UNIVERSITA' DEGLI STUDI DI PADOVA

Facoltà di Scienze Statistiche

Corso di Laurea in

Statistica e Gestione delle Imprese

Tesi di Laurea

Teorie e Tecniche del Recupero Multimediale delle Informazioni

Relatore: Ch.mo Prof. Massimo Melucci

Laureando: Fabio Cangianiello

ANNO ACCADEMICO 2004/2005

INDICE

Introduzione

Nel corso degli ultimi anni, lo sviluppo del desktop publishing[1], la progressiva convergenza verso la tecnologia digitale di numerosi dispositivi di produzione di immagini analogici di utilizzo oramai quotidiano (fotocamere, videocamere), la sempre maggiore accessibilità di strumenti di digitalizzazione di immagini analogiche un tempo riservati ai professionisti (scanner) e più in generale il progressivo aumento di potenza e capacità di storage dei moderni computer nonché il loro sempre minore costo, hanno prodotto un'enorme quantità di informazioni digitali.

La tecnologia di Internet e lo spazio del World Wide Web permettono di distribuire e rendere reperibili questi dati ad un numero sempre maggiore di persone, originando quelli che potremmo concepire come veri e propri database distribuiti di documenti multimediali sempre più estesi. Il testo è il tipo di oggetto mediale cui si riesce ad accedere nella maniera migliore.

Per una gestione semplificata di questa mole di informazioni, si è ricorsi a delle nuove e diverse tecniche di ricerca e catalogazione del materiale. Nell'ultimo decennio il campo dei database non solo si è arricchito solo con nuovi metodi e tecnologie per la gestione dei dati testuali, ma anche con nuovi tipi di dati, come quelli visivi e audiovisivi.

Nell'odierna situazione culturale e tecnologica dovrebbero apparire evidenti i diversi limiti del continuare a operare nei termini di un generico Information Retrieval, dove, nella sua pratica tradizionale, ogni tipo di ricerca documentale è effettuata tramite linguaggio testuale. E' oggi necessario, invece, definire un più ampio criterio di MultiMedia Information Retrieval, dove ogni genere di documento elettronico possa essere trattato e ricercato tramite gli elementi di linguaggio più adatti alla sua natura di documento multimediale.

[1] Il Desktop Publishing (DTP) è l'insieme delle procedure di creazione, impaginazione e produzione di materiale stampato dedicato alla pubblicazione (come libri, giornali, riviste o depliant), usando un computer.

Nei database multimediali risultano poco efficaci e troppo riduttivi i metodi di indicizzazione e di ricerca basati sulle annotazioni terminologiche, che si rivelano, al contrario, ben utili nei sistemi di reperimento di informazione testuale.

Negli archivi dove il contenuto dei documenti è sostanzialmente un testo appare ovvio e appropriato che le chiavi che ne consentono l'accesso siano termini e frasi estratti dall'interno di quel contenuto stesso. Negli archivi di immagini o di suoni, invece, si rivela troppo semplificativo e impreciso attribuire, dall'esterno, una descrizione testuale a tali contenuti.

Il metodo del MultiMedia Information Retrieval sperimenta, in sostanza, la possibilità di ricercare le immagini tramite gli appropriati mezzi del linguaggio visivo stesso, i documenti sonori con i mezzi del linguaggio dei suoni, e i video attraverso le forme di rappresentazione audiovisive. Ad esempio dovrebbe apparire dispersivo cercare una fotografia di paesaggio, rappresentante un tramonto con certe tinte, tramite una complicata descrizione "a parole" delle tonalità di colore desiderata, anziché sottoporre a un apposito sistema di ricerca un campione delle tinte stesse.

L'innovazione apportata dal MMIR è fondata sui presupposti del content-based information retrieval (CBIR): il trattamento dei documenti multimediali viene strutturato tramite tecniche di archiviazione e recupero che operano direttamente sul contenuto visivo, sonoro o audiovisivo degli oggetti digitali di un database, in opposizione ai tradizionali sistemi di indicizzazione e ricerca term-based, basati sulla descrizione testuale di tale contenuto.

Questo lavoro vuole illustrare come i sistemi di MMIR consentano di recuperare immagini, documenti audiovisivi e brani sonori tramite gli attributi dei linguaggi specifici di ogni documento: in base ai criteri della similarità e dei rapporti di misure e valori si possono, ad esempio, utilizzare come chiavi di ricerca la forma, il colore, la struttura, i suoni, i singoli fotogrammi, contenuti nei documenti.

