

# CAPÍTULO 2

## Visión Computacional

---

### 2.1 Introducción

La visión computacional, también conocida como visión artificial o visión técnica, es un subcampo de la inteligencia artificial. El propósito de la visión artificial es programar una computadora para que "entienda" una escena o las características de una imagen, en otras palabras, hacer que las computadoras "vean".

Los objetivos típicos de la visión artificial incluyen:

- La detección, segmentación, localización y reconocimiento de ciertos objetos en imágenes (ej. caras humanas),
- La evaluación de los resultados (ej. segmentación, registro),
- Registro de diferentes imágenes de una misma escena u objeto, hacer concordar un mismo objeto en diversas imágenes,
- Seguimiento de un objeto en una secuencia de imágenes,
- Mapeo de una escena para generar un modelo tridimensional de la escena; tal modelo podría ser usado por un robot para navegar por la escena.
- Estimación de las posturas tridimensionales de humanos,
- Búsqueda de imágenes digitales por su contenido.

Estos objetivos se consiguen por medio de reconocimiento de patrones, aprendizaje estadístico, geometría de proyección, procesamiento de imágenes, teoría de gráficos y otros campos. La visión artificial cognitiva está muy relacionada con la psicología cognitiva y la computación biológica. (Fig. 2.1)

El objetivo de la visión por computadora (VC) es tomar decisiones útiles acerca de los objetos físicos reales del mundo (de escenas) en base a imágenes adquiridas digitalmente. Por lo tanto, la tarea de la VC es la construcción de descriptores de la escena en base a características relevantes contenidas en una imagen. "La visión humana es una tarea de procesamiento de información, es el proceso de describir con las imágenes lo que está presente en el mundo real y dónde se encuentra." El cerebro es capaz, de alguna manera, de representar esta información en color, forma, belleza, movimiento y detalle.

El estudio de la visión incluye no sólo la tarea de cómo extraer de las imágenes los aspectos varios del mundo que nos son útiles, sino también, incluye investigar la naturaleza de las representaciones internas y hacerlas disponibles como base para la toma de decisiones sobre nuestros pensamientos y acciones.

# CAPÍTULO 2

## Visión Computacional

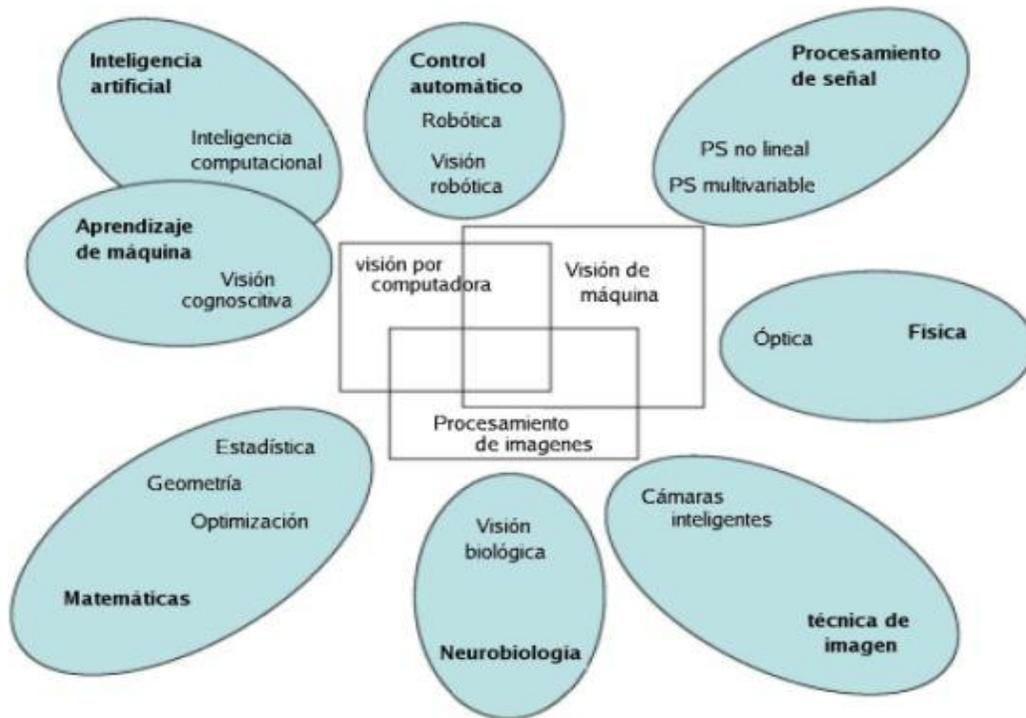


Fig 2.1 Esquema de relaciones entre visión por computadora y otras áreas afines. [21]

