

ESTUDIO DE ALGORITMOS DE CÁLCULO DE DISPARIDAD PARA EL CONTROL DE CONVERGENCIA EN SISTEMAS ESTEREOSCÓPICOS

Rafael Aracil, Luis M. Jiménez, Oscar Reinoso, César Fernández

DCEYT, Universidad "Miguel Hernández", Campus de Elche 03212 Alicante, Spain, Tel. +34-96-665-8655, Fax. +34-96-6658798, E-mail: r.aracil@umh.es

Abstract: En el marco de un proyecto de teleoperación, se analiza el problema de reconstrucción 3D a partir de un sistema estereoscópico. El artículo contempla las distintas fases del proceso y propone procedimientos efectivos y poco costosos computacionalmente. Adicionalmente, se discuten los resultados ofrecidos por diversos algoritmos encontrados en la literatura.

Palabras Clave: Control de Convergencia, Visión Estéreo, Cálculo de Disparidad.

1. INTRODUCCIÓN

Este artículo describe un algoritmo de control de convergencia para un sistema estereoscópico de reconstrucción 3D, desarrollado dentro del marco del proyecto ROBTET en colaboración con las empresas IBERDROLA S.A. Y COBRA S.A.

En la solución del problema de reconstrucción 3D mediante visión estereo binocular, muchos autores asumen que los ejes ópticos de las cámaras son paralelos y están en un plano ortogonal a los planos de imagen. Esta simplificación reduce la complejidad en la búsqueda de la correspondencia, al convertir la restricción epipolar a una línea horizontal, y la reconstrucción 3D al establecer una sencilla ecuación entre la disparidad horizontal y la profundidad. A pesar de ello, esta simplificación es difícil de lograr en la práctica en un sistema mecánico real. Generalmente se introduce la técnica de rectificación para re proyectar las imágenes originales sobre un sistema de ejes paralelos. Esta re proyección plantea el problema de cuantificación de las imágenes y no supone una disminución del coste computacional ya que este se traslada al proceso de rectificación.

La utilización de sistemas binoculares para extracción de profundidad de los objetos de una escena, exige establecer la geometría de los ejes de proyección de los sensores de acuerdo a la naturaleza de la escena. Esta adecuación viene impuesta por la necesidad de fijar objetos de

interés de la escena, estableciendo el solape necesario para la correspondencia estereo. La convergencia de los ejes focales constituye una herramienta básica para obtener unos resultados correctos en los algoritmos de correspondencia estereo al disminuir la disparidad sobre objetos de interés de la escena, pudiendo trabajar de forma adecuada con objetos situados a diferentes distancias del observador.

Por otra parte, la convergencia establece también una medida de distancia sobre el punto de fijación (punto intersección de los ejes focales sobre un objeto de la escena) a través de la triangulación obtenida por la intersección de los ejes focales. Por tanto el proceso de convergencia establece un mecanismo de adaptación de la geometría del sistema sensor a la escena analizada, proporcionando una estimación de profundidad puntual. Así mismo el control de convergencia actúa manteniendo la región de interés del análisis tridimensional sobre el punto de fijación.

Esta relación permite el planteamiento de algoritmos cooperativos junto a otra técnica de extracción de profundidad como la estereoscopia o el enfoque. Abbott y Ahuja [Ahuja-93], Krotkov [Krotkov-89] Jimenez [Jiménez-95] utilizan este tipo de interrelaciones. De este modo la convergencia permite al sistema concentrarse en áreas de interés para obtener una mayor precisión y confianza en la medida de profundidad, presentando una serie de ventajas como la simplificación matemática al estar los puntos del objeto de interés próximos al eje óptico

de ambas cámaras, permitiendo simplificaciones en el modelo geométrico del par estereoscópico (perspectiva débil). Así mismo facilita la correspondencia estéreo ya que el punto de fijación tiene disparidad nula, y las zonas vecinas tendrán disparidades pequeñas (figura 1). Esto disminuye el número de candidatos en la correspondencia y por tanto disminuye la posibilidad de falsas correspondencias, especialmente cuando existen patrones repetitivos en la escena [Nishijara-84ab][Olson-91a]. Los estudios de Julesz [Julesz-60,71] sobre la visión binocular humana corroboran esta apreciación, ya que la fusión estéreo solo se produce dentro de un rango máximo de disparidades (límite de fusión de Panum).



