

CORRELACION DE PARES DE IMAGENES PARA MEDICION DE SOLIDOS POR FENOMENOS ESTEREO

*J. Sánchez**, *J. A. Christen⁺*, *B. Valera**

** Centro de Instrumentos UNAM, Circuito Exterior C.U 04510, D.F. Tel
622 8603, Fax 622 8603, sachezj@aleph.cinstrum.unam.mx,
valerab@aleph.cinstrum.unam.mx*

*+ Instituto de Matemáticas UNAM, Morelia Mich. Nicolás Romero 150
Centro. CP 58000, Tel (43) 13 83 64, Fax (43) 12448,
jac@matem.unam.mx*

Resumen

Se hace una descripción óptico-geométrica, de fenómenos estereo, empleados para obtener información 3D, a partir de pares de información 2D. El enfoque principal se centra en los problemas de correlación inteligente de segmentos de fotografías, tomadas sin restricciones. La correlación automática de esta información es muy útil en las tareas de medición de sólidos, donde es deseable alta exactitud, así como el ahorro o reducción de asistencia humana. Se reportan soluciones propuestas en el taller de Industria y Matemáticas, así como otras complementarias del dominio público, empleadas para edición de imágenes.

Abstract

It is made a description of optical principles, which are used to obtain 3D information from 2D information. The main interest is focused, around the problems for the intelligent matching between segments of photographs, which are taken without constrains. The automation of such a matching task, is so useful in measuring solids, where the high accuracy, and also the avoidance of human aid are desirable. Solutions, which were proposed in the Industry and Mathematics Workshop, and procedures of public domain for editing photographs, are proposed as complementary.

Introducción.

El fenómeno estereoscópico era ya explicado por Euclides (280 AC), al señalar que la sensación de profundidad era debida a la diferencia de imágenes que cada ojo percibe al observar un mismo objeto (paralaje

binocular). La psicofisiología actual, ha encontrado más de diez efectos que se complementan en la producción de sensaciones tridimensionales, de las que destacan la sensación de paralaje binocular, la acomodación del cristalino para enfocar en función de la distancia y la convergencia de ejes oculares para establecer la triangulación a diferentes distancias. Los fenómenos de paralaje binocular y convergencia, fueron utilizados por Giovanni Bautista della Porta (1600), quién dibujó escenarios con exactitud, observados desde diferentes ángulos, y produjo las sensaciones de profundidad esperadas (Tekenori¹). El empleo de la fotografía en el siglo XIX, sólo vino a confirmar los viejos descubrimientos de la perspectiva.

En la recopilación de Tekenori se distinguen las imágenes binoculares estereoscópicas, de las imágenes espaciales tridimensionales, siendo el atributo principal que las distingue, la naturaleza de su información; así, el par estereoscópico es único, mientras que la imagen tridimensional cambia continuamente al cambiar el punto de observación. El holograma es un ejemplo de esta imagen tridimensional, pero ha habido otros intentos como arreglos matriciales de fotografías de alta densidad, observadas con arreglos similares de lentes.

La imagería actual ha tenido un auge principalmente en la producción de efectos de video para fines publicitarios y de entretenimiento. Una de sus técnicas emplea objetos reales con trazos, que son digitalizados con cámaras solas o en pares, y luego son reconstruídos y editados por modelos tridimensionales², añadiéndoles propiedades de textura, e introducidos dentro de escenarios computacionales con fuentes de iluminación. La tecnología actual permite navegar estereoscópicamente dentro de tales ambientes, y se le conoce como realidad virtual.

Antecedente inmediato de aplicación industrial de principios estéreo, es la aerofotogrametría, empleada en la reconstrucción de objetos geodésicos. Sus principios se basan en la observación de imágenes fotográficas que sustituyen a objetos reales, al mismo tiempo que unos trazadores recorren el campo visual, produciendo curvas de nivel cuando se les desliza evitando que se impacten o floten sobre los sólidos virtuales. Su exactitud depende de la calidad de la fotografía, de los mecanismos de acoplamiento, y de la calidad de visión de los técnicos.