### 2.2 Procesamiento de imágenes

El procesamiento de imágenes tiene como objetivo mejorar el aspecto de las imágenes y hacer más evidentes en ellas ciertos detalles que se desean hacer notar.

#### 2.2.1 Restauración de una imagen:

Toma una imagen distorsionada y devuelve la imagen original sin ruido. Permite restaurar imágenes que con el tiempo han sido dañadas fig. 2.2.

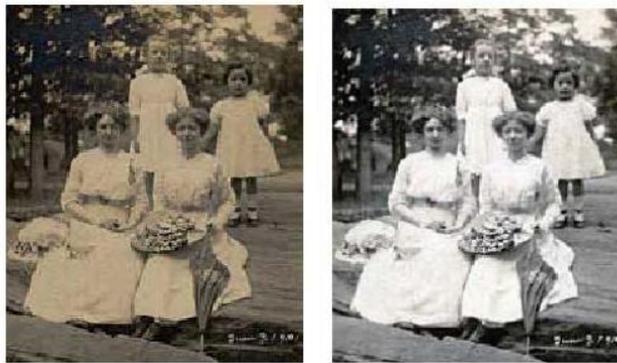


Fig. 2.2 Comparación de una imagen con ruido y una imagen restaurada. [21]

### 2.2.2 Realce de una imagen:

Altera una imagen para hacerla más clara a los observadores. A menudo se usa incrementando el contraste en imágenes muy oscuras o muy claras fig. 2.3.



Fig. 2.3 Comparación de una imagen normal y después de aplicar realce. [21]

### 2.2.3 Comprensión de una imagen:

Imita el sistema visual humano al extraer una imagen. Incluye diferentes tareas:

- Segmentación: Identifica objetos en una imagen.
- Clasificación: Asigna etiquetas a los objetos o píxeles individualmente.
- Interpretación: Extrae la imagen en su totalidad.

## 2.3 Características.

Una característica es un atributo usado para hacer decisiones respecto a objetos en la imagen. Algunos atributos son naturales y se definen mediante la apariencia visual de la imagen, en tanto otros, los artificiales, son el resultado de operaciones realizadas a la imagen.

El poder realizar mediciones sobre las imágenes generalmente requiere que sus características estén bien definidas, los bordes bien delimitados y tanto el color como el brillo sean uniformes. El tipo de mediciones a realizar para cada característica específica es un factor importante para poder determinar los pasos apropiados para su procesamiento.

## 2.4 Filtros de imágenes.

Se define como ruido cualquier entidad en las imágenes, datos o resultados intermedios que no son interesantes para la computación que se pretende llevar a cabo [22]. El filtrado modifica los píxeles en una imagen basado en alguna función de una vecindad local de píxeles.

# CAPÍTULO 2

## Visión Computacional

Las técnicas de filtraje son transformaciones de la imagen píxel a píxel, que no dependen solamente del nivel de gris de un determinado píxel, sino también del valor de los niveles de gris de los píxeles vecinos en la imagen original. El proceso de filtraje se realiza utilizando matrices denominadas máscaras, que son aplicadas sobre la imagen. La imagen resultante de la aplicación de un filtro es una nueva imagen. Los filtros sirven para suavizar o realzar detalles de la imagen, o inclusive minimizar efectos de ruido.

En la tabla 2.1, se ilustra el ejemplo de una máscara 3x3. La máscara se centra sobre el píxel de interés de modo que el nuevo valor del píxel depende de los píxeles que cubre la máscara.