Figura 0 Algoritmo cooperativo de correspondencia estéreo en un sistema convergente.

2. EL PROBLEMA DE CONVERGENCIA

El problema de convergencia de un sistema binocular consiste en mantener los dos ejes focales coincidentes sobre un punto de un objeto denominado punto de fijación (figura 2). El ángulo de convergencia es el ángulo formado por ambos ejes focales. El ángulo de convergencia junto a la línea base y el ángulo de mirada establecen geoméricamente el punto de fijación en el espacio.

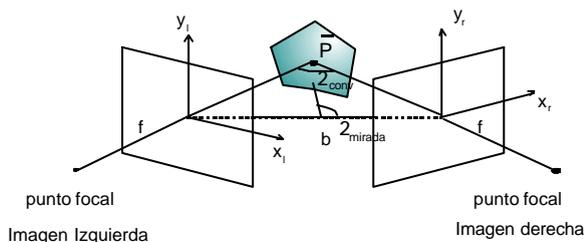


Figura 2 Sistema estereoscópico. P punto de fijación 2_{conv} : ángulo de convergencia. 2_{mirada} : ángulo de mirada

El ángulo de convergencia establece por tanto la distancia desde la cámara al punto de fijación, de este modo se puede considerar el problema de convergencia como el control de la distancia del objeto a la cámara. Se plantean por tanto dos subproblemas: el primero es seleccionar un punto de fijación (un objetivo) en la escena, y en segundo lugar estimar el ángulo de convergencia para ese punto de fijación.

El primer problema conlleva un análisis de la escena bidimensional en busca de propiedades que permitan segmentar los objetos de la misma, como pueden ser contornos, texturas, regiones, etc. Esta información permite seleccionar la región de interés. Se trata por tanto de un problema más general.

El segundo problema es más específico del control de convergencia suponiendo la estimación del ángulo que establece la intersección de los dos ejes focales sobre el punto de fijación en el espacio, perteneciente al objetivo. En este caso cualquier información de profundidad permite calcular el ángulo de convergencia, o al contrario, a partir del cálculo del ángulo de convergencia se puede estimar la profundidad.

La convergencia de un sistema binocular puede establecerse mediante mecanismos independientes en cada cámara, o ser dirigida de forma coordinada teniendo en cuenta, no solo la información bidimensional, sino también la información tridimensional que se genera a partir del proceso de convergencia.

Generalmente no se tiene información de profundidad, o ésta es grosera, ya que el proceso de convergencia es una etapa previa al de extracción de profundidad apoyando a métodos como la estereoscopia. Algunos sistemas (figura 1) implementan una realimentación entre ambos procesos (cálculo de convergencia y cálculo de profundidad) permitiendo en sucesivas iteraciones mejorar los resultados de cada método actuando por separado.

En todo caso el proceso de convergencia puede considerarse en dos situaciones:

1. **Estática:** para un punto de fijación determinado establecer el ángulo de convergencia que mantenga la zona de interés de la adquisición

sobre dicho punto. Produce saltos bruscos en el ángulo de convergencia

2. *Dinámica*: Mantener el ángulo de convergencia sobre un punto de fijación móvil sobre la escena. Las variaciones de posición del objetivo producen generalmente pequeñas variaciones del ángulo de convergencia.

Un sistema de convergencia completo debe considerar ambos aspectos: conseguir la convergencia y mantenerla sobre el objetivo.