CAPITOLO 1
Sistemi di ricerca e di indicizzazione

Per comprendere al meglio le tecniche del recupero di informazioni multimediali dobbiamo prima spiegare come avvengono le ricerche e le indicizzazioni sui dati, divise essenzialmente tra un sistema più tradizionale, chiamato Term-Based Retrieval (ricerca basata sui termini) e un sistema di recupero più innovativo come il Content-Based Retrieval (recupero basato sul contenuto).

1.1 I sistemi tradizionali

I sistemi di indicizzazione e ricerca tradizionali sono basati sui *termini* (Term-Based Retrieval), sulla logica delle parole chiave, che ci si propone naturalmente, data la nostra propensione naturale ad esprimere e rappresentare verbalmente e testualmente qualunque oggetto di conoscenza e informazione.
In una logica di archiviazione e recupero **term-based** la query viene espressa testualmente, e il processo di ricerca individua i descrittori più pertinenti alla richiesta. I descrittori dei documenti multimediali assumono la forma di termini di indicizzazione, di titoli o didascalie. A questi, nel database, è collegato il documento archiviato, che viene automaticamente collegato alla figura, al suono, o alla relativa anteprima, come in una banca dati testuale le parole chiave fanno emergere il full-text o il suo abstract.
Tutto viene tradotto "in parole": il documento, il suo contenuto, le chiavi d'accesso che lo identificano nella registrazione, anche la formulazione e la struttura della query nella ricerca.

La rappresentazione linguistica lascia però emergere molti problemi. Le queries tradizionali, espresse terminologicamente, sono inadeguate alle nuove e più complesse esigenze della società dell'informazione multimediale, almeno quando gli utenti propongono certi livelli di ricerca. Sarebbe più utile un sistema dove la formulazione di richiesta possa essere inviata così come nasce, tramite forme, colori, movimenti, sonorità, e così come spontaneamente prodotta, in caratteri immediatamente visivi o sonori, possa essere dal sistema afferrata e soddisfatta. Se non è più ammesso un monopolio dell'informazione scritta, è contraddittorio che permanga il monopolio del metodo specifico di reperimento di tale informazione: la nuova società culturale richiede un sistema flessibile alla multimedialità per la ricerca di un'informazione multimediale.

Negli archivi di immagini di musei e gallerie, ad esempio, per la ricerca del materiale visivo è essenziale la conoscenza di una precisa terminologia tesaurale. Il vocabolario degli studiosi, però, risulta spesso poco intuitivo e di non facile uso non solo per gli utenti medi; anche per gli specialisti esso può rivelarsi limitativo quando si vuole andare oltre gli schemi tradizionali. Se in tali archivi si vuole sviluppare l'utilità del servizio informativo sia verso altre utenze sia per quelle già presenti, dunque, a poco servirà affrontare il complesso e costoso lavoro di assegnare termini a ogni immagine secondo i metodi tradizionali: il problema, si deduce, va affrontato cambiando la struttura di base del sistema.

Serve dunque un metodo più flessibile di ricerca, non vincolato da chiavi e metodi di classificazione decisi e imposti da qualcun altro.

1.2 I sistemi di recupero innovativi

I sistemi di recupero più innovativi richiedono una riflessione più approfondita, basata sulla capacità di intendere l'importanza e la funzione di altri linguaggi nel trattare oggetti diversi; ma tale riflessione si è tradotta in pratica di ricerca da poco più di un decennio, e un ruolo molto importante bisogna tuttora riconoscere ai sistemi di reperimento term-based.

Il sistema di archiviazione e recupero **content-based** si trova in una fase di sperimentazione piuttosto avanzata e tecnologicamente molto impegnativa. Può apparire evidente la sua maggiore efficacia, nell'essere un sistema che si basa di fatto sulla ricerca del contenuto del documento visivo, sonoro o audiovisivo, e non di un contenuto "nominale", costituito da termini linguistici, ma del contenuto "sostanziale" vero e proprio, composto di strutture e forme, suoni e colori.

