



CONTENIDO

Unidad IV. Visión Artificial	2
4.1 Digitalización	3
4.2 Procedimiento de bajo nivel	7
4.3 Recuperación de información tridimensional	8
4.4 Algoritmo de Waltz	10
4.5 Visión activa	10
4.6 Reconocimiento de objetos y escenas, y descripción de imágenes	12



Unidad IV. Visión Artificial

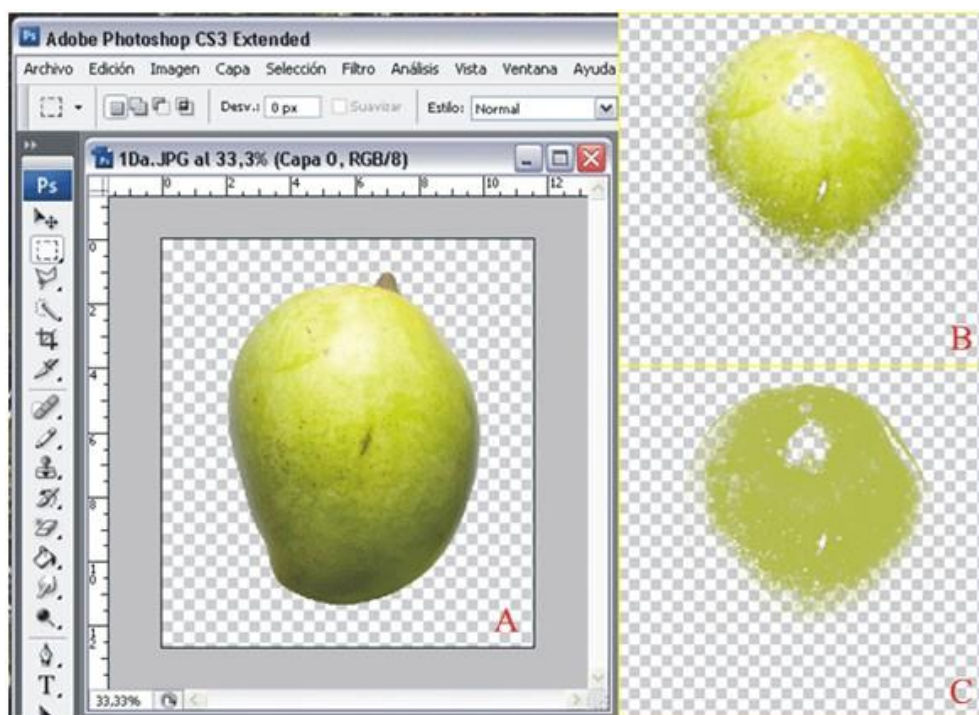


Ilustración 1. Visión Artificial. [flickr.com](https://www.flickr.com/photos/14910410@N06/10051111111/). CCO

La Visión Artificial es una de las áreas de investigación dentro de la Inteligencia Artificial de gran aplicación en la actualidad. Uno de los desafíos de la Visión Artificial es interpretar y comprender el mundo real a partir de una imagen. Para ello, se deben realizar una serie de procesos para el análisis de las imágenes. En el desarrollo de esta unidad se explicará con mayor detalle cada uno de estos procesos. Se empezará con el proceso de digitalización. Luego, se explicará en qué consiste el procedimiento de bajo nivel. Seguidamente, se presenta la recuperación de información tridimensional y como se puede efectuar la misma. De igual forma, se presenta el algoritmo de Waltz y su importancia en esta temática. Finalmente, se explican conceptos importantes como lo son la visión activa y el reconocimiento.



La visión artificial es un campo de inteligencia artificial que entrena a las computadoras para interpretar y comprender el mundo visual. El objetivo es extraer características de una imagen para su descripción e interpretación por la computadora. Lo que se busca en Visión es obtener descripciones útiles para cada tarea a realizar, en donde, la misma necesitará modificar ciertos atributos que le permita identificar los objetos en la imagen.

El verano del año 1966, Seymour Papert y Marvin Minsky del grupo de Inteligencia Artificial del MIT comenzaron un proyecto titulado Proyecto de Visión de Verano el cual sentó las primeras bases para construir un sistema que fuera capaz de analizar una escena e identificar objetos en ella. En la década de los 70 se avanza un poco más en el área con las contribuciones de David Marr, neurocientífico del MIT, quien establece los componentes básicos de la visión por computadora moderna. Por sus aportes en esta área se le conoce como el padre de la visión por computadora moderna. De igual forma, se establecieron los primeros conceptos de muchos de los algoritmos de visión que se utilizan en la actualidad, algunos de ellos son: el etiquetado de líneas, la representación de objetos como interconexiones de estructuras más pequeñas, el modelado no poliédrico y poliédrico, flujo óptico, la extracción de bordes de imágenes, y estimación del movimiento.