3. ESTRATEGIAS PARA EL CONTROL DE CONVERGENCIA.

El proceso de convergencia se puede dividir en dos fases: la primera consiste en seleccionar el objetivo que puede realizarse mediante un filtrado para detectar los contornos de las superficies, o bien tomando una cámara como dominante. La segunda fase consiste en variar el ángulo de convergencia en función de un estimador de error hasta que éste se anule. Para poder decidir el ángulo de convergencia para poder alcanzar el punto de fijación es necesario una estimación del error de convergencia. El objetivo es calcular este error a partir de la información visual y ejecutar una acción de control que lo haga nulo, es decir, lograr que la intersección de los dos ejes focales esté situada sobre la superficie del objeto. En la literatura se pueden encontrar diferentes técnicas para dar una estimación de este error [Ahuja-93], [Coombs-92], [Olson-91b]:

* Una primera aproximación es utilizar una medida de profundidad obtenida a partir de la **disparidad binocular** entre los centros focales de las imágenes pero requiere resolver el problema estéreo previamente, lo cual es difícil y costoso si la disparidad es elevada. La ventaja es que la relación entre el ángulo de convergencia y la disparidad es simple y directa, e incluso no es necesario utilizar valores absolutos de ángulo de convergencia. El inconveniente es que los estimadores de disparidad estéreo suelen estar optimizados para obtener una buena localización y densidad de características pero no para ser robusto y capaz de ejecutarse en tiempo real al necesitar una etapa de optimización global.

* Otro tipo de técnicas son las denominadas "registration" que operan sobre el espacio de la imagen trasladando la imagen hasta que las

regiones solapadas son similares. La operación de correlación se puede ver como una convolución. La correlación se puede representar como la siguiente fórmula:

$$r = \sum_{i=1}^N I_i M_i$$

Desafortunadamente con la correlación ordinaria, los resultados son erróneos cuando las imágenes presentan niveles de gris altos (255). La solución es algo más compleja al utilizar una versión normalizada de la función de correlación:

$$r = \frac{N \sum IM - \sum I \sum M}{\sqrt{\left[N \sum I^2 - \left(\sum I \right)^2 \right] \left[N \sum M^2 - \left(\sum M \right)^2 \right]}}$$

La **correlación cruzada** entre ventanas de ambas imágenes es utilizada como medida de disparidad.



Figura 3 Correlación cruzada entre imágenes tomadas por un par estereoscópico.

Este tipo de medidas se utilizan también en correspondencia estéreo cuando las variaciones de intensidad son relativamente continuas y suaves (generalmente esto es cierto en imágenes

aéreas)[Ahuja-93]. Estas medidas tienen problemas cuando el gradiente de las superficies es elevado o cuando hay oclusión del punto de fijación en cada una de las imágenes. El fenómeno de oclusión está ligado con el límite del gradiente de disparidad [Mayhew-83][Faugeras-93], problema que subyace también en la visión estereoscópica.

* **La transformada de Fourier** resulta útil para realizar esta estimación de error debido a que permite detectar el eco espacial entre las dos imágenes. Las dos imágenes convergentes serán similares y con un cierto desplazamiento que se traducirá en un eco en el espectro de frecuencias. el problema radica el coste computacional de la transformada de Fourier. Una de las implementaciones que utilizan esta transformada es la del filtro **Cepestal** (Yeshurum y Schwartz) [Yeshurum-89]. El "*cepestrum*" de una señal es la transformada de Fourier del logaritmo del espectro de potencia de la señal. Fue inicialmente desarrollado por Bogert [Bogert-63] para analizar señales con eco.

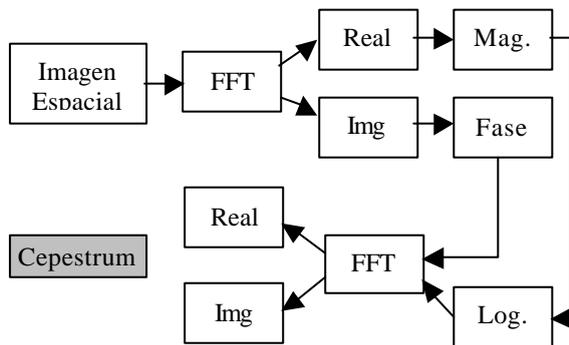


Figura 4 Formación del Filtro Cepestal.

