Visión por computador. Cuestionario 3.

José Carlos Martínez Velázquez

10 de diciembre de 2016

1.(Op. 6 puntos) En clase se ha mostrado una técnica para estimar el vector de traslación del movimiento de un par estéreo y solo ha podido estimarse su dirección. Argumentar de forma lógica a favor o en contra del hecho de que dicha restricción sea debida a la técnica usada o sea un problema inherente a la reconstrucción.

La restricción **se debe a la técnica usada**. Veamos por qué:

El algoritmo visto, en sus tres primeros pasos dice los siguiente:

1. Calcular la matriz esencial $E = [T]_x R$
2. Calcular $E^T E = R^T [T][T]_x R$. De aquí obtenemos la siguiente matriz:

$$E^T E = \begin{pmatrix} T_y^2 + T_z^2 & -T_x T_y & -T_x T_z \\ -T_y T_x & T_z^2 + T_x^2 & -T_y T_z \\ -T_z T_x & -T_z T_y & T_x^2 + T_y^2 \end{pmatrix}$$

3. Normalizar con la traza y estimar la dirección de T. Si normalizamos (dividir por $\sqrt{Traza(E^T E)/2}$), obtenemos la siguiente matriz:

$$E^T E_{norm} = \begin{pmatrix} 1 - \hat{T}_x^2 & -\hat{T}_x \hat{T}_y & -\hat{T}_x \hat{T}_z \\ -\hat{T}_y \hat{T}_x & 1 - \hat{T}_y^2 & -\hat{T}_y \hat{T}_z \\ -\hat{T}_z \hat{T}_x & -\hat{T}_z \hat{T}_y & 1 - \hat{T}_z^2 \end{pmatrix}$$

En algoritmo no puede precisar más datos sobre el vector de translación en pasos posteriores. Como se puede apreciar, los términos $\hat{T}_x$, $\hat{T}_y$ y $\hat{T}_z$ aparecen elevados al cuadrado, lo que significa que hay ambigüedad con respecto a su signo. Esto radica en que son posibles dos vectores de translación (T_1 y $T_2 = -T_1$). Por culpa de esto también obtenemos dos posibles matrices de rotación (R_1 y R_2), conformando cuatro configuraciones posibles:

- $[R_1|T_1]$
- $[R_1|T_2]$
- $[R_2|T_1]$
- $[R_2|T_2]$

Para cada pareja de $[R|T]$ resolvemos el sistema siguiente:

$$\underbrace{\begin{bmatrix} x'_1 \times Rx_1 & 0 & 0 & 0 & \dots & x'_1 \times T \\ 0 & x'_2 \times Rx_2 & 0 & 0 & \dots & x'_2 \times T \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & x'_n \times Rx_n & x'_n \times T \end{bmatrix}}_{M\Lambda} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \gamma \end{bmatrix} = 0$$

Entonces, obtendremos cuatro soluciones distintas:

- Solución para $[R_1|T_1]$: $(\lambda_1\ \lambda_2\ ...\ \gamma)_{11}$
- Solución para $[R_1|T_2]$: $(\lambda_1\ \lambda_2\ ...\ \gamma)_{12}$
- Solución para $[R_2|T_1]$: $(\lambda_1\ \lambda_2\ ...\ \gamma)_{21}$
- Solución para $[R_2|T_2]$: $(\lambda_1\ \lambda_2\ ...\ \gamma)_{22}$

La pareja $[R|T]$ correcta es aquélla que devuelve en su solución el menor número de λ_i negativas. Esto es porque las λ_i ofrecen información sobre la profundidad. Esta debe ser positiva para garantizar que los puntos estaban delante de la cámara. Idealmente debería haber una pareja $[R|T]$ para la que todas las λ_i fuesen positivas.

Dado que hemos visto que existe un modo de solucionar la ambiguedad, mediante el método descrito anteriormente, vemos cómo la restricción que sufre el vector de translación **es debida a la técnica usada**. Con la pareja correcta de $[R|T]$, podemos enfrentarnos a la reconstrucción.

2.(Op. 6 puntos) Verificar matemáticamente que se deben de cumplir las ecuaciones $Fe = 0$ y $F^T e' = 0$.