Tabla 2.1 Ejemplo de máscara 3x3

$w_{1,1}$	$w_{1,2}$	$w_{1,3}$
$w_{2,1}$	$w_{2,2}$	$w_{2,3}$
$w_{3,1}$	$w_{3,2}$	$w_{3,3}$

### 2.5. Filtros de suavizado:

El objetivo de los filtros de suavizado es eliminar ruidos o pequeños detalles que no sean de interés. Esto corresponde a un filtro paso-bajo en el dominio de la frecuencia, es decir que se reducen o eliminan las altas frecuencias fig. 2.4.

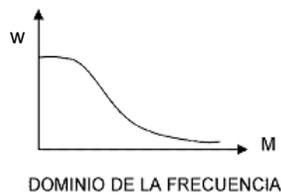


Fig. 2.4. Filtro pasa-bajos en frecuencia

Existen varios tipos de filtros de suavizado, los más comunes son:

- Promedio o media aritmética: Obtiene el promedio de los píxeles vecinos ( $w=1$ ); es decir, todos los valores de la máscara son 1.
- Mediana: Substituye el valor del píxel central por el de la mediana de los valores contenidos en el vecindario.
- Gaussiano: Este filtro implementa máscaras que intentan imitar la forma de una Gaussiana:  $G(x, y) = e^{-(x+y)^2} / 2\sigma^2$ , donde  $x, y$  son las coordenadas de la imagen y sigma una desviación estándar de la probabilidad de distribución asociada.

# CAPÍTULO 2

## Visión Computacional

---

La desviación estándar sigma es el único parámetro del filtro Gaussiano y es proporcional al tamaño del área de los vecinos más cercanos en que opera este filtro. Los píxeles más lejanos del centro del operador tienen menor influencia y los píxeles más allá de tres valores de sigma a partir del centro tienen una influencia insignificante.

El filtro Gaussiano, en general, da mejores resultados que un simple promedio o media y se argumenta que la vista humana hace un filtrado de este tipo. El filtro Gaussiano “normal” tiene la desventaja de suavizar las orillas o discontinuidades, haciendo que se borren. El algoritmo de mediana es efectivo con imágenes de poco ruido. Su efectividad decrece drásticamente con imágenes ruidosas.

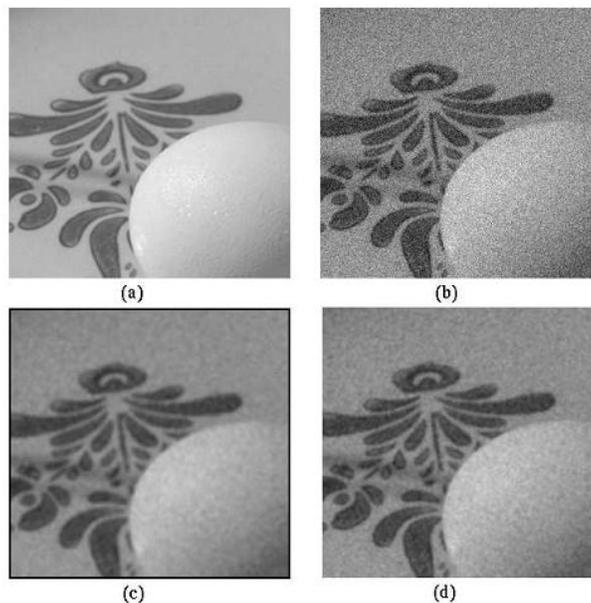


Fig 2.5 Filtros pasa-bajos; a) imagen original, b) imagen con ruido, c) resultado de aplicar un filtro promedio con máscara de 5x5, d) resultado de filtro gaussiano,  $\sigma = 1.0$ .

### 2.6 Filtros de acentuamiento

El objetivo de los filtros de acentuamiento es intensificar los detalles y cambios bruscos de intensidad mientras atenúa las bajas frecuencias. El resultado es un acentuamiento de las orillas (edge-sharpening). Se conocen como filtros pasa-altos porque dejan pasar las altas frecuencias y eliminan las bajas frecuencias, en forma inversa al filtro pasa-bajo fig 2.6.