La visión artificial la componen un conjunto de procesos destinados a realizar el análisis de imágenes. Estos procesos son: captación de imágenes, memorización de la información, procesado e interpretación de los resultados.

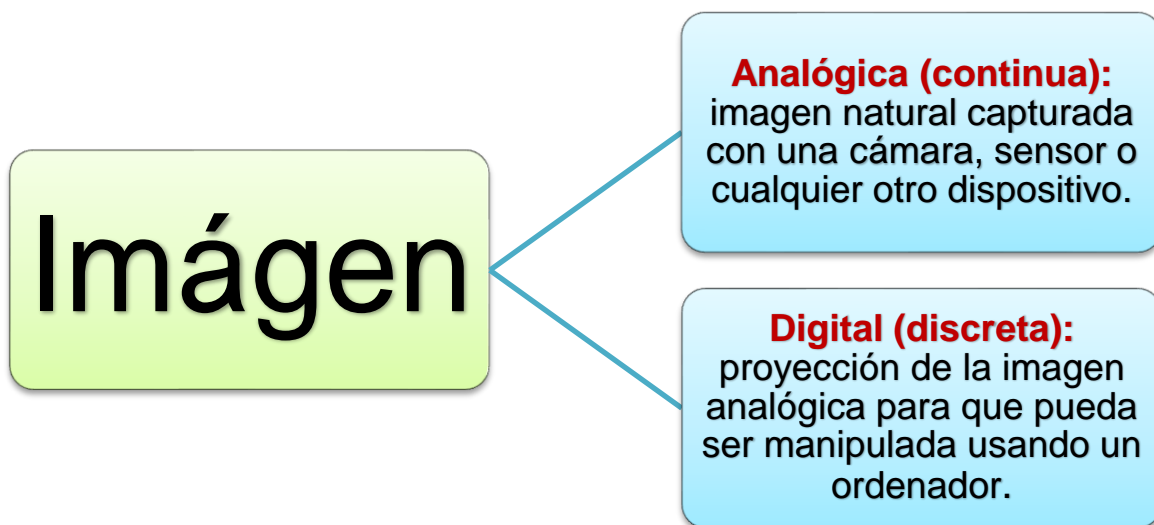
4.1 Digitalización

La digitalización es el proceso de convertir información a un formato digital. En el formato digital la información se organiza en unidades de datos discretas (llamadas bits) que se pueden abordar por separado (generalmente en grupos de múltiples



bits llamados bytes). Estos son los datos binarios que las computadoras y muchos dispositivos con capacidad informática como las cámaras digitales pueden procesar.

En la temática que vamos a tratar en esta unidad relacionada con la Visión Artificial nos enfocaremos en la transformación de una imagen analógica a otra digital. Esto constituye el primer paso en cualquier aplicación de procesamiento de imágenes digitales. Para comprender mejor esta etapa de la visión artificial es necesario definir los conceptos de imagen analógica e imagen digital.



Lo primero que debemos hacer para digitalizar las imágenes es el muestreo, en donde, se convierte la imagen en una matriz discreta de $M \times N$ píxeles. Estos píxeles están dispuestos en una cuadrícula regular (o ráster) de filas y columnas, por lo que se suele trabajar como una matriz.

