

Abstract. . . . En este paper se aborda sobre la recuperación de imágenes, una de las ramas de la recuperación de información, sus principales métodos como el basado en keywords (keyword based image retrieval) o el basado en contenido (content based image retrieval), entre otros. Las ventajas y desventajas de estos métodos y que aplicaciones reales tiene la recuperación de imágenes .

Keywords: content-based image retrieval, keyword-based image retrieval, image retrieval,

Table of Contents

Características del problema.....	3
<i>David Orlando De Quesada Oliva, Javier Domínguez</i>	
Diferencias entre Image Retrieval y Content-based image retrieval	4
<i>David Orlando De Quesada Oliva, Javier Domínguez</i>	
Principales métodos para la recuperación de imágenes	5
<i>David Orlando De Quesada Oliva, Javier Domínguez</i>	
Evaluación del sistema.....	12
<i>David Orlando De Quesada Oliva, Javier Domínguez</i>	
Aplicaciones	13
<i>David Orlando De Quesada Oliva, Javier Domínguez</i>	
Ventajas y desventajas.....	15
<i>David Orlando De Quesada Oliva, Javier Domínguez</i>	

Características del problema

David Orlando De Quesada Oliva, Javier Domínguez

MATCOM, Universidad de La Habana,
d.quesada2@estudiantes.matcom.uh.cu,
j.dominguez@estudiantes.matcom.uh.cu,

La Recuperación de imágenes es el campo que se encarga de buscar y obtener imágenes digitales de una base de datos. Debido a la cantidad creciente de imágenes digitales alrededor del mundo, desde 1970 este campo ha estado bien activo. Un sistema de recuperación efectivo y rápido de imágenes necesita operar sobre una colección de imágenes y devolver las imágenes relevantes basadas en la consulta, la cual se realiza lo más cercana posible a la percepción humana. Los investigadores de este campo poco a poco, han ido mejorando e implementando varios tipos de sistemas de recuperación de imágenes, de los sistemas basados en **palabras claves**, pasando por los sistemas basados en el contenido (características) de una imagen, y finalmente llegando a los sistemas de recuperación semánticos, con el objetivo de reducir el vacío semántico que existe entre la representación de características de bajo nivel (color, textura, forma, etc) y la semántica de alto nivel en las imágenes.

Desarrollar un motor de búsqueda de imágenes omnipotente, capaz de satisfacer a todos los usuarios requiere entender y caracterizar la interacción y la búsqueda de imágenes desde el punto de vista del usuario y del sistema. Desde el punto de vista del usuario, claridad en lo que desea, donde lo quiere buscar y de que forma quiere realizar su consulta. Desde el punto de vista del motor de búsqueda, como desea el usuario que le sean presentados los resultados de su consulta, dónde desea buscar el usuario y cual es la naturaleza de la consulta del usuario.

Diferencias entre Image Retrieval y Content-based image retrieval

David Orlando De Quesada Oliva, Javier Domínguez

MATCOM, Universidad de La Habana,
`d.quesada2@estudiantes.matcom.uh.cu`,
`j.dominguez@estudiantes.matcom.uh.cu`,

La recuperación de imágenes (Image Retrieval) se refiere a la rama de la recuperación de información que se encarga de dada una consulta que puede realizar un usuario de disímiles formas, devolver un ranking de las imágenes que más se asemejen a la consulta hecha por el usuario, teniendo en cuenta alguna medida de similitud. La recuperación de imágenes basada en contenido(Content-based image retrieval) se refiere a un tipo específico de método de recuperación de imágenes que se caracteriza por usar el contenido de las imágenes (color, forma, textura, etc) para realizar el proceso de consulta.

Principales métodos para la recuperación de imágenes

David Orlando De Quesada Oliva, Javier Domínguez

MATCOM, Universidad de La Habana,
d.quesada2@estudiantes.matcom.uh.cu,
j.dominguez@estudiantes.matcom.uh.cu,

Keyword Based Image retrieval:

En un **image retrieval system convencional**, los **keywords** son usados como descriptores para indexar y recuperar una imagen. Las palabras claves (key words) no transmiten mejor que el contenido de una imagen el significado de esta. Antes de que las imágenes sean almacenadas en la base de datos, son examinadas manualmente y se les asigna una palabra clave (**keyword**) para describir su contenido. Estos keywords son almacenados como parte de los atributos asociados a la imagen. En el proceso de hacer una consulta, el sistema aceptará del usuario una o varias **keywords** que serán el criterio de búsqueda. Luego se realiza un proceso para encontrar las imágenes que cumplen con el criterio de búsqueda. Las técnicas de text based image retrieval usan texto para describir el contenido de una imagen lo que a menudo crea ambigüedad e insuficiencia en el procesamiento de la query y el rendimiento en una búsqueda de la base de datos de imágenes. El proceso de asignación de meta datos con captions o keywords a una imagen digital es conocido como anotación automática de imágenes (automatic image annotation). Este tipo de text based information retrieval está motivado léxicamente en lugar de conceptualmente, lo que lleva a resultados de búsqueda irrelevantes en la recuperación de información.

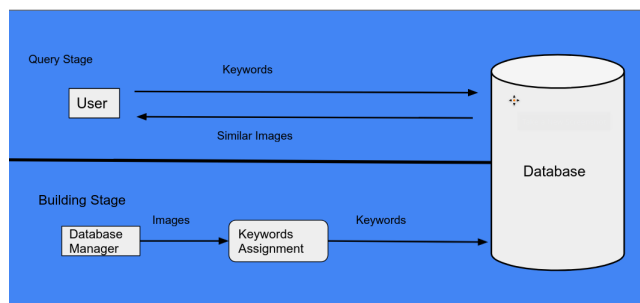


Fig. 1. Proceso general de un keyword based image retrieval

0.1 Text Based Image Retrieval:

Las técnicas básicas de recuperación de documentos pueden ser usadas para la recuperación de imágenes basadas en metadatos sin modificación. En un keyword based image retrieval, los metadatos que describen las imágenes pueden ser categorizados en 2 partes. Una parte se refiere a las herramientas usadas en el proceso de creación de la imagen, estilo de arte de la imagen, artista, precio, y otras propiedades explícitas de la imagen. La otra parte describe lo que realmente hay en la imagen, las propiedades implícitas que pueden entenderse al percibir la imagen en si. En el contexto actual de la recuperación, el texto plano anotado en imágenes responde de manera similar al texto plano en documentos, debido a que ambos contienen texto, lo cual permite que sean explotados por las técnicas convencionales de text-bases information retrieval. La recuperación de información basada en texto genérica se realiza de tal manera que inicialmente el usuario realiza una consulta(query) que tiene de 1 a m **keywords**. En los sistemas de recuperación basados en metadatos(metadata based information retrieval), el buscador compara los keywords con un conjunto de imágenes recopiladas de una base de datos y les da prioridad a los valores. Por ejemplo, si el keyword es **book**, y la imagen A contiene 2 ocurrencias de book y la imagen B solo una ocurrencia, entonces A tiene una prioridad mayor. Las imágenes con palabras claves anotadas son mostradas al usuarios en el orden de reducción de la prioridad. Imágenes irrelevantes son recuperadas y el usuario tiene que gastar tiempo en el filtrado de la información, usualmente navegando a través de los resultados de búsqueda.

0.2 Field Based Image Retrieval:

Field based retrieval es una extensión del text based retrieval donde solo un campo(field) es usado en anotación y recuperación. El enfoque basado en el campo (field based) describe y recupera artículos usando uno o más pares de valores del campo. Regularmente un esquema de metadatos es descrito por un conjunto de campos y pocas indicaciones sobre el tipo de valores que puede ser elegidos por un campo particular. La plantilla (template) de metadatos y esquemas ampliamente utilizada para describir documentos online en general es la **Dublin Core(DC)**. Los campos de la DC version 1.1 son rights, coverage, relation, language, source, identifier, format, type, data, contributor, publisher, description, subject, creator y title. Versiones calificadas de DC han sido creadas para dominios particulares como la descripción de piezas de arte en museos.

0.3 Structure Based Image Retrieval:

El paradigma de recuperación basado en estructuras. En este método, se utiliza un enfoque basado en el campo(field) que principalmente utiliza una estructura de pares de valores atributo. Este método permite descripciones más complejas implicando relaciones. Por ejemplo, una definición de una parte de un auto puede incluir especificaciones de esos componentes. Cada elemento del objeto se

puede especificar de nuevo usando varios atributos como la forma, el tamaño y el material. Los elementos pueden incluso tener elementos ellos mismos, por ejemplo, una mesa tiene patas, y sus subelementos pueden moverse hasta el nivel donde un elemento no puede obtener un subelemento más particular.