2.6. Filtro pasa-alto en frecuencia

Existen también varias formas de implementar este tipo de filtrado. Una forma típica de un filtro pasa-alto es una máscara del tipo de la fig.2.7.

-1	-1	-1
-1	8	-1
-1	-1	-1

Fig. 2.7. Máscara 3x3 para un filtro pasa-altos simple

Para este filtro la suma de los pesos es cero, de modo que una región de intensidad constante resultaría en un valor 0. A diferencia del filtro de suavizado, los pesos de los vecinos son negativos, este efecto substractivo genera la acentuación de los cambios de intensidad. Otra forma de implementar este tipo de filtrado es restando a la imagen original el resultado de un filtro pasa-bajo:

$$PA = \text{original} - PB$$

Donde PA representa la imagen resultante de aplicar un filtro pasa-alto y PB es el resultado de aplicar un filtro pasa-bajos a la imagen original.

### 2.7 Realce de contraste

La técnica de realce de contraste tiene como objetivo mejorar la calidad de las imágenes bajo ciertos criterios subjetivos del ojo humano. Normalmente esta técnica es utilizada como una etapa de pre-procesamiento para sistemas de reconocimiento de patrones.

El contraste entre dos objetos se puede definir como la razón entre sus niveles de gris medios. La manipulación de contraste consiste en una transferencia radiométrica en cada píxel, con el objetivo de aumentar la discriminación visual entre los objetos presentes en la imagen. La operación es realizada punto a punto. Esta transformación radiométrica se realiza con la ayuda de histogramas manipulados para obtener el realce deseado.

# CAPÍTULO 2

## Visión Computacional

El histograma es una de las formas más comunes de representar la distribución de los niveles de gris de una imagen, y también la más utilizada en procesamiento de imágenes. El histograma provee la información de cuántos píxeles poseen un determinado nivel de gris en la imagen en un intervalo definido entre 0 (negro) y 255 (blanco) para una imagen de 8 bits, ofreciendo datos importantes como la intensidad media y la dispersión de los valores de niveles de gris, siendo esta última la medida de contraste de la imagen.

Cuanto mayor es la dispersión a lo largo del eje que representa los niveles de gris, mayor es el contraste de la imagen fig.2.8.

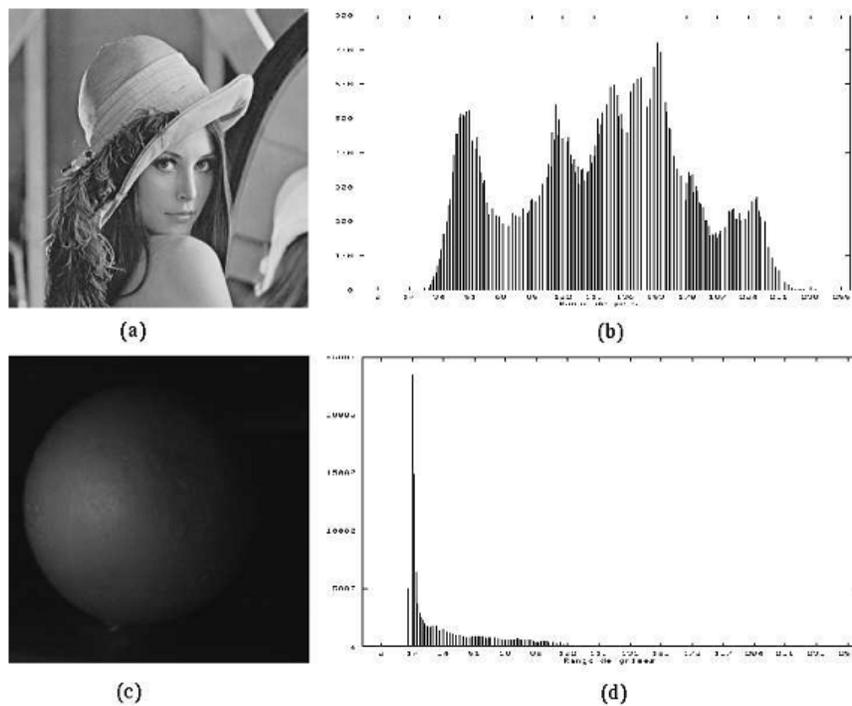


Fig. 2.8. Ejemplos de histogramas; a) Imagen con varias intensidades, b) Su histograma mostrando un rango amplio de grises o alto contraste, c) Imagen oscura, d) Su histograma con un rango de grises reducido o menor contraste. [22]