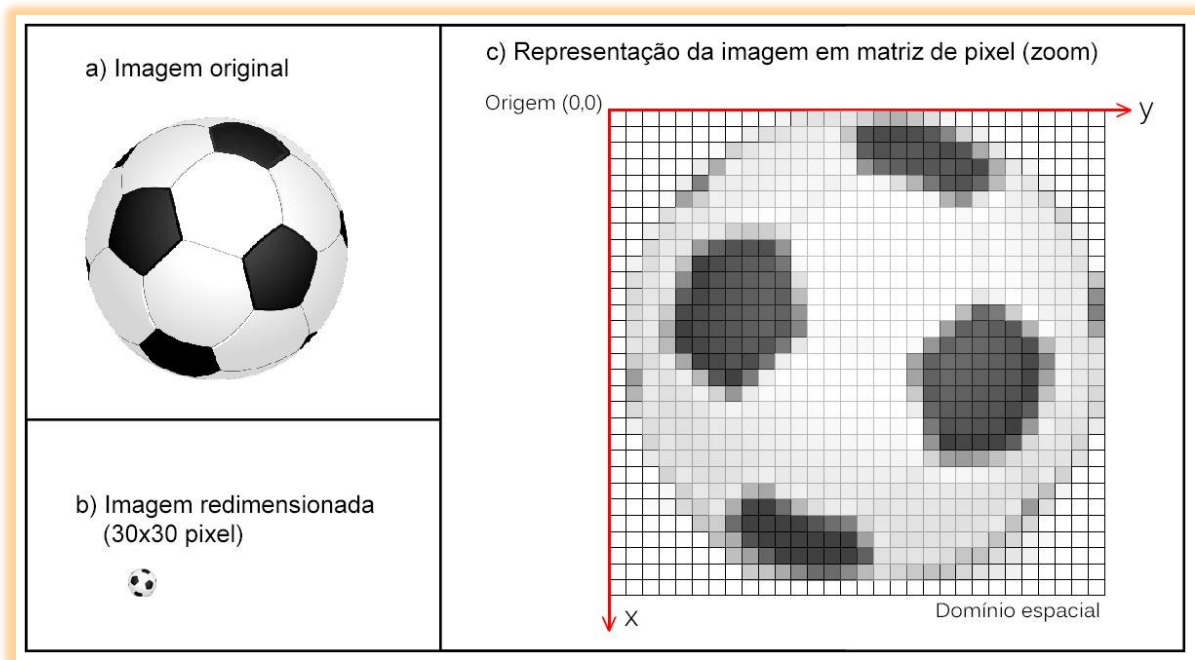


Ilustración 2. Imagen representada en matriz de píxeles. commons.wikimedia.org. CCO

Cada píxel en una imagen en escala de grises (también llamada imagen de intensidad) es un entero sin signo de 8 bits, lo que significa que puede tener un valor entero entre 0 y 255. Un valor de 0 corresponde al tono negro, un valor de 255 a blanco puro. Los valores entre estos extremos producen varios niveles de gris entre blanco y negro.

Una imagen en color también se almacena como una trama de píxeles. Cada píxel ahora está representado por tres valores enteros entre 0 y 255: uno para el rojo, otro para el verde y otro para el azul. Estas tres intensidades primarias se agregan para reproducir un determinado color en la pantalla, y esta forma comúnmente utilizada de representar el color se denomina esquema de color RGB.

El muestreo está relacionado con la resolución espacial de la imagen. La imagen capturada se expresa como una función continua $f(x, y)$ de dos coordenadas en el



plano, donde x e y son las coordenadas de un plano que contiene todos los puntos de la imagen, y $f(x,y)$ es la amplitud en el punto (x, y) a la cual se le llama intensidad o nivel de gris de la imagen en ese punto.

En una imagen digital las coordenadas x e y , así como, los valores de intensidad de la función f son discretos y finitos. Si representamos estas funciones gráficamente, se tienen 3 dimensiones: dos que corresponden a las coordenadas de la imagen y la tercera a la función de intensidad. Se deben responder dos preguntas en relación con el muestreo de la función de imagen:

1. Se debe determinar el período de muestreo: la distancia entre dos puntos de muestreo vecinos en la imagen
2. Se debe establecer la disposición geométrica de los puntos de muestreo (cuadrícula de muestreo).

Factores importantes que considerar al muestrear la imagen para obtener una representación digital

El intervalo de muestreo (resolución)

El patrón espacial de los puntos de muestreo



La digitalización incluye la cuantificación del valor de la función de intensidad y el muestreo de las dos dimensiones espaciales. La cuantificación de la imagen asigna a cada muestra continua un valor entero. El rango continuo de la función de imagen $f(x, y)$ se divide en K intervalos. Cuanto más fino sea el muestreo (es decir, cuanto mayor sea M y N) y la cuantificación (cuanto mayor sea K), mejor será la aproximación de la función de imagen continua $f(x, y)$.

4.2 Procedimiento de bajo nivel

En el procedimiento de bajo nivel se trabaja directamente con las imágenes para extraer propiedades como discontinuidades u orillas, gradiente y profundidad, textura, color, entre otros, de tal forma que sean de mayor utilidad para los siguientes niveles de visión. De esta forma, se obtiene una nueva imagen por cada característica que se extraiga de la imagen original. Para realizar esta tarea, es necesario pasar por distintas etapas de procesamiento y filtrado donde se analiza la imagen y se adecua para cierta aplicación específica.

