Capítulo 2

Trabajo relacionado

Los multimedios pueden estudiarse desde varias perspectivas: por un lado existe el problema de la manipulación física de los documentos, donde se estudian técnicas de compresión, transmisión y almacenamiento en bases de datos; por otro lado, el del análisis semántico de los contenidos para la creación de índices y anotaciones y la realización de búsquedas a través de sus contenidos. Los proyectos discutidos a continuación forman parte de esta última tendencia y son fuente de inspiración de este trabajo de tesis.

2.1 Informedia [Wactlar 2000]

Sin duda, este es uno de los proyectos de recuperación de información multimedios más complejo realizado hasta el momento. Utilizando técnicas de procesamiento de voz (con un corpus de noticias de 60,000 palabras), sonidos, texto, imágenes y lenguaje natural, se lograron segmentar programas de noticias en pequeñas historias de un solo contexto. La biblioteca digital cuenta con 1500 horas de programación descompuesta automáticamente en 40,000 segmentos independientes. Este proyecto realizó investigación en las siguientes áreas:

- Integración de procesamiento de voz, lenguaje e imágenes para generar abstracciones multimedios, segmentado de video en historias y elaboración de presentaciones basadas en el contexto.
- Procesamiento de texto: generación de encabezados, clasificación de texto en tópicos y recuperación de información.

?? Procesamiento de audio: reconocimiento de voz, segmentado de diálogos hablados y detección de silencios.

?? Procesamiento de imágenes: reconocimiento de caras y comparación de imágenes de acuerdo a regiones, colores y texturas.

?? Procesamiento de video: selección de tramas clave y texto OCR.

Actualmente se busca extender la funcionalidad descrita a bibliotecas en diferentes idiomas y realizar resúmenes automáticos de los segmentos de noticias.

Informedia realiza análisis espacial y temporal en el procesamiento de sus consultas, ofreciendo resultados con dependencia del tiempo y analizando búsquedas que tienen relación con lugares geográficos.

Fig. 2.1 Interfaz principal del proyecto Informedia mostrando el resultado de una búsqueda sobre "El niño", con 12 documentos encontrados junto con algunas abstracciones de los mismos creadas automáticamente. Imágenes copyright ? CNN Figura copyright ? Carnegie Mellon University. Figura incluida con permiso de Informedia Digital Video Library Project, 29/Abril/2002.

Las aportaciones de este proyecto son:

- Remarcar la importancia de investigaciones multidisciplinarias en el análisis de los contenidos multimedios.
- Ofrecer amplia disponibilidad a eventos históricos registrados en video.
- Analizar nuevas capacidades para entender cómo evolucionan los eventos en el tiempo y como se relacionan geográficamente.
- Incrementar el uso del video en dominios previamente dominados por el texto y la voz.

2.2 Consulta de imágenes por contenido (QBIC) [Maybury 1997]

Este es un sistema desarrollado para recuperar imágenes basado en sus características visuales (color, textura, forma). Con este método, las facilidades de recuperación especifican parámetros de color y descripción de formas y texturas. Para manejo de video, se desarrollaron funciones de detección de escenas analizando cambios en el histograma y movimientos de cámara así como la identificación de imágenes representativas que se utilizan para la creación de índices.

La organización perceptual -el proceso de agrupar características de las imágenes en objetos significativos y asociar descripciones semánticas a escenas a través de comparaciones- es un proceso no resuelto de entendimiento computacional de imágenes. A las personas les es muy fácil extraer descripciones semánticas de las imágenes, algo muy difícil para las computadoras, sin embargo, las computadoras son mejores que los humanos para medir las propiedades de las imágenes.

Uno de los principios guías del sistema QBIC es dejar a las computadoras hacer lo que les es más fácil (mediciones cuantificables) y dejar que los humanos definan el contenido semántico de las imágenes. Así, QBIC propone un lenguaje visual en el cual las búsquedas son definidas dibujando y seleccionando formas y colores.

Durante el almacenamiento de los objetos (imágenes y videos), éstos son procesados para extraer características que describen su contenido –colores, texturas, formas, objetos en movimiento- y estas características son guardadas en la base de datos. En las búsquedas el usuario compone una consulta gráfica que es procesada para extraer sus propiedades y el resultado es provisto a un proceso que efectúa comparaciones y encuentra imágenes o videos en la base de datos con características similares.

El resultado de esta investigación es un producto comercial incluido en el manejador de bases de datos DB2.

2.3 Correo de video [Jones and Foote 1997]

Para realizar búsquedas en documentos multimedia en donde el mayor contenido de información se encuentra en el canal de audio, estos investigadores combinaron técnicas de reconocimiento de voz (basado en modelos estadísticos) para hacer transcripciones y métodos de recuperación de información (asignación de pesos a los términos más frecuentes) para generar índices precisos que agilicen las búsquedas de temas específicos en un sistema de correo electrónico de videos.

Así, al usuario se le presenta un sistema de correo electrónico que en realidad opera analizando el texto generado a partir de los contenidos de video en la aplicación.

El objetivo fundamental de esta investigación fue como integrar métodos de recuperación de texto con tecnología de reconocimiento automático de voz. Como el procesamiento de voz es muy costoso computacionalmente hablando, la única manera práctica de proveer un rápido acceso a la información fue crear índices de los documentos de video junto con su tiempo de ocurrencia. De esta manera para recuperar información se busca en los índices para encontrar documentos potencialmente relevantes, aunque esto es sólo una parte de la solución: el usuario necesita una interfaz para seleccionar y visualizar el documento referido.