### 2.8 Movimiento y seguimiento

Al considerar movimiento, puede ser que los objetos se muevan, o que la cámara se mueva, o ambos. Sin embargo, todos los casos se pueden agrupar que existe un movimiento relativo entre la cámara y los objetos. De esta forma se obtiene una secuencia de imágenes entre las que existen pequeñas diferencias debidas a dicho movimiento relativo.

El análisis de una secuencia de imágenes se puede ver desde dos puntos de vista, que corresponden a los dos enfoques principales para imágenes en movimiento:

- Continuo: se considera la secuencia de imagen como un flujo de intensidades cambiantes a la se denomina flujo óptico.
- Discreto: se considera la secuencia de imágenes como un conjunto de diferentes imágenes estáticas.

### 2.9 Flujo óptico

Diversos experimentos han demostrado que la vista humana responde directamente al movimiento, que se puede considerar como uno de los aspectos básicos de la visión humana. Para esto se considera que el movimiento produce cambios diferenciales en la imagen que son percibidos como una especie de flujo de pixeles en el espacio. Esto se puede ver como un arreglo de vectores de velocidad que se le denomina flujo óptico y puede ser obtenido de la secuencia de imágenes.

A partir del flujo óptico se pueden obtener otras características, como el movimiento global, información de profundidad y orillas.

#### a) Obtención del flujo óptico

Para estimar el flujo óptico, se considera una secuencia de imágenes en movimiento como una función continua en tres dimensiones,  $X$ ,  $Y$ , y tiempo ( $t$ ):  $f(X, Y, t)$ . Es decir, al integrar la secuencia de imágenes en diferentes tiempos, la intensidad de cada uno depende de su posición en la imagen ( $X$ ,  $Y$ ), y de la imagen en el tiempo  $t$ , como se muestra en la fig. 2.9.

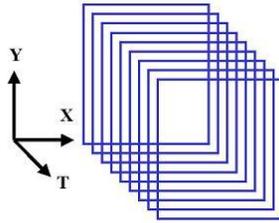


Fig. 2.9. Secuencia de imágenes.

### b) Utilización del flujo óptico

Una vez estimado el flujo óptico, este puede ser utilizado para obtener información adicional de las imágenes. Se aplica para obtener información acerca de la profundidad mediante el cálculo del foco de expansión. También se puede utilizar como base para la detección de bordes y la segmentación.

### c) Foco de expansión

Si consideramos que el observador se mueve y los objetos son estáticos, todos los vectores de velocidad parecen unirse en un punto de la imagen. A dicho punto se le denomina el foco de expansión o FOE. Si existen varios objetos con diferentes movimientos, a cada uno corresponde un foco de expansión. Esto se puede entender mejor con la fig. 2.10.