Il metodo del Content-Based Retrieval risulta dunque l'**unico veramente efficace** per cogliere in pieno l'obiettivo del MultiMedia Information Retrieval: restituire l'oggetto che esattamente si cerca, al di là di ogni classificazione, decisa, la maggior parte delle volte, da un'altra persona. Bisogna ricordare però che si deve comunque parlare di mediazione tra la fisicità originaria dell'oggetto – tela, marmo, pellicola, nastro magnetico – e la sua messa a disposizione informativa – ossia la sua trasformazione in un documento digitale. La ricerca non può avvenire sulla fisicità della tela, o della pellicola, ma avverrà sul loro diretto corrispondente in valori elettronici, direttamente nel dominio degli effettivi valori spaziali, formali e sonori.

CAPITOLO 2
MultiMedia Information Retrieval

Fin dagli anni ‘60 sono state sviluppate diverse tecniche che aiutano nell’indicizzazione e recupero di testi elettronici. Date queste premesse, il primo approccio all’indicizzazione di immagini e informazioni digitali fu quello di annotare le immagini o i dati stessi tramite parole-chiave e di utilizzare dunque gli strumenti già in possesso nell’ambito testuale per operare su queste annotazioni in fase di recupero. Fu però presto chiaro come l’utilizzo di query testuali per questi tipo di file fosse inadeguato, e di quanto si rendessero sempre più necessari sistemi di ricerca appropriati allo specifico dell’elemento da recuperare.

Per meglio comprendere le differenze, ma anche l’integrazione stessa, tra i vari sistemi di recupero delle informazioni digitali, bisogna analizzare le quattro grandi categorie sulle quali basiamo la nostra ricerca e cioè, la ricerca sulle immagini (Visual Retrieval, nel quale i documenti visivi sono cercati e recuperati tramite dati visivi), sui video (Video Retrieval, dove per il recupero di documenti audiovisivi si utilizza il linguaggio audiovisivo), sul testo (Text Retrieval o Information Retrieval, basato su informazioni testuali per la ricerca di documenti testuali) e sui file audio (Audio Retrieval, nel quale l’informazione sonora è ricercata in misure di suoni), studiando tecniche e applicazioni pratiche.

Tutto questo complesso di tecniche costituisce la sostanza di una nuova strategia di ricerca dell’informazione che, al di sopra delle metodologie tradizionali term-based, tenta di risolvere il problema del reperimento content-based dell’informazione, nonchè delle architetture necessarie nei nuovi grandi database multimediali.

Gli oggetti multimediali, per via della loro complessa struttura non sono efficacemente rappresentabili con la logica dei sistemi term-based, essendo documenti multimediali ripresi dal mondo reale, estratti anche con

strumenti di registrazione diretta e tradotti in rappresentazioni dell'oggetto reale. Con le tecniche attuali i sistemi elettronici riescono a identificare propriamente gli oggetti reali, estraendo determinate informazioni dal corrispondente oggetto multimediale.

Tutte queste informazioni sono contenute nel **data model** dell'oggetto multimediale, un insieme di dati meno complesso e voluminoso di quello dell'oggetto che rappresenta, e vengono definite *features* (caratteristiche) e possono rappresentare la struttura dettagliata dei vari oggetti multimediali, le loro proprietà, le operazioni definite su di esse, le relazioni tra gli oggetti multimediali e quelli del mondo reale.

I data model, l'insieme di dati digitali identificativi del documento, possono essere quindi trattati dai sistemi informativi più avanzati.

L'intero processo, che porta gli oggetti della realtà quotidiana al loro trattamento digitale, viene definito "multimedia data modelling", ed è il nucleo teorico e tecnico del Multimedia Information Retrieval.

La differenza tra una banca dati testuali e un database multimediali implica anche procedure di ricerca molto diverse, con una continua interazione con l'utente. Una particolare importanza viene assunta dal *browsing*, mentre i risultati delle nostre *query* sono basati su certi gradi di similitudine tra dati inviati e quelli ritrovati, oltre al fatto che le stringhe di interrogazioni possono contenere direttamente valori del data model.

Nel procedimento più tipico la ricerca inizia con la preliminare selezione di un archivio e di una sua parte attraverso una richiesta tipicamente testuale, per ridurre la mole di tutto il materiale potenzialmente interrogabile, attraverso l'uso di termini, titoli e nomi. Si procede poi con un browsing d'esplorazione, muovendosi tra tanti oggetti che assomigliano a quello cercato e selezionando con il mouse un oggetto o la sua anteprima direttamente sul display. In questo modo si inviano all'elaboratore diversi data model, contenenti i dati caratteristici che dovranno essere rintracciati negli oggetti dell'archivio per estrarli come risultati della nostra query multimediale.