La matriz F es la matriz fundamental, que codifica información acerca de la geometría epipolar entre dos imágenes. Sean pues, dos puntos en correspondencia x (perteneciente a la primera imagen) y x' (perteneciente a la segunda), podemos calcular una recta l' en la primera imagen que pasa por x y por el epipolo e'. Del mismo modo, podemos calcular una recta l en la segunda imagen que pasa por x' y por el epipolo e. Estas rectas se calculan como sigue:

$$l' = Fx$$
$$l = F^T x'$$

Del mismo modo, sabemos que el producto de una recta por un punto por el que pasa es igual a 0. Dicho de otro modo, si p es un punto que pasa por una recta r, entonces $r^T p = 0$. Como sabemos que l' pasa por e' y l pasa por e, aplicando esta ecuación, tenemos que:

$$l'^T e' = 0$$
$$l^T e = 0$$

Si sustituímos l y l' por su expresión equivalente, tenemos:

$$(Fx)^T e' = 0 \Rightarrow x^T F^T e' = 0 \Rightarrow \underline{F^T e' = 0}$$
$$(F^T x')^T e = 0 \Rightarrow x'^T F e = 0 \Rightarrow \underline{Fe = 0}$$

3.(Op. 10 puntos) Verificar matemáticamente que cuando una cámara se desplaza las coordenadas retina de puntos correspondientes sobre la retina deben de verificar la ecuación $x' = x + Kt/Z$

Vamos a suponer un punto de mundo Q y una cámara P. Por simplicidad, supondremos que la cámara es $P = K[I|0]$. Si esto es así, podemos proyectar el punto Q con la cámara P como sigue:

$$q = K[I|0] \underbrace{\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}}_{Q} = \begin{pmatrix} K_{11} & K_{12} & K_{13} & 0 \\ K_{21} & K_{22} & K_{23} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} K_{11}X + K_{12}Y + K_{13}Z \\ K_{21}X + K_{22}Y + K_{23}Z \\ Z \end{pmatrix} = KQ$$

Dado que q es una proyección de Q a través de la cámara P, debería aparecer en coordenadas homogéneas. Denotaremos el punto q en coordenadas homogéneas como:

$$q_h = \frac{1}{Z}q = \frac{1}{Z}KQ$$

Ahora supongamos una cámara P' que se trasladó arbitrariamente con respecto a P (ojo, la cámara no rota). Entonces, tenemos que $P = K[I|t]$. Si proyectamos el mismo punto Q con esta nueva cámara, tenemos lo siguiente:

$$q' = K[I|t]\underbrace{\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}}_{Q} = \begin{pmatrix} K_{11} & K_{12} & K_{13} & (K_{11}t_1 + K_{12}t_2 + K_{13}t_3) \\ K_{21} & K_{22} & K_{23} & (K_{21}t_1 + K_{22}t_2 + K_{23}t_3) \\ 0 & 0 & 1 & t_3 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} =$$

$$= \begin{pmatrix} K_{11}X + K_{12}Y + K_{13}Z + (K_{11}t_1 + K_{12}t_2 + K_{13}t_3) \\ K_{21}X + K_{22}Y + K_{23}Z + (K_{21}t_1 + K_{22}t_2 + K_{23}t_3) \\ Z + t_3 \end{pmatrix} = KQ + Kt$$

Recapitulando las cosas relevantes, hemos obtenido que:

$$q_h = \frac{1}{Z}KQ$$
$$q' = KQ + Kt$$

Lo que buscamos es conocer las coordenadas homogéneas de q', esto es, q'_h. Entonces podemos sustituir en la expresión de q' lo siguiente:

$$q' = Z\underbrace{\frac{1}{Z}KQ}_{q_h} + Kt \Rightarrow q' = Zq_h + Kt$$

Si queremos poner q' en función de las coordenadas homogéneas de q, esto es q_h, sólo tenemos que dividir por Z en ambos miembros:

$$\frac{q'}{Z} = \frac{Zq_h + Kt}{Z} \Rightarrow q'_h = \frac{\cancel{Z}q_h}{\cancel{Z}} + \frac{Kt}{Z} \Rightarrow q'_h = q_h + \frac{Kt}{Z}$$

Finalmente si renombramos q_h como x y q'_h como x' tenemos lo que se pide:

$$x' = x + \frac{Kt}{Z}$$

4.(Op. 6 puntos) Dar una interpretación geométrica a las columnas y filas de la matriz P de una cámara.

La cámara es una matriz P de dimensiones 3×4, esto es, tiene tres filas y cuatro columnas:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix}$$

La primera columna, formada por p_{11}, p_{21} y p_{31} codifica el punto de fuga en la imagen para el eje X. Del mismo modo, la segunda columna y la tercera codifican los puntos de fuga en la imagen para los ejes Y y Z respectivamente. La cuarta columna es la imagen del origen de coordenadas. Debemos entender la imagen de un punto como el conjunto de vectores que se pueden obtener al aplicar la transformación geométrica desde dicho punto.