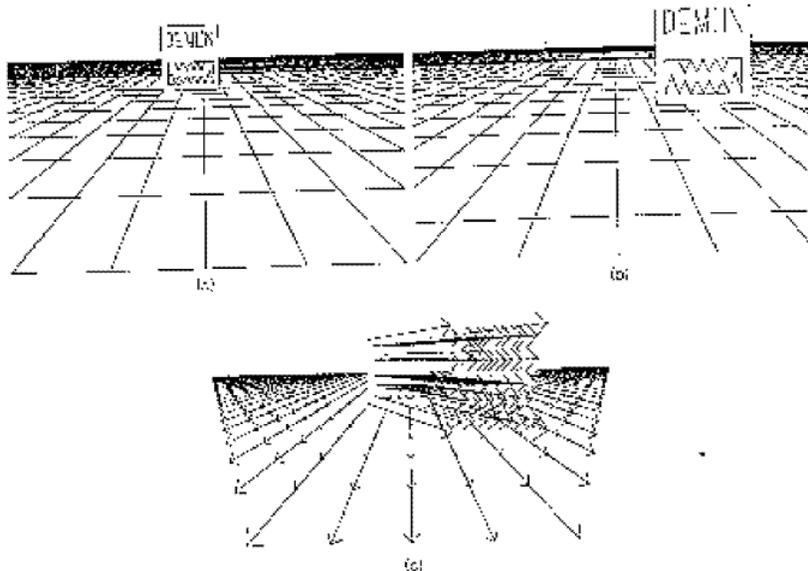


Fig. 2.10. Foco de expansión. En a) se muestra la imagen en tiempo inicial y en b) en un tiempo después. c) Los vectores de flujos ópticos. d) Profundidad, e) Profundidad relativa y orillas

# CAPÍTULO 2

## Visión Computacional

---

A partir del flujo óptico también es posible estimar la profundidad relativa o gradiente de la superficie y detectar los cambios bruscos que corresponden a las orillas. Podemos calcular la normal en cada punto de la superficie que corresponde a su profundidad relativa. Las discontinuidades en la superficie se refieren a las orillas. La ventaja de estas orillas, respecto a orillas obtenidas directamente de la intensidad de la imagen, es que corresponden a cambios de profundidad que representan las fronteras de los objetos o cambios bruscos en la superficie.

### 2.10 Múltiples imágenes

Un enfoque alternativo al flujo óptico que considera la secuencia de imágenes como un continuo, es de considerarlas de forma discreta. En este caso el movimiento se analiza a partir de un conjunto de imágenes. Al considerar múltiples imágenes estáticas, el problema principal se convierte en el apareamiento de puntos entre las imágenes. Para simplificarlo, podemos considerar que el movimiento entre imágenes consecutivas es pequeño, es decir, que imágenes consecutivas son similares. Con estas consideraciones, se aplican las siguientes cinco heurísticas de movimiento:

- a) Distancia máxima: un punto tiene una velocidad máxima  $V$  y se mueve una distancia máxima, donde;  $t$  es el tiempo que transcurre entre la toma de una imagen y la siguiente.
- b) Cambios de velocidad: la velocidad de un punto de una imagen a la siguiente es similar, es decir, existen pequeños cambios.
- c) Movimiento común: regiones de puntos cercanos en la imagen tienen el mismo movimiento o un movimiento muy similar.
- d) Consistencia: un punto en una imagen corresponden a un solo punto en la imagen siguiente.
- e) Movimiento conocido: en ocasiones se tiene conocimiento a priori del tipo de movimiento de los objetos y/o del observador.

El problema de análisis de imágenes en movimiento, se enfoca a analizar las imágenes individuales, segmentándolas en atributos u objetos relevantes, para a partir de estos buscar similitudes y diferencias entre imágenes y encontrar información del movimiento.

### 2.11 Seguimiento

Un uso común del análisis de una secuencia de imágenes en movimiento es el seguimiento de objetos en las imágenes. Esto tiene múltiples aplicaciones prácticas, como el seguimiento de las partes del cuerpo de una persona para reconocer actividades o ademanes, y el seguimiento de vehículos.

El seguimiento en una secuencia de imágenes consiste en determinar la posición y velocidad de un punto en una imagen, dada su posición y velocidad en una secuencia anterior de una o más imágenes.

El seguimiento se puede realizar en base a diferentes atributos de la imagen, en particular se pueden distinguir las siguientes clases de objetos.

- Modelos rígidos bidimensionales y tridimensionales de objetos
- Modelos deformables
- Regiones
- Características de la imagen

Se han desarrollado diversas técnicas para el seguimiento de objetos en imágenes, entre las principales se pueden mencionar:

- Filtros de Kalman,
- Técnicas de simulación estocástica, como el algoritmo de condensación,
- Técnicas heurísticas, que aprovechan las heurísticas de movimiento ya mencionadas.

