

Sensores CMOS para robótica e industria: Sensor retínico espacio variante y visión activa

Fernando Pardo, Francisco Vegara, Jose A. Boluda, Santiago Felici

Instituto de Robótica
Universidad de Valencia
C/ Hugo de Moncada, 4 Entlo. 46010-Valencia SPAIN
Email: Fernando.Pardo@uv.es

27 de Abril de 1995

Resumen

La utilización de la representación espacio-variante, o la posibilidad de realizar sensores de imágenes con tecnología CMOS, son todavía conceptos poco conocidos en entornos industriales. En este artículo se muestra la transformación log-polar como un paso más en dotar a los robots móviles de percepción visiva, asemejándose cada vez más a sus creadores. Por otro lado se presenta la tecnología CMOS como una alternativa más económica que la tecnología CCD para la realización de cámaras, teniendo unas ventajas que la hacen especialmente interesante para su uso industrial en aplicaciones específicas. Por último, se funden ambos conceptos y se presenta un sensor de imágenes, basado en tecnología CMOS, que capta las imágenes siguiendo una distribución log-polar, de la misma manera que lo hace el ojo humano, presentando además, una respuesta logarítmica frente a la intensidad luminosa.

1 Introducción

El problema del procesamiento de imágenes en tiempo real tiene una especial relevancia en sistemas controlados por visión (como la navegación de robots en entornos poco estructurados). Una forma de alcanzar el entendimiento del entorno en tiempo real [1] es la reducción selectiva de la información visual que llega al sistema, consiguiéndose de esta manera una reducción de los datos a procesar y, por lo tanto, del tiempo requerido para llevar a cabo una determinada tarea. Un ejemplo de reducción selectiva de la información redundante se en-

cuentra en el propio ojo humano. En la retina, los fotorreceptores se encuentran distribuidos de manera que existe una gran concentración en el centro o fovea, concentración que va disminuyendo al alejarse hacia la periferia. De esta manera, se consigue una alta resolución en las zonas interesantes de la escena, normalmente en el centro, mientras que se mantiene un amplio campo visual sin que la cantidad de datos a procesar de la imagen completa sea excesivamente grande.

El uso de transformaciones espacio variantes para la adquisición de imágenes no es nuevo y se ha venido estudiando especialmente en el terreno de aplicaciones en robótica [2]. Sin embargo, en el momento actual, esta investigación alcanza una nueva dimensión con la posibilidad de la realización de sensores o cámaras de estado sólido, que directamente captan la imagen con una de estas especiales representaciones.

La principal tecnología utilizada en la realización de sensores en cámaras de estado sólido ha sido siempre la CCD (Charge Coupled Device). Sin embargo, aunque esta tecnología es la que más calidad ofrece para sensores basados en matrices cartesianas de pixels, no es la más adecuada para la realización de sensores espacio-variantes [3]. La tecnología CMOS, que en principio viene indicada para el diseño de circuitos digitales de aplicación específica o ASICs, ha demostrado ser también útil para la realización de sensores de imágenes [4]. Su calidad no es tan buena, pero su alta versatilidad la hacen idónea para este tipo de diseños, gracias a las numerosas ventajas que se describirán más adelante.

2 Representación log-polar

Una de las ventajas de la representación espacio variante, donde la resolución no es uniforme, como en el ojo humano por ejemplo, es la reducción selectiva de la información a procesar. Esta ventaja es común a casi cualquier representación espacio-variante que se elija, sin embargo, hay algunos casos particulares donde la transformación puede presentar, además, interesantes propiedades matemáticas. Una de estas transformaciones espacio-variantes es la representación log-polar.

La representación log-polar es la transformación de puntos del plano polar o *plano retínico* (r, θ) , al plano cartesiano o *plano cortical*¹ (u, v) .

Matemáticamente cualquier punto en un sistema de coordenadas polares se puede escribir de la forma:

$$P = r \exp i\theta \quad (1)$$

donde r es la distancia radial desde el origen de coordenadas y θ es el ángulo. La transformación al plano cartesiano se puede escribir como:

$$P' = \log P = \log r + i\theta \quad (2)$$

de donde las coordenadas corticales serán:

$$\begin{cases} \xi = \log r \\ \gamma = \theta \end{cases} \quad (3)$$

En la figura 1 podemos apreciar la transformación log-polar de una forma gráfica. De esta figura también se puede vislumbrar la estructura básica de pixels que un sensor que realice la transformación log-polar debería tener.

