo conseguiremos con la matriz:

$$M_e = \begin{pmatrix} e_x & 0 & 0 \\ 0 & e_y & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

  Si multiplicamos la matriz por un punto, obtenemos lo que buscamos:

$$\begin{pmatrix} e_x & 0 & 0 \\ 0 & e_x & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x \cdot e_x \\ y \cdot e_y \\ 1 \end{pmatrix}$$

- Rotación: esta transformación quizás sea la más compleja de explicar de todas. Dado un punto $P = (x_{inicial}, y_{inicial}, 1)$, queremos convertirlo en $P' = (x_{final}, y_{final}, 1)$ rotado en un ángulo α. Vamos a tratar de explicarlo con la siguiente figura:

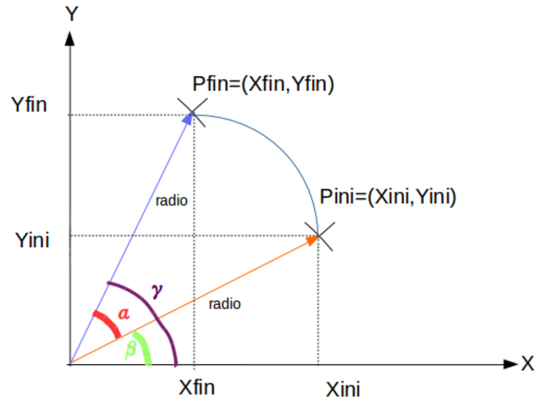

Figura 1. Proceso de rotación de un punto.

Como se aprecia en la figura, tenemos tres ángulos, α, β y γ cuya relación es $\gamma = \alpha + \beta$. Las coordenadas de P' serán:

$$x_{final} = radio \cdot cos\gamma = radio \cdot cos(\alpha + \beta)$$

$$y_{final} = radio \cdot sen\gamma = radio \cdot sen(\alpha + \beta)$$

Aplicando que $sen(a + b) = sen(a)cos(b) + cos(a)sen(b)$ y $cos(a + b) = cos(a)cos(b) - sen(a)sen(b)$, tenemos que

$$x_{final} = radio \cdot cos(\alpha)cos(\beta) - radio \cdot sen(\alpha)sen(\beta)$$

$$y_{final} = radio \cdot sen(\alpha)cos(\beta) + radio \cdot cos(\alpha)sen(\beta)$$

Sabemos que $x_{inicial} = radio \cdot cos(\beta)$ y que $y_{final} = radio \cdot sen(\beta)$. Si lo aplicamos, nos queda que:

$$x_{final} = x_{inicial} \cdot cos(\alpha) - y_{inicial} \cdot sen(\alpha)$$

$$y_{final} = x_{inicial} \cdot sen(\alpha) + y_{inicial} \cdot cos(\alpha)$$

Ahora que hemos entendido la transformación, nuestro objetivo es pues, transformar un punto $P = (x, y, 1)$ en $P' = (x \cdot cos(\alpha) - y \cdot sen(\alpha), x \cdot sen(\alpha) + y \cdot cos(\alpha), 1)$. Esto se consigue con la matriz:

$$M_r = \begin{pmatrix} cos(\alpha) & -sen(\alpha) & 0 \\ sen(\alpha) & cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Si multiplicamos la matriz por un punto, obtenemos lo que buscábamos:

$$\begin{pmatrix} cos(\alpha) & -sen(\alpha) & 0 \\ sen(\alpha) & cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x \cdot cos(\alpha) - y \cdot sen(\alpha) \\ x \cdot sen(\alpha) + y \cdot cos(\alpha) \\ 1 \end{pmatrix}$$