El advenimiento de la fotografía en los ambientes computacionales, permite un acomodo más libre entre cámaras, y entre éstas y el objeto. Los principios empleados en la conversión de objetos 3D en imágenes 2D, y su reconstrucción 3D, está basada en transformaciones de coordenadas. Los rayos de un conjunto de puntos de un objeto del mundo real (world), con coordenadas $\bar{x}_w = (x_w^1, x_w^2, x_w^3)$, producirán una imagen "volumétrica" en un arreglo óptico, en su región de máximo enfoque: $\bar{x}_{im} = (x_{im}^1, x_{im}^2, x_{im}^3)$. Ver figura 1.

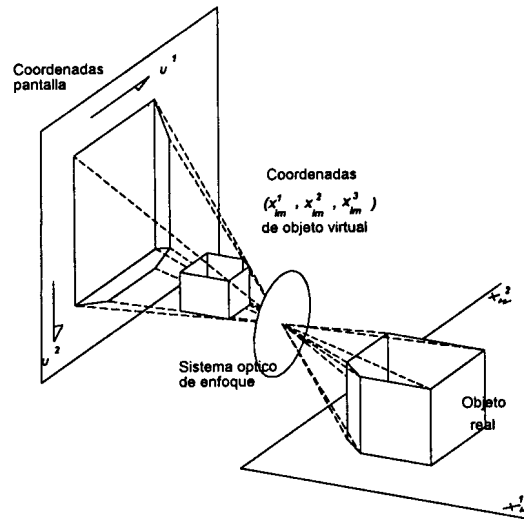


Figura 1. Transformación de coordenadas objeto en coordenadas imagen 3D, y en coordenadas pantalla 2D.

La región de máximo enfoque no es plana, sino que variará de acuerdo a la relación:

$$\frac{1}{do} + \frac{1}{di} = \frac{1}{f} \quad (1)$$

donde do , di , f son las distancias lente-objeto, lente-imagen y distancia focal del lente respectivamente, sobre rayos que viajan en línea recta.

A fin de capturar imágenes sobre pantallas planas, se pueden elegir lentes diafragmadas o de gran profundidad de campo, cuyo análisis se aproxima a

la lente de orificio o "pin hole". En esta cámara idealizada, los puntos objeto, orificio e imagen estarán alineados, y enfocados para cualquier distancia.

Los puntos imagen $\bar{u}=(u_1, u_2)$ de los puntos objeto $\bar{x}_w=(x_w^1, x_w^2, x_w^3)$ se relacionarán en un sistema de coordenadas homogéneas por:

$$\begin{bmatrix} su_1 \\ su_2 \\ s \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & 1 \end{bmatrix} \begin{bmatrix} x_w^1 \\ x_w^2 \\ x_w^3 \\ 1 \end{bmatrix} \quad \text{ó} \quad s\bar{u} = M\bar{x}_w \quad (2)$$

Donde la 2 x 3 submatriz que va de m_{11} a m_{23} contiene términos de rotación y escalamiento, m_{14} y m_{24} son los términos de desplazamiento, y el renglón m_{31} a m_{34} produce los efectos de perspectiva o de escalamiento por distancia. A la matriz $[M]$ se le conoce como matriz de calibración de cámaras, y describe el acoplamiento lineal entre el objeto real y su imagen en películas o detectores.

Para fines de medición precisa, Tsai³, considera que las aberraciones principales tienen una distribución radial, y transforma las coordenadas distorcionadas $\bar{x}_d=(x_d^1, x_d^2)$ en las teóricas $\bar{u}=(u_1, u_2)$, a través de las correcciones no lineales:

$$\begin{aligned} u_1 &= x_d^1 + x_d^1 (k_1(r)^2 + k_2(r)^4), & u_2 &= x_d^2 + x_d^2 (k_1(r)^2 + k_2(r)^4), \\ r &= \sqrt{(x_d^1)^2 + (x_d^2)^2} \end{aligned} \quad (3),$$

donde k_1, k_2 son propiedades internas de las cámaras, y se determinan por un proceso de calibración en bancos ópticos. Nombrando a $m_{i,j}$ y $n_{i,j}$, como los términos de las matrices de calibración de las cámaras izquierda y derecha respectivamente, y a $\bar{u}=(u_1, u_2)$, y $\bar{v}=(v_1, v_2)$, los datos corregidos de sendos detectores, la reconstrucción de puntos $\bar{x}_w=(x_w^1, x_w^2, x_w^3)$ puede obtenerse resolviendo el sistema (4):

