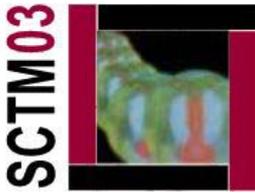


Matemáticas y Visión por Ordenador



Luis Álvarez León

Catedrático de Ciencias de la Computación e Inteligencia Artificial
Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran Canaria

<http://serdis.dis.ulpgc.es/~lalvarez>

Introducción

En los últimos años, los enormes avances tecnológicos en el campo de los ordenadores, fotografía y video digital, etc., han permitido el desarrollo de nuevas disciplinas científicas como es el caso de la visión por ordenador. El sistema de visión humano es muy eficiente y nos suministra información útil y muy variada sobre nuestro entorno. Somos capaces, por ejemplo, de identificar y reconocer fácilmente objetos y formas, también podemos detectar y seguir con facilidad objetos en movimiento en nuestro campo de visión, gracias a la visión binocular podemos estimar la distancia que nos separa de los objetos presentes en nuestro entorno, etc.. Todas estas capacidades de la visión humana y muchas otras que no mencionamos aquí son extremadamente complejas de modelizar y formalizar desde el punto de vista matemático. La visión por ordenador es una disciplina científica de reciente desarrollo que estudia la modelización e implementación en ordenador de procesos propios de la visión. La visión por ordenador empezó siendo una disciplina esencialmente tecnológica donde se creía que todos los problemas se irían resolviendo fácilmente en base al incremento de potencia de cálculo de los ordenadores y las mejoras en la calidad de los dispositivos de adquisición como son las cámaras digitales. Pero pronto se descubrió que el tipo de problemas que aparecen son muy complejos y difíciles de analizar. De tal forma que para muchos de ellos no existe actualmente una solución plenamente satisfactoria. Esta complejidad de los problemas ha llevado a un esfuerzo científico muy importante en la disciplina donde las Matemáticas juegan un papel fundamental. De hecho, actualmente, el perfil de un investigador en la disciplina de visión por ordenador es un perfil mucho más científico que técnico y con una sólida formación matemática.

El objetivo de este seminario es presentar algunas reflexiones sobre la importancia del papel de las Matemáticas en el campo de la visión por ordenador. La exposición se organizará en base a diferentes disciplinas matemáticas y sus relaciones con diferentes problemas de visión por ordenador. Concretamente trataremos las siguientes materias:

- La Geometría Proyectiva.
- Las Transformadas Integrales.
- El Cálculo Variacional.
- Los Modelos Probabilísticos.
- Las Ecuaciones en Derivadas Parciales Geométricas.
- La Optimización y el Análisis Numérico.

Esta lista de disciplinas matemáticas no pretende ser exhaustiva y sólo representa algunos temas con los que quien suscribe está familiarizado. La organización del seminario será la siguiente: iremos visitando cada una de las disciplinas matemáticas mencionadas arriba y veremos algunas de sus aplicaciones en el campo de la visión por ordenador.

Respecto a la bibliografía descrita al final del texto, he optado por citar unas pocas referencias [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11] que, aunque no cubren por completo el espectro de conocimientos presentado en esta memoria (lo que obligaría a llenar varias páginas de referencias), ilustran bastante bien algunas de las materias presentadas.

1. La geometría proyectiva

El modelo de cámara proyectivo es el más sencillo y el que más se utiliza habitualmente. Viene definido por un plano de proyección que representa el plano imagen y un foco. La proyección de un punto 3D en la imagen viene dada por la intersección entre la recta que une dicho punto y el foco, y el plano de proyección. En la Figura 1, podemos observar una ilustración de cómo se aplica este modelo en pintura. El pintor mira a través de un foco una escena 3D y ha intercalado entre él y la escena una cuadrícula que representa el plano de proyección. Apoyándose en la proyección de la escena 3D sobre esta cuadrícula va pintando la escena.



Figura 1. Ilustración de Albrecht Dürer, pintor alemán del Renacimiento (año 1527).

En principio, las coordenadas en las que se representan los puntos son coordenadas euclídeas, esto es, 3 coordenadas para un punto 3D y 2 coordenadas para un punto en el plano. Sin embargo, cuando buscamos la expresión matemática que determina cómo se proyecta un punto 3D en el plano, resulta mucho más conveniente trabajar en los correspondientes espacios proyectivos, de tal manera que un punto 3D se identifica con 4 coordenadas $\mathbf{X} = (x, y, z, v)$; cuando v es distinto de 0, recuperamos las coordenadas euclídeas del punto haciendo $(x/v, y/v, z/v)$. Trabajar en el espacio proyectivo nos permite manejar correctamente puntos en el infinito ($v = 0$). Además, la aplicación que determina cómo se proyecta un punto 3D en el plano, que en coordenadas euclídeas es no-lineal, en coordenadas proyectivas es lineal y viene dada por una matriz $P \in M_{3 \times 4}$. De tal manera que dado un punto 3D \mathbf{X} en coordenadas proyectivas, su correspondiente proyección \mathbf{x} en el plano viene dada por la expresión

$$\mathbf{x} = P\mathbf{X}$$

La matriz P depende de la posición y orientación del plano de proyección en el espacio 3D y del sistema de referencia elegido en el plano de proyección. Un problema fundamental que surge con mucha frecuencia en visión por ordenador es el de calibrar una cámara, que consiste

básicamente en encontrar la matriz P . La primera pregunta que surge es si cualquier matriz P no nula determina una proyección admisible en el sentido de que corresponda físicamente a una configuración real del modelo de proyección. La respuesta a esta pregunta es negativa. Para que P determine una proyección admisible es necesario que sus coeficientes verifiquen una cierta relación algebraica. El estudio de las relaciones algebraicas que aparecen de forma natural al plantear el problema de calibración de cámaras involucra técnicas de álgebra computacional para su resolución.

Existen muchas formas distintas de plantear el problema de calibración en función del tipo de información y número de cámaras que queramos calibrar. La manera más simple es el uso de un calibrador. Un calibrador es un objeto como el representado en la imagen de la Figura 2, del cual conocemos con exactitud las coordenadas 3D de algunos puntos. Por ejemplo, en este caso, conocemos (porque lo hemos medido físicamente) las coordenadas 3D de los puntos que corresponden a las esquinas de los rectángulos negros. Por otro lado, para cada esquina 3D X_i calculamos en la imagen las coordenadas 2D (en pixels) de su proyección x_i en el plano imagen. A partir de las correspondencias $X_i \longleftrightarrow x_i$ es posible calcular P resolviendo un cierto sistema de ecuaciones lineales.