Aparte de la reducción selectiva de información, hay dos ventajas matemáticas directas, que son la facilidad con que se pueden tratar rotaciones y escalados. De las ecuaciones presentadas anteriormente, es fácil derivar que una rotación en el plano retínico se transforma en una traslación respecto al eje γ del plano

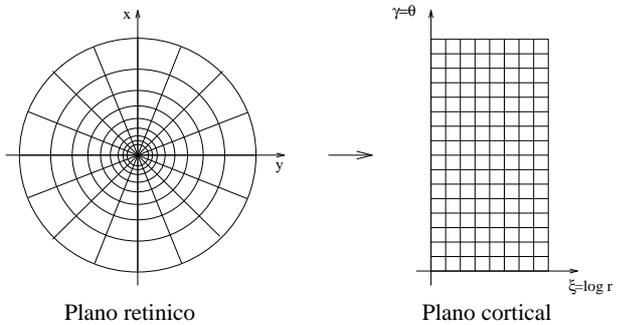


Figura 1: Representación gráfica de la transformación

cortical, mientras que una operación de escalado en el plano retínico se transforma en una traslación respecto al eje ξ del plano cortical. Esto quiere decir que operaciones difíciles de tratar matemáticamente en el plano cartesiano, como rotaciones y escalados, se transforman en simples operaciones de traslación en el plano cortical. Estas ventajas tienen un impacto inmediato sobre algunos algoritmos de imágenes, entre los que destacan el cálculo del flujo óptico y cálculo del tiempo al impacto o colisión [5]. El cálculo del flujo óptico en objetos que se alejan o acercan al sensor es inmediato ya que sólo interviene una componente, disminuyendo los grados de libertad y simplificando los cálculos. También en el caso del cálculo del tiempo al impacto hay una gran simplificación ya que son sólo los objetos que se aproximan al sensor los que tienen especial interés, también en este caso los grados de libertad se simplifican disminuyendo el número de ecuaciones necesarias para este cálculo.

3 Cámaras CMOS

La representación log-polar es muy interesante para aplicaciones en robótica, es por esto que la realización de un sensor específico, que tenga esta especial representación, está plenamente justificada. El siguiente paso consiste en decidir qué proceso tecnológico, para la realización de sensores de estado sólido, es el más adecuado para el diseño de un sensor de imágenes espacio-variante.

La mayoría de los esfuerzos tecnológicos en la investigación de sensores de imágenes, se han dirigido a mejorar la calidad de cara al observador humano sin prestar demasiada atención

¹Los términos *retinal* y *cortical* derivan de la observación de que la transformación log-polar se asemeja mucho a la transformación de la imagen retínica en el cortex visual del ojo humano.

al tipo de información que puede ser extraída de las imágenes y comunicada por otros tipos de canales no humanos. Este direccionamiento de la investigación ha llevado consigo un gran avance en dispositivos como cámaras de televisión, cámaras portátiles, etc. sin embargo, este tipo de cámaras o sensores no tienen por qué ser los más adecuados para sistemas industriales, de vigilancia, o también robots. La tecnología más utilizada para estas cámaras, dirigidas al consumo humano, ha sido la tecnología CCD, ya que ofrece una buena calidad de imagen, entendiéndola como un parámetro subjetivo de los seres humanos. Calidad de imagen para un robot o un sistema industrial es, en cambio, un concepto diferente, de manera que se puede decir que una buena calidad de imagen, para un sistema automático, es aquella a partir de la cual el sistema puede realizar la tarea que tiene encomendada.