$$\begin{bmatrix} m_{11} - m_{31}u_1 & m_{12} - m_{32}u_1 & m_{13} - m_{33}u_1 \\ m_{21} - m_{31}u_2 & m_{22} - m_{32}u_2 & m_{23} - m_{33}u_2 \\ n_{11} - n_{31}v_1 & n_{12} - n_{32}v_1 & n_{13} - n_{33}v_1 \\ n_{21} - n_{31}v_2 & n_{22} - n_{32}v_2 & n_{23} - n_{33}v_2 \end{bmatrix} \begin{bmatrix} x_w^1 \\ x_w^2 \\ x_w^3 \end{bmatrix} = \begin{bmatrix} u_1 - m_{14} \\ u_2 - m_{24} \\ v_1 - n_{14} \\ v_2 - n_{24} \end{bmatrix} \quad (4)$$

Aspectos de trascendencia industrial.

A fin de emplear competitivamente los recursos estereoméricos, en sustitución de máquinas de medición por coordenadas (MMCs), éstos deben ser exactos, voloces, y prescindir de la asistencia humana, o reducirla a un mínimo. El estado del arte reporta^{4, 5} exactitudes de 1/1000 de la dimensión mayor, para accidentes o trazos identificados con códigos luminosos, obtenidos con cámaras formato PAL de 625 x 625 pixeles. La identificación luminosa es necesaria para guiar los algoritmos de búsqueda y apareamiento de pares punto izquierdo $\vec{u}=(u_1, u_2)$ punto derecho $\vec{v}=(v_1, v_2)$. Aunque no es despreciable la bondad del método para algunos objetos que por su diseño o proceso de fabricación cuenten con trazos fácilmente identificables, el espectro de objetos a medir se amplía si en vez de procesar trazos, se procesan propiedades naturales, como: manchas, sombras, texturas, patrones proyectados, etc, que residan en ambas imágenes.

Propósitos de este proyecto.

Este proyecto se propone la correlación automática de pares de imágenes, sin trazos, explotando sólo sus propiedades o accidentes naturales, o patrones proyectados, a fin de obtener los beneficios: a) ganancia de exactitud, pues es más repetible la posición de un centro de gravedad, límites de sombras o intersección de superficies, que la posición de trazos puntuales, debido a la burda discretización en pixeles que hacen los detectores, b) Ahorro de asistencia humana, pues una vez establecida la correlación entre puntos (o zonas) de imagen izquierda con sus equivalentes de la derecha, la identificación de tales accidentes sobre objeto o pares de imágenes para la correlación no es necesaria. Ver figura 2.

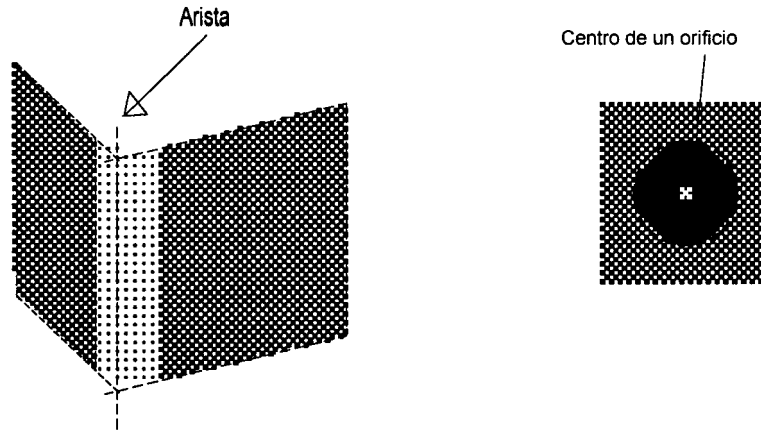


Figura 2. La indefinición de límites en objetos, así como la burda discretización de imágenes en píxeles, hace necesaria la medición de elementos geométricos por procesamiento indirecto.

Problemas típicos presentes en la correlación automática de imágenes, son: a) la definición de contornos de análisis, para ambas imágenes, que aseguren que todos los puntos interiores de una de ellas, tienen su equivalente dentro de la otra, y b) la discontinuidad que una imagen puede presentar sin que se presente en la otra, como en el caso de prominencias del objeto que ocultan regiones para sólo unos ángulos de vista.

Soluciones propuestas.