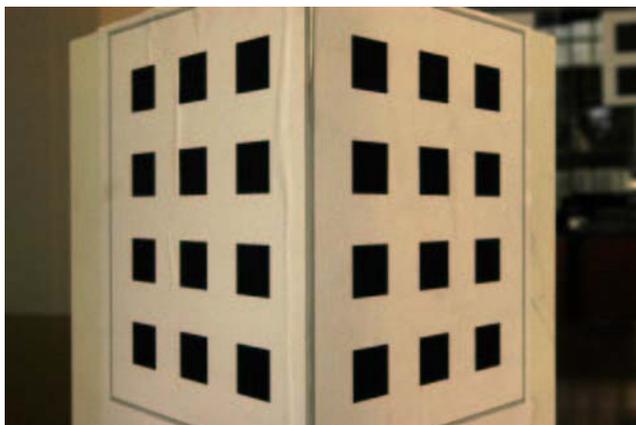


Figura 2. Imagen de un calibrador.

En muchas ocasiones no es posible la utilización de calibradores para calibrar las cámaras. En general, el estudio del problema de la calibración de un sistema de varias cámaras que están observando una misma escena 3D requiere de un fuerte aparato matemático basado principalmente en la geometría y el álgebra. Del hecho de que las cámaras están observando una misma escena se pueden derivar muchos tipos de relaciones, en su mayoría algebraicas entre las matrices de proyección de las diferentes cámaras.

2. Las transformadas integrales

Las transformadas integrales son una herramienta muy poderosa que posee múltiples utilidades en visión por ordenador. En esta sección veremos algunas aplicaciones concretas relacionadas con el análisis multiescala lineal, el reconocimiento de formas planas y la compresión de imágenes.

2.1. El análisis multiescala lineal. La transformada integral por excelencia viene dada por la convolución de una función de entrada con un cierto núcleo de convolución. En nuestro caso, la función de entrada será una imagen bidimensional que formalmente definimos como una aplicación $I : \mathbb{R}^2 \rightarrow \mathbb{R}$, donde $I(x,y)$ representa el nivel de gris de la imagen en dicho punto. Por simplicidad en la exposición supondremos que el dominio donde está definida la imagen es todo \mathbb{R}^2 y que la imagen es de niveles de gris (y no de color, en cuyo caso tendría 3 canales). En una misma imagen coexisten informaciones a diferentes escalas sobre la escena que estamos observando. Por ejemplo, en una panorámica de un bosque podemos observar el bosque en su conjunto, los árboles individuales, e incluso, si la resolución de la imagen es suficientemente buena, los detalles de las hojas de los árboles. La existencia de una información tan variada y compleja en una imagen hace muy complicado su análisis. Para intentar simplificar la información presente en una imagen y poder analizarla más fácilmente, una herramienta muy utilizada consiste en extraer de la imagen original una secuencia de imágenes, que representan la imagen a escalas diferentes. La manera más sencilla de hacer esto es convolucionar la imagen original con un núcleo de convolución gaussiano, es decir, construimos la secuencia

$$I_\sigma(x, y) = \int_{\mathbb{R}^2} G_\sigma(x - u, y - v) I(u, v) du dv$$

donde

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

Aquí, σ representa la escala del análisis y cuanto mayor sea, mayor simplificación se produce en la imagen original.

2.2. El reconocimiento de formas planas. La transformada de Fourier es una herramienta muy útil en Matemáticas, tanto desde el punto de vista teórico como práctico. Desgraciadamente, en general, en las Facultades de Matemáticas no se dedica en los planes de estudio el tiempo suficiente para enseñar correctamente el interés y alcance de esta herramienta, o, al menos, esto era así cuando yo estudié. La transformada de Fourier de una función $f(x)$ en el intervalo $[a,b]$ viene dada por la expresión

$$\hat{f}(w) = \int_a^b f(x) e^{-iwx} dx.$$

A continuación veremos una sencilla aplicación de la transformada de Fourier a la descripción y reconocimiento de formas planas. Consideraremos formas sencillas definidas por un conjunto en el plano cuya frontera es una única curva cerrada como se muestra en la Figura 3. Las formas de este tipo vienen caracterizadas por la geometría de su contorno. Con objeto de

utilizar una representación del contorno adecuada para comparar contornos entre sí, vamos a asociar a la curva cerrada que delimita el contorno la función unidimensional $\theta(s)$ que para cada punto de la curva determina su orientación, es decir, el ángulo que forma la normal a la curva en dicho punto con el eje x ; en la Figura 3 puede verse tal representación. Por tanto, para comparar dos formas lo que haremos será comparar las funciones de orientación asociadas.

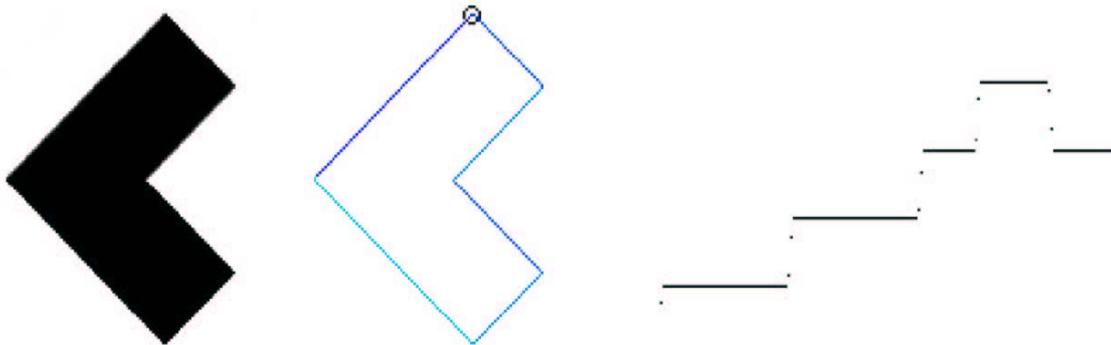


Figura 3. Ilustración de una forma y la función de orientación de su contorno tomando como punto inicial la esquina marcada.

Ahora bien, queremos que dicha comparación sea independiente del tamaño, posición y orientación de los objetos y por supuesto, independiente del punto inicial que se ha tomado en la curva para obtener su función de orientación. Dado que estamos comparando funciones periódicas, la transformada de Fourier resulta una herramienta idónea y realizamos la comparación de las funciones de orientación de 2 formas utilizando su representación en el espacio de Fourier. Comparando los coeficientes de Fourier podemos asociar a 2 formas una medida de similitud entre ellas, de tal forma que cuanto menor sea esta medida mayor parecido habrá entre las formas. En la Figura 4 se ilustra esta técnica.