Con questa tecnica si potrà quindi, ad esempio:

- trovare l'intera opera o altre opere simili semplicemente con l'immagine di un particolare di un opera di un pittore;
- trovare le fotografie di un criminale attraverso un identikit;
- con le note di un motivo trovare diverse composizioni che lo hanno sviluppato.

La misura della similitudine tra due oggetti multimediali è indicata come variabile tra 0 (completamente differente) e 1 (esattamente corrispondente), indicando il grado di corrispondenza e similitudine tra i parametri della stringa richiesta e quelli dell'oggetto restituito come risultato.

Un'altra importante questione è quella relativa all'indicizzazione (vista ovviamente in senso più ampio rispetto alla sua accezione comune). Nella nostra ricerca l'indice viene creato impostando come collegamenti di accesso ai documenti i dati costitutivi del loro stesso contenuto multimediale: forme, colori, suoni e altri elementi simili.

Vista la complessità e diversità tra le tecniche sopra descritte vediamo come ci si comporta specificatamente nei casi di documenti specificamente sonori, visivi o audiovisi.

2.1 Visual Retrieval

La disponibilità sempre maggiore di immagini e archivi video ha fatto aumentare gli sforzi nello sviluppo di mezzi per l'effettiva ricerca nel contenuto dei visual data based.

Per le immagini i miei dati (data model) possono contenere informazioni come il formato, la risoluzione, il numero di pixel, i valori dei colori, i valori caratteristici della struttura, le forme, le relazioni spaziali, le texture, eccetera. E' essenziale allora sviluppare tali modelli rappresentativi, in quanto praticamente tutte le operazioni relative agli oggetti del database multimediale si basano su di essi.

E’ fondamentale, per attuare l’indicizzazione e le successive operazioni di recupero dei documenti visivi (in particolare per il trattamento automatico dei valori relativi ai loro elementi), produrre un certo livello di astrazione del documento, estraendo alcuni elementi del contenuto concreto dell’immagine.

Il nostro data model sarà il risultato di una serie di operazioni di identificazione, isolamento, valutazione ed estrazione dei valori rappresentativi delle immagini digitali.

Sono possibili cinque possibili modalità per indicizzare, archiviare, ricercare e recuperare i documenti visivi digitali, definibili come modi di astrazione dei materiali. Esse si possono indicare come:

- modalità semantica
- modalità formale
- modalità strutturale
- modalità coloristica
- modalità parametrica

La modalità *semantica* è il metodo più tradizionale, ma anche il più problematico nel campo delle immagini. Si basa sulla definizione di etichette testuali, descriventi nomi, caratteristiche, classi o concetti, da attribuire con precisione a un’immagine, le quali dovranno essere conosciute e richiamate per consentire il recupero del documento associato.

La modalità *formale* si basa sulla capacità dell’elaboratore di attuare un confronto tra forma o contorno, estratti dalla figura archiviata e quelli estratti dal modello con cui si definisce la query, il quale è messo a disposizione dal sistema oppure può essere proposto anche dall’esterno. Il recupero del documento potrà avvenire se l’elaboratore valuterà un certo grado di vicinanza tra i valori dei dati rappresentativi delle immagini confrontate.

Il modo di astrazione *strutturale* si basa, invece, sulla scomposizione delle immagini dell’archivio in sezioni; il sistema stimerà poi la somiglianza della composizione strutturale di tali immagini con la struttura delle sezioni di una figura modello, le quali faranno dunque da chiavi di ricerca. Il recupero

di un'immagine sarà effettuato, allora, in base alla similitudine a tali chiavi di alcune delle sezioni che la compongono.

L'astrazione *coloristica* consiste nel rappresentare le immagini estraendo da esse i diversi colori, o le scale dei grigi, che le costituiscono. Le operazioni di archiviazione e recupero si baseranno, di conseguenza, sul trattamento e il confronto dei valori dei dati relativi a tali proprietà coloristiche della figura.

Infine, la modalità *parametrica* è fondata sulla determinazione dei valori dei parametri costitutivi della forma, della struttura e del colore dell'immagine. Il sistema potrà recuperare un'immagine tramite il confronto tra i valori dei vari parametri immessi nella query, attraverso una figura modello o compilando un'apposita griglia, e quelli posseduti dalle immagini archiviate.