La tecnología CMOS no ofrece todavía la calidad que las CCD, sin embargo, sí que ofrece la suficiente calidad para la mayoría de sistemas industriales actuales donde se emplea visión. Además, la tecnología CMOS ofrece muy interesantes características, frente a la CCD y otras, que la hacen especialmente idónea para su uso en industria y robótica. Estas ventajas que ofrecen se pueden resumir en los siguientes puntos:

- Muy simple interface con el mundo exterior. La información es extraída como si de una ROM se tratara, por lo que puede ser conectado directamente con un procesador.
- Acceso aleatorio a cada pixel en el sensor. Esto puede permitir ahorrar tiempo de procesamiento ya que se pueden leer sólo los pixels que interesen.
- Inexistente lógica de control. Las señales de control de un chip de este tipo son todos niveles TTL (5V) y prácticamente no existen, por lo que la cámara es de muy reducido tamaño. Esto es una ventaja frente a los dispositivos CCD que precisan de diferentes tensiones y a veces complicadas señales de sincronización.
- Bajo consumo. Al reducirse la lógica de control el consumo también disminuye.
- Facilidad de realización de sensores hechos a medida. Esto permite diseñar fácilmente sensores de cualquier resolución, e incluso con diferentes tamaños de pixel, por lo que son especialmente interesantes para la realización de sensores espacio-variantes.
- Fácil integración de hardware de procesamiento en el propio chip. Esto permite la realización, en un único chip, de todo un sistema que incluye sensorización, procesamiento y salida. Con tecnología CCD esto es prácticamente imposible.
- Bajo coste. La tecnología CMOS es la más económica actualmente. Además, el hecho de que se pueda integrar todo un sistema en el propio chip, hace que se reduzcan costes para grandes producciones.

Se puede apreciar que la única desventaja, comparada con la tecnología CCD, es una menor calidad de la imagen. Sin embargo, esta menor calidad tiene poca importancia para un sistema de vigilancia donde quizá sólo se pretende detectar movimiento, o en sistemas de control de tráfico [6] donde sólo se pretende definir el contorno de un coche, o para un robot donde quizá la única tarea a realizar sea detectar obstáculos y esquivarlos. Por lo tanto, la tecnología CMOS es la más adecuada para ser usada en entornos industriales y en robótica [7].

Existen dos métodos fundamentales para la realización de sensores utilizando la tecnología CMOS. El primero, más extendido, es muy parecido al que utilizan los CCDs para captar la luz. Este primer método [8] consiste en capturar los electrones generados por la acción de la luz dentro de un pozo de potencial durante un tiempo determinado. La tensión de salida será proporcional a la carga almacenada, es decir, será proporcional a la luz incidente y al tiempo durante el cual se está recogiendo la luz. Aunque la forma de captar la luz es idéntica tanto en este método como en CCD, es la forma en que la carga es leída lo que diferencia una tecnología de la otra. En CCD la carga es transportada, mediante registros de desplazamiento de cargas (que son precisamente los que reciben el nombre de CCDs) hacia un amplificador común en la salida que realiza la transformación de carga a tensión. Por el contrario, en la tecnología CMOS, esta transformación tiene

lugar directamente en cada pixel, por lo que lo que se lee en el exterior no son cargas, sino tensiones. Debido a esto, el sensor es menos inmune al ruido. Por otro lado, existen no-uniformidades en la salida al ser la conversión carga-tensión diferente para cada célula. Todo esto hace que los sensores CMOS ofrezcan una peor calidad. En cualquier caso, el hecho de tener tensiones en vez de cargas, hace que sea sencillo el direccionar cada célula dentro del sensor como si de una ROM se tratara. Además la información circula por hilos, y por tanto, no entorpecen la captación de luz, mientras que en los sensores CCD en vez de hilos hay registros asociados a las células receptoras, para el transporte de cargas, que no presentan problemas en disposiciones cartesianas de los pixels pero sí en cualquier otro tipo de topología espacio-variante.

El segundo método para la realización de sensores utilizando la tecnología CMOS [9] es mediante la transformación directa de la luz incidente en corriente. Esta operación se realiza directamente sobre un transistor, es decir, la corriente provocada por la luz atraviesa un transistor polarizándolo a una determinada tensión de salida. La diferencia fundamental con los anteriores es que no es necesario esperar un tiempo de integración durante el cual la célula se va cargando con la luz incidente. La desventaja es que esta corriente generada es muy baja y es bastante sensible al ruido. Esta corriente generada se transforma en tensión directamente en el transistor, transformación que produce el efecto más interesante es esta forma de captar la luz, y es que la tensión tiene una dependencia logarítmica con la corriente (siempre que el transistor MOS trabaje en su zona de inversión débil, lo cual ocurre siempre dada la baja corriente que produce la luz incidente).