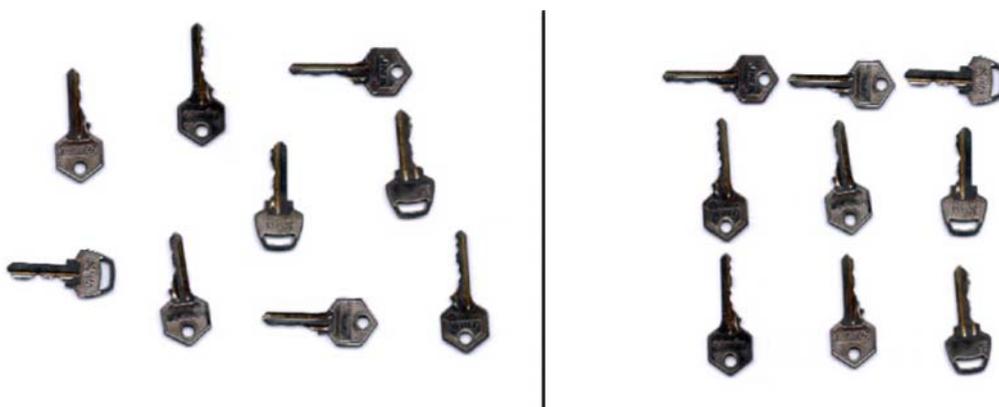


Figura 4. Reconocimiento de formas: a la izquierda observamos un grupo de llaves, a la derecha su clasificación por criterios de similitud. Las llaves idénticas se encuentran ordenadas por columnas.

2.3. La compresión de imágenes. La transformada de Fourier es una herramienta muy útil cuando analizamos fenómenos ondulatorios y periódicos, pero no es adecuada para analizar todo tipo de funciones debido principalmente a que en el espacio de Fourier se pierde totalmente la información espacial de la señal, en el sentido de que la información suministrada por la transformada de Fourier es global, y es imposible extraer información del carácter local sobre el comportamiento de la función. En los últimos años se ha desarrollado la teoría del análisis multiresolución y las *ondelettes* que permiten una representación muy interesante de una función, incluyendo la localización tanto en espacio como en frecuencia. Un análisis multiresolución viene dado por 2 funciones $\psi(s)$ (la ondelette madre) y $\phi(s)$ (la ondelette padre), tales que el conjunto

$$\psi_{m,n}(s) = 2^{m/2} \psi(2^m(s - 2^{-m}n)) \quad m, n \in \mathbb{Z}$$

representa una base ortonormal de $L^2(\mathbb{R})$; y para todo $m \in \mathbb{Z}$ el conjunto de funciones $\{\psi_{m,n}(s)\}_{n \in \mathbb{Z}} \cup \{\phi_{m,n}(s)\}_{n \in \mathbb{Z}}$ es una base ortonormal del espacio $\{\phi_{m+1,n}(s)\}_{n \in \mathbb{Z}}$.

Los denominados coeficientes de ondelettes de una función $f(\cdot)$ vienen dados por:

$$f_\psi(m, n) = \int_{-\infty}^{\infty} \psi_{m,n}(s) f(s) ds, \quad f_\phi(m, n) = \int_{-\infty}^{\infty} \phi_{m,n}(s) f(s) ds.$$

El nombre de análisis multiresolución viene del hecho de que podemos considerar los coeficientes $\{f_\phi(m, n)\}_{n \in \mathbb{Z}}$ como el análisis de la función $f(\cdot)$ a la resolución m , y $\{f_\psi(m, n)\}_{n \in \mathbb{Z}}$ representa la información de detalles necesaria para pasar de la resolución m a la resolución $m+1$. En dimensión 2, las ondelettes de base se definen como:

$$\Psi_1^{m,n,k,l}(x, y) = 2^{m/2} \phi(2^m x - n) 2^{k/2} \phi(2^k y - l)$$

$$\Psi_2^{m,n,k,l}(x, y) = 2^{m/2} \phi(2^m x - n) 2^{k/2} \psi(2^k y - l)$$

$$\Psi_3^{m,n,k,l}(x, y) = 2^{m/2} \psi(2^m x - n) 2^{k/2} \phi(2^k y - l)$$

$$\Psi_4^{m,n,k,l}(x, y) = 2^{m/2} \psi(2^m x - n) 2^{k/2} \psi(2^k y - l)$$

Si consideramos que una imagen discreta $I(n, l)$ representa los coeficientes de ondelettes de una función $I(x, y)$ para la ondelette $\Psi_1^{1,n,1,l}(x, y)$ a partir de $I(n, l)$ podemos descomponer la imagen calculando los coeficientes para las 4 ondelettes $\Psi_i^{0,n,0,l}(x, y)$. En la Figura 5 se ilustra el resultado de tal descomposición. Como puede observarse, muchos de los coeficientes de ondelettes son nulos o muy pequeños (el tono de gris blanco representa un valor nulo), por tanto si sólo almacenamos los coeficientes de ondelettes no nulos podemos ahorrar memoria al almacenar la imagen.

3. El cálculo variacional

La modelización de la solución de un problema como el mínimo de un cierto funcional de energía es una herramienta de gran utilidad en el campo de la visión por ordenador. En esta sección vamos a ver algunas aplicaciones de esta técnica a problemas como la eliminación de ruido en una imagen, el seguimiento de objetos en una secuencia video o la reconstrucción 3D de objetos a partir de un par estéreo de imágenes.