Yeshurum y Schwartz [Yeshurum-89] han aplicado el filtro cepestal bidimensional como estimador de disparidad. Toma ventanas $(h \times w)$ en cada imagen y las une lado a lado formando un imagen $(h \times 2w)$. Coombs [Coombs-92] presenta un estimador derivado del filtro cepestal caracterizado por un comportamiento similar pero con un coste computacional menor. Este estimador está relacionado con la correlación de fase y se basa en un estudio del efecto producido por el filtro cepestal. El espectro de potencia, al ser la transformada de Fourier de la función de autocorrelación, es real y par, y por tanto la transformada de Fourier y su inversa son equivalentes. El término logarítmico es lo que lo

diferencia de la autocorrelación, actuando como un filtrado no lineal.

Este filtro reduce la contribución de señales de banda estrecha producida por patrones periódicos y objetos suaves y grandes que tienen pobres características para correlar, y deja las señales de banda ancha casi inalteradas.

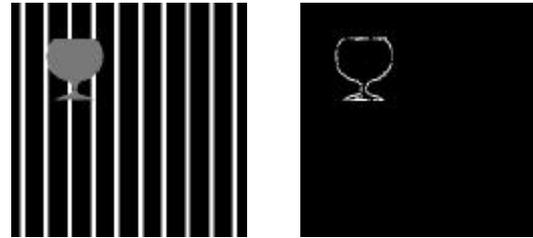


Figura 5 Filtro Cepestal: Eliminación de patrones periódicos que dificultarían la correlación.

* **Correlación de fase.** La disparidad entre dos señales que están desfasadas puede modelarse como una variación de fase en la transformada de Fourier. Partiendo del análisis del filtro cepestal se derivan los filtros de correlación de fase. En este caso se debe utilizar la información de fase que se pierde en el espectro de potencia. Kuglin y Hines [Kuglin-75] y Person [Person-77] describen una implementación de este algoritmo. Este método requiere menor coste computacional al procesar ventanas de dimensión $(h \times w)$ y no $(h \times 2w)$ del filtro cepestal.

RESULTADOS

La implementación realizada está formada por un algoritmo basado en la correlación de fase para el mantenimiento de la convergencia sobre el objeto de interés. Los cambios bruscos de convergencia producidos por la variación de los objetos de interés es dirigida mediante una estimación de la profundidad generada por un sistema de correspondencia estéreo [Jiménez-95].

Para ciertos elementos de la escena que tienen una colinealidad con el plano epipolar se producen grandes errores tanto en la estimación de profundidad, como en la correlación de fase, presentándose falsas correspondencias o puntos de fijación alejados de los objetos de la escena. Este tipo de situaciones se presentan con los cables anclados a la torre (objetos estrechos con una alineación próxima a la horizontal). Para asegurar una convergencia y extracción de profundidad correcta se ha introducido un sistema

de iluminación láser que simplifica la selección del punto de fijación y por tanto la medida de profundidad [Aracil-97].

BIBLIOGRAFÍA

[Ahuja-93] Ahuja, N., Abbot, A.L. "ACTIVE STEREO: INTEGRATING DISPARITY, VERGENCE, FOCUS, APERTURE, AND CALIBRATION FOR SURFACE ESTIMATION", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 15, no. 10, pp.1007-1029, October 1993.

[Aracil-95a] Aracil, R. y col. "3D DYNAMIC NON STRUCTURED ENVIRONMENT MODELLING BY MEANS ARTIFICIAL VISION, FOR ROBOT TRAJECTORY PLANNING", Technical Report, 1995.

[Aracil-95b] Aracil, R. y col. "TELEOPERATED SYSTEM FOR LIVE POWER LINES MAINTENANCE", Proc. of SPIE-The International Society for Optical Engineering. (International Symposium on Intelligent Systems and Advanced Manufacturing.) Vol.2590. pp. 40-49, Octubre 1995.

[Aracil-97] Aracil, R. y col. "OBSTACLE AVOIDANCE FOR TELEOPERATED ROBOTS FOR LIVE POWER LINES MAINTENANCE USING ARTIFICIAL VISION", Proc. Int. IFAC Conf. on Autonomous Vehicles, Sevilla, Spain, 1997

[Bajcsy-88] Bajcsy, R. "ACTIVE PERCEPTION", Proc. of the IEEE, vol. 76, n° 8, pp.996-1005, August 1988.