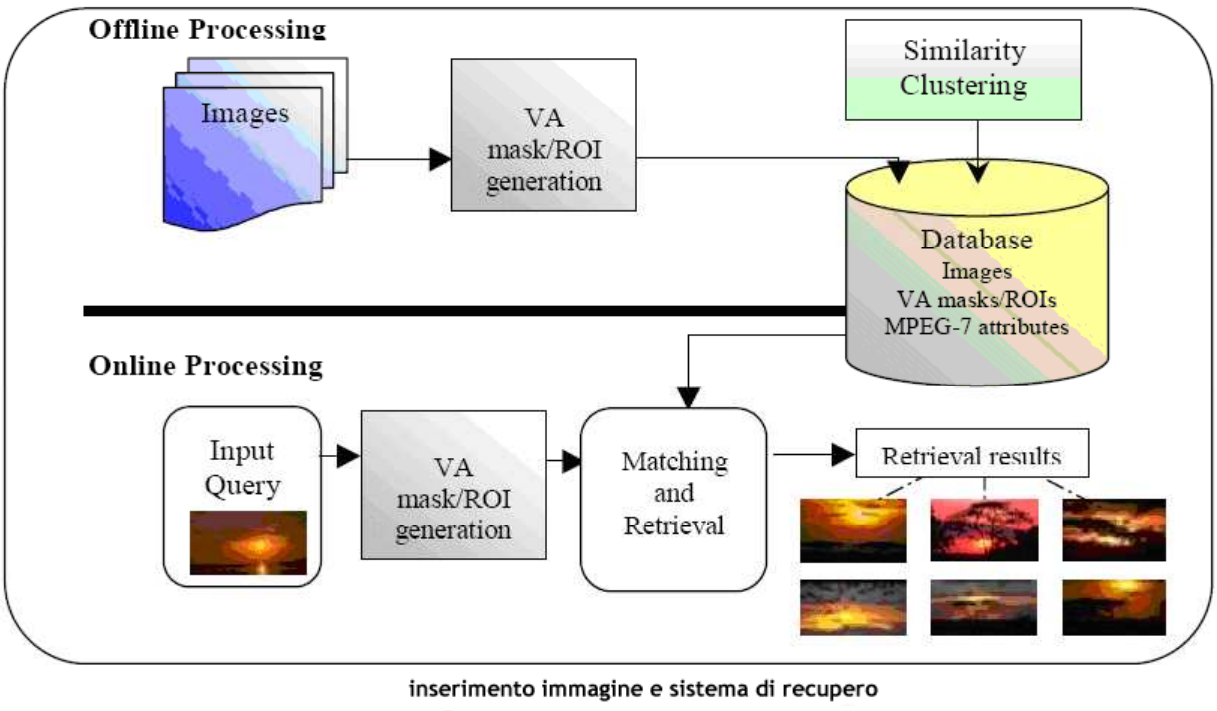

Figura 1: Inserimento dell'immagine nel nostro archivio e recupero dei dati.

I livelli di astrazione delle caratteristiche visive sono divisi in proprietà "globali" e proprietà "locali": le proprietà globali vengono indicate nel

“colore” e nella “struttura” mentre le proprietà locali riguardano la “forma” e le “relazioni spaziali”.

Il tradizionale sistema term-based mantiene un valore di analisi iniziale, con la precisa utilità nello sfoltire il database di ampie porzioni dell’archivio che non si rivelerebbero di alcun interesse in base all’obiettivo della ricerca.
Per la nostra ricerca la soluzione ideale per l’interrogazione della banca dati è ovviamente quella content-based, che esplora direttamente la natura figurativa dell’immagine, tanto attraverso l’impiego di modelli e campioni quanto di griglie di valori, in base ai quali il sistema cercherà documenti di simile contenuto.
In tal modo nelle stringhe della query visiva useremo elementi realmente appropriati, ad esempio:
- immagini intere (per query relative alla struttura o alla forma);
- dettagli e icone (nelle query basate sulle relazioni spaziali);
- bozze esemplificative (per query sulle forme);
- griglie di valori (nelle query basate sui colori).