Las filas representan planos. La última fila de P representa el plano principal de la cámara, mientras que la primera y la segunda representan planos en el espacio que contienen al centro de la cámara. El plano que representa la primera fila se corresponde con la línea de la imagen $x = 0$, mientras que el plano representado por la segunda fila se corresponde con la línea de la imagen $y = 0$.

5.(Op. 6 puntos) Suponga una matriz A(3x3) de números reales. Suponga rango(A)=3. ¿Cuál es la matriz esencial más cercana a A en norma de Frobenius? Argumentar matemáticamente la derivación.

La matriz esencial E tiene una restricción en su descomposición en valores singulares y es que dos de sus valores singulares deben ser idénticos y el tercero debe ser 0. De modo formal, diremos que:

$$SVD(E) = UDV^T \ Donde\ D = diag(v_s, v_s, 0)$$

Si tenemos una matriz A cualquiera con rango 3 y realizamos su descomposición en valores singulares, generalmente obtendremos que:ventajas y desventajas de usar un algoritmo

$$SVD(A) = UDV^T \ Donde\ D = diag(a, b, c)$$

Para satisfacer la restricción de la matriz diagonal de la matriz esencial, elegiremos como el valor v_s la media de los valores de la diagonal D, como sigue:

$$D' = diag(\frac{a+b+c}{3}, \frac{a+b+c}{3}, 0)$$

En este momento, podemos obtener la matriz esencial E más cercana a A como:

$$E = UD'V^T$$

Si nos fijamos en cómo se calcula la norma de Frobenius de una matriz M:

$$||M||_F = \sqrt{\sum_{i=1}^{nfilas} \sum_{j=1}^{ncols} M_{ij}^2} = \sqrt{traza(M^2)} = \sqrt{\sum_{i=1}^{min\{nfilas,ncols\}} \sigma_i^2}$$

Vemos cómo en la última parte de la igualdad, el cálculo de la norma depende de los valores singulares de la matriz en cuestion (σ_i). La media entre tres valores **minimiza la distancia** a cada uno de ellos. Como la matriz D es la única que cambia y el resto de información permanece inalterada (las matrices U y V), podemos estar seguros que esta forma de calcular la matriz esencial nos devuelve aquélla más cercana a la matriz original.

6.(Op. 6 puntos) Discutir cuales son las ventajas y desventajas de usar un algoritmo de reconstrucción Euclídea que calcule la profundidad de varios puntos a la vez en lugar de uno a uno.

La reconstrucción de puntos debe realizarse uno a uno por la sencilla razón de que cada punto tenía una profundidad con respecto a la cámara en el momento en que se tomó la fotografía. Si se desea realizar una reconstrucción lo más precisa posible, la reconstrucción de varios puntos a la vez no cabe en nuestros planes, pues aunque varios puntos estuvieran a la misma profundidad en el momento de la fotografía, no podemos saberlo a priori, de modo que si reconstruímos varios puntos a la vez podríamos estar colocando algunos puntos en profundidades incorrectas, lo que no quiere decir que el método punto a punto no cometa errores. En definitiva, la reconstrucción punto a punto, aunque computacionalmente más pesada, es más precisa.

7.(Op. 10 puntos) Deducir la expresión para la matriz Esencial $E = [t]_x R = R[R^T t]_x$. Justificar cada uno de los pasos

8.(Op. 6 puntos) Dada una pareja de cámaras cualesquiera, ¿existen puntos del espacio que no tengan un plano epipolar asociado? Justificar la respuesta

Todo punto del espacio siempre tendrá asociado un plano epipolar. Esto es así, porque cualesquiera tres puntos en el mundo definen un plano, ahora falta demostrar que este plano sea epipolar.

