

CONTROL DE ENFOQUE DE UN SISTEMA ACTIVO DE VISIÓN ESTEREOSCÓPICA

L.M. Jiménez⁽¹⁾, C. Fernández⁽¹⁾, R. Puerto⁽¹⁾, N. García⁽¹⁾, J.M. Sabater⁽¹⁾, F. Torres⁽²⁾

⁽¹⁾ Dpto. de Ingeniería, Universidad Miguel Hernández, Avda. del Ferrocarril s/n, Elche (Alicante) España
E-mail: luis.jimenez@umh.es

⁽²⁾ Dpto. de Física Aplicada, Ingeniería de Sistemas y Teoría de la Señal, Universidad de Alicante

RESUMEN.

Se presenta un sistema de control de enfoque para un sistema de visión tridimensional activo, dentro de un marco de integración de las técnicas de visión estereoscópica y del control activo de los parámetros de convergencia y enfoque de un cabezal estereoscópico. Se analizan diferentes implementaciones de medida de desenfoque, y se presenta un algoritmo de control de enfoque automático, analizando las interrelaciones con la reconstrucción tridimensional estereoscópica.

Palabras clave: *Visión activa, control de enfoque.*

1. INTRODUCCIÓN

La visión artificial constituye en la actualidad una de las técnicas más atractivas e innovadoras, siendo objeto de investigación en los principales centros del mundo y poseyendo importantes aplicaciones desarrolladas. La mayor parte de dichas aplicaciones presentan como característica común el trabajar sobre imágenes bidimensionales, bien porque sean objetos planos, o bien porque la información que se quiere obtener esté suficientemente condensada en una o varias proyecciones. Esto supone una fuerte restricción en el amplio campo de posibilidades de la visión artificial. Para paliar esta limitación, se han desarrollado durante la última década estructuras que permiten obtener la información tridimensional de una escena, siendo una de las más utilizadas la constituida por un par estereoscópico de cámaras, a semejanza de la visión humana. Sin embargo son bastantes los problemas que aún no han sido resueltos plenamente y que dan lugar tanto a una pérdida en la precisión de los resultados obtenidos, como a un incorrecto análisis de aquellas escenas que por su complejidad no permiten la adecuada correspondencia entre las imágenes del par estereoscópico. Uno de los problemas que se plantean es la dependencia que se establece entre la distancia de las cámaras al objeto, que es uno de los resultados finales que se pretende conseguir, y la nitidez de las imágenes, que requiere el conocimiento previo de dicha distancia. Para solventarlo es necesario disponer de algoritmos que permitan estimar de forma iterativa cuando una imagen es nítida, a partir de los datos existentes en la misma.

2. RECUPERACIÓN DE INFORMACIÓN TRIDIMENSIONAL

La posibilidad de fijar la mirada en un determinado motivo de la escena, móvil o no, o la de variar rápidamente la atención de una zona a otra, aptitudes inspiradas en los órganos visuales del ser humano, constituyen áreas de creciente interés en la comunidad científica [1][2][3]. Un sistema sensorial visual de tales características disminuye notablemente la complejidad del tratamiento e interpretación de los datos visuales.

Dentro de las técnicas de recuperación tridimensional de la escena se puede realizar una primera clasificación en técnicas *pasivas* y *activas* [4]. Las primeras se basan en información captada de forma pasiva sin modificar las condiciones de captación de la escena. Por contra, las técnicas *activas* modifican las condiciones del entorno con el fin de extraer información de su estructura tridimensional. Cada una de ellas presenta características específicas en todos los niveles del proceso de interpretación visual, desde la etapa de formación de la imagen, hasta el análisis e interpretación de la misma, requiriendo tanto de equipos como de algoritmos específicos. Dentro de cada uno de los grupos anteriores se identifican un conjunto de técnicas basadas en diferentes principios [5][6] como son la *visión estereoscópica*, *cámara móvil*, *enfoque*, *textura*, *sombreado*, *luz estructurada*, *telemetría*...