1 Content Based Image Retrieval(CBIR):

El Content Based Image Retrieval(CBIR) es uno de los métodos de visión por computadoras para la recuperación de imágenes, lo que significa que para poder recuperar es necesario imágenes digitales de una base de datos de imágenes. La búsqueda basada en contenido(Content based search) realizará el análisis con el contenido real de la imagen, en lugar de metadatos como etiquetas(tags), palabras clave(keywords), o descripciones anotadas con la imagen. La palabra contenido aquí puede referirse a formas, color, texturas o algún otro detalle que se puede obtener dentro de la propia imagen. El motor de búsqueda de imágenes relacionadas con la web se basa en metadatos, por lo que genera una gran cantidad de resultados basura. Por lo tanto CBIR es deseable en este caso. Dándole palabras clave (keywords) de forma manual a las imágenes de búsqueda en una larga base de datos puede obtener resultados incorrectos. Además el proceso es costoso y puede que no identifique todas las palabras clave(keywords) que especifican la imagen y, por tanto, es ineficiente. Al proporcionar una buena técnica de indexación basada en el contenido real de las imágenes, se puede recuperar y producir resultados precisos.

1.1 Low-Level Image Feature:

Para poder realizar el CBIR las características de bajo nivel de la imagen (low-level image feature) deben ser extraídas primero. La extracción de características puede hacerse en toda la imagen o solo en una región de interés. La técnica simple usada en la recuperación de imágenes depende de las características globales. La percepción humana coincide estrechamente con la representación de imágenes a nivel de región. Para realizar la recuperación de imágenes basada en regiones el paso más importante es la segmentación de imágenes. De la región segmentada, las características de bajo nivel como textura, el color, la forma o la ubicación espacial se pueden extraer fácilmente. Basado en las características de la región, se puede encontrar fácilmente la coincidencia entre dos imágenes

1.2 Image Segmentation

El proceso automático de la realización de la segmentación de una imagen es una tarea difícil. Las técnicas académicas usadas en la segmentación de imágenes son curva de difusión de energía(curve energy diffusion), evolución(evolution) y particionamiento de grafos(graph partitioning). La mayoría de los métodos pueden ser apropiados solo para imágenes que tienen regiones con colores similares, como los métodos de direct clustering en el espacio de color. Tales métodos pueden

adaptarse para la recuperación de sistemas que funcionen con colores. Pero las escenas naturales contienen tanto colores como texturas. Aplicar segmentación en texturas resulta difícil. Incluso en la segmentación basada en texturas la estimación del parámetro del modelo de textura es difícil. Para superar esto el algoritmo 'JSEG' es usado. Otro algoritmo llamado segmentación Blobworld es ampliamente utilizado. Algunos algoritmos de segmentación hacen uso de segmentación basada en color, en textura o en ambas. Estos algoritmos usan k-means para propósitos de clasificación. Los bloques de una misma clase se agrupan dentro de una misma región. El algoritmo k-means con restricción de conectividad (KMCC) es un trabajo de segmentación para segmentar objetos en las imágenes. Esta utilización del algoritmo se basa en la confianza en la necesidad del sistema y el uso del conjunto de datos. Es difícil determinar que algoritmo proporciona mejores resultados. El resultado del JSEG es la textura y el color de regiones similares, pero el resultado de KMCC produce objetos que son diferentes. El algoritmo KMCC es computacionalmente mucho más exhaustivo que el JSEG. Por tanto, Blobworld y JSEG son principalmente los algoritmos usados.