A veces es necesario mejorar la imagen para resaltar aspectos que sean de nuestro interés, agudizar y/o contrastar determinados aspectos de la imagen, así como también, eliminar aspectos no deseados (también conocidos como ruido). Las mejoras se realizan mediante las técnicas de mejora de la imagen y se debe realizar previo a la obtención de las características de la imagen antes mencionadas.

A continuación, se explica brevemente algunos de los procedimientos de bajo nivel:

Transformaciones de suavizado: es utilizada para suavizar los bordes de una imagen, reducir los picos de ruido o simplemente hacer menos bruscos los cambios de intensidad en la imagen.

Transformaciones de perfilado: realizan la operación contraria al suavizado, es decir, destaca y hace más visibles las variaciones y bordes de la imagen. De esta forma, se mejora la apariencia difuminada de las imágenes.

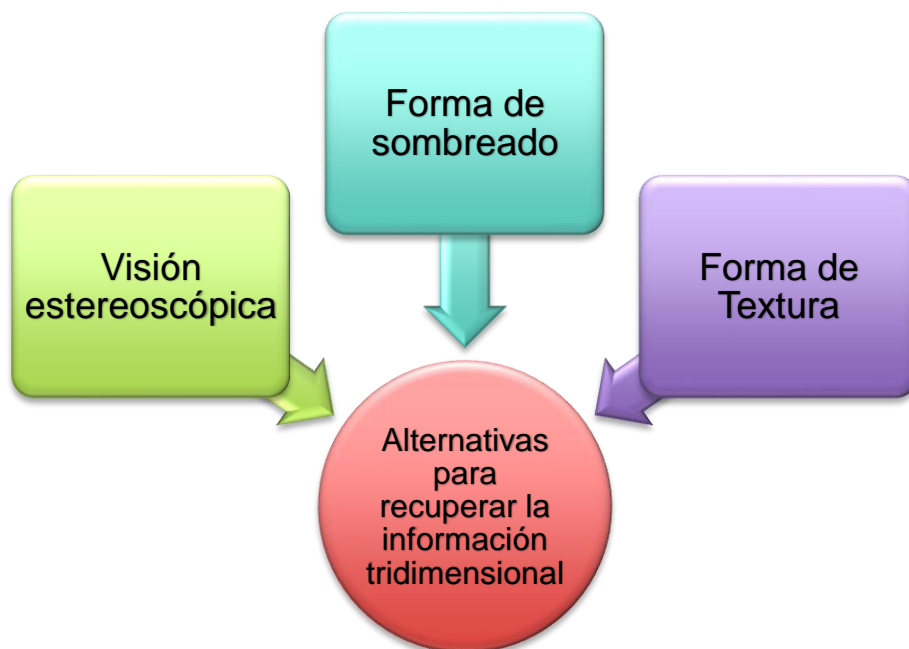


Detección de bordes: detecta las zonas de cambios bruscos de intensidad y las realza.

4.3 Recuperación de información tridimensional

Cuando se proyecta un objeto del mundo tridimensional (3-D) a una imagen bidimensional (2-D) perdemos información en dicha transformación ya que una línea en 3-D se convierte en un punto en la imagen.

Con la finalidad de recuperar la tercera dimensión que se perdió en el proceso de proyección algunas alternativas se han desarrollado como lo son usar dos imágenes mediante visión estereoscópica, usar propiedades de la imagen como sombreado o textura para obtener un estimado de la profundidad, o al menos de la profundidad relativa (gradiente). También se puede utilizar información de una secuencia de imágenes (forma de movimiento).





Visión estereoscópica: se utilizan dos o más cámaras para determinar la geometría de la escena. Estas cámaras deben estar colocadas en posiciones distintas y a una distancia conocida para obtener dos imágenes de cada punto de donde se pueda recuperar su posición en 3-D. Luego de identificar los puntos correspondientes, se utiliza la triangulación para determinar las dos líneas en 3D en las que está el punto. Después, se intersecan las líneas para obtener el punto en 3D.