La propuesta obtenida en el taller consiste en partir de un accidente perfectamente localizado para ambas imágenes, y suponer que en su vecindad, su superficie puede aproximarse a un plano, o que se proyecta a un plano tangente en dicho accidente, con poco error. Enseguida, suponer que el plano se comporta como un espejo, que proyecta el accidente de la imagen izquierda en la derecha, siguiendo leyes de simetría al rededor de la normal al plano. Ver figura 3.

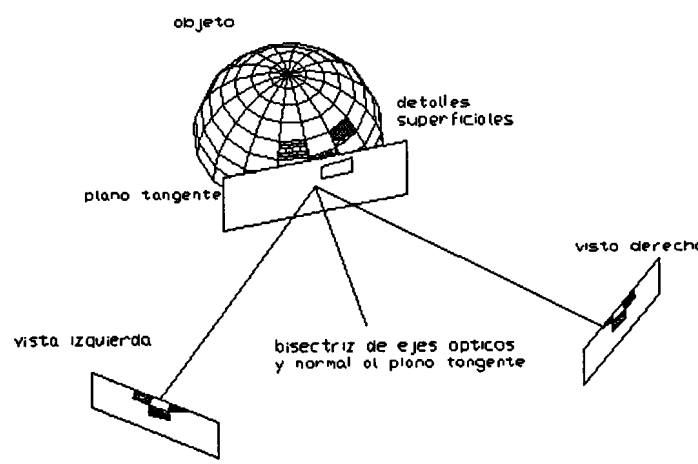


Figura 3. Obtención de la reflexión de un parche planar, sobre un "espejo" normal a la bisectriz de ejes de cámaras.

En el proceso de correlación, se elegirá un segmento plano rectangular o trapezoidal de la imagen izquierda, que se estime no contiene discontinuidades, y por un proceso montecarlo, se buscarán las orientaciones de la normal que mejor correlacione los brillos del parche planar izquierdo con los del derecho obtenido por la reflexión. La máxima correlación puede hacerse por la minimización de:

$$E = \sum_i^n [I'(\bar{x}'_i) - I(\bar{x}_i)]^2 \quad (5)$$

donde I e I' representan los brillos de las posiciones de la imagen real y proyectada de los puntos $\bar{x}_i = (x_i^1, x_i^2)$ y $\bar{x}'_i = (x_i'^1, x_i'^2)$ respectivamente, para todos los n puntos de interés o para todos los puntos o píxeles de esa región.

Cuando se alcanza una buena minimización con (5), se pueden seleccionar los i pares de información $\bar{x}_i = (x_i^1, x_i^2)$ y $\bar{x}'_i = (x_i'^1, x_i'^2)$ que más ventajas ofrezcan para la tarea de medición 3D. Después del proceso de correlación de un trapecio, se pueden elegir cantos o esquinas de éste, para iniciar un nuevo proceso de correlación, hasta completar un mosaico. En los procesos recursivos de correlación, en que se han incluido criterios de escalamiento de muestras (ventanas) por analizar, se pueden incluir criterios

adicionales para interpretar que cuando una de ellas se ha escalado a cero sin obtener correlación, se pudo haber llegado a condiciones de discontinuidad.

El segundo intento ahora en desarrollo, se basa en la técnica de Szeliski⁶, usado para editar fotografías traslapadas. La región de traslape produce fenómenos estereo, ya que detalles en el extremo derecho de la fotografía izquierda se deformán de manera diferente en el extremo izquierdo de la fotografía derecha; sin embargo, las fotografías planas pueden aproximarse a segmentos cilíndricos, que serían la proyección de un escenario continuo a grán distancia. Este cilindro puede, a su vez, desdoblarse, permitiendo establecer traslapes con transformaciones planares de R^2 en R^2 por medio de:

$$\begin{bmatrix} s'x^{1'} \\ s'x^{2'} \\ s' \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 \end{bmatrix} \begin{bmatrix} sx^1 \\ sx^2 \\ s \end{bmatrix} \quad \text{ó} \quad s'\mathbf{x}' = M s\mathbf{x} \quad (6)$$

y la minimización de (5). Este proceso a su vez se puede dirigir empleando la sensibilidad del error de correlacion E con respecto a cada valor m_j :

$$m_j^{(k)} = m_j^{(k-1)} \frac{\partial E}{\partial m_j} \quad (7)$$

donde $k = 1, 2, \dots$ representa el paso del proceso recursivo. El proceso de edición se completa transformando un segmento en el otro de forma suave, por un proceso de mezcla proporcional, que parte de los extremos sin traslapar, hasta los completamente inmersos en la fotografía adyacente. Esta técnica se emplea también para producir sensaciones tridimensionales a partir de fotografías únicas deformadas, presentadas por pares.