4 Sensor retínico o foveal

En el caso particular de sensores espacio-variantes para robótica, existe el precedente de un sensor diseñado con tecnología CCD [10]. Este sensor tenía numerosos problemas que incluían una baja calidad de imagen a pesar de ser tecnología CCD. La conclusión de este primer intento en la realización de un sensor de este tipo, basado en tecnología CCD, es que esta tec-

nología no es adecuada [3] para diseños donde el tamaño de los pixels no es uniforme y la distribución de estos no sigue una trama cartesiana.

Sólo existen dos sensores más que presentan una distribución log-polar de sus pixels. Ambos han sido diseñados utilizando tecnología CMOS, y dada la tan reciente realización de ambos, es difícil compararlos entre sí. Uno de ellos ha sido diseñado en la Universidad McGill en Canadá [11], y el otro, más reciente, ha sido diseñado por miembros del Instituto de Robótica de la Universidad de Valencia [12] en IMEC (Interuniversity MicroElectronics Center, Bélgica). A continuación se presenta el segundo sensor diseñado por el Instituto de Robótica.

El sensor propuesto posee dos partes principales, como cualquier otro sensor de este tipo, una parte llamada retina, que sigue la representación log-polar expuesta anteriormente; y la otra zona llamada fovea, que es la parte central y que en el caso particular de este sensor es diferente a cualquiera de sus precursores. En anteriores realizaciones de este tipo de sensores, la fovea era siempre una matriz cuadrada de pixels con resolución constante. Esta no es la mejor solución para esta zona ya que no existe continuidad entre la retina y la fovea, por lo que la aplicación de algoritmos de procesamiento de imágenes, se hace muy complicada en la frontera entre la fovea y la retina. La solución propuesta es la realización de anillos concéntricos cuya resolución disminuye al acercarse al centro. Esto permite una continuidad entre la retina y la fovea por lo que es fácil aplicar algoritmos de procesamiento directamente sobre la representación log-polar.

El trazado final del sensor se observa en la figura 2, mientras que el detalle de la zona foveal se muestra en la figura 3, donde la separación entre circunferencias es constante (y no logarítmica como en la retina) y el número de pixels por circunferencias disminuye hasta un único pixel en el centro. La resolución en la zona retínica es de 128x56, es decir, 56 circunferencias cada una de ellas formada por 128 pixels. Es la retina de mayor resolución diseñada hasta ahora, ya que las dos anteriores presentaban una resolución de 64x30 para la CCD, y 64x16 para la CMOS de Canadá.