[Ballard-89] Ballard, D.H., "REFERENCE FRAMES FOR ANIMATE VISION", Proc. Int. Joint Conference on Artificial Intelligence, 1989.

[Bogert-63] Bogert B.P., M.J.R. Healy, J.W. Tukey "THE FREQUENCY ANALYSIS OF TIME SERIES FOR ECHOES: CEPESTRUM, PSEUDO-AUTOCOVARIANCE, CROSS-CEPESTRUM, AND SHAPE CRACKING", Proc. Symp. Time Series Analysis, pp. 209-243, New York, 1963.

[Coombs-92] Coombs, J. D. "REAL-TIME GAZE HOLDING IN BINOCULAR ROBOT VISION", Phd Thesis. University of Rochester, 1992.

[Faugeras-93] Faugeras, O. "THREE-DIMENSIONAL COMPUTER VISION", The MIT Press, 1993

[Jiménez-95] Jiménez, L.M. Aracil, R., Sebastian J.M., Peñín L.F., Ferre M., Torres F. "MODELADO DE ENTORNOS 3D PARA PLANIFICACIÓN DE TRAYECTORIAS DE ROBOTS MEDIANTE VISIÓN ESTEREOSCÓPICA Y CONTROL ACTIVO DE PARÁMETROS ÓPTICOS", Proc. XVI Jornadas de Automática (Grupo de Trabajo Visión Artificial), San Sebastián, Septiembre 1995.

[Julesz-60] Julesz, B. "BINOCULAR DEPTH PERCEPTION OF COMPUTER-GENERATED PATTERNS", Bell Syst. Tech. J., vol. 39, pp.1125-1162, 1960.

[Julesz-71] Julesz, B. "FOUNDATIONS OF CYCLOPEAN PERCEPTION", Univ. of Chicago Press, 1971.

[Krotkov-89] Krotkov, E.P., "ACTIVE COMPUTER VISION BY COOPERATIVE FOCUS AND STEREO", Springer-Verlag, New-York, 1989.

[Kuglin-75] Kuglin, C.D., Hines, D.C. "THE PHASE CORRELATION IMAGE ALIGNMENT METHOD", Proc. IEEE Int. Conference on Cybernetics and Society, pp.163-165, 1975.

[Mayhew-81] Mayhew, J.E.W. and Frisby, J.P. "PSYCHOPHYSICAL AND COMPUTATION STUDIES TOWARDS A THEORY OF HUMAN STEREOPSIS", Artificial Intelligence, 17, pp.349-386, 1981.

[Nishihara-84a] Nishihara, H.K. "PRACTICAL REAL-TIME IMAGING STEREO MATCHER", Opt. Eng., vol. 23, pp.536-545, Sept. 1984.

[Nishihara-84b] Nishihara, H.K. and Poggio, T. "STEREO VISION FOR ROBOTICS", R. Paul and M. Brady, editors, Robotics Research, The First International Symposium, pp.489-505, MIT Press, 1984.

[Olson-91a] Olson, T.J., "STEREOPSIS FOR FIXATING SYSTEMS", Proc. IEEE Int. Conf. On Systems, Man and Cybernetics, Charlottesville, Virginia, USA. October, 1991

[Olson-91b] Olson, T.J., Coombs, D.J. "REAL-TIME VERGENCE CONTROL FOR BINOCULAR ROBOTS", Int. Journal of computer Vision, vol.7, no.1, pp.67-89, November, 1991

[Person-77] Person J.J., Hines D.C., Golosman S. Kugli C.D., "VIDEO-RATE IMAGE CORRELATION PROCESSOR", SPIE v.119, Applications of Digital Image Processing, pp.197-205, San Diego, 1977.

[Yeshurun-89] Yeshurun Y., Schwartz E. "CEPESTRAL FILTERING ON A COLUMNAR IMAGE ARCHITECTURE: A FAST ALGORITHM FOR BINOCULAR STEREO SEGMENTATION", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. PAMI-11, no.7, pp.759-767, July 1989.