Representación gráfica de transformaciones planares y no planares.

Las tomas presentadas en la figura 4, coinciden con los ensayos de Szeliski, que con transformaciones planares pueden convertir o aproximar cada segmento traslapado en el otro. Sin embargo, aunque la región traslapada ha sufrido rotación y traslación, las variaciones de perspectiva son pequeñas, y los objetos volumétricos han sufrido sólo transformaciones planares.

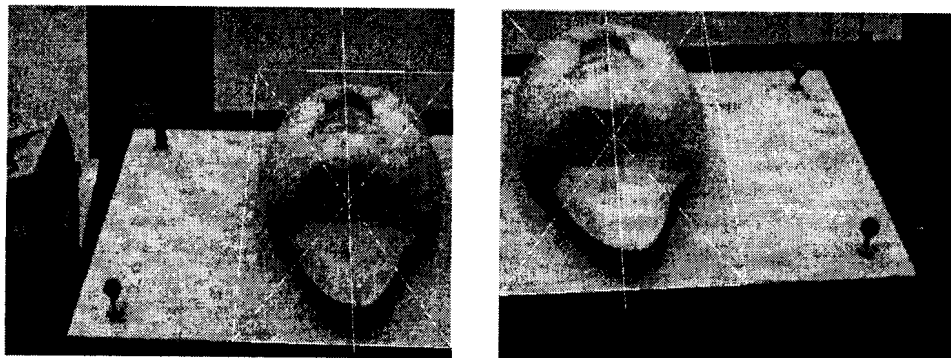


Figura 4. Las fotografías tomadas desde cámaras que pivotan, son editables en las áreas que comparten, y por lo tanto correlacionables con técnicas de transformación planar.

Como puede observarse de la figura 5, para grandes traslaciones de cámaras, las imágenes no equivalen a traslapes de una en otra, por lo que no le son aplicables los procesos de Szeliski. Finalmente, el inicio de un proceso de correlación puede iniciar por un proceso de reflexión y ajuste montecarlo, que permita mapear las áreas correlacionables, y delimitar subregiones interiores equivalentes a "áreas traslapadas", para continuar con procesos donde el ajuste planar sobre ellas produzca poco error.

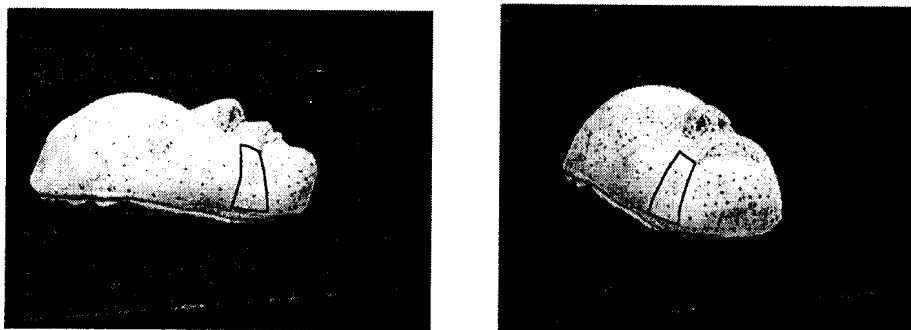


Figura5. El cambio de curvatura de los límites de un parche de superficie curva, revela la imposibilidad de transformar planarmente ambos segmentos de imágenes, pero trapecios interiores pueden ser transformados con poco error.

Referencias

1. Three Dimensional Imaging Techniques. Tekenori Okoshi. Academic Press 1976.
2. Jim Landers et al, 3D Studio Versión 4.0, de Autodesk, Prentice Hall. 1996
3. R.Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. IEEE Computer society conference on computer vision. (Florida 1986), pp 364-374.
4. Aguilar JJ et al. Calibración de cámaras ccd para su uso en estereometría. Quinto Congreso Internacional de Metrología Industrial. Zaragoza España, Noviembre 1991. Proceedings pp 287-296.
5. J. Weng. Calibration of stereo cameras using a no linear distorsion model. Pattern Recognition 16-20 June 1990. pp 246-253.
6. Richard Szeliski. Video mosaics for virtual environments. IEEE Computer Graphics and applications. V 16, N 2, March 96.