Forma de sombreado: se trata de obtener información de profundidad (3-D) a partir de los cambios de intensidad o sombreado del objeto. Este es un problema complejo ya que existen múltiples soluciones. Se han propuesto diversos algoritmos para lograr una solución única. En general, a estos algoritmos se les puede ubicar en alguna de las tres categorías siguientes:

- Uso de múltiples fuentes de iluminación (estéreo fotométrico)
- Uso de restricciones entre elementos (relajación)
- Uso de información local (algoritmo diferencial)

Forma de Textura: nos permite determinar el gradiente o profundidad relativa de una superficie. Para obtener el gradiente a partir de textura se puede usar alguna de las siguientes técnicas:

- Razón de cambio máximo de las primitivas de textura
- Forma del elemento de textura
- Puntos de desvanecimiento



4.4 Algoritmo de Waltz

El algoritmo de filtrado de Waltz analiza dibujos lineales asignando una etiqueta a cada uno de los bordes. La etiqueta de línea se selecciona de un conjunto finito de posibles etiquetas. Cada etiqueta proporciona información semántica específica sobre la naturaleza del borde y las regiones en cada uno de sus lados.

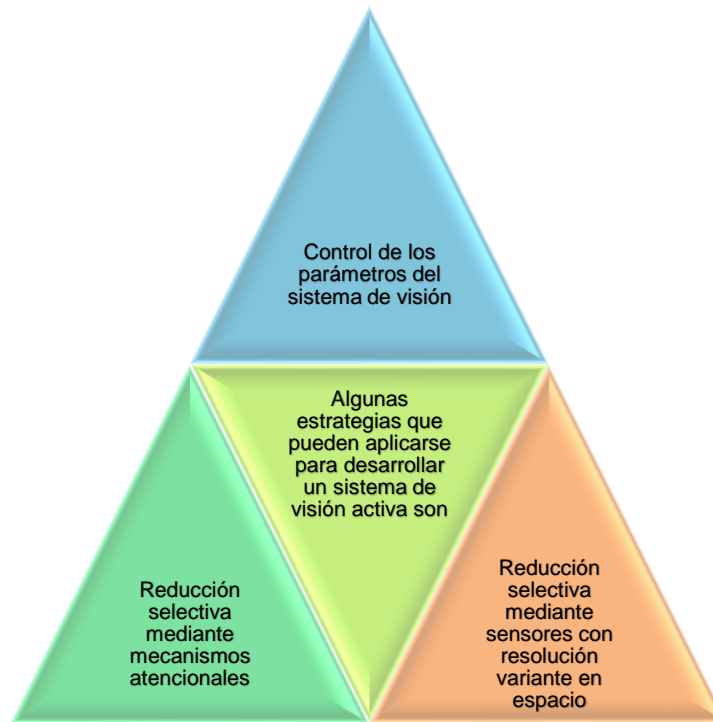
Waltz propuso inicialmente el algoritmo de filtrado como una forma de reducir la combinatoria asociada con el etiquetado lineal de escenas tridimensionales. El algoritmo popularizó la técnica de propagación de restricciones y se conoce como Algoritmo de filtrado de Waltz. Este algoritmo reduce sustancialmente la búsqueda al podar el espacio de búsqueda desde el principio utilizando restricciones específicas del dominio.

Para el problema del etiquetado de líneas, el espacio de búsqueda contiene todas las combinaciones posibles de etiquetas de línea potenciales para una escena. Para las escenas complicadas, este espacio de búsqueda puede ser bastante enorme. Waltz descubrió que el filtrado puede eliminar un espacio de búsqueda y, a menudo, eliminar la necesidad de búsqueda en total.

El algoritmo de filtrado de Waltz se implementó originalmente en Micro-Planner y Lisp.

4.5 Visión activa

Consiste en la elaboración de estrategias de control para ajustar los parámetros de entrada con la finalidad de mejorar el conocimiento que se tiene del entorno. Algunas **estrategias** que pueden aplicarse para desarrollar un sistema de visión activa son:



Control de los parámetros del sistema de visión: utilizado para adaptar las imágenes a la tarea que se va a realizar. El objetivo es adquirir múltiples vistas de la escena y extraer de cada una pocas características con algoritmos simples mediante el control de ciertos parámetros como pueden ser: la resolución y las condiciones de iluminación, el ángulo de vergencia, el zoom, el foco y apertura de la lente, el movimiento y posición de la cámara.

Reducción selectiva mediante mecanismos atencionales: se enfoca en obtener la información relevante para cada una de las tareas visuales que se deben realizar. Determinar qué parte de la información es la de utilidad y cuál se puede obviar para cada tarea constituye el mayor desafío para utilizar esta estrategia.