- Cizalla: esta transformación parece más compleja de lo que es. Podemos hacer cizalla en la dirección del eje X o en la dirección del eje Y. Para el primer caso, aumentaremos x en función de un escalado de y, quedando la coordenada y igual que estaba, esto es, dado un punto $P = (x, y, 1)$, se transforma en $P' = (x + ky, y, 1)$, siendo k una constante. Del mismo modo, para una cizalla en la dirección del eje y, un punto $P = (x, y, 1)$, se transforma en $P' = (x, y + kx, 1)$. Vamos a conseguir este resultado con la matriz:

$$M_{c_x} = \begin{pmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ ó } M_{c_y} = \begin{pmatrix} 1 & 0 & 0 \\ k & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Para probar que funciona, primero multiplicamos la matriz de cizalla en la dirección del eje X por un punto:

$$\begin{pmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x + ky \\ y \\ 1 \end{pmatrix}$$

Ahora probamos la matriz de cizalla en la dirección del eje Y:

$$\begin{pmatrix} 1 & 0 & 0 \\ k & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ kx + y \\ 1 \end{pmatrix}$$

En efecto, funciona.

- Proyectivo: Una matriz de proyección ya no es una transformación afín y realiza cambios en la tercera coordenada, es decir, multiplica el punto por un determinado coeficiente. Si la última fila de la matriz es distinta de $(0, 0, 1)$, podemos estar seguros de que la última coordenada del punto resultante será distinta de 1 y, por tanto, todo el punto quedará multiplicado por una constante. Una matriz que **solamente** aplica proyección debe ser igual a la matriz identidad excepto la última fila, es decir, así:

$$M_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & b & c \end{pmatrix}$$

Conociendo el aspecto de las matrices de transformación afín, vamos a comentar entonces qué transformaciones haríamos con las composiciones que se nos plantean:

- H_1: las matrices que forman H_1 se pueden corresponder fácilmente por los tipos explicados anteriormente. De derecha a izquierda, se trata de una transformación de **cizalla en la dirección del eje X**, seguido de un **escalado** (la mitad del tamaño original con respecto al eje X, y aproximadamente un tercio del tamaño original con respecto al eje Y), seguida de una **translación** en el sentido positivo del eje X y una translación en el sentido positivo del eje Y.

- H_2: Las matrices que forman H_2 no pueden identificarse directamente con los tipos explicados antes, pero es fácil descomponerlas en dichos tipos resolviendo ecuaciones matriciales. La matriz de la izquierda se puede descomponer como:

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

  Que corresponden a transformaciones de translación y rotación. La matriz de la derecha se puede descomponer como:

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

  Que corresponden a transformaciones de cizalla y de escalado respectivamente. Por tanto, de derecha a izquierda, H_2 aplica un **escalado** al doble en ambos ejes junto con una transformación de **cizalla** con respecto al eje Y. A continuación se aplica una **rotación** y una **translación**.

- H_3: La matriz de la derecha, dado que es la identidad con la última fila cambiada, se puede identificar directamente con una matriz de proyección. Para identificar la matriz de la izquierda es necesario descomponerla. Resolviendo las ecuaciones matriciales oportunas, se puede descomponer en:

$$\begin{pmatrix} 1 & 0.25 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

  Que se corresponde con un escalado en el eje Y y una cizalla en ambos ejes. Por tanto, de derecha a izquierda, la transformación H_3 aplica una **proyección**, a continuación una **cizalla** en el sentido positivo de X e Y junto con un **escalado** al doble sólo en el eje Y.

- H_4: La matriz H_4 tiene la última fila distinta de $(0, 0, 1)$, por lo tanto aplica una transformación de proyección. Vamos a tratar de aislar la proyección y ver qué más transformaciones está aplicando. Para aislar la proyección basta con cambiar a la matriz identidad la última fila por la última fila de la proyección, entonces, la transformación H_4 quedaría como sigue:

$$H_4 = \begin{pmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

Para encontrar los valores a, b, c, d, e y f, basta con multiplicar ambas matrices, e igualar a los elementos de la matriz H_4 original, de donde obtenemos el siguiente sistema:

$$\begin{cases} a \cdot 1 + b \cdot 0 + c \cdot 0 = 2 \Rightarrow \underline{a = 2} \\ a \cdot 0 + b \cdot 0 + c \cdot 2 = 3 \Rightarrow \underline{c = \frac{3}{2}} \\ a \cdot 0 + b \cdot 1 + c \cdot 1 = 0 \Rightarrow b + c = 0 \Rightarrow \underline{b = -\frac{3}{2}} \\ d \cdot 1 + e \cdot 0 + f \cdot 0 = 0 \Rightarrow \underline{d = 0} \\ d \cdot 0 + e \cdot 0 + f \cdot 2 = -1 \Rightarrow \underline{f = -\frac{1}{2}} \\ d \cdot 0 + e \cdot 1 + f \cdot 1 = 2 \Rightarrow e + f = 2 \Rightarrow \underline{e = \frac{5}{2}} \end{cases} \tag{1}$$

La matriz que no aplica proyección resulta en:

$$\begin{pmatrix} 2 & -\frac{3}{2} & \frac{3}{2} \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

Que no puede ser identificada con alguna de las transformaciones explicadas. Vamos a descomponerla hasta que se puedan reconocer (resolviendo ecuaciones matriciales). Las transformaciones que realiza H_4 quedan recogidas en el siguiente árbol de descomposición:

$$\begin{tikzcd}[column sep=small]
& H_4 = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \arrow[dl, dash] \arrow[dr, dash] & & & \\
\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}}_{Proyección} & & \begin{pmatrix} 2 & -\frac{3}{2} & \frac{3}{2} \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \arrow[dl, dash] \arrow[dr, dash] & & \\
& \underbrace{\begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{5}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{Escalado} & & \begin{pmatrix} 1 & -\frac{3}{4} & \frac{3}{4} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix} \arrow[dl, dash] \arrow[dr, dash] & \\
& & \underbrace{\begin{pmatrix} 1 & 0 & \frac{3}{4} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix}}_{Translación} & & \underbrace{\begin{pmatrix} 1 & -\frac{3}{4} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{Cizalla}
\end{tikzcd}$$

De derecha a izquierda, pues, la transformación H_4 realiza una transformación de **cizalla** en el eje X, una transformación de **translación**, un **escalado** (del doble en el eje X y del doble y medio en el eje Y) y una **proyección**.

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

Una homografía define una relación entre dos planos, luego estamos hablando de una relación en un espacio proyectivo. Si esto es así, todos los puntos tendrán coordenadas homogéneas y, por tanto, tendrán tres coordenadas. Para transformar un punto cuya dimensión es $P_{3\times 1}$, **necesitamos una matriz $H_{3\times 3}$. Esta es la primera propiedad. La segunda propiedad** tiene que ver el sentido en el que se puede aplicar una homografía. Una homografía transforma un punto de un plano I_1 a un sistema de referencia relativo a un plano I_2 y su inversa transforma un punto I_2 a un sistema de referencia relativo a I_1. El hecho de que una homografía deba ser una matriz invertible, convierte en condición indispensable **para ser homografía que dicha matriz H tenga determinante distinto de 0**. Las homografías pueden agruparse en clases de equivalencia, por lo que no necesariamente dos matrices distintas que cumplan estas propiedades provocan transformaciones distintas.

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

Una homografía general engloba a las transformaciones afines y las proyecciones, así que quedaremos acotados por el caso más destructivo y este caso es la proyección. Tal y como vimos en la primera pregunta, en el espacio afín se mantenía el paralelismo, quedando este destruído en una transformación proyectiva. Aunque se pierde el paralelismo y las lineas paralelas pasan a cortarse en un punto de fuga, todos los puntos que formaban parte de una recta, siguen formando parte de ella, aunque han sufrido transformaciones importantes. Por tanto, **se mantiene la colinealidad de los puntos.** Además se mantiene lo que se denomina *cross-ratio* de cuatro puntos, que se calcula, dados cuatro puntos A, B, C, D, como la siguiente proporción:

$$C_r(A, B; C, D) = \frac{AC \cdot BD}{BC \cdot AD}$$

7. ¿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.