1.3 Varias características de bajo nivel de las imágenes:

En las diversas categorías de algoritmos muy pocos se pueden aplicar en la recuperación de imágenes en tiempo real con semántica de alto nivel que son:

■ Color:

El color es la más común de las características adoptadas en la recuperación de imágenes. Varios espacios de color son usados para definir colores. Esos espacios de color son usados dependiendo de las diferentes aplicaciones. Los espacios de color más usados son RGB, LAB, LUV, HSV (HSL), YCrCb, y el hue-min-max-difference (HMMD). La covarianza del color, el histograma de color, y los momentos de color (color moments) son principalmente las características de color usadas en RBIR (Region Based Image Retrieval). El color principal (leading color), el color escalable (scalable color) y el diseño de color (color layout) son las características de color que se utilizan principalmente en MPEG-7. Con el origen las características de 3 colores, el par matiz-matiz y matiz se construyen las invariantes de color. La semántica de alto nivel no está directamente relacionada con las características de color mencionadas anteriormente. Para mapear los colores de una región a nombres de colores en semántica de alto nivel, el promedio de color de todos los píxeles en una región podría usarse como su característica de color. Si la segmentación es errónea terminará porque la región original es visualmente diferente al color promedio. Dependiendo de los resultados de la segmentación solo se seleccionan las características de color. El color promedio no es una opción deseable si la segmentación da como resultado objetos que no tienen colores similares. En la mayoría de los trabajos CBIR, las imágenes en color no están preprocesadas. Los filtros de color adecuados son esenciales para mejorar la eficiencia de recuperación debido a que el color en las imágenes siempre está dañado por el ruido.

■ **Textura:** Pocos sistemas no utilizan la textura como el color en la recuperación de imágenes. La textura es una característica importante para describir la semántica de alto nivel en la recuperación de imágenes porque proporciona detalles esenciales en un catálogo de imágenes, ya que define el contexto de muchas imágenes del mundo real como nubes, ladrillos, árboles y telas. El resultado de aplicar **la transformada de Wavelet** o **el filtrado de Gabor**, medidas estadísticas confinadas así como las seis características de textura de Samura, son las características de textura más utilizadas en el proceso de recuperación de imágenes. Las características de textura de Samura son:

- ▶ La regularidad
- ▶ Semejanza de línea
- ▶ La rugosidad
- ▶ La direccionalidad
- ▶ El contraste
- ▶ La aspereza

De estas características la aspereza, la direccionalidad y la regularidad son las más importantes. Estas tres están relacionadas con otras que son menos eficaces con respecto a la descripción de la textura. Los descriptores de navegación de texturas son obtenidos desde **MPEG-7**. Estos son regularidad, aspereza, direccionalidad. Se ha encontrado que la textura de Brodatz funcionará de manera excelente con características de palabras como aleatoriedad, direccionalidad y periodicidad. Las características de Tamura no funcionan para múltiples resoluciones que se consideren para la medición. Las características globales se ven afectadas por distorsiones de la imagen como diferencias de orientación debido a la distorsión del punto de vista y la escala. Si las regiones de textura en la imagen no están organizadas y son similares, se produciría una respuesta de recuperación deficiente para imágenes de escenas naturales. El estudio de la visión humana puede coincidir bastante con las características de Wavelet y Gabor en la mayor parte de la recuperación de imágenes. Pero el diseño actual del filtro de Gabor y la transformada de Wavelet solo está destinado a imágenes rectangulares. Pero en un RBIR la región de la imagen tiene formas erráticas. Por tanto en tal caso las características de la textura se utilizan eficazmente. Pero para la representación de imágenes naturales el descriptor de histogramas de borde (EHD) es el más adecuado y eficaz.

■ **Forma:** Uno de los conceptos más distintivos es la forma. Esta característica tiene un límite consecutivo de segmentos, una relación de aspecto, descriptores de Fourier, circularidad e invariantes de momento. El color y la textura son más útiles en imágenes particulares de dominio, como objetos artificiales. Aún así las características de la forma son características esenciales pero no tienen tanta popularidad los RBIR como las características de textura y color para explorar los beneficios inherentes de RBIR, algunos sistemas podrían utilizar las características de forma como evaluadores. Por ejemplo, las características de orientación y excentricidad se utilizan para este propósito. La forma de una imagen puede definir la configuración de superficie características de un

objeto, un contorno. Esto permite distinguir a un objeto de su entorno por su contorno. La representación de la forma puede generalmente dividirse en dos categorías: basado en los límites y basados en regiones. La representación de la forma basada en los bordes solo usa los bordes exteriores de la forma. Esto se hace describiendo la región considerada usando sus características externas, como los píxeles a lo largo del límite del objeto. Pero la representación de la forma basada en la región es totalmente diferente al método anterior. Utiliza la región de forma completa describiendo la región considerada sus características internas; es decir los píxeles contenidos en esa región.

