Sobre un problema de optimización no-lineal en visión estereoscópica. Aspectos Computacionales

Luis Alvarez León[1] Javier Sánchez Pérez[1]

Resumen

Presentamos un método no lineal para la estimación de la geometría 3-D de una escena a partir de 2 imágenes esteroscópicas. El problema principal consiste en calcular la posición relativa de las 2 cámaras a partir de un número de puntos que se corresponden en ambas cámaras. La posición relativa de las 2 cámaras viene dada por un vector de 7 parámetros: $-X = (s, l, m, n, t_x, t_y, t_z)$–. Para calcular estos parámetros hay que minimizar una energía no-lineal del tipo $E(X) = \|Aq(X)|^2$ donde A es una matriz $9x9$ y $q(X)$ es un vector función de X. En este trabajo presentamos un algoritmo para la busqueda de mínimos locales de $E(X)$ basado en una modificación del método de gradiente de paso óptimo. Presentamos algunas experiencias comparativas con otros métodos clásicos.

Introducción

En los últimos años se han investigado diferentes técnicas que permiten determinar la estructura $3-D$ a partir de dos imágenes proyectivas. Para sentar las bases del problema, presentamos en la figura 1 un modelo proyectivo utilizado normalmente en trabajos sobre imágenes en perspectiva (Ver [?],[?]y [?]). Consideramos un conjunto de puntos $3-D$ que se proyectan en cada cámara. Para cada cámara se utiliza un sistema de coordenadas distinto. Denotamos por (x, y, z) las coordenadas $3-D$ de un punto y por (u, v) las coordenadas de la imagen del punto proyectado en una cámara, y denotamos por (x', y', z') las coordenadas $3-D$ de un punto y por (u, v) las coordenadas de la imagen del punto en la otra cámara. Consideramos que se conocen los parámetros intrínsecos de las dos cámaras. Esto significa, en particular, que el sistema de coordenadas de la imagen puede estar normalizado de tal manera que el origen de cada cámara está localizado en el punto de la imagen correspondiente a la interseccción del punto focal con el plano de la imagen; el punto focal está en dirección del eje z (eje z' en la otra cámara), y los vectores unitarios $\hat{u}, \hat{v}$ ($\hat{u}', \hat{v}'$ resp.) están alineados y con la misma magnitud que los vectores unitarios $\hat{x}, \hat{y}$ ($\hat{x}', \hat{y}'$ resp.). En el sistema de referencia $3-D$, podemos asumir también, que se conocen las distancias entre los focos y los planos de las imágenes, (denotamos por D y D' estas distancias). Con esta normalización, obtenemos que el sistema de coordenadas de la imagen de un punto $3-D$ (x, y, z) proyectado en la imagen, viene dado por $(u, v) = (Dx/z, Dy/z)$ ($(u', v') = (D'x'/z', D'y'/z')$ respect.).

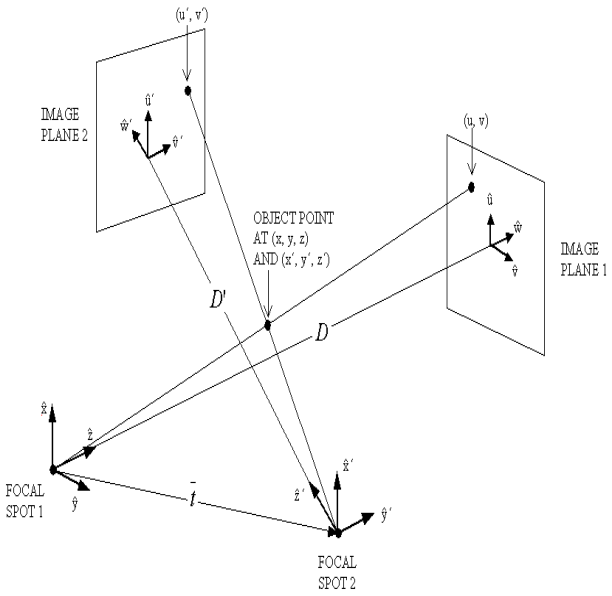

Figura 1: Modelo Proyectivo.