Las deformaciones más fuertes son provocadas por la transformación de proyección. Como hemos visto en anteriores cuestiones, la proyección destruye ángulos, paralelismo y proporciones. Propiedades que las transformaciones afines sí mantenían. Si nos fijamos en las características de una matriz de proyección, es la que menos restricciones matemáticas tiene (sólo debe cumplir que el tamaño sea 3×3 y su determinante distinto de 0), ya que deja de ser necesario que la última fila sea $(0, 0, 1)$. Esto redunda en que puede modificar más propiedades de la imagen que cualquier otra transformación.

8. ¿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 **usa la información de la intensidad de gradiente.** Consiste en pasear una ventana por la imagen en busca de cambios significativos en la intensidad del gradiente. Para que un punto sea detectado por el detector debe presentar cambios fuertes en ambas direcciones (horizontal, eje X y vertical, eje Y). Una vez que se han detectado los cambios mencionados, hay que dilucidar si son lo suficientemente fuertes para el detector, para lo que se usa el criterio Harris, dado por la expresión:

$$V_{Harris} = \lambda_1 \lambda_2 - \alpha(\lambda_1 + \lambda_2)$$

Donde λ_1 y λ_2 son matrices que contienen los valores propios de cada píxel. α es una constante en el intervalo $[0.04, 0.06]$. También hay que definir un umbral tal que, si el valor calculado supera el umbral, entonces el punto harris es suficientemente fuerte y es seleccionado.

Por su forma de funcionar (buscar cambios de intensidad de gradiente en X y en Y), el detector de Harris va a detectar esquinas, por lo tanto, está diseñado para **detectar patrones geométricos**.

9. ¿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

No. Todo buen descriptor debe proporcionar una robustez suficiente. Si el descriptor sólo se compusiese de los valores de los píxeles de su región de soporte y transformásemos la misma, al buscar puntos en correspondencia con otra imagen con puntos en común, este descriptor no daría información correcta en general. **Sólo en caso de que la transformación sea una translación obtendríamos los resultados esperados.** En cualquier otro caso podría ser un desastre. Si hiciéramos una rotación, por ejemplo, al carecer de información sobre la orientación del píxel en cuestión, lo que con la imagen en su posición original era un emparejamiento correcto, ahora no, pues estamos considerando la misma orientación con distintos valores. Se debe añadir información de la gradiente, tal como su orientación y magnitud.

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

El descriptor SIFT define un tamaño de ventana o de vecindario de 16×16 píxeles. A su vez, tomamos los 16 sub-bloques resultantes de dividir este vecindario en ventanas de 4×4. De cada sub-bloque extrae **información sobre la orientación y magnitud del vector gradiente, representándola en un histograma** de 8 intervalos en los que se dividen todos los posibles ángulos de orientación: Esto es calculado mediante los valores de los píxeles del vecindario. Para establecer la orientación y magnitud del bloque grande de 16×16, hace la media de los datos obtenidos en cada sub-bloque 4×4. Esto ofrece fiabilidad en los resultados, pues no están basados en un sólo criterio, como en el supuesto de la anterior cuestión.

11. 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

Dada una característica f_1 en una imagen I_1, se puede obtener la mejor correspondencia f_2 en I_2 bajo dos criterios:

- Mínima distancia entre dos características (vecino más cercano): se calcula la norma L2 (o distancia L2) entre ambas:
$$D_{f_1,f_2} = ||f_1 - f_2||$$
Este criterio puede dar buenas puntuaciones a puntos incorrectos. Esto ocurre en imágenes con patrones que se repiten a lo largo de la imagen, luego este critero no es demasiado bueno.