Le espressioni di ricerca possono essere solo approssimazioni, formulazioni parziali, o individuali, dell’effettivo contenuto visivo che l’utente desidera recuperare, questo perché non vi sarà mai un matching esatto tra il data model inviato per la ricerca e quello archiviato, e quindi da un punto di vista tecnico non ci sarà necessità di essere troppo rigorosi nel comporre la stringa di query.
Il recupero dei dati avviene per similitudine, essendo diverse la valutazione che può dare un essere umano rispetto ad un elaboratore. Nella ricerca per forme ci saranno differenze e variazioni dei contorni a cui il computer è sensibile in modo diverso dall’utente, quindi il risultato del recupero può essere inaspettato. Per i colori e le strutture il giudizio di similitudine è ancora più legato a fattori propriamente sensibili della visione umana, e può essere molto più distante da quello della macchina.

Bisognerà quindi progettare e programmare un sistema di recupero (attraverso calcoli geometrici e matematici) legato alla sensibilità dell'uomo nel vedere le immagini.

La ricerca e il reperimento delle immagini si baserà molto sull'interattività con l'utente, attraverso le fasi di browsing e di visualizzazione, permettendo così il perfezionamento della query.
Il processo di ricerca inizierà quindi con una selezione terminologica di massima di una parte dell'archivio e la visualizzazione di una certa quantità di immagini rappresentative di un insieme.
Il processo avverrà tramite la selezione di una figura d'esempio che sarà inviata come stringa di query, oppure scegliendo un modello sulla base della forma o del colore o ancora con l'immissione di un'immagine campione dall'esterno.
Ovviamente, se l'utente lo ritiene necessario, potrò continuare ad utilizzare parametri testuali per perfezionare la ricerca, senza limitarsi alla sola fase preliminare.
I sistemi term-based e content-based possono infatti essere integrati, al fine di sviluppare ricerche di maggiore raffinatezza, laddove i pregi indiscutibili di un sistema compensano le lacune tecniche dell'altro. In ogni modo, l'interazione con l'utente deve guidare tutto il processo di ricerca, e i sistemi elettronici devono fornire gli strumenti necessari per elaborare e adattare i risultati parziali che man mano vengono recuperati.

Content Based Image System

Il Content Based Image Retrieval System (CBIRS) si basa su concetti presi in prestito da ambiti quali la computer vision, il pattern matching, la psicologia cognitiva e molti altri, al fine di estrapolare dalle immagini caratteristiche ad esse, proprie e intrinseche, funzionali all'indicizzazione e a processi di recupero automatizzati.
Un ambito di utilizzo tipico è quello pubblicitario, dove ci si trova spesso a dover cercare rapidamente tra migliaia un'immagine che corrisponda a

determinate caratteristiche, oppure in ambito giornalistico, dove ancora una volta si deve reperire in maniera veloce ed efficace un'immagine pregnante ad un determinato contesto.

La maggior parte dei file di immagine sono mappature di entità appartenenti al mondo reale in una forma binaria. Questi file contengono due tipi base di informazioni: quelle che si riferiscono ad attributi e relazioni appartenenti ad entità del mondo reale (i contenuti) e quelle che sono specifiche alle rappresentazioni binarie di tali entità (la loro codifica). Qualunque sia la rappresentazione binaria (la codifica) che noi scegliamo, il contenuto di un'immagine dovrebbe rimanere lo stesso. Al tempo stesso il contenuto di un'immagine non dipende dalla sua risoluzione. Una bassa qualità o una bassa risoluzione potrebbero rendere più difficile estrapolare caratteristiche descrittive dell'immagine stessa: ciò che è importante sottolineare è che il formato dell'immagine e il tipo di compressione ad essa applicata, così come la risoluzione e i colori, sono elementi specifici alla codifica binaria dell'immagine stessa e non al suo contenuto. I CBIRS operano a basso livello sulla codifica delle immagini per consentire all'utente di recuperare in maniera veloce ed efficiente il loro contenuto. Date una serie di immagini queste vengono processate operando su di esse determinati algoritmi: questi algoritmi estrapolano dalle immagini determinate informazioni, definite firma (signature) o caratteristiche (features) dell'immagine.

In base all'algoritmo implementato queste informazioni possono dirci quali colori sono presenti in una data immagine, dove si trovano nell'immagine e in quale misura e percentuale, quali contorni presenta l'immagine, quali linee tendono ad essere presenti in maggiore percentuale all'interno dell'immagine, quale la loro direzione, ecc.