Las características y limitaciones de cada proceso de extracción de profundidad de la escena llevan a la necesidad de integración, no solo de medidas, sino también de técnicas que faciliten la *fusión estereoscópica*. Los estudios sobre la visión humana permiten observar ese

proceso de integración. Por ejemplo, la fusión binocular humana no se produce si la disparidad es superior al límite de Panum [7][8], lo que supone que el movimiento ocular sea importante en la fusión estereoscópica. La integración de técnicas ha sido objeto de un gran tratamiento en los últimos años, así Krotkov [2] presenta un sistema estéreo utilizando la información de estereo, enfoque y convergencia de forma cooperativa. Abbot y Ahuja [9] integran enfoque, estereo, convergencia e interpolación de superficies para ir extendiendo el mapa del entorno. Cada uno de los métodos descritos presenta unas características específicas en tres aspectos diferentes [9]: características de las imágenes requeridas, capacidad y complejidad, y precisión.

La *estereoscopia* basada en características (bordes o segmentos de borde) requiere una localización de detalles de alta frecuencia lo que supone una escena bien enfocada en los objetos principales de la misma. Esto implica que la profundidad de campo del sistema de formación de imagen debe incluir todos los objetos de interés, lo cual supone un valor elevado de la misma. Esta característica se contrapone a la de un sistema de enfoque que precisa una pequeña profundidad de campo para obtener una precisión aceptable. La segunda interrelación importante entre *estereoscopia* y *enfoque* es el carácter monocular de la segunda que le permite obtener una estimación de profundidad más robusta en escenas donde la falta de visibilidad de cierta característica se produce en alguna de las imágenes, o cuando en las imágenes existen zonas con gradiente de disparidad elevado en las cuales los algoritmos estereo tienden a producir falsas correspondencias debido a la imposición de las restricciones de continuidad de la escena. Este efecto es más acusado en los objetos huecos formados por estructuras finas, donde se suele incumplir la condición de orden y pueden existir bordes muy próximos pero con disparidades significativamente diferentes.

3. SISTEMA DE CONTROL DE ENFOQUE DE UNA ÓPTICA MOTORIZADA

Una tarea fundamental dentro de un sistema de visión activa es mantener enfocados los objetos de interés dentro de la superficie sensora. En fisiología se denomina "*acomodación*" a la acción del cristalino para enfocar la imagen en la retina. Este fenómeno de acomodación, mantiene enfocados los objetos de primer plano ó interés con el movimiento de la cabeza y constituye en sí mismo una fuente de información de la distancia de los objetos. Las tareas que debe resolver un sistema de enfoque en visión activa son dos. Dada la proyección P' de un punto $P=(X,Y,Z)^T$ (Z desconocido) perteneciente a la superficie de un objeto de la escena:

- 1.-Sin información acerca de Z , ¿cómo se puede determinar la posición del motor de enfoque para lograr la mejor definición de P' ?
- 2.-Dada la posición del motor de enfoque que da la proyección P' más contrastada, ¿cómo se puede calcular la componente Z de P ?

Este proceso se denomina en la literatura "*profundidad a partir del enfoque*" y ha sido tratado por varios investigadores. Krotkov [2] discute varios criterios para determinar el mejor enfoque de la imagen de la escena y el cálculo de la profundidad a partir del motor de enfoque. Pentland [10] y Grossmann [11] ha tratado el problema desde un punto de vista diferente, considerando la extracción de profundidad a partir del desenfoque, dando una estimación de la profundidad de la escena a partir de la borrosidad de cada zona de la imagen.

3.1 Análisis del desenfoque.

La figura 1 presenta un modelo general del proceso de formación de imagen por un sistema óptico. Al ser el mismo medio (aire) el existente entre el sensor y las lentes, y las lentes y la escena, la expresión que relaciona d_{ext} y d_{int} viene indicada por la ecuación 1.

La superficie sensora (situada en un plano normal al eje óptico) se encontrará a una distancia d_{sen} , medida desde el plano principal imagen (P_2). Es obvio que la imagen del objeto estará perfectamente enfocada en el plano del sensor cuando coincidan d_{sen} y d_{int} . La distancia conjugada de d_{sen} (plano sensor), situada enfrente de la lente, se denomina distancia de enfoque.