- Diferencia entre dos características: cuando hay dudas sobre si un emparejamiento es correcto, se usa este criterio. Consiste en calcular un ratio de diferencia con el primer mejor matching, f_2, y el segundo mejor, f_2'. Esto es, si f_2 es el mejor matching para f_1 calculado mediante distancia L2, se obtiene también f_2', el segundo punto con mas puntuación de matching. A continuación se calcula el ratio de las distancias como:
$$r_d = \frac{||f_1 - f_2||}{||f_1 - f_2'||}$$

Cuando esta operación resulta muy cercana a 1, quiere decir que no hay diferencias sustanciales entre f_2 y f_2' y no se sabe cuál es el correcto. Normalmente, cuando esto ocurre, se desecha la posibilidad de encontrar una pareja a f_1. Hay que definir un umbral (o margen de error) alrededor de 1 para decidir si las diferencias son sustanciales o no. Como en el caso anterior, en una imagen con patrones repetitivos nos haría perder una gran cantidad de puntos.

Existen algunas propuestas para solucionar los problemas que aparecen. Una de ellas es el algoritmo cross-checking, cuya versión básica consta de los siguientes pasos:

1. Encontrar el mejor matching para f_1 en la imagen 2. Esto es f_2.
2. Encontrar el mejor matching para f_2 en la imagen 1. Esto es f_1'.
3. Si $f_1 = f_1'$, entonces nos quedamos con la pareja de puntos.
4. Si no, se desecha.

Cross-check nos hace perder una cantidad importante de puntos, pero los que sobreviven, tienen un porcentaje alto de acierto.

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

El algoritmo RANSAC (RANdom SAmple Consensus) es usado en visión por computador para encontrar una homografía entre dos planos. Dada una nube de n puntos, donde visualmente podemos definir una tendencia (recta), RANSAC nos devuelve esta recta. Podríamos decir que esto es lo que consigue regresión lineal, pero no es así. La diferencia esencial con regresión lineal es que RL se ve influenciada por los n puntos y trata de devolver la recta que minimiza la distancia a todos los puntos. Entonces, si entre los n puntos hay señales ruidosas, con regresión lineal dichos puntos afectan al resultado final, tal y como muestra la siguiente figura:

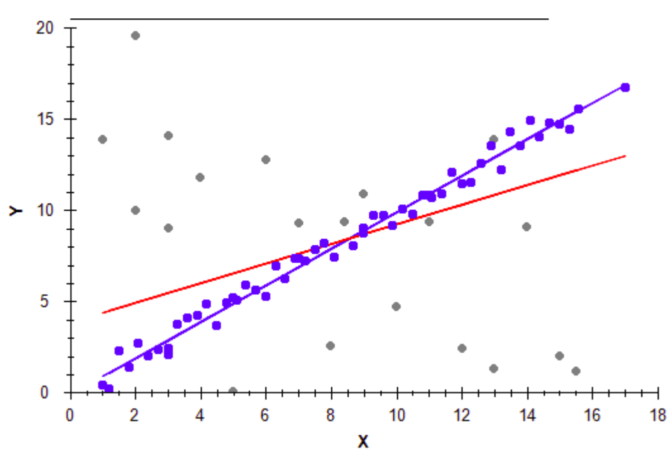

Figura 2. RANSAC vs. Regresión lineal.
(RANdom Sample Consensus (RANSAC) in C# - César Souza)

La linea roja es el resultado del algoritmo de regresión lineal, mientras que el algoritmo RANSAC devuelve la recta de color azul, que claramente es lo que cabría esperar.

Para conocer cómo funciona RANSAC, debemos definir dos conceptos: inliers (o puntos que "están de acuerdo") y outliers (o puntos que "están en desacuerdo"). Más concretamente, los inliers son aquéllos puntos que pueden ser explicados mediante un modelo matemático y los outliers son aquéllos que no se ajustan al mismo. También se debe definir un margen de error que permita flexibilidad a la hora de clasificar un punto como inlier u outlier. RANSAC es un algoritmo iterativo no determinista, esto es, debemos definir un número de iteraciones suficiente para obtener una respuesta de calidad razonable. Evidentemente, aunque el numero de iteraciones sea igual en dos experimentos, no tiene por qué devolver la misma recta, pues es un algoritmo aleatorio, pero sí muy parecida, pues se basa en la tendencia que describen la mayoría de los puntos. En cada iteración, el algoritmo RANSAC sigue los siguientes pasos:

1. Elige un subconjunto arbitrario de k puntos de entre los n puntos.
2. Se cuentan los inliers y los outliers y se comparan con los del mejor modelo encontrado hasta ahora.
3. Si el número de inliers es mayor que el mayor número de inliers encontrado, el mejor modelo encontrado es el que acabamos de encontrar. En caso contrario, todo permanece igual.