Consideramos un conjunto de puntos $3-D$ que denotamos por (x_i, y_i, z_i) en un sistema de referencia $3-D$ y por (x'_i, y'_i, z'_i) en el otro sistema de referencia $3-D$. La transformación entre los dos sistemas de referencia $3-D$ está definido por una traslación y una rotación rígida, y se puede expresar como

$$\begin{pmatrix} x'_i \\ y'_i \\ z'_i \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \tag{1}$$

donde los r_{ik} son elementos de una matriz de rotación R, y el vector $t = (t_x, t_y, t_z)$ representa la traslación del primer foco al segundo. Asumimos que para cualquier punto $3-D$, $z_i, z'_i > 0$. Asignamos $(\xi_i, \eta_i) = (u_i/D, v_i/D)$ y $(\xi'_i, \eta'_i) = (u'_i/D', v'_i/D')$. Utilizando la ecuación anterior, obtenemos 4 ecuaciones que envuelven las coordenadas escaladas de la imagen (ξ_i, η_i) y (ξ'_i, η'_i). Resolviendo estas ecuaciones se obtiene una sola ecuación que se puede expresar como

$$(\xi'_i, \eta'_i, 1)\, Q \begin{pmatrix} \xi_i \\ \eta_i \\ 1 \end{pmatrix} = 0 \tag{2}$$

donde Q se puede expresar como

$$\begin{pmatrix} r_{13}t_y - r_{12}t_z & r_{11}t_z - r_{13}t_x & r_{12}t_x - r_{11}t_y \\ r_{23}t_y - r_{22}t_z & r_{21}t_z - r_{23}t_x & r_{22}t_x - r_{21}t_y \\ r_{33}t_y - r_{32}t_z & r_{31}t_z - r_{33}t_x & r_{32}t_x - r_{31}t_y \end{pmatrix} \tag{3}$$

Por lo tanto, para cada par de puntos en correspondencia (ξ_i, η_i) y (ξ'_i, η'_i) obtenemos una ecuación dada por (2). Con N puntos localizados en ambas imágenes, el sistema

de ecuaciones resultante se puede escribir como

$$A \begin{pmatrix} q_{11} \\ q_{12} \\ . \\ . \\ q_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{4}$$

donde A es una matriz $Nx9$. La ecuación anterior es homogénea y, por lo tanto, su solución sólo se puede calcular en función de una constante multiplicativa indeterminada. Cuando $N \geq 8$ y los puntos $3-D$ no están en alguna configuración geométrica especial, la solución de la ecuación (4) se puede encontrar minimizando la siguiente energía

$$E(q) = \|Aq\|^2 = q^T A^T A q \tag{5}$$

con la condición $\|q\| = 1$, donde $q = (q_{11}, q_{12}, ..., q_{31})$ es un vector $9x1$ con los elementos de la matriz Q. Ya se sabe que la solución del problema de minimización anterior viene dado por el autovector asociado al autovalor más pequeño de la matriz $B = A^T A$.

Una vez que se ha obtenido Q, se utiliza alguna técnica estándar para calcular, a partir de Q, la matriz de rotación R y el vector de traslación t (en función de un parámetro de escala). Por ejemplo, se puede obtener t como el autovector asociado al autovalor más pequeño de la matriz $Q^T Q$, y la matriz de rotación R se puede calcular utilizando la siguiente expresión

$$\begin{aligned} r_1 &= \frac{q_1 \times t + q_2 \times q_3}{\|t\|^2} \\ r_2 &= \frac{q_2 \times t + q_3 \times q_1}{\|t\|^2} \\ r_3 &= \frac{q_3 \times t + q_1 \times q_2}{\|t\|^2} \end{aligned} \tag{6}$$

donde $r_i = (r_{i1}, r_{i2}, r_{i3})$ y $q_i = (q_{i1}, q_{i2}, q_{i2})$. El principal problema con esta aproximación es que cuando existe ruido en las coordenadas de la imagen, la expresión anterior no genera, normalmente, una matriz de rotación. Esto significa que R no es una matriz ortonormal. En tal caso, se pueden realizar algunas operaciones adicionales para transformar R en una matriz de rotación real. Estas operaciones son de tipo algebraico y no tienen en cuenta la precisión en la ecuación que pone en correspondencia los puntos (2). En [**?**] los autores proponen un método que utiliza algunas técnicas de geometría algebraica que proveen una caracterización de las matrices fundamentales, forzando la restricción de rigidez.