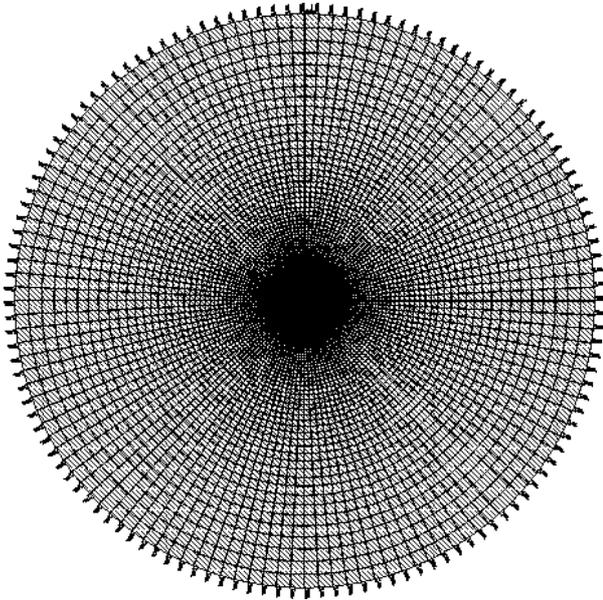


Figura 2: *Layout completo del sensor mostrando la retina*

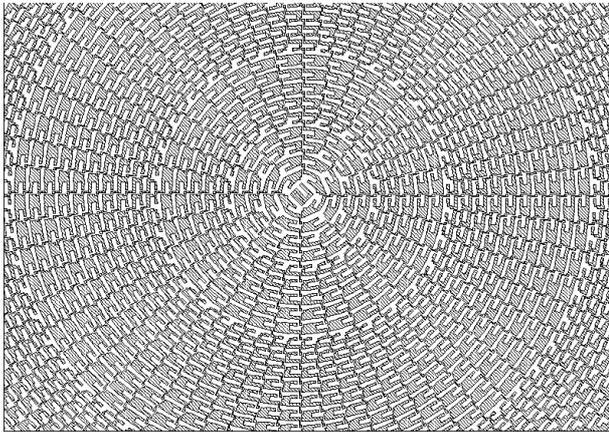


Figura 3: *Detalle de la fóvea del sensor*

Desde el punto de vista tecnológico, la realización de este tipo de sensores presenta numerosos problemas, principalmente debido a que tanto el proceso tecnológico como las herramientas de diseño han sido diseñados para la realización de circuitos “ortogonales” por lo que la inclusión de arcos o circunferencias en el circuito no es nada trivial. El diseño final está siendo fabricado en tecnología CMOS de $0.7 \mu m$.

En el apartado anterior se habían presentado dos métodos para la captación de luz utilizando tecnología CMOS. Después de evaluar los dos diferentes métodos se eligió el segundo por adecuarse más a las características de un sensor espacio-variante. La principal caracte-

terística de este sensor es que la respuesta frente a la luz es logarítmica en vez de lineal como en las cámaras convencionales. Esta respuesta logarítmica tiene múltiples ventajas frente a la lineal [13]. En primer lugar el margen de diferentes luminosidades que puede tratar es muy amplio, hasta 5 órdenes de magnitud frente a los escasos 2 órdenes de las cámaras de respuesta lineal. Esto hace que en muchas aplicaciones no sea necesario el uso de un iris o diafragma para controlar la luz, además de poder hacer frente a imágenes con un alto contraste como las que se dan, por ejemplo, en tráfico, donde coexisten en la misma imagen, zonas con sombra y zonas con sol. Además, dado que este sensor pretende ser una emulación de un ojo biológico, era más lógico disponer de una respuesta logarítmica que lineal.

Las aplicaciones de un sensor de este tipo van más allá de las académicas y se meten de lleno en la industria de consumo. Un claro ejemplo de esto es el proyecto europeo IBIDEM, que tiene como objetivo el desarrollo de un video-teléfono para la lectura de labios y señalización y conversación con manos y dedos [14]. La idea es que este video-teléfono permita la comunicación telefónica entre sordomudos. La cámara de este video-teléfono es precisamente el sensor expuesto en este artículo. La razón se ha explicado ya, y es la reducción selectiva de la información. En principio, sólo los labios son interesantes para poder leer en ellos, pero alguna información adicional es necesaria alrededor para poder realizar un seguimiento de la persona, etc. Con la cámara log-polar esto se consigue fácilmente ya que el centro es de alta resolución y permite ver los labios, mientras que la periferia es de baja y permite un seguimiento ahorrando información. El hecho de utilizar representación espacio-variante tiene su justificación dado el limitado ancho de banda de las líneas telefónicas convencionales. Con solo unos 12 Kbytes por imagen (esta es la información contenida en una imagen tomada con este sensor sin comprimir) es posible enviar imágenes por el cable telefónico a una velocidad de unas 16 imágenes por segundo, que se han demostrado suficientes para un entendimiento de signos visuales.

5 Navegación basada en el sensor espacio-variante

[15]

6 Conclusiones

Se ha presentado la representación log-polar como una forma de captar el entorno especialmente interesante para aplicaciones industriales y robóticas, gracias a la reducción selectiva de información. También se ha presentado la tecnología CMOS como la más adecuada para la realización de sensores de imágenes de aplicación en industria y robótica, por su bajo coste, alta versatilidad, fácil diseño, etc. Como ejemplo, se ha explicado el sensor espacio-variante basado en tecnología CMOS diseñado por miembros del Instituto de Robótica de la Universidad de Valencia. También se muestra una posible aplicación de un robot móvil que utiliza un sensor retínico espacio-variante como base para reconocer el entorno y navegar en entornos poco estructurados.