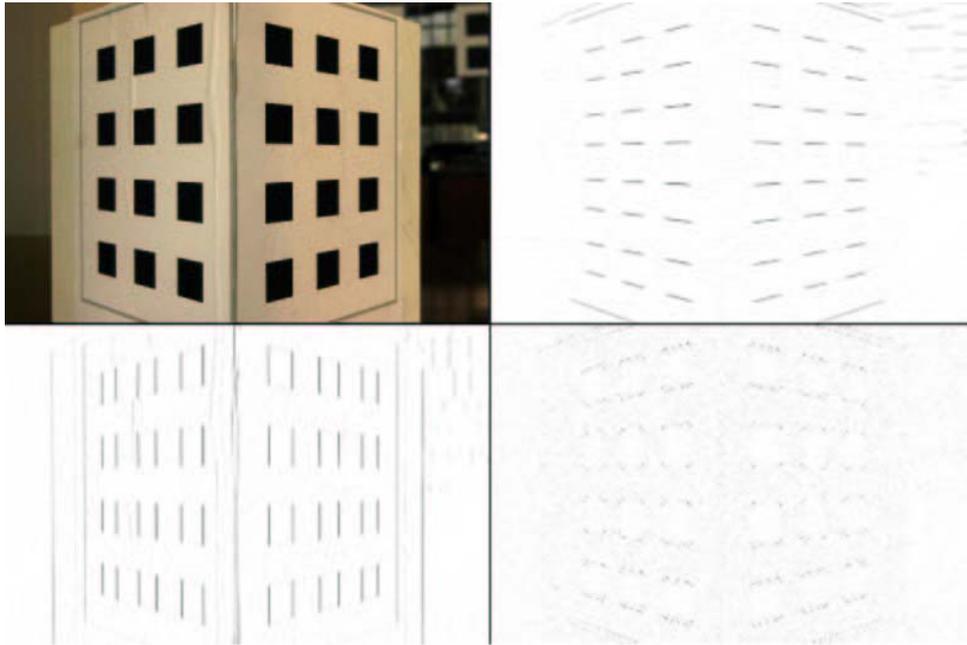


Figura 5. Descomposición por ondelettes de la imagen del calibrador.

3.1. La eliminación de ruido en una imagen. Normalmente, existe siempre un cierto ruido en las imágenes producido por los dispositivos de adquisición, manipulaciones posteriores, etc. El ruido en una imagen se presenta habitualmente en forma de irregularidades de la función imagen. Formularemos el problema de la eliminación de ruido de la siguiente forma: Dada una imagen $I(x,y)$, pretendemos encontrar otra imagen $I'(x,y)$ tal que por un lado sea una función más regular que $I(x,y)$, y por otro lado se parezca a $I(x,y)$ lo más posible. Existen muchas formas de expresar matemáticamente estas condiciones. Una forma sencilla consiste en buscar $I'(x,y)$ minimizando el funcional

$$E(I') = \int_{\mathbb{R}^2} (I' - I)^2 + \alpha \int_{\mathbb{R}^2} \Phi(\|\nabla I'\|)^2$$

donde $\Phi(\cdot)$ es una función positiva y α representa el balance entre el término que ajusta I a I' y el término que suaviza la función I' minimizando globalmente su gradiente, módulo la función $\Phi(\cdot)$.

Para calcular los posibles mínimos de esta energía se iguala a cero la derivada del funcional de energía, lo que da lugar a la denominada ecuación en derivadas parciales de Euler-Lagrange del funcional. A título de ejemplo, en el caso del funcional anterior, la ecuación de Euler-Lagrange resultante es

$$(I' - I) - \alpha \operatorname{div} \left(\frac{\Phi'(\|\nabla I'\|)}{\|\nabla I'\|} \nabla I' \right) = 0.$$

Por tanto, observamos cómo la minimización del funcional se reduce a resolver una ecuación en derivadas parciales. En la Figura 6 se ilustra el resultado de aplicar esta técnica.

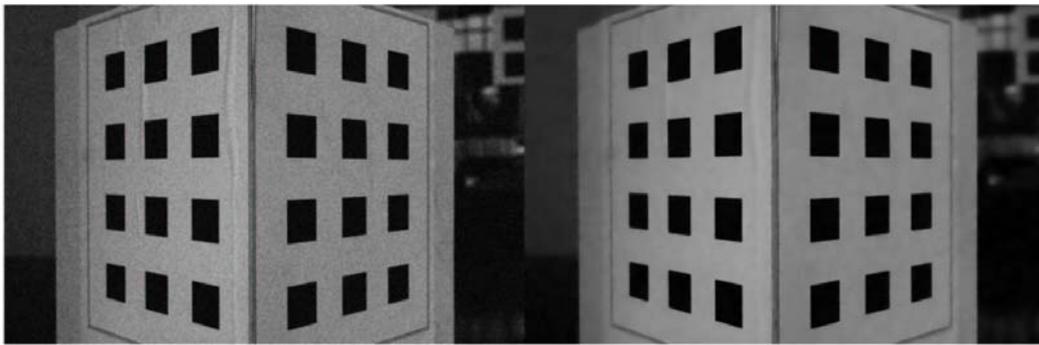


Figura 6. A la izquierda, imagen del calibrador a la que hemos añadido ruido artificialmente; a la derecha, la imagen restaurada.

3.2. Seguimiento de objetos en una secuencia video. Cuando filmamos una secuencia de imágenes, se producen movimientos debidos al propio movimiento de la cámara o al desplazamiento de objetos en la escena. Dadas 2 imágenes consecutivas de la secuencia que denotaremos por $I_1(x,y)$ e $I_2(x,y)$ consideramos el problema de encontrar una función $h(x,y) = (u(x,y), v(x,y))$ denominada flujo, que determina el movimiento de cada punto entre la imagen I_1 e I_2 . Utilizaremos como hipótesis de base que el nivel de gris de un punto no varía cuando el punto cambia de posición, es decir: $I_1(x,y) = I_2((x,y) + h(x,y))$. Ahora bien, esta hipótesis no es suficiente para definir $h(x,y)$, pues, en general, para cada punto en I_1 , hay múltiples combinaciones de puntos en I_2 que verifican la hipótesis. Para poder resolver el problema, añadiremos como condición que el flujo $h(x,y)$ sea regular. Ello nos lleva a minimizar el funcional

$$E(h) = \int_{\mathbb{R}^2} (I_1(x,y) - I_2((x,y) + h(x,y)))^2 + \alpha \int_{\mathbb{R}^2} \operatorname{traza}(\nabla h^T D(\nabla I_1) \nabla h).$$

El término de regularización es un poco especial, porque regulariza el flujo h pero permitiendo discontinuidades en los bordes de los objetos de I_1 . En la Figura 7 se ilustran algunos resultados obtenidos con este tipo de técnicas.



Figura 7. En la parte inferior observamos 2 frames de una secuencia video. En la parte superior representamos las dos componentes del desplazamiento encontrado. El nivel de gris medio indica desplazamiento 0. Un tono claro indica desplazamiento en dirección positiva, y un tono oscuro desplazamiento en dirección negativa.

3.3. Reconstrucción 3D de objetos a partir de un par estéreo de imágenes. Una vez calibrado un sistema de cámaras, el principal problema a resolver es la puesta en correspondencia de puntos en ambas imágenes. El problema se puede formular en los mismos términos que el cálculo del flujo en una secuencia video; la única diferencia es que podemos reducir la complejidad del problema utilizando la información suministrada por la denominada Geometría Epipolar. En un par estéreo, dado un punto en una imagen, su correspondiente en la otra imagen debe estar sobre una recta denominada recta epipolar. Si las cámaras están calibradas, la recta epipolar que corresponde a cada punto es conocida y por tanto ello simplifica considerablemente la complejidad del problema. En la Figura 8 se muestra un resultado de aplicar este tipo de técnicas a la reconstrucción 3D.