En este artículo, presentamos una nueva aproximación al problema de la recuperación de la matriz de rotación R y del vector de traslación t. Utilizaremos cuaterniones para representar una matriz de rotación R, se conoce (ver, por ejemplo, [**?**]) que, al utilizar cuaterniones, la matriz de rotación R se puede escribir como

$$\begin{pmatrix} s^2+l^2-m^2-n^2 & 2(lm-sn) & 2(nl+sm) \\ 2(lm+sn) & s^2-l^2+m^2-n^2 & 2(mn-sl) \\ 2(nl-sm) & 2(mn+sl) & s^2-l^2-m^2+n^2 \end{pmatrix}$$

donde $s^2+l^2+n^2+m^2=1.$, el vector (l,m,n) representa el eje de rotación y $s=\cos(\alpha/2)$, α representa el ángulo de rotación. Usando la ecuación anterior y (3), deducimos que el vector q se puede expresar como

$$q(X)=\begin{pmatrix} t_y(2nl+2sm)-t_z(2lm-2sn) \\ t_z(s^2+l^2-m^2-n^2)-t_x(2nl+2sm) \\ t_x(2lm-2sn)-t_y(s^2+l^2-m^2-n^2) \\ t_y(2mn-2sl)-t_z(s^2-l^2+m^2-n^2) \\ t_z(2lm+2sn)-t_x(2mn-2sl) \\ t_x(s^2-l^2+m^2-n^2)-t_y(2lm+2sn) \\ t_y(s^2-l^2-m^2+n^2)-t_z(2mn+2sl) \\ t_z(2nl-2sm)-t_x(s^2-l^2-m^2+n^2) \\ t_x(2mn+2sl)-t_y(2nl-2sm) \end{pmatrix}$$

donde $X=(s,l,m,n,t_x,t_y,t_z)$. Utilizando esta formulación, reescribimos el problema de optimización de energía (5) en función de X

$$E(X)=\|Aq(X)\|^2 =^T q(X)^T Bq(X) \tag{7}$$

donde $B=A^TA$, con las restricciones $s^2+l^2+n^2+m^2=1$ y $t_x^2+t_y^2+t_z^2=C^2$. (donde C representa la distancia entre los focos de ambas cámaras, en el caso en que no se conozca C, fijamos $C=1$ y obtenemos sus coordenadas $3-D$ en función de un factor de escala). Nótese que con esta formulación las variable son s,l,m,n,t_x,t_y,t_z, así que tenemos 7 variables en vez de 9 (en el caso de tomar q). y 2 condiciones en vez de 1. Más aún, con esta formulación la matriz de rotación obtenida es perfecta utilizando la ecuación que pone en correspondencia los puntos (2)

Los cuaterniones ya han sido utilizados en [**?**] con el fin de calcular la matriz de rotación R, pero en este caso, los cuaterniones se utilizan para obtener una matriz de rotación real a partir de la matriz Q, sin tener relación con la energía (7).

En el caso en que los parámetros intrínsecos de las cámaras sean desconocidos, la situación es más compleja, sin embargo, podemos llegar a una formulación similar a la presentada en (2), pero en este caso, la matriz Q depende también de los parámetros intrínsecos. En este caso la matriz Q se denomina matriz fundamental, que proporciona la geometría epipolar de las cámaras.

Búsqueda de un mínimo local de la energía no-lineal E(X).

Consideremos el problema de optimización no-lineal (7)

$$E(X) =^T q(X)^T Bq(X)$$

con las condiciones $s^2+l^2+m^2+n^2=1$ y $t_x^2+t_y^2+t_z^2=C^2$.

Primero, calculamos las derivadas de la función $q(X)$ que se pueden calcular fácilmente gracias a que los componentes de $q(X)$ son polinomios. Denotamos por $\nabla E(X)$ el vector gradiente del funcional $E(X)$. Al ser B simétrica, $\nabla E(X)$ se puede escribir como

$$\nabla E(X) = 2^T q(X) \cdot B \cdot Jq(X) \tag{8}$$

donde $Jq(X)$ es el Jacobiano $9x7$ de la función $q(X)$.

Para encontrar un mínimo local de la energía $E(X)$ aplicamos un método de descenso por gradiente. Esto significa que utilizando una primera estimación X^0 del mínimo (en muchas ocasiones se puede suministrar una estimación "a priori" sobre la posición de las dos cámaras a partir de algún tipo de cálculo aproximado), aproximamos el mínimo local más cercano de $E(X)$ como el estado asintótico del esquema iterativo:

$$X^{n+1} = X^n - \rho_n d^n \tag{9}$$

donde X^n representa la aproximación del mínimo local de $E(X)$ en el paso n. d^n es la dirección de descenso en el paso n y ρ_n es un parámetro. Por lo tanto, para calcular X^{n+1} a partir de X^n, necesitamos determinar en cada paso el valor de d^n y ρ_n.

En un problema de minimización sin restricciones, la elección natural de d^n es $\nabla E(X^n)$. Para poder incluir la información de las restricciones en la dirección de descenso d^n, se proyecta el vector $\nabla E(X^n)$ en la intersección del espacio tangente a las superficies $s^2 + l^2 + m^2 + n^2 = 1$ y $t_x^2 + t_y^2 + t_z^2 = C^2$. De esta manera se minimiza la distorsión con respecto a las restricciones de la nueva estimación X^{n+1}. En el siguiente lema, se demuestra cómo se puede calcular esta proyección.