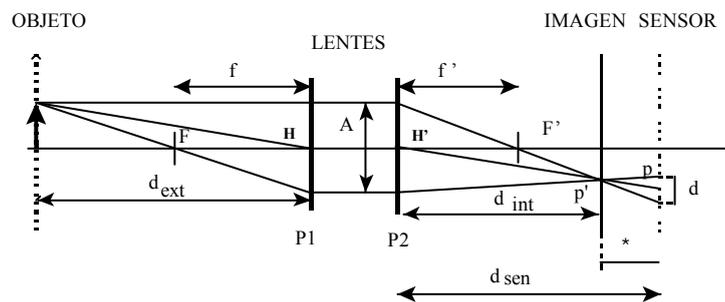


Figura 1 Modelo de lente física, P_1 , P_2 son los planos objeto e imagen F , F' son los planos focales, primario y secundario. A es apertura de las lentes, d es diámetro del círculo de confusión.

Existen dos formas de abordar el proceso de formación de imágenes [12]: *Óptica Geométrica*, se basa en el trazado de rayos y sus resultados son una aproximación de primer orden; *Óptica Física*, se basa en la teoría de difracción y sus resultados son más exactos. En el presente estudio se utiliza la primera técnica, alternativa utilizada en la mayoría de los trabajos que abordan este problema, debido al límite de precisión del resto de los elementos de proceso de formación de imagen. Si las lentes tienen una apertura A y el plano sensor está desplazado una distancia $\delta = |d_{int} - d_{sen}|$, entonces la imagen formada sobre él será un círculo de radio r (ecuación 2).

$$\frac{1}{d_{ext}} + \frac{1}{d_{int}} = \frac{1}{f'} \quad (\text{Ecu.1})$$

$$r = \frac{A \delta}{2} \left(\frac{1}{f'} - \frac{1}{d_{ext}} \right) \quad (\text{Ecu.2})$$

El proceso de formación de imagen de un sistema de lentes produce distorsiones que puede modelarse mediante una función de transferencia (**MTF**) [12]. Esta distorsión se traduce en la supresión de frecuencias espaciales a partir de una cierta frecuencia de corte, la atenuación de las componentes frecuenciales y una pérdida de contraste, debidas fundamentalmente a la difracción y las aberraciones. Esto supone que la intensidad en el plano sensor no es uniforme.

Suponiendo que hasta 1^{er} orden la intensidad es uniforme, el círculo de confusión puede ser aproximado matemáticamente mediante la convolución con la función de distribución de intensidad descrita en la ecuación 4. Un modelo más exacto, que incluye los fenómenos de difracción, es el modelado mediante una distribución de intensidad Gaussiana. El efecto de desenfoque se puede modelar, por tanto, mediante la convolución de un filtro con la imagen original.

$$I_b(x,y) = I_g(x,y) * circ_r(s) \quad circ_r(s) = \begin{cases} \frac{1}{\pi r^2}, & s = \sqrt{x^2 + y^2} \quad 0 < s < r \\ 0 & \text{en otro caso} \end{cases} \quad (\text{Ecu. 4})$$

$$I_g(x,y) \rightarrow \text{Imagen ideal}$$

$$F \{ f_\alpha \} = 2\pi r^2 \left[\frac{J_1(rf_\alpha)}{rf_\alpha} \right], \quad f_\alpha = \sqrt{f_x^2 + f_y^2} \quad (\text{Ecu. 5})$$

$J_1 \rightarrow$ Función de Bessel de 1^a clase de primer orden

La transformada de Fourier de la función círculo se puede expresar con la función de Bessel de primer orden [13] (ecuación 5, figura 2). La transformada es simétrica circularmente con un punto central y anillos concéntricos de amplitud decreciente. Actúa como un filtro paso-bajo (Ecuación 4). De la función de desenfoque se pueden extraer las siguientes conclusiones:

- Cuando r aumenta, la integral del espectro de potencia decrece y el primer cero de $J_1(rf_\alpha)$ se aproxima al origen (se concentra más la energía en la parte central).
- Cuando δ crece, r crece, por consiguiente más se atenúan las altas frecuencia espaciales y más borrosa será la imagen.