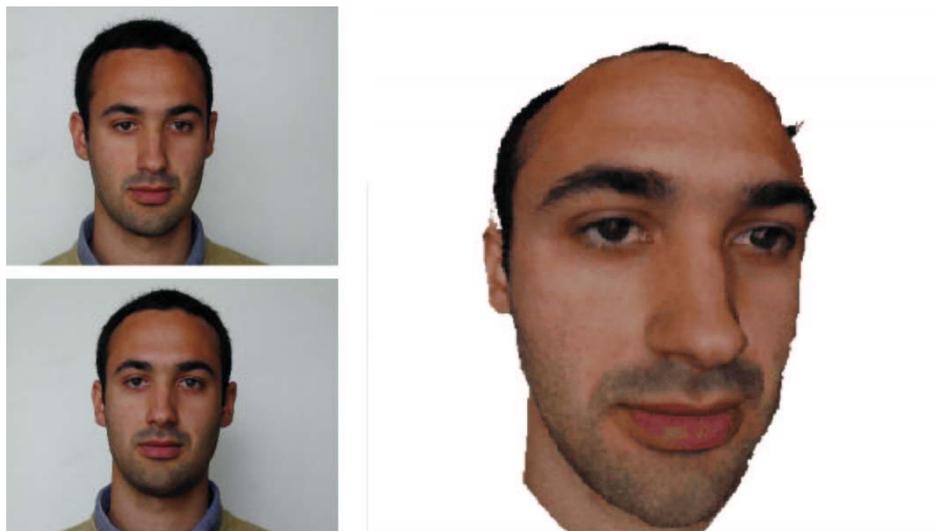


Figura 8. A la izquierda, un par estéreo de una cara humana; a la derecha, la reconstrucción 3D obtenida.

4. Los modelos probabilísticos

Sin duda, los modelos probabilísticos son de una gran utilidad en visión por ordenador y existe toda una comunidad de investigadores en visión que utilizan como base dichos modelos. Una imagen puede modelizarse como la realización de una cierta variable aleatoria, y muchos filtros y transformaciones en imágenes pueden modelizarse como procesos estocásticos. En esta sección vamos a ilustrar cómo se utilizan los modelos probabilísticos para modelizar la formación de escenas naturales.

En concreto veremos el denominado “modelo de hojas muertas” que modeliza la estructura resultante de la superposición de objetos que se acumulan de forma aleatoria; por ejemplo, y de ahí su nombre, la estructura de hojas caídas en un bosque en otoño. La estructura del objeto que va cayendo (la hoja) se modeliza a través de los denominados conjuntos aleatorios. El objetivo que se plantea es estudiar las características geométricas de las partes visibles (denominadas celdas) de la estructura resultante después de superponer aleatoriamente las hojas. Existen resultados realmente interesantes como que la longitud media de los segmentos resultantes de intersectar una recta con las celdas es exactamente la mitad de la longitud media de los segmentos resultantes de interceptar las hojas originales. En la Figura 9 se ilustra un modelo de hojas muertas donde la hoja esta compuesta por rectángulos de tamaño y orientación uniformemente distribuidos.



Figura 9. Ilustración de un modelo de hojas muertas.

5. Las ecuaciones en derivadas parciales geométricas

Una de las características más importantes de nuestro sistema de visión es que la percepción que tenemos de nuestro entorno es invariante respecto a múltiples transformaciones. Somos capaces de reconocer el mismo objeto cuando cambiamos la luminosidad, su orientación posición o tamaño. Intentar que los modelos de visión artificiales posean también de alguna forma este tipo de invariantes es algo realmente complicado. Por ejemplo, si buscamos filtros

que sean invariantes por el tipo de transformaciones mencionadas nos encontramos con que dichos filtros están asociados a un tipo muy especial de ecuaciones en derivadas parciales no lineales que tienen la expresión general

$$\frac{\partial G}{\partial t} = G \left(\operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right), t \right) \|\nabla u\|,$$

donde $G(s,t)$ es una función creciente respecto a su primer argumento. Este tipo de ecuaciones se denominan geométricas porque su evolución sólo depende de la geometría de la frontera de los conjuntos de nivel de u .

6. La optimización y el análisis numérico

La mayor parte de los modelos formulados para analizar los diferentes problemas que aparecen en la visión por ordenador requieren una fase de validación a través de su implementación en ordenador. Para realizar esta implementación es necesario hacer un uso intensivo de técnicas de optimización y análisis numérico. En concreto, resulta necesario con frecuencia abordar los siguientes problemas:

- Interpolación de funciones.
- Resolución de grandes sistemas de ecuaciones lineales.
- Cálculo de autovalores y autovectores.
- Descomposición SVD de matrices.
- Análisis numérico de ecuaciones en derivadas parciales.
- Estimación robusta de parámetros.
- Ajuste de parámetros utilizando Levenberg-Marquardt.

A título de conclusión de este seminario, podemos constatar que la investigación en visión por ordenador requiere de una sólida formación matemática y que el número de problemas abiertos es muy amplio. Además, dado el interés tecnológico de este campo es de prever en el futuro un esfuerzo considerable de las administraciones públicas y empresas privadas en términos de financiación de la investigación. Un joven licenciado en matemáticas con cierta sensibilidad hacia los problemas aplicados y la experimentación es un perfil de investigador idóneo en visión por ordenador. Para iniciarse como investigador en este campo lo más adecuado es seguir un programa de cursos de doctorado específico. En este sentido, los mejores programas se imparten en el extranjero. Por ejemplo, en Francia, que es el país que mejor conozco, los cursos de doctorado se hacen todos en un único año, están bien organizados y un alumno puede dedicarse en exclusiva a seguir los cursos sin aburrirse; una vez terminados los cursos de doctorado el alumno decide si quiere continuar con la tesis o no. En España, los cursos de doctorado se imparten por bienios, son mucho más dispersos en el tiempo, no están diseñados para que el alumno se dedique en exclusiva a ellos, y la realización de la tesis se empieza a hacer muchas veces en paralelo con los cursos.

Por último, me gustaría terminar esta memoria con una invitación a todos aquellos que tuvieran interés en iniciarse en este campo, en especial a jóvenes licenciados o alumnos de último curso de Matemáticas, a contactar conmigo, y gustosamente, dentro de mis limitados conocimientos, les intentaría orientar sobre los pasos a seguir.