Lema 1 *La proyección del vector $\nabla E(X^n)$ en la intersección del espacio tangente a las superficies $s^2 + l^2 + m^2 + n^2 = 1$ y $t_x^2 + t_y^2 + t_z^2 = C^2$ viene dado por*

$$d^n = (Id - p \otimes p - q \otimes q)\, \nabla E(X^n) \tag{10}$$

donde p y q son vectores unitarios

$$p =^T (s^n, l^n, m^n, n^n, 0, 0, 0)$$
$$q = \frac{^T(0, 0, 0, 0, t_x^n, t_y^n, t_z^n)}{C}$$

y $p \otimes p$, $q \otimes q$ representan la matriz $7x7$ $p^T p$ y $q^T q$.

Demostración: *Por un lado, la dirección normal en el punto X^n de la superficie $s^2+l^2+m^2+n^2 = 1$ es el vector p, y la dirección normal en el punto X^n de la superficie $t_x^2+t_y^2+t_z^2 = C^2$ viene dada por el vector q. Observemos que si el vector Y, es ortogonal a p y q, es decir, Y es la intersección de los espacios tangentes, entonces*

$$(Id - p \otimes p - q \otimes q)\, Y = Y$$

por otro lado, p y q son ortogonales y $\|p\| = \|q\| = 1$

$$(Id - p \otimes p - q \otimes q)\, p = p - p = 0$$
$$(Id - p \otimes p - q \otimes q)\, q = q - q = 0$$

y, por lo tanto, $(Id - p \otimes p - q \otimes q)$ *representa la matriz de proyección en la intersección de los espacios tangentes a* $s^2 + l^2 + m^2 + n^2 = 1$ *y* $t_x^2 + t_y^2 + t_z^2 = C^2$.

Una vez que se calcula la dirección de descenso d^n, se elige ρ_n minimizando la función

$$\Phi(\rho) = E(X^n - \rho d^n) \tag{11}$$

la condición de extremo de la función $\Phi(\rho) = E(X^n - \rho d^n)$ es

$$\Phi'(\rho) = \langle \nabla E(X^n - \rho d^n), d^n \rangle = 0 \tag{12}$$

Para calcular el valor óptimo de ρ se utiliza una aproximación de primer orden de $\nabla E(X)$, esto es

$$\nabla E(X^n - \rho d^n) \cong \nabla E(X^n) - HE(X^n)\rho d^n$$

donde $HE(X)$ es la matriz Hessiana del funcional $E(X)$. Por lo tanto, si sustituimos la función anterior en la ecuación (12), obtenemos:

$$\rho_n = \frac{\langle \nabla E(X^n), d^n \rangle}{\langle d^n, HE(X^n) d^n \rangle} \tag{13}$$

La matriz Hesiana $HE(X^n)$ se puede calcular fácilmente utilizando la expresión

$$HE(X) = 2^T Jq(X) \cdot B \cdot Jq(X) + 2^T q(X) \cdot B \cdot Hq(X) \tag{14}$$

donde $q(X) \cdot B \cdot Hq(X)$ es la matriz $7x7$ dada por

$$(q(X) \cdot B \cdot Hq(X))_{ij} = q(X) \cdot B \cdot \frac{\partial^2 q(X)}{\partial X_i \partial X_j}$$

Nota: *Al ser* d^n *la proyección de* $\nabla E(X^n)$ *en el espacio ortogonal de* p *y* q *entonces existen* λ, μ *tal que*

$$\nabla E(X^n) + \lambda p + \mu q = d^n$$

entonces $\langle \nabla E(X^n), d^n \rangle = \langle d^n, d^n \rangle$, *de tal forma que el numerador en el cálculo de* ρ_n *es igual a cero sí y solo sí* $d^n = 0$ *y, por lo tanto en este caso,* λ *y* μ *representan los multiplicadores de Lagrange de un mínimo local del funcional* $E(X)$ *con las restricciones* $s^2 + l^2 + n^2 + m^2 = 1$ *y* $t_x^2 + t_y^2 + t_z^2 = C^2$ *obtenidos por la técnica de los multiplicadores de Lagrange. En otras palabras, si* $d^n = 0$, *la solución asociada* X^n *satisface la condición del mínimo local suministrado por la técnica de Lagrange.*

Resumiendo, el algoritmo completo para encontrar el mínimo local del funcional $E(X)$ se puede expresar en los siguientes pasos:

1. Se elige un valor inicial para X^0 (por ejemplo, la rotación y traslación obtenidas por el método lineal)

2. Hasta la convergencia de X^n

(a) Se calcula $\nabla E(X^n)$ utilizando ec. (8)

(b) Se calcula d^n utilizando ec. (10)

(c) Se calcula $HE(X^n)$ utilizando ec. (14)

(d) Se calcula ρ_n utilizando ec. (13)

(e) Se calcula $X^{n+1} = X^n - \rho_n d^n$.