Por tanto desenfoque un sistema óptico atenúa las altas frecuencias espaciales. Esta característica se puede utilizar para medir el desenfoque de la imagen.

3.2 Funciones de medida de la calidad del enfoque.

Se debe definir una métrica de agudeza visual o enfoque. La idea general es utilizar el contenido de altas frecuencias, de forma que disponga de unas buenas características en cuanto a precisión, no presente máximos locales que hagan difícil la búsqueda, y sea robusta frente al ruido y texturas de la escena. Existen diferentes métodos que permiten medir el desenfoque, a continuación se presenta un resumen de estos métodos con sus características principales [2].

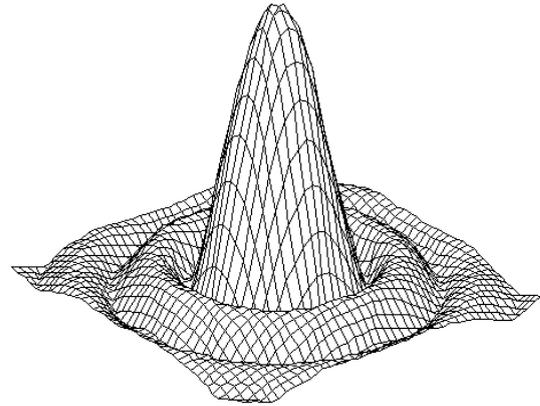


Figura 2 Transformada de Fourier de la Función de desenfoque considerando una distribución uniforme de intensidad

Así se pueden distinguir en métodos basados en la utilización de la información de alta frecuencia de la transformada de Fourier, la relación entre características de borde y enfoque utilizando la magnitud de gradiente (Tenenbaum [14] y Schlag [15]), o el filtrado paso-alto mediante el operador Laplaciano [14]. Otro tipo de técnicas utilizan patrones texturados, o la entropía del histograma (en una imagen desenfocada el histograma tiende hacia una distribución uniforme), la varianza del nivel de gris (maximizar la integral del espectro de potencia de la distribución de intensidad) o la diferencia módulo-suma (Jarvis [16][17]).

3.3 Obtención del mejor enfoque de la imagen.

La selección del estimador de enfoque más adecuado se realiza en función de la curva de respuesta en el rango de posiciones del motor de enfoque. El objetivo en un sistema de auto-enfoque es calcular la posición del motor de enfoque que da un valor máximo en la función de respuesta, pero estas curvas de respuesta están sometidas a perturbaciones provenientes por un lado de variaciones temporales en la iluminación, el proceso de digitalización, ruido o interferencias. Estos problemas se deben tener en cuenta ya que varias de las funciones anteriores calculan derivadas y amplifican los efectos del ruido. Otro elemento que influye en el tipo de curva de respuesta obtenida es el tamaño de la ventana sobre la que opera el estimador. Una imagen puede contener la proyección de puntos situados a diferente distancia, esto produce que se presenten en la función criterio varios picos. Por tanto para obtener un enfoque correcto debemos seleccionar el área que se debe enfocar, esta selección es tarea del sistema de convergencia en un par estereoscópico que sitúa los ejes focales de los captadores sobre la parte de la escena a tratar. Otro problema que se presenta es que la distancia de enfoque cambia con el aumento de la lente (zoom) y los puntos característicos pueden cambiar de coordenadas. Así mismo se produce un efecto reducción de brillantez al variar el estado de las lentes. Esto se puede compensar normalizando los valores de intensidad (media en la ventana de procesamiento).

Todos estos efectos dan lugar a que la curva de respuesta presente máximos locales, varios de los cuales pueden ser pronunciados. Krotkov [2] presenta un análisis comparativo de los métodos anteriores obteniendo como resultado que el método de Tanenbaum (magnitud del gradiente) es el que presenta una curva de respuesta con un máximo muy pronunciado dando un comportamiento más robusto. Este resultado es corroborado por otros autores y por la experimentación realizada en la cual se han comparado las medidas de la magnitud del gradiente, entropía del histograma, laplaciana y varianza de los niveles de gris. Este comportamiento monótonico permite utilizar algoritmos de búsqueda basados en el gradiente ascendente que permiten encontrar el máximo en pocas iteraciones. La búsqueda de Fibonacci permite seleccionar el incremento, fijando el máximo número de iteraciones para la resolución del accionador del motor de enfoque.