Bibliografía

- L. Alvarez, R. Deriche, J. Sánchez, J. Weickert: Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. *Journal of Visual Communication and Image Representation* **13** (2002), 3-21.
- L. Alvarez, F. Guichard, P.L. Lions, J.M. Morel: Axioms and fundamental equations of image processing. *Arch. Rat. Mech. Anal.* **123** (1993), 199-257.
- G. Aubert, P. Kornprobst: *Mathematical Problem in Image Processing*. Springer, 2002.
- O. Faugeras, Q.T. Long, T. Papadopoulos: *The Geometry of Multiple Images*. MIT Press, 2001.
- O. Faugeras, Q.T. Long, T. Papadopoulos: *Three-Dimensional Computer Vision, a Geometric Viewpoint*. MIT Press, 2001.
- R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- T. Lindeberg: *Scale Space Theory in Computer Vision*. Kluwer, 1994.
- Y. Meyer: *Ondelettes et Operateurs I*. Hermann, Paris, 1990.
- J.M. Morel, S. Solimini: *Variational Methods in Image Segmentation*. Birkhauser, 1995.
- Bart M.ter Haar Romeny (editor): *Geometry-Driven Diffusion in Computer Vision*. Kluwer, 1994.
- J. Weickert: *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

En Internet

- <http://serdis.dis.ulpgc.es/~lalvarez/ami/index.html>
Grupo de Análisis Matemático de Imágenes
Universidad de Las Palmas de Gran Canaria.
- <http://www.cmla.ens-cachan.fr/Cmla/index.html>
CMLA
Centre de Mathématiques et de Leurs Applications.



Las Matemáticas y La Visión por Ordenador

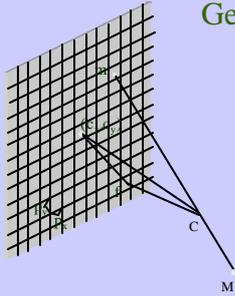
Luis Alvarez León

Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran Canaria

Contenido

- Geometría proyectiva
- Las transformadas integrales
- El cálculo variacional
- Los modelos probabilísticos
- Ecuaciones en derivadas parciales geométricas
- La optimización y el análisis numérico

Geometría Proyectiva



Parámetros Intrínsecos

(c_x, c_y) Proyección ortogonal del foco C en la imagen

$$a_x = \frac{p_x}{f} = \frac{\text{tamaño horizontal pixel}}{\text{distancia focal}}$$

$$a_y = \frac{p_y}{f} = \frac{\text{tamaño vertical pixel}}{\text{distancia focal}}$$

Parámetros Extrínsecos

$t=(t_x, t_y, t_z)$ vector \overline{CO} de traslación entre los sistemas de referencia

$R=(r_{ij})$ matriz de rotación entre los sistemas de referencia

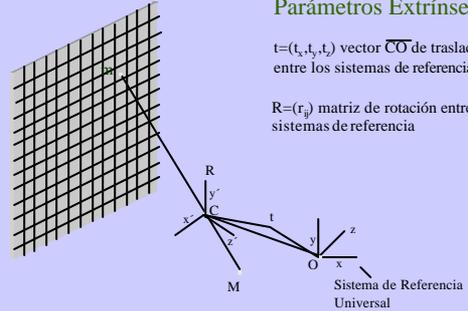


Ilustración Pintor Alemán del Renacimiento (Albrecht Dürer, 1527)



Los espacios proyectivos P^4 y P^3

$$M = (M_1, M_2, M_3) \in R^3 \rightarrow \tilde{M} = (M_1, M_2, M_3, 1) \in P^4$$

$$m = (m_1, m_2) \in R^2 \rightarrow \tilde{m} = (m_1, m_2, 1) \in P^3$$

Puntos en el infinito

$$\tilde{M} = (M_1, M_2, M_3, 0) \in P^4$$

Intersección rectas paralelas
(vector director recta)

Matriz de proyección P asociada a una cámara

Teorema (Faugeras). Sea $\tilde{M}=(M_1, M_2, M_3, M_4) \in P^4$ y $\tilde{m}=(m_1, m_2, m_3) \in P^3$ su proyección en la imagen, entonces:

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{a}_x r_1 + c_x r_3 & \mathbf{a}_x t_x + c_x t_z \\ \mathbf{a}_y r_2 + c_y r_3 & \mathbf{a}_y t_y + c_y t_z \\ r_3 & t_z \end{pmatrix}}_P \begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \end{pmatrix}$$

Donde r_1, r_2 y r_3 son los vectores columna de la matriz R.

Teorema (Caracterización de P): P se puede expresar en función de un conjunto de parámetros intrínsecos y extrínsecos si sólo si

$$\begin{cases} p_{31}^2 + p_{32}^2 + p_{33}^2 = 1 \\ \left(\begin{pmatrix} p_{11} \\ p_{12} \\ p_{13} \end{pmatrix} \wedge \begin{pmatrix} p_{31} \\ p_{32} \\ p_{33} \end{pmatrix} \right) \cdot \left(\begin{pmatrix} p_{21} \\ p_{22} \\ p_{23} \end{pmatrix} \wedge \begin{pmatrix} p_{31} \\ p_{32} \\ p_{33} \end{pmatrix} \right) = 0 \end{cases}$$

Información Euclídea a partir de P

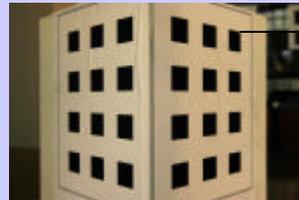
Foco $\tilde{C}=(C_1, C_2, C_3, 1) \rightarrow P\tilde{C}=0$

vector director recta $V=mC$

$$\hat{V}=(V_1, V_2, V_3, 0) \rightarrow P\hat{V}=\tilde{m}$$

Utilización de un calibrador para encontrar los pares

$$(\tilde{M}, \tilde{m})$$



$$\begin{cases} \tilde{M}=(-12.4, 0.8, 4) \text{ en cm} \\ \tilde{m}=(180, 20) \text{ en pixels} \end{cases}$$

Cada par (\tilde{M}, \tilde{m}) determina 2 ecuaciones lineales homogéneas en los coeficientes de P.

Problema de Calibración Fuerte: Encontrar P a partir de un conjunto N de correspondencias $(\tilde{M}_i, \tilde{m}_i)$

Si $N > 5$ obtenemos $2N > 11$ ecuaciones lineales en los coeficientes de P. Es decir $AP=0$ donde A es una matriz de $(2N \times 12)$.

$$P = \min_{\|P\|=1} \|AQ\|^2 = \text{auvector de } A^T A \text{ del autovalor más pequeño.}$$

Si $N=5$ obtenemos 10 ecuaciones lineales en los coeficientes de P, el núcleo de A tiene dimensión > 1 y el método anterior no funciona.