■ **Localización espacial:** No solo la textura, el color, o la forma son características importantes sino también la localización espacial en la catalogación de la región. Por ejemplo una imagen que contenga árboles con césped en el suelo podría tener características de color y textura similares pero su localización espacial ser diferente, normalmente las hojas de los árboles aparecen en la parte superior de una imagen, mientras que las hojas que caen en la parte inferior. Las ubicaciones espaciales se definen intuitivamente como **:izquierda(left)**, **derecha(right)**, **top(arriba)**, **bottom(abajo)**, según el lugar de la región en la imagen. El rectángulo delimitador mínimo (minimum bounding rectangle) y el centroide de la región se utilizan para encontrar la ubicación espacial.

Semantic Based Image Retrieval:

Ni una, ni la combinación de varias características visuales o de bajo nivel (color, textura, forma, relación espacial) pueden capturar completamente los conceptos de alto nivel de las imágenes. Además, debido a que el rendimiento de la recuperación de imágenes basado en características de bajo nivel no es satisfactorio, hay una necesidad de que la investigación vaya dirigida hacia la recuperación de imágenes basada en el significado semántico, tratando de utilizar el concepto cognitivo del ser humano para traducir esas características de bajo nivel a conceptos semánticos de alto nivel (vacío semántico). Este acercamiento permite a los usuarios acceder a imágenes a través de consultas por texto, la cual es más intuitiva, fácil y preferida por los usuarios para expresar su deseo. Como mencionamos previamente en el CBIR, las imágenes pasan por un proceso de extracción de características de bajo nivel que se almacena junto con las imágenes, ahora aquí, se hace necesario traducir esas características de bajo nivel a conceptos de alto nivel que no es capaz de comprender una máquina. Esta traducción usualmente se lleva a cabo utilizando herramientas de aprendizaje supervisado o no-supervisado, para asociar las características de bajo nivel con conceptos de alto nivel, los cuáles serán apuntados con palabras, durante el proceso de anotación de imágenes. Como podemos ver la recuperación de imágenes basada en la semántica utiliza técnicas de los 2 mecanismos que surgieron previamente, la extracción de características de bajo nivel utilizada en CBIR y la anotación de imágenes utilizada en el mecanismo basado en palabras claves, para almacenar en la imagen palabras **keywords** que se obtengan del proceso de conversión

de características de bajo nivel a conceptos de alto nivel propio de este nuevo mecanismo.

Evaluación del sistema

David Orlando De Quesada Oliva, Javier Domínguez

MATCOM, Universidad de La Habana,
d.quesada2@estudiantes.matcom.uh.cu,
j.dominguez@estudiantes.matcom.uh.cu,

La recuperación de imágenes es esencialmente un problema de recuperación de información. Por tanto, las métricas de evaluación han sido adoptados de forma bastante natural a partir de la investigación de recuperación de información. Dos de las Las medidas de evaluación más populares son:

- Precisión: el porcentaje de imágenes recuperadas que son relevantes para la consulta.
- Recobrado: el porcentaje de todas las imágenes relevantes en la base de datos de búsqueda que se recuperan.

Es importante tener en cuenta que cuando la consulta en cuestión es una imagen, la relevancia es extremadamente subjetiva. La investigación sobre recuperación de información ha demostrado que la precisión y el recobrado siguen una relación inversa. La precisión cae mientras que la recobrado aumenta a medida que el número de imágenes recuperadas, a menudo denominadas alcance, aumenta. Por lo tanto, es típico que tengan un valor numérico alto tanto para la precisión como para la recuperación. Tradicionalmente, los resultados se resumen como curvas de recuperación de precisión o curvas de alcance de precisión. Una crítica por la precisión se deriva del hecho de que se calcula para todo el conjunto recuperado y esta no se ve afectada por las clasificaciones respectivas de las entidades relevantes en la lista recuperada. Una medida que aborda el problema anterior y es muy popular en la comunidad de recuperación de imágenes, es la precisión media (AP). En una lista clasificada de entidades recuperadas con respecto a una consulta, si la precisión se calcula en la profundidad de cada entidad relevante obtenida, la precisión promedio se da como la media de toda la precisión individual. Como es obvio, esta métrica está muy influenciada por elementos relevantes de alto rango y no tanto por los que se encuentran al final de la lista recuperada.