Dada la localización del objeto o región de interés en una imagen, se pueden utilizar las heurísticas de velocidad máxima y cambios de velocidad para delimitar la región de búsqueda en la siguiente imagen de la secuencia. Esto es particularmente útil si se consideran imágenes con una separación temporal mínima. La fig. 2.11 ilustra la aplicación de este principio en el seguimiento de una mano en una secuencia de imágenes. En el siguiente ejemplo, la región de la mano es segmentada en la base al color de piel, y posteriormente se hace un seguimiento en una ventana alrededor de su posición en la imagen previa, utilizando la heurística de velocidad máxima.

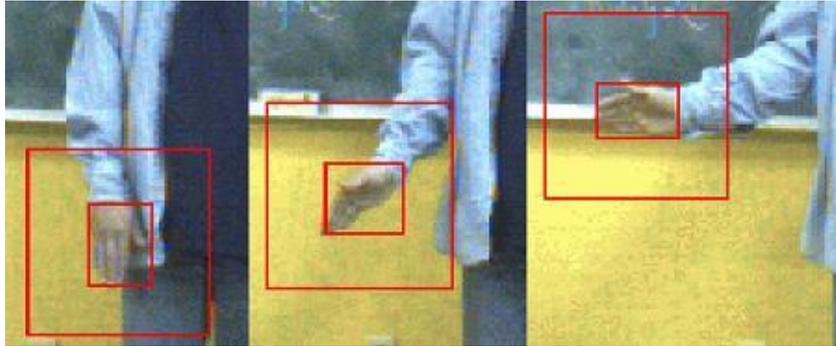


Fig. 2.11 Seguimiento. Se ilustra el proceso de seguimiento de una región, el cuadro exterior define la región de búsqueda.

### 2.12 Clasificación general de los métodos de seguimiento visual

Atendiendo a las técnicas utilizadas para caracterizar y localizar en la imagen los objetos que son seguidos, es posible establecer la siguiente clasificación general.

#### 2.12.1 Seguimiento de “blobs”

Esta técnica consiste en realizar en cada cuadro una segmentación previa del objeto seguido, caracterizándolo como un agregado de píxeles de acuerdo a algún determinado criterio. Este criterio puede ser el color, la textura, la intensidad o la presencia de movimiento, entre otros. El “blob”, o región de píxeles que corresponden al objeto, se parametriza mediante su centroide, de forma que el objetivo de seguimiento será ofrecer a cada cuadro la nueva posición de este lugar geométrico.

Frecuentemente se asume que sólo existe un objeto en movimiento en la imagen y que el objeto mantiene un tamaño invariable a lo largo de la secuencia. El método es atractivo por su simplicidad, pero el éxito del seguimiento dependerá fuertemente de los resultados de la segmentación, que pueden no ser satisfactorios en imágenes relativamente complejas.

#### 2.12.2 Seguimiento mediante patrones

Se trata de buscar una correspondencia píxel a píxel entre una región de píxeles de un cuadro y los cuadros sucesivos. Habitualmente se empleará alguna técnica de correlación de patrones. Las posiciones de máxima correlación encontradas para el patrón en cada cuadro proporcionan la trayectoria seguida por el objeto.

# CAPÍTULO 2

## Visión Computacional

---

En unos casos se asume que la apariencia del objeto no cambia demasiado de un cuadro a otro. En otros se plantean métodos de correlación que consideran algún tipo de transformación del patrón (geométrica o de luminosidad). Una solución consiste en plantear la búsqueda como un problema de optimización local que considere transformaciones geométricas más complejas que una simple traslación. La transformación afín es la alternativa más utilizada.

En general, estos métodos requieren mayor esfuerzo de cómputo que el anterior, ya que la correlación debe evaluarse para cada posible localización del patrón. Si se contemplan transformaciones geométricas distintas de la traslación la complejidad aumenta por la necesidad de realizar la transformación de la imagen. Se denominará seguimiento de puntos característicos cuando el objeto de interés se caracterice por varias pequeñas regiones que son seguidas de forma más o menos autónoma mediante patrones.