Transformadas Integrales. Convolución con una Gaussiana.

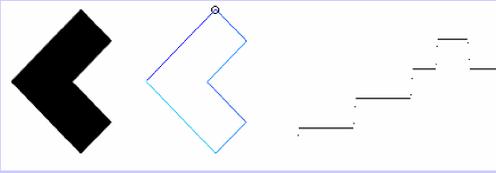
$$I_s(x, y) = \frac{1}{2\pi s^2} \int e^{-\frac{(x-u)^2 + (y-v)^2}{2s^2}} I(u, v) dudv$$



Transformadas Integrales . Transformada de Fourier.

$$\hat{f}(w) = \int_a^b f(x)e^{-iwx} dx$$

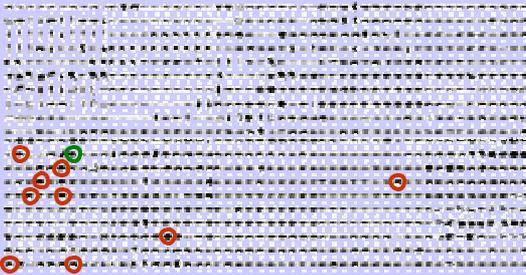
Reconocimiento de formas planas



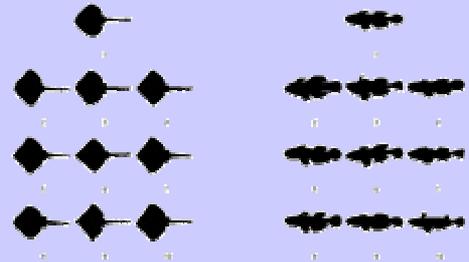
Clasificación de objetos



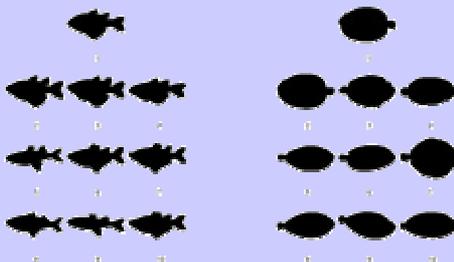
Clasificación de objetos



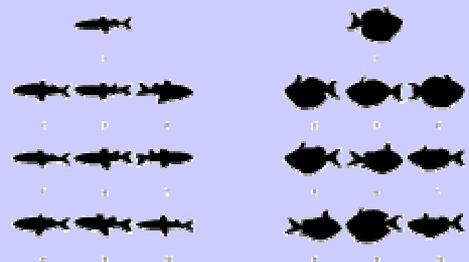
Clasificación de objetos



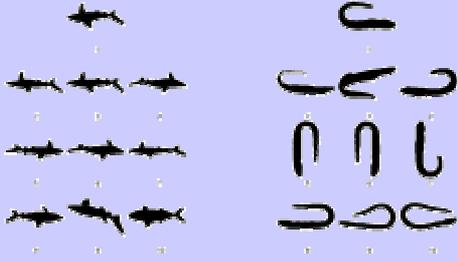
Clasificación de objetos



Clasificación de objetos



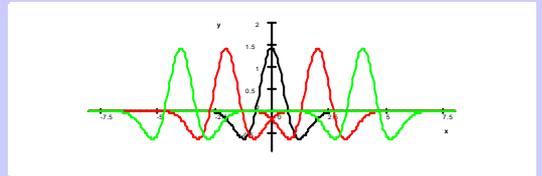
Clasificación de objetos



Transformadas Integrales . Transformada en Ondelettes

$$y_{m,n}(s) = 2^{\frac{m}{2}} y(2^m(s - 2^{-m}n)) \quad m, n \in \mathbb{Z} \quad \text{base } L^2(\mathbb{R})$$

m=01



Análisis MultiResolución .

$$f_y(m,n) = \int f(s) y_{m,n}(s) ds \quad f_f(m,n) = \int f(s) f_{m,n}(s) ds$$

$$\{y_{m,n}(s)\}_{n \in \mathbb{Z}} \cup \{f_{m,n}(s)\}_{n \in \mathbb{Z}} \quad \text{base de } \{f_{m+1,n}(s)\}_{n \in \mathbb{Z}}$$

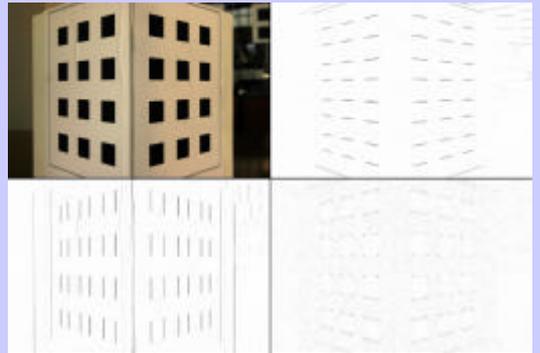
Dimensión 2

$$\left\{ \begin{array}{l} \{y_{m,n}(x) f_{m,l}(y)\} \quad \{f_{m,n}(x) f_{m,l}(y)\} \\ \{y_{m,n}(x) y_{m,l}(y)\} \quad \{f_{m,n}(x) y_{m,l}(y)\} \end{array} \right\}_{n,l \in \mathbb{Z}}$$

base de

$$\{f_{m+1,n}(x) f_{m+1,l}(y)\}_{n,l \in \mathbb{Z}}$$

Codificación de Imágenes



Cálculo Variacional. Eliminación de ruido

$$E(I') = \int (I(x) - I'(x))^2 + \alpha \int \Phi(\|\nabla I'\|)$$

Derivada del funcional $E(I')$

$$\lim_{t \rightarrow 0} \frac{E(I'+tI'') - E(I')}{t} = \langle \nabla E(I'), I'' \rangle = \int \nabla E(I')(\bar{x}) I''(\bar{x})$$

Ecuaciones de Euler-Lagrange

$$\nabla E(I') = -\text{div} \left(\frac{\Phi'(\|\nabla I'\|)}{\|\nabla I'\|} \nabla I' \right) + (I - I') = 0$$