4. RESULTADOS

Los algoritmos de control de enfoque se han implementado sobre un sistema basado en una tarjeta de procesamiento de imágenes en tiempo real (DATACUBE MV-200) que procesa la información óptica y cinemática del cabezal estereoscópico RTG-400, constituido por una base binocular con control de acimut, inclinación, convergencia, enfoque, apertura y zoom. El sistema óptico está formado por dos

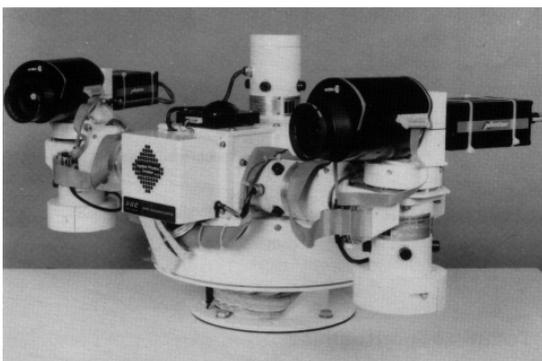


Figura 3 Cabezal RTG-400 con óptica motorizada Ernitec M10Z10.

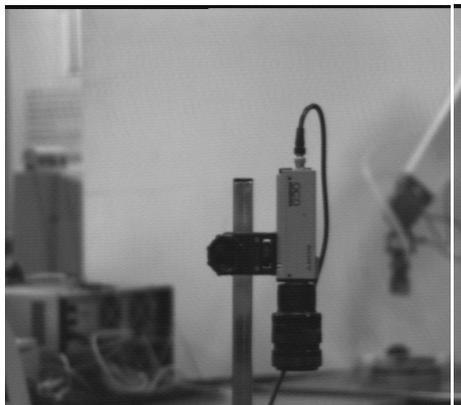
cámaras CCD de mediana resolución SONY-AVDC7 que ofrecen una superficie sensora de 768x580 píxeles. Llevan montadas una óptica zoom motorizada ERNITEC M10Z10 equipada con potenciómetros de posición para obtener una medida para el sistema de control (figura 3).

Sobre este equipo se muestran se han realizado varios experimentos con dos objetos de interés a dos distancias de enfoque diferentes. La selección del área de enfoque es crítica para la convergencia del algoritmo de enfoque por lo que se calcula en base al campo de velocidades obtenido a partir del desplazamiento de la cámara. El tamaño de la ventana de procesamiento utilizada es de 30x30 píxeles, valor que permite una elevada inmunidad al ruido de la función criterio.

Se muestran en las figuras 4(a1,a2,b1,b2) las imágenes estereoscópicas de una escena de trabajo con dos objetos de interés a dos distancias de enfoque diferentes. La figura 5 muestra la variación de la función criterio

para el recorrido del motor de enfoque (con un muestreo de 10 pasos por imagen) para las dos distancias de trabajo. El proceso de búsqueda del enfoque óptimo (máximo de la función criterio) puede ser acelerado de este modo obteniendo el resultado en pocas iteraciones.

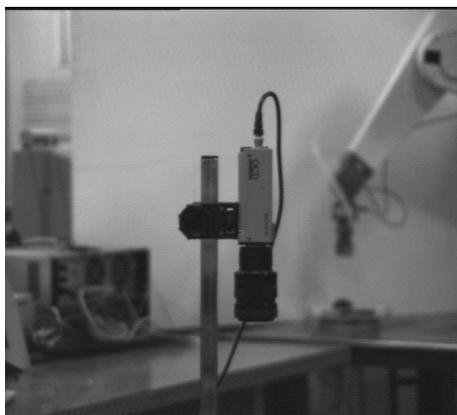
En las imágenes obtenidas se observa las limitaciones físicas de las ópticas comerciales, presentando un acoplamiento importante entre los motores de enfoque y zoom, así como variaciones en el centro focal de la imagen. El buen comportamiento de la función de medida en estos casos es fundamental para lograr una correcta convergencia del algoritmo.