### 2.12.3 Seguimiento de contornos

Esta técnica fue introducida originalmente por Kas y Witkin con un planteamiento de minimización de energía y posteriormente se han desarrollado diversas variantes. Se trata de ajustar de forma dinámica (a cada cuadro) un modelo del contorno del objeto de imagen seguido. Este planteamiento tiene la ventaja de que puede ser menos costoso computacionalmente que los anteriores ya que sólo requiere de búsquedas de bordes a lo largo de líneas normales al contorno y estimación del nuevo contorno por mínimos cuadrados. Sin embargo el proceso de ajuste de los contornos ha demostrado ser problemático si la imagen no presenta bordes bien definidos.

### 2.12.4 Seguimiento basado en reconocimiento de formas

En algunos trabajos se recurre a caracterizar el elemento de imagen a seguir a partir del conocimiento a priori del que se dispone sobre su apariencia. Una posibilidad consiste en utilizar un conjunto de muestras del objeto de imagen para introducir técnicas de reconocimiento que permitan su identificación y localización en la imagen en los sucesivos cuadros.

Podrían también clasificarse los seguidores que hacen uso de operadores de localización especializados, que se diseñan a partir de la información de que se dispone sobre la apariencia del elemento a seguir. Se trata de pequeños algoritmos que buscan formas concretas, como una pupila de un ojo caracterizada

# CAPÍTULO 2

## Visión Computacional

---

por una forma circular oscura rodeada de un área blanca, o los límites de los labios a partir de los extremos de un segmento de recta extraído de una imagen de bordes.

Estos algoritmos pueden resultar ventajosos frente a otros métodos de localización, como el empleo de patrones, en algunas circunstancias. Por ejemplo, cuando existen variaciones de iluminación, estos operadores pueden desarrollarse para responder más al contraste que a valores absolutos de píxeles y resultar así más robustos. También es ventajoso su uso cuando hay alta variabilidad en la forma buscada, porque no buscan una forma exacta sino alguna disposición característica de variaciones de luminosidad.

La visión computacional es mejor conocida como visión artificial, cuyo propósito es programar una computadora para que determine una escena o las características de una imagen. La visión computacional tiene objetivos clave los cuales se consiguen por medio de reconocimiento de patrones, aprendizaje estadístico, geometría de proyección, procesado de imágenes, teoría de gráficos y otros campos, vimos varios métodos para el procesamiento de estas imágenes.

La visión computacional en la industria es la utilización de dispositivos ópticos para recibir e interpretar automáticamente una imagen de una escena real o para obtener informaciones o controlar máquinas. Así también es un dominio de la computación cuyo objetivo es modelar el mundo real o reconocer objetos en imágenes digitales.

La visión computacional nos sirve para inspeccionar, medir, contar, identificar, verificar, localizar, comparar, clasificar, leer, controlar, monitorear y reconocer información adquirida usando vídeo, cámaras red infrarrojo, cámaras digitales, radares, o transductores especializados.

Un sistema de visión debe crear una imagen con calidad suficiente para permitir la extracción de las informaciones deseadas de la imagen. Existe una variedad de factores que contribuyen para la calidad de la imagen: resolución, contraste, profundidad de campo, errores de perspectiva y errores geométricos (distorsión), todos o cada uno de ellos para intensificar los detalles de las imágenes u objetos. Cada método o filtro tiene sus características, cada uno se usa de acuerdo a las necesidades. Para nuestro caso es considerado el método de realce de contraste ya que es la más utilizada en el procesamiento de imágenes además que el contraste entre dos objetos se puede definir como la razón entre sus niveles de

# CAPÍTULO 2

## Visión Computacional

---

gris medios. Sin embargo no debemos olvidar que el objetivo de la visión computacional es capturar una imagen, tratarla y analizarla y finalmente hacer algo con los resultados; es decir dar alguna aplicación a los resultados obtenidos.

Ahora una vez que hemos visto cómo interpretar las imágenes y los dispositivos o instrumentos para adquirirlas, que en nuestro caso es una cámara web veremos cómo manipular dicha cámara mediante un motor a pasos junto con su driver y hacer un “zoom” para obtener una imagen lo más nítida posible o para tener una imagen enfocada, para lo cual ahora estudiaremos en el siguiente capítulo los tipos de motor a pasos y sus características.