7 Agradecimientos

El desarrollo del sensor presentado ha sido financiado por el proyecto Europeo IBIDEM dentro del programa TIDE. El consorcio está formado por UNITEK, IMEC, Thomson, Scuola Superiore S. Anna, Scienza Machinale Srl, Knossos Technologies S.A., DIST Università di Genova, Instituut voor Doven y AFA Centro REUL. Parte del proyecto también ha sido financiado por el programa Human Capital and Movility de la Comunidad Europea.

Un reconocimiento especial al profesor Giulio Sandini de la Universidad de Génova por su contribución y su experiencia en procesamiento de imágenes usando representación espacio-variante. A los doctores Bart Dierickx y Nicco Riquier del IMEC por sus grandes aportaciones al diseño final del sensor.

Referencias

- [1] Jack Shandle. Retina-like image sensor promises real-time systems. *Electronic Design*, pages 33–34, May 1993.
- [2] M. Tistarelli and G. Sandini. Dynamic aspects in active vision. *CVGIP: Image Understanding*, 56 No.1:108–129, 1992.
- [3] Fernando Pardo and Enrico Martinuzzi. Hardware environment for a retinal CCD visual sensor. In *EU-HCM SMART Workshop: Semi-autonomous Monitoring and Robotics Technologies*, Ispra, Italy, April 1994.
<http://faeton.eleinf.uv.es/decv/decv.html>.
- [4] D. Rensaw, P.B. Denyer, G. Wang, and M. Lu. ASIC vision. In *IEEE Custom Integrated Circuits Conference*, pages 7.3.1–7.3.4, 1990.
- [5] M. Tistarelli and G. Sandini. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Trans. on PAMI*, PAMI-15, No. 4:401–410, 1991.
- [6] C.L. Wan, K.W. Dickinson, and T.D. Binnle. A cost-effective image sensor system for transport applications utilising a miniature CMOS single chip camera. In *7th IFAC/IFORS Symposium on Transportation Systems: Theory and Application of Advanced Technology*, pages 158–163, Tianjin, China, August 1994.
- [7] S. Anderson, W.H. Bruce, P.B. Denyer, D. Renshaw, and G. Wang. A single chip sensor & image processor for fingerprint verification. *IEEE Custom Integrated Circuits Conference*, pages 12.1.1–12.1.4, 1991.
- [8] Christer Jansson, Per Ingelthag, Christer Svensson, and Robert Forchheimer. An addressable 256x256 photodiode image sensor array with an 8-bit digital output. *Analog Integrated Circuits and Signal Processing*, pages 37–49, 1993.
- [9] Nico Riquier and Bart Dierickx. Addressable imager with a logarithmic response for machine vision. In *ISIROM*. Brussels, 1993.
- [10] I. Debusschere, E. Bronckaers, C. Claeys, G. Kreider, J. Van der Spiegel, G. Sandini, P. Dario, F. Fantini, B. Bellutti, and

- G. Soncini. A retinal CCD sensor for fast 2D shape recognition and tracking. *Sensors and Actuators*, pages 456–460, 1990.
- [11] Robert Wodnicki, Gordon W. Roberts, and Martin D. Levine. A foveated image sensor in standard CMOS technology. In *Custom Integrated Circuits Conference*, Santa Clara, California, May 1995.
- [12] Fernando Pardo. Development of a retinal image sensor based on CMOS technology. Technical Report LIRA-TR 6/94, University of Genoa, Dept. of Computing, LIRA Lab, October 1994. <http://faeton.eleinf.uv.es/decv/decv.html>.
- [13] Nico Ricquier and Bart Dierickx. Pixel structure with logarithmic response for intelligent and flexible imager architectures. *Microelectronic Engineering*, 19:631–634, 1992.
- [14] F.Ferrari, J.Nielsen, P.Questa, and G.Sandini. Space variant imaging. *Sensor Review*, 15(2):17–20, 1995.
- [15] F.Panerai, C.Capurro, and G.Sandini. Space variant vision for an active camera mount. In *SPIE AeroSense95*, Orlando, Florida USA, April 1995.