Aplicaciones

David Orlando De Quesada Oliva, Javier Domínguez

MATCOM, Universidad de La Habana,
d.quesada2@estudiantes.matcom.uh.cu,
j.dominguez@estudiantes.matcom.uh.cu,

El campo de la recuperación de imágenes ha demostrado ser uno necesario para los tiempos en que vivimos, donde tantas imágenes se generan diariamente, desde satélites orbitando la Tierra, misiones espaciales en lo más lejos del cosmos, animales en la naturaleza, paisajes, cámaras de seguridad y mucho más. Las imágenes son una fuente de información muy útil en la actualidad, y por tanto almacenarlas y organizarlas es una tarea necesaria, para facilitar el acceso a la información. De ahí que este campo tenga una gran cantidad de aplicaciones en la actualidad, a continuación mencionamos algunas de las más importantes:

■ El problema de la anotación automática de imágenes.

El propósito principal de un sistema de recuperación de imágenes basado en contenido es descubrir imágenes que pertenecen a algún concepto, en la ausencia de meta-datos, todos los intentos de automatizar el proceso de creación de estos meta-datos tiene ese objetivo. La anotación de imágenes puede facilitar la búsqueda de imágenes utilizando texto. Si el mapeo resultante imagen-palabras clave es confiable, la búsqueda de imagen basada en texto puede tener semánticamente más sentido que buscar en la ausencia de texto. Uno de los métodos para resolver este problema de la anotación automática es usar aprendizaje supervisado para categorizar las imágenes. La detección de conceptos simples como: paisaje, ciudad, animales, etc, alcanza una alta precisión.

■ El arte y la cultura siempre han sido importantes en la vida del ser humano. A lo largo de la historia, los museos y galerías de arte del mundo se han encargado de preservar nuestra diversa herencia cultural para utilizarlos como fuentes educación y aprendizaje. Es por esto que recientemente se ha expresado la preocupación por digitalizar todos los materiales antiguos, históricos y culturales, para la posteridad. Esto es muy importante por dos razones, primero las computadoras se han convertido en el principal medio de aprendizaje y se supone que así sea durante los próximos años, por tanto la representación digital de los artefactos culturales y las imágenes es algo que facilitará su popularidad, además de que sería accesible desde cualquier rincón del mundo, y segundo al contrario de la información almacenada de forma digital, los artefactos y pinturas antiguas están a merced a la degradación con el paso del tiempo, a los desastres y al vandalismo.

■ Las interacciones entre CBIR y la seguridad de la información ha sido prácticamente nulo, hasta que recientemente ciertas perspectivas han emergido para unir ambos campos, las pruebas de interacción humana (HIPs por sus siglas en inglés) y el

cumplimiento de la protección de los derechos de autor. Mientras por un lado constantemente estamos ampliando las fronteras de la ciencia para diseñar sistemas que pueda imitar las capacidades humanas, no podemos negar los riesgos de seguridad inherentes asociados con programas extremadamente "inteligentes". Uno de dichos riesgos es cuando un sitio web o algún servidor es atacado por programas maliciosos que solicitan servicios a escalas masivas. Pueden ser escritos programas que consuman una gran cantidad de recursos web o que influyan en los resultados de votaciones. En este caso los HIPs también conocidos como CAPTCHAs, son la solución. Estas interfaces están diseñadas para diferenciar entre humanos o programas, basados en la respuesta a algunas preguntas.

Ventajas y desventajas

David Orlando De Quesada Oliva, Javier Domínguez

MATCOM, Universidad de La Habana,
d.quesada2@estudiantes.matcom.uh.cu,
j.dominguez@estudiantes.matcom.uh.cu,

Desventajas:

Keyword Based Image Retrieval: En el keyword based image retrieval el contenido de una imagen es mucho más rico de lo que cualquier conjunto de palabras pueda expresar, solo usando texto para describir el contenido de una imagen a menudo causa ambigüedad e insuficiencia en el rendimiento de la búsqueda de imágenes en una base de datos y el procesamiento de las query de los usuarios. Este problema se debe a la dificultad de especificar términos y frases exactas al describir el contenido de una imagen.

Los sistemas basados en Content Based Image Retrieval(CBIR) sufren de lo que se denomina brecha semántica. Es una brecha entre la captación de imágenes de alto nivel de una mente humana y la estimación de imágenes de bajo nivel de las computadoras.