(f) Se normaliza X^{n+1} para cumplir con las restricciones.

Experiencias numéricas

La simulación que realizamos consistió en lo siguiente: Elegimos 5 puntos $(3-D)$ de un cubo inscrito en la esfera de centro $(0,0,2)$ y radio 1. Situamos el foco de la primera cámara en el origen y su plano proyectivo, tangente a la esfera de centro $(0,0,2)$ y radio 2, en el punto $(0,0,4)$. El foco de la segunda cámara se sitúa en el punto $(2,0,2)$ sobre la misma esfera y su plano proyectivo, tangente a la misma, en el punto $(-2,0,2)$. Proyectamos los 5 puntos $(3-D)$ en ambas cámaras. En este caso los valores de los parámetros, que determinan la posición de la segunda cámara con respecto al de la primera, vienen dados por $(s,l,m,n,t_x,t_y,t_z) = (\frac{1}{\sqrt{2}}, 0.0, \frac{1}{\sqrt{2}}, 0.0, 2.0, 0.0, 2.0)$, que sería el vector X que minimiza la energía $E(X)$. Para realizar las experiencias numéricas tomamos como aproximación inicial X_0 una perturbación de (X) añadiendole un ruido uniformemente distribuido.

Para esta aproximación inicial (X_0), aplicamos los siguientes métodos de optimización no-lineal:

1. El método propuesto.

2. El método de gradiente paso óptimo (normalizando, en cada iteración, los vectores (s,l,m,n) y (t_x,t_y,t_z)).

3. El método de gradiente paso alterno, en el que en cada iteración se toma de forma alternada las direcciones $d^i = (0,\ldots,\underbrace{1}_{i},\ldots,0)$.

Realizamos 10.000 pruebas distintas para la configuración anterior teniendo en cuenta dos casos distintos: En el primero añadimos un ruido uniformemente distribuido entre $[-0.1, 0.1]$ sobre el vector X, y en el segundo un ruido entre $[-0.25, 0.25]$. En las siguientes tablas mostramos los resultados obtenidos para estos dos casos. Calculamos la media del número de iteraciones necesarias para converger y su desviación estándar para cada método.

Método	Media	Desviación
Método propuesto	4814	2.178
Gradiente paso óptimo	4989	2.233
Gradiente paso alterno	12.770	2.078

Ruido 0.1 sobre X

Método	Media	Desviación
Método propuesto	6.409	2.709
Gradiente paso óptimo	6.695	2.693
Gradiente paso alterno	13.842	2.543

Ruido 0.25 sobre X

De los resultados obtenidos se deduce que el método propuesto converge, de forma general, más rápidamente hacia el resultado final que los otros dos métodos.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la acción integrada Hispano-Francesa HF98-0098 y el proyecto español PB95-1225 de la D.G.I.C.Y.T.

Referencias

[1] O.Faugeras, "3-D computer vision. A geometric viewpoint," *MIT Press*, 1993.

[2] O.Faugeras y S.Maybank "Motion from point matches: multiplicity of solutions," *International Journal of Computer Vision*, Vol. 4(3) pp 225-246, 1990.

[3] K.Hoffmann,C.Metz y Y.Chen "Determination of 3-D imaging geometry and object configurations from two biplane views: An enhancement of the Metz-Fencil technique.," *Med. Phys.*, Vol. 22(8) pp 1219-1227, 1995.

[4] H.C.Longuest-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature* , Vol. 293, 133, 1981.

[5] C.Metz y L.Fencil, "Determination of three-dimensional structure in biplane radiography without prior knowledge of the relationship between the two views: Theory," *Med. Phys.* , Vol. 16(1), pp. 45-51, 1989.

[6] J.Weng,T.S.Huang y N.Ahuja, "Motion and structure from two perspective views: algorithms, error analysis, and error estimation, "*IEEE Trans. Pattern Anal. Machine Intel. PAMI,*, Vol.11 451(1989)

1. Departamento de Informática y Sistemas. Universidad de Las Palmas de Gran Canaria. Campus de Tafira. 35017 Las Palmas. e-mail: {lalvarez/jsanchez}@dis.ulpgc.es, http://serdis.dis.ulpgc.es/ lalvarez