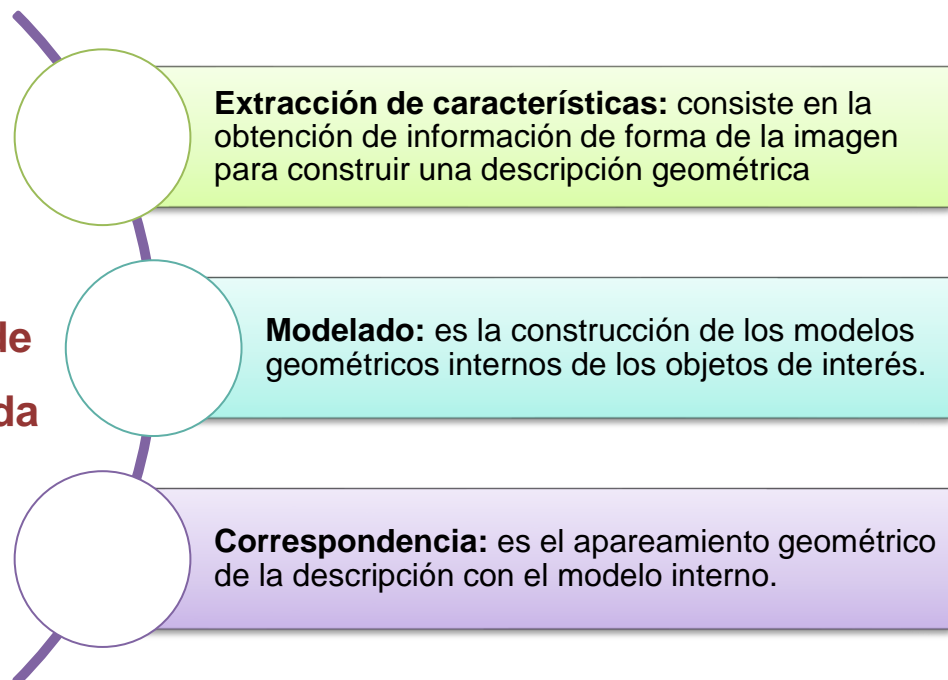


Reducción selectiva mediante sensores con resolución variante en espacio: efectúan una reducción selectiva en espacio, resolución y tiempo. Esto les permite discriminar entre los datos contenidos en las imágenes de resolución uniforme para obtener la información relevante para efectuar una actividad.

4.6 Reconocimiento de objetos y escenas, y descripción de imágenes

Para reconocer los objetos cuya descripción se ha obtenido de la imagen se utiliza la visión basada en modelos. Este reconocimiento se logra mediante la utilización de una serie de modelos geométricos predefinidos.

Componentes principales de la visión basada en modelos





Tipos principales de sistemas basados en modelos:

Modelos en dos dimensiones (2-D): utilizan modelos geométricos en 2-D para modelar y reconocer objetos en dos dimensiones. Se puede representar objetos en base a contornos y regiones. En los contornos el objeto se representa en base a su borde o contorno mientras que en las regiones el objeto se representa en base a la región que el objeto define.

Modelos en tres dimensiones (3-D): utilizan una representación tridimensional de los objetos. Se asume que los objetos son sólidos y rígidos. Se puede representar objetos mediante modelos basados en la superficie del objeto o en base en una representación volumétrica del sólido. Algunos modelos utilizados son: poliedros planos, cilindros generalizados, geometría sólida constructiva.

Para llevar a cabo el reconocimiento es necesario realizar una interpretación de los datos de entrada mediante el establecimiento de una correspondencia (match) entre los atributos obtenidos de la(s) imagen(es) a través de los procesos antes explicados y los modelos representados en la computadora. Las técnicas para correspondencia dependen del tipo de representación y las podemos dividir en:

Reconocimiento estadístico de patrones: busca dentro de un espacio paramétrico, la clase (modelo) más cercana a la descripción del objeto en la imagen. Cuando un objeto es desconocido se debe encontrar a que grupo (clase) pertenece partiendo de los parámetros dados.

Optimización paramétrica: buscan la correspondencia entre modelos paramétricos y representaciones de bajo nivel (por ejemplo, encontrar la correspondencia entre una serie de orillas y una curva).



Algoritmos basados en teoría de gráficas: cuando tenemos una representación relacional, tanto de los modelos internos como de la descripción de la imagen podemos aplicar los algoritmos basados en teoría de grafos. Se considera que tanto de los modelos internos como de la descripción de la imagen están representadas en forma de un grafo (nodos y relaciones). Encontrar la correspondencia entre dichos grafos, desde el punto de vista de la teoría de grafos, se refiere al problema de isomorfismo entre grafos. Este consiste en encontrar una relación 1:1 entre arcos y nodos de ambos grafos, considerando que no están etiquetados.