Ogni firma di ogni immagine viene inserita in un database; quando un utente interroga il database presentando al sistema un'immagine di riferimento, questa stessa immagine subisce i processi precedentemente applicati a quelle già presenti nel database, ovverosia indicizzate: anche per essa viene dunque estrapolata una firma che verrà confrontata con

quelle già presenti nel database, permettendo così all'utente di recuperare l'immagine che più assomigli a quella presentata.

Per il recupero di immagini in base al contenuto si identificano di norma le seguenti tipologie di richiesta (query):
– Diretta: l'utente sa esattamente che cosa sta cercando e conosce quali chiavi il sistema utilizza per identificare un particolare oggetto.
– Per similitudine (Query-by-example): l'utente seleziona uno o più documenti o parte di un documento che sa essere simili al tipo di documento che sta cercando.
– Per Prototipo (Query-by-sketch): questa tecnica è correlata con la precedente. Il prototipo in questione potrebbe essere un disegno improvvisato dall'utente al momento di effettuare la query o un oggetto interpretato dal sistema per produrre una particolare rappresentazione.
Per quanto concerne gli algoritmi con i quali vengono processate le immagini gli approcci sono essenzialmente due: da un lato l'utilizzo di funzionalità quali color histogram e caratteristiche relative alla forma come circularity e orientamento degli assi principali di regioni dell'immagine, così come combinazioni di queste tecniche, dall'altro i wavelet transform.

2.2 Video Retrieval

I database di audiovisivi sono oggi sempre più utili e diffusi e sono indispensabili in diversi campi di applicazione: si passa dagli studi televisivi e cinematografici, fino ad ambiti ingegneristici o aerospaziali. Di conseguenza trovare una metodologia di ricerca ottimale per i documenti video è diventato sempre più indispensabile.
Un video è essenzialmente un flusso continuo di immagini, e questo rende piuttosto complessa la rappresentazione e la ricerca all'interno del suo contenuto.

Negli ultimi tempi però c'e' stata una maggiore tendenza alla realizzazione dei documenti video in formato digitale che rende possibile il *video data processing*, cioè l'estrazione di informazioni digitali per la strutturazione del database e di conseguenza lo sviluppo di un sistema di recupero content-based.

In questo modo potranno essere estratti, anche da sequenze molto lunghe, i video clips – brevi frammenti utilizzabili come chiavi di archiviazione e ricerca per l'intero filmato –, operando sulla base della "struttura temporale" del film e sulla base del suo "contenuto semantico".

Un esempio di tali chiavi può essere la presenza, nel video, di un oggetto che oltre ad essere nella sua figura un attributo semantico statico, diventa un attributo dinamico apparendo, scomparendo e relazionandosi ad altri oggetti nel tempo del filmato.

Nell'astrazione dei dati video l'ideale sarebbe utilizzare un sistema di riconoscimento e prelevamento automatico dei dati caratteristici del video, ma la tecnica attuale consente solo sistemi impostati manualmente e computer-assisted.

Fondamentale però è che si riesca a identificare le componenti sintattiche e semantiche del video, che consentono di definire il data model delle proprietà strutturali e semantiche del documento. Il modello dei dati, infatti, sarà il centro del query processing, e dovrà contenere tutti i dati che si possono associare agli elementi che supportano la query.

Questi dati sono elencati come:

– "dati bibliografici", quali genere, soggetto, titolo, produttore, regista, periodo, ecc.;

– "dati strutturali", ad esempio inquadratura, sequenza, scena, brano, pezzo di montaggio;

- "dati contenutistici", il ricco insieme di elementi visivi e sonori, di immagini sia statiche sia in movimento, di frasi scritte nei fotogrammi come parole e altri suoni.

Si può intendere quindi come i dati di tipo testuale, nonché i dati riguardanti l'immagine fissa presa in sé, abbiano un'importanza

determinante, in combinazione con i dati più specificamente audiovisivi, nell'impostare, condurre e precisare la video query.

Modalità di video query

Il processo di ricerca è attuato attraverso una serie di stadi concatenati, i primi dei quali agiscono da filtri per i successivi, riducendo sempre più la mole dei documenti fino a individuare quelli che soddisfano le necessità informative dell'utente.

L'utente potrà dunque rivolgersi al database avendo anche solo una vaga idea di ciò che desidera – la memoria di una scena, del volto di un personaggio, delle battute di una trasmissione – e in seguito a un processo fondato sull'interazione con il sistema, trovare il video desiderato.

Il primo stadio, definito *navigating*, ha