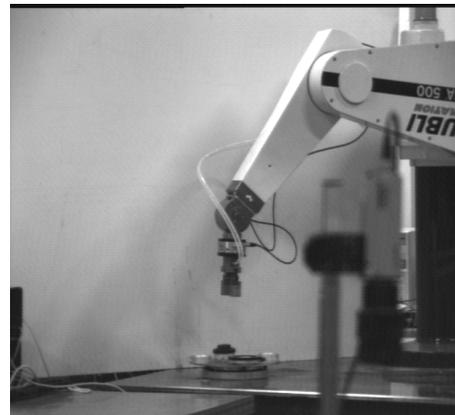
(a1)



(a2)

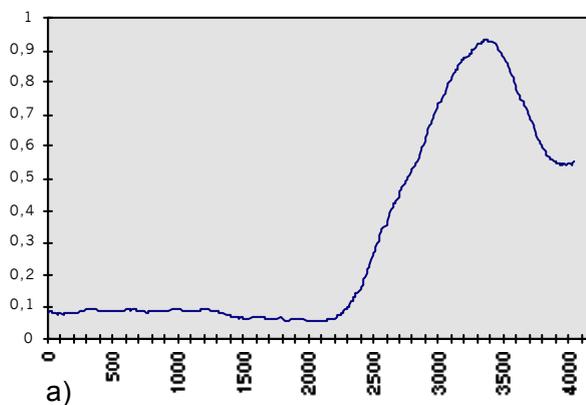


(b1)

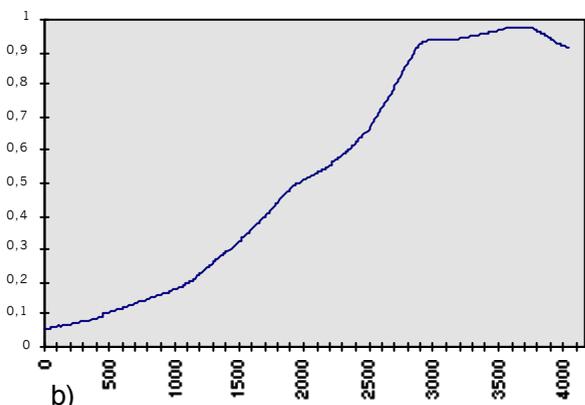


(b2)

Figura 4. Imágenes del cabezal estereoscópico RTG-400 obtenidas con control de enfoque y convergencia para dos distancias de trabajo: a) imágenes izquierda y derecha del sistema con fijación en la cámara; b) imágenes izquierda y derecha del sistema con fijación en el extremo del brazo robot.



a)



b)

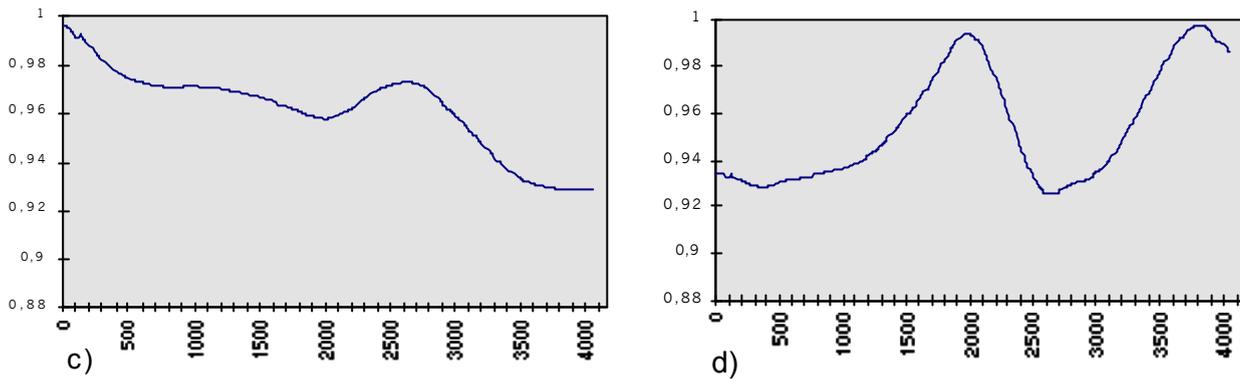


Figura 5. Evolución de la función criterio del gradiente para las diferentes posiciones del motor de enfoque (imagen izquierda) a) sistema con fijación en la cámara; b) sistema con fijación en el extremo del brazo robot. Evolución de la función criterio entropía del histograma para las diferentes posiciones del motor de enfoque (imagen izquierda) c) sistema con fijación en la cámara; d) sistema con fijación en el extremo del brazo robot.