Eliminación de Ruido

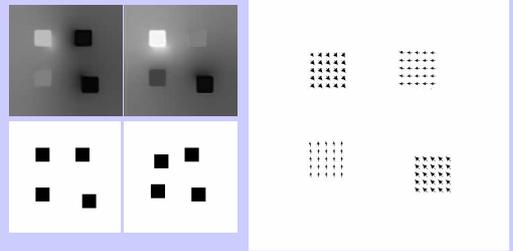


Cálculo Flujo Óptico

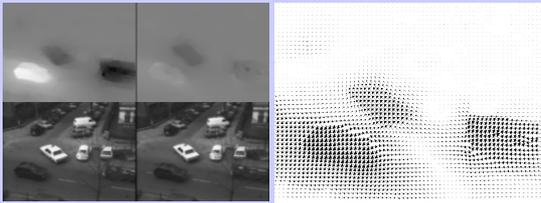


$$E(\vec{h}) = \int_{\Omega} (I_1(x) - I_2(x + \vec{h}))^2 + \alpha \int_{\Omega} \Phi(|\nabla \vec{h}|) \cdot d\omega$$

Flujo óptico



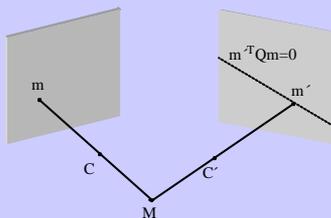
Cálculo flujo óptico



Visión Stereo



Geometría Epipolar



$$Q = \begin{pmatrix} r_{13}f_x - r_{12}f_z & r_{11}f_x - r_{13}f_z & r_{12}f_x - r_{11}f_z \\ r_{23}f_x - r_{22}f_z & r_{21}f_x - r_{23}f_z & r_{22}f_x - r_{21}f_z \\ r_{33}f_x - r_{32}f_z & r_{31}f_x - r_{33}f_z & r_{32}f_x - r_{31}f_z \end{pmatrix}$$

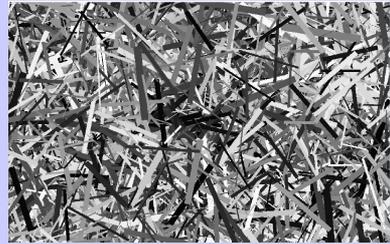
Reconstrucción de la Geometría 3D



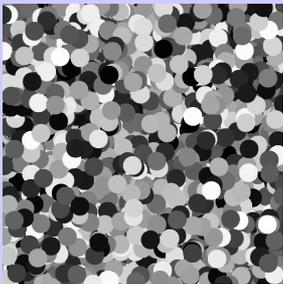
Reconstrucción de la Geometría 3D



Modelos Probabilísticos. El modelo de hojas muertas



Modelos Probabilísticos. El modelo de hojas muertas



Modelos Probabilísticos. El modelo de hojas muertas



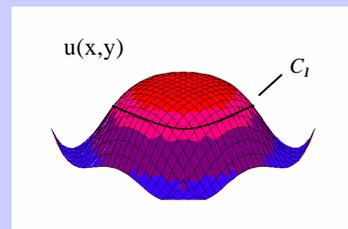
Modelos Probabilísticos. El modelo de hojas muertas

Longitud media segmentos partes visible de las celdas

||

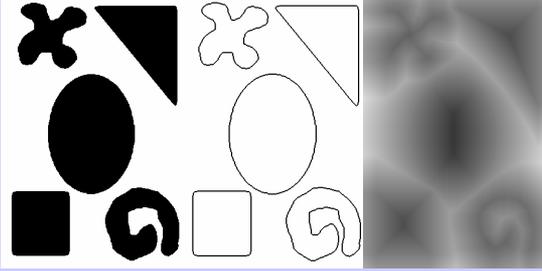
$0.5(\text{Longitud media segmentos grano})$

Ecuaciones diferenciales geométricas. Lineas de nivel



línea de nivel $C_1 \stackrel{\text{def}}{=} \partial\{(x, y) : u(x, y) > I\}$

Líneas de Nivel



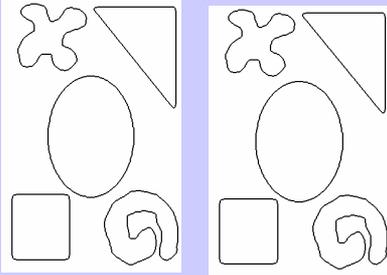
Ecuaciones en derivadas parciales geométricas

$$\frac{\partial u}{\partial t} = G\left(t, \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right)\right) \|\nabla u\|$$

$$\operatorname{curv}(u) = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right)$$

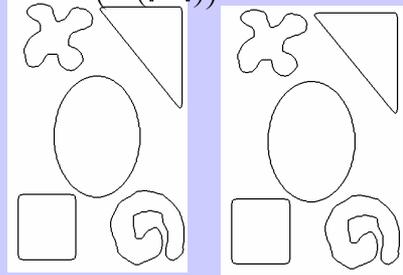
Ecuación de la Curvatura Media

$$\frac{\partial u}{\partial t} = \|\nabla u\| \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \leftrightarrow \frac{\partial C_L}{\partial t} = -k\bar{N}$$



Ecuación Invariante Afin

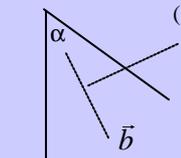
$$\frac{\partial u}{\partial t} = \|\nabla u\| \left(\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \right)^{\frac{1}{3}} \leftrightarrow \frac{\partial C_L}{\partial t} = -k^{\frac{1}{3}}\bar{N}$$



Detector morfológico de esquinas

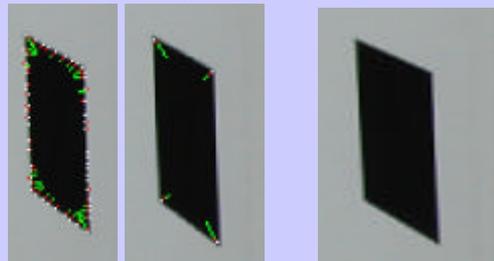
$$\frac{\partial u}{\partial t} = \|\nabla u\| \left(\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \right)^{\frac{1}{3}} \leftrightarrow \frac{\partial C_L}{\partial t} = -k^{\frac{1}{3}}\bar{N}$$

(x_0, y_0)



$$(x(t), y(t)) = (x_0, y_0) + \frac{t}{\sqrt{\tan \frac{\alpha}{2}}} (b_x, b_y)$$

Aplicación del detector morfológico de esquinas (1)



$t_0=0$

$t_0=6$

Aplicación del detector morfológico de esquinas (2)



La optimización y el análisis numérico

- Interpolación de funciones
- Resolución de grandes sistemas de ecuaciones
- Cálculo de autovalores y autovectores
- Descomposición de SVD de matrices
- Análisis Numérico de Ecuaciones en Derivadas Parciales
- Estimación robusta de parámetros
- Ajuste de parámetros (Levenberg-Marquardt)
- Estimadores robustos