Un plano es epipolar si y sólo si contiene a la recta base. La recta base es aquella recta definida por los puntos donde se encuentran las dos cámaras. Veámoslo más claro en la siguiente figura:

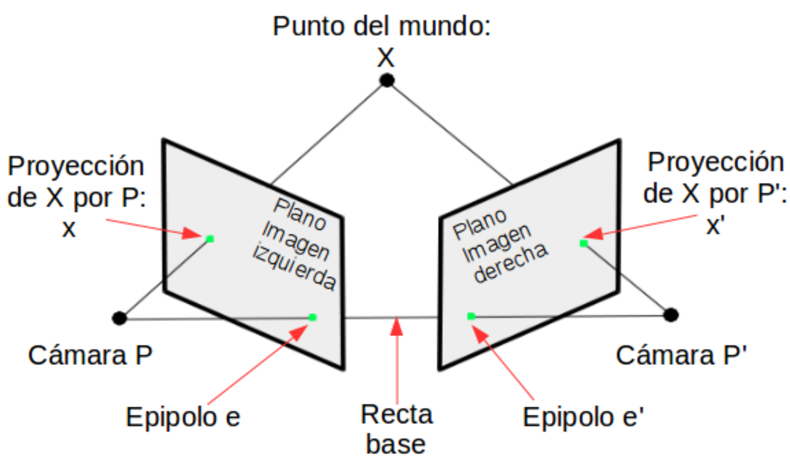

Figura 1. Triangulación de un punto y concepto de recta base.

Dicho esto, no queda mucho que añadir: Si elegimos siempre los dos puntos donde están las cámaras y cualquier punto del mundo, dichos puntos **siempre** forman un plano y éste **siempre** es epipolar ya que **siempre** contiene a la recta base. Respondiendo a la pregunta, **NO existen puntos del espacio que no tengan un plano epipolar asociado**.

9.(Op. 6 puntos) Si nos dan las coordenadas de proyección de un punto escena en la cámara-1 y nos dicen cuál es el movimiento relativo de la cámara-2 respecto de la cámara-1, ¿es posible reconstruir la profundidad el punto si las cámaras están calibradas?. Justificar la contestación

No es posible reconstruir la profundidad. Veamos por qué. Comencemos suponiendo que sí se puede e intentemos conocer las coordenadas del otro punto en correspondencia. Vamos a llegar a una situación de bloqueo.

Llamemos al punto que conocemos proyectado por la cámara-1 x. El hecho de que ambas cámaras estén calibradas implica conocer las matrices de parámetros intrínsecos de ambas cámaras. K y K'. En este momento, podemos calcular la matriz esencial que relaciona ambas imágenes como:

$$E = K'TF^K$$

La matriz E codifica la misma información que la matriz fundamental, pero en este caso, es llamada esencial porque las cámaras están calibradas. Esto significa que E tiene información sobre la geometría epipolar. La relación entre dos puntos en correspondencia con la matriz esencial es la siguiente:

$$\hat{x}'^T E \hat{x} = 0$$

Donde $\hat{x}$ indica que se representa x en coordenadas normalizadas. Para normalizar un punto, basta con multiplicarlo por la izquierda por la inversa de la matriz de parámetros intrínsecos de la cámara por la que fue tomado:

$$\hat{x} = K^{-1}x$$

Pues bien, si tenemos la matriz esencial, y el punto x que conocemos normalizado, esto es, $\hat{x}$, podemos plantear la ecuación siguiente:

$$\hat{x}'^T E \hat{x} = 0$$

En este momento, hemos llegado a una situación de bloqueo, pues independientemente de las coordenadas de $\hat{x}$, y de la matriz esencial E, el único punto que satisface esta restricción es siempre $\hat{x}^T = (0\ 0\ 0)$ y eso no es cierto. Dado que hemos llegado a esta situación, podemos concluir que **no se puede reconstruir la profundidad del punto**, para hacerlo, serían necesarias las coordenadas de ambas proyecciones.

10.(Op. 10 puntos) Suponga que obtiene una foto de una escena y la cámara gira para obtener otra foto ¿Cuál es la ecuación que liga las coordenadas de la proyecciones en ambas imágenes, de un mismo punto escena, en términos de los parámetros de las cámaras. Justificar matemáticamente

11.(Op. 10 puntos) Suponga una cámara Afín. Discutir cuales son los efectos de la proyección ortogonal sobre los parámetros intrínsecos y extrínsecos de la cámara.

12.(Op. 10 puntos) Dadas dos cámaras calibradas, $P = K[I|0]$ y $P' = K'[R|t]$. Calcular la expresión de la matriz fundamental en términos de los parámetros intrínsecos y extrínsecos de las cámaras. Todos los pasos deben ser justificados

Referencias

1. *Multiple View Geometry in Computer Vision.* R. Hartley A. Zisserman. 2004. http://cvrs.whu.edu.cn/downloads/ebooks/Multiple%20View%20Geometry%20in%20Computer%20Vision%20(Second%20Edition).pdf

2. *Norma matricial.* Wikipedia. https://es.wikipedia.org/wiki/Norma_matricial