BIBLIOGRAFÍA

- [1] Bajcsy, R, "ACTIVE PERCEPTION", Proc. of the IEEE, vol. 76, n° 8, pag. 996-1005, Aug. 1988.
- [2] Krotkov, E.P., "ACTIVE COMPUTER VISION BY COOPERATIVE FOCUS AND STEREO", Springer-Verlag, NY, 1989.
- [3] Coombs, J. D. "REAL-TIME GAZE HOLDING IN BINOCULAR ROBOT VISION". Phd Thesis. Univ. of Rochester. 1992
- [4] Dhond U.R, Aggarwal J.K., "STRUCTURE FROM STEREO - A REVIEW", IEEE Trans. on Systems, Man, and Cybernetics, Vol. 19, No.6, Nov/Dec. 1989.
- [5] Jimenez, L.M. "MODELADO DE ENTORNOS 3D NO ESTRUCTURADOS MEDIANTE VISION ARTIFICIAL PARA PLANIFICACIÓN DE TRAYECTORIAS DE ROBOTS". 1994 Informe Técnico. Dep. de Ingeniería de Sistemas y Automática. (E.T.S.I.I) Univ. Politécnica de Madrid.
- [6] Jimenez, 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". Resúmenes XVI Jornadas de Automática (Grupo de Trabajo Visión Artificial) 1995, pp. 179-197.
- [7] Julesz, B. "BINOCULAR DEPTH PERCEPTION OF COMPUTER-GENERATED PATTERNS", Bell syst. Tech. J., vol. 39, pp.1125-1162, 1960.
- [8] Julesz, B. "FOUNDATIONS OF CYCLOPEAN PERCEPTION", Univ. of Chicago Press, 1971.
- [9] Ahuja, N. "ACTIVE STEREO: INTEGRATING DISPARITY, VERGENCE, FOCUS, APERTURE, AND CALIBRATION FOR SURFACE ESTIMATION", IEEE Trans. on Pattern Ana. & Mach. Intell. Vol. 15, No. 10, October 1993
- [10] Pentland, A., "A NEW SENSE FOR DEPTH OF FIELD", Proc. Intl. Joint Conf. Artificial Intelligence, Los Angeles, pp. 988-994 (August, 1985).
- [11] Grossmann, P. "DEPTH FROM FOCUS", Pattern Recognition Letters, Vol.5, pp 63-69.
- [12] Casas, J. "ÓPTICA", Librería General Zaragoza, 1994.
- [13] Hetch E., Zajac, A., "OPTICS", Addison-Wesley Massachusetts 1974.
- [14] Tenenbaum, J. M., "ACCOMMODATION IN COMPUTER VISION", PhD. Dissertation, Stanford University (November, 1970).
- [15] Schlag, J. F., A. C; Sanderson, C. P. Neumann, and F. C. Wimberly, "IMPLEMENTATION OF AUTOMATIC FOCUSING ALGORITHMS FOR A COMPUTER VISION SYSTEM WITH CAMERA CONTROL", Technical Report CMU-RF-TR314, Carnegie Mellon University (August, 1983).
- [16] Jarvis, R. A., "FOCUS OPTIMIZATION CRITERIA FOR COMPUTER IMAGE PROCESSING", Microscope 24(2), pp. 163-180 (1976).
- [17] Jarvis, R. A., "A PERSPECTIVE ON RANGE FINDING TECHNIQUES FOR COMPUTER VISION" IEEE Trans. Pattern Analysis and Machine Intelligence PAMI-5(2), pp. 122-139 (March, 1983).