Es muy interesante anotar como los investigadores planearon el sistema de reconocimiento de voz: en la primera etapa del proyecto, trabajaron con quince locutores diferentes que entrenaron al sistema de reconocimiento y el número de palabras reconocidas era definido de antemano (10 categorías de 35 palabras cada una). Además, la captura de voz se efectuaba en un ambiente libre de ruido. Para la segunda etapa, se eliminó la restricción de la identificación del locutor y ciertas restricciones de ruido ambiental de la primera etapa y se atacaron problemas de mala pronunciación y espontaneidad en la señal vocal. Muchas palabras, particularmente nombres propios, fueron difíciles de reconocer ya que no pertenecían al dominio del vocabulario. El rendimiento alcanzado por el sistema de reconocimiento fue de un 90% para dictados y un 40% para conversaciones informales.

El sistema de recuperación de información presenta también algunos problemas: existe un gran número de falsas alarmas (términos que el reconocedor creyó encontrar pero que en realidad se trató de una similaridad fonética) y palabras perdidas (no detectadas por el reconocedor). Estos problemas se vuelven críticos solo cuando los términos en las búsquedas son pocos.

2.4 La biblioteca de música [Witten et al. 1999]

Es un sistema basado en el reconocimiento de tonos. Una entrada acústica permite que los usuarios canten o silben una melodía para hacer búsquedas en un archivo de más de 10,000 canciones tradicionales de Norte América, Irlanda, Inglaterra, Alemania y China. Los resultados pueden ser seleccionados para ser reproducidos o ser sólo consultados visualmente. Los usuarios pueden buscar en toda la base de datos o limitar sus búsquedas a países individuales.

Con los avances en el procesamiento digital de señales y las técnicas de representación musical, se ha vuelto posible transcribir melodías automáticamente con la entrada de un micrófono. Esta capacidad para realizar búsquedas en bases de datos con música es un componente importante de las bibliotecas digitales del futuro. Con ello, los investigadores podrán analizar la música de ciertos compositores para encontrar temas recurrentes o frases musicales duplicadas, y los músicos podrán recuperar composiciones basados en pasajes musicales que se recuerdan solo vagamente.

La primera parte del proceso comienza con la trascripción automática de melodías: la señal musical analógica es muestreada y la frecuencia de cada nota es identificada y etiquetada con su tono y su valor rítmico (con la notación MIDI). Este proceso incluye la detección del inicio y final de la nota encontrada utilizando algoritmos basados en la transformada de Fourier. Los tonos detectados son normalizados a la frecuencia LA-440 para que sean independientes del usuario. Finalmente se realiza un proceso de búsqueda de cadenas para la recuperación de música de la base de datos. Este proceso no busca subcadenas exactas y utiliza algunos algoritmos de aproximación para efectuar las búsquedas.

Actualmente la base de datos cuenta con 9600 canciones tradicionales y utiliza la identificación de 12 notas para efectuar las búsquedas. El proyecto

incluye un estudio de cómo las personas recuerdan música y cómo es la estructura musical de canciones típicas.

2.5 Entendiendo el comportamiento humano en video [Pentland 2000]

Con base a métodos de reconocimiento de siluetas, reconocimiento de expresiones faciales y funciones de reconocimiento de comportamientos (basado en modelos estocásticos similares a los utilizados en reconocimiento de voz), el autor construye un modelo de estimación para reconocer expresiones en rostros humanos y poder determinar probabilísticamente lo que la persona *está haciendo* en una secuencia de video.

Este trabajo ha derivado en aplicaciones de inteligencia perceptual en computadoras, donde las máquinas son conscientes de su entorno y son sensitivas a las personas que interactúan con ellas. La investigación se basa en la idea de comportamiento adaptivo logrado con un aparato perceptual que clasifica las situaciones inmediatas y reacciona con respuestas aprendidas del usuario. De esta manera se está en contraste con las teorías cognitivas que sugieren que el comportamiento adaptivo es resultado de mecanismos complejos de razonamiento [Pentland 2000].

Podemos entonces hacer clasificaciones del "estado del arte" de las técnicas de búsquedas basadas en contenidos [Maybury 1997]:

- Recuperación de imágenes basado en sus características visuales
- Recuperación de audio basado en sus características tonales
- Acceso a video basado en representaciones semánticas
- Procesamiento de voz y lenguaje natural para acceso a video
- Técnicas auxiliares de procesamiento de texto, caracteres y gráficas

A continuación se muestra un resumen de los proyectos presentados.

Tabla 2.1 Proyectos de acervos multimediales que efectúan búsquedas basadas en contenido.

Proyecto	Lugar, Autor	Objetos del acervo	Tecnología empleada para generar meta-datos
Informedia	Carnegie Mellon. Wactar et al.	Video (noticieros y documentales)	Reconocimiento de voz. Procesamiento de texto, imágenes y lenguaje natural. Detección facial y de texto OCR con RNAs.
Consulta de imágenes por contenido	IBM. Flickner et al.	Imágenes, video (de propósito general)	Procesamiento de imágenes con RNAs en base a características visuales. Detección de escenas en video basado en histogramas e información espacial.
Correo de video	Cambridge. Jones et al.	Video (mensajes de personas)	Reconocimiento de voz y métodos de recuperación de información
Biblioteca de música	Waikato. Witten et al.	Audio (música)	Trascripción de música a texto en formato MIDI usando procesamiento digital de señales.
Comportamiento humano en video	MIT. Pentland et al.	Video (personas en actividad)	Reconocimiento de formas y expresiones faciales con RNAs. Modelado estocástico de sus variaciones.

El siguiente capítulo describe la concepción técnica de los componentes básicos de Video U-DL-A.