Aunque es un algoritmo aproximado, RANSAC tiene la ventaja de que los datos ruidosos apenas afectan, a diferencia de regresión lineal. Esto permite encontrar homografías de buena calidad si el número de iteraciones es lo suficientemente grande y justifica totalmente su uso para este cometido.

13. ¿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

Necesitamos 24 puntos (12 parejas en correspondencia) para construirlo, veamos por qué:

Para construir un mosaico con dos imágenes, esta claro que necesitamos 8 puntos (o cuatro parejas de puntos), es decir, cuatro de la primera imagen y cuatro del siguiente. Esta sería la solución para formar un mosaico con 1-2. Dado que tenemos las relaciones 1-2, 2-3 y no tenemos 1-3, está claro que los puntos que solapan 1-2 y 2-3 no son los mismos. Entonces necesitamos otros 8 puntos para obtener una homografía entre 2-3. Del mismo modo se razona que se necesitan otros 8 puntos para relacionar 3-4. Entonces, si por cada par de imágenes necesitamos 8 puntos, **necesitamos 24 puntos (o 12 parejas de puntos en correspondencia)** para formar el mosaico completo.

Sería algo así:

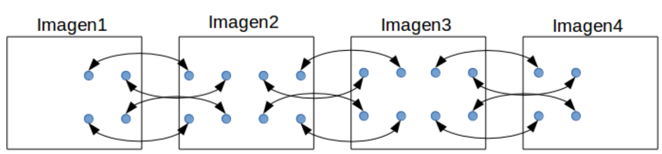

Figura 3. Número de correspondencias necesarias.

14. 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

Estas deformaciones aparecen porque las imagenes tomadas, generalmente, no pertenecen al mismo plano. Si comenzamos a construir un mosaico rectangular de izquierda a derecha o de derecha a izquierda, la homografía que coloca la primera imagen sólo hace una operación de translación en el mosaico, luego la deformación sufrida es 0. Conforme vamos colocando imágenes sobre el lienzo, las homografías ya no son tan simples, pasando a ser estas de proyección y deformando cada vez más la imagen. **Entonces, las deformaciones son las que realiza una homografía de proyección** y los errores acumulados serían mayores cuanto más lejos estamos de la primera imagen que se coloca, tal y como vemos en la siguiente figura:

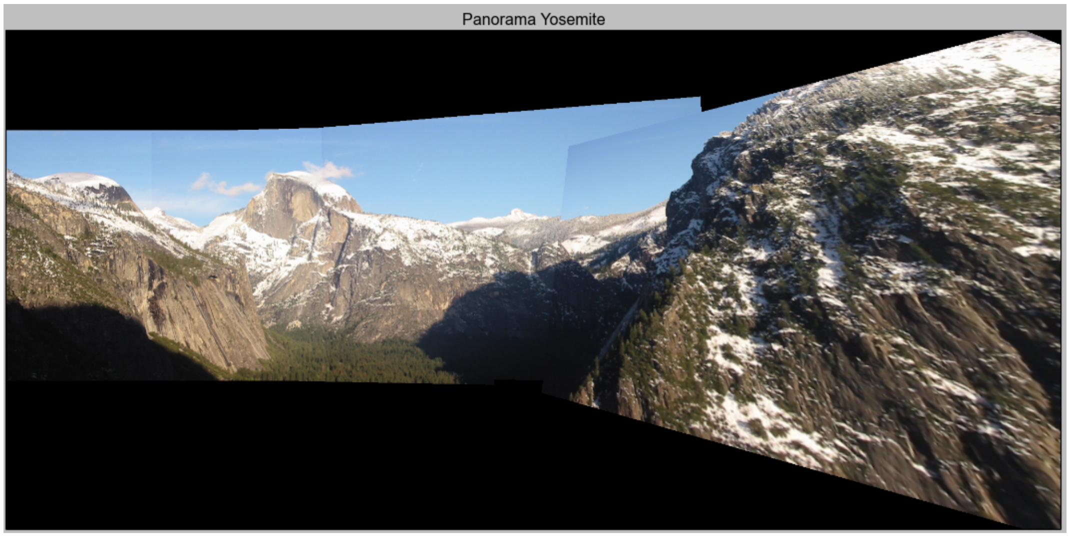

Figura 4. Acumulación de deformaciones de extremo a extremo.

La figura de la derecha ha sufrido una gran deformación que resulta demasiado evidente a la vista. Aunque no es fácil construir un mosaico sin deformaciones, se puede mitigar este efecto comenzando a colocar las imágenes desde el centro hacia los lados. Así, el error acumulado en los extremos es la mitad de lo que se acumularía en en caso anterior. Podemos compararlo con la siguiente figura:

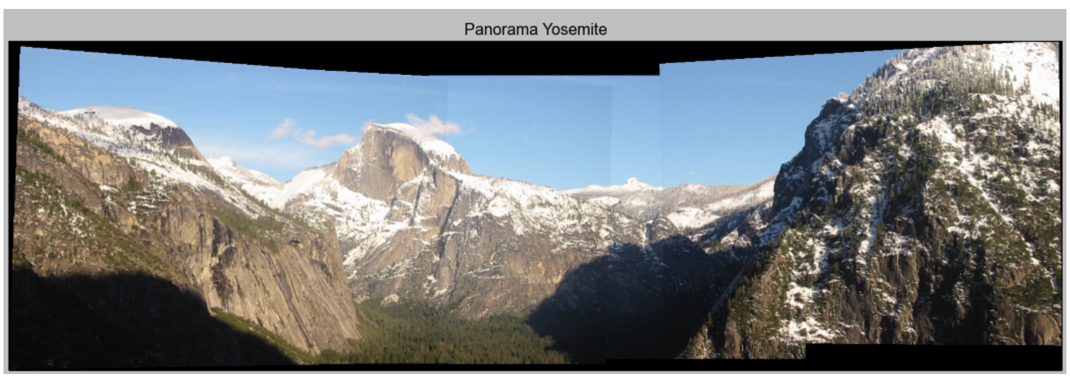

Figura 5. Acumulación de deformaciones comenzando por el centro.

Las deformaciones no son tan evidentes a la vista y, por tanto, los errores acumulados no son tantos como en el primer caso.

Para construir un mosaico sin deformaciones, todas las imágenes tomadas deben pertenecer al mismo plano y, si no se puede mover la cámara, un mosaico no tendrá deformaciones si todas las imágenes que forman el mosaico son la misma, o lo que es lo mismo, el mosaico sólo está formado por una imagen.

Referencias

1. *Transformación afín.*
   Wikipedia
   https://es.wikipedia.org/wiki/Transformaci%C3%B3n_af%C3%ADn

2. *Matriz de rotación de un punto en el plano X-Y*
   CEIAR.UR - Youtube
   https://www.youtube.com/watch?v=IcxIqoFMiWM

3. *Cross-ratio*
   Wikipedia
   https://en.wikipedia.org/wiki/Cross-ratio

4. *Scale Invariant Feature Transform*
   Wikipedia
   https://es.wikipedia.org/wiki/Scale-invariant_feature_transform#Descriptor_de_punto_de_inter.C3.A9s

5. *RANdom Sample Consensus (RANSAC) in C#*
   César Souza
   http://crsouza.com/2010/06/02/random-sample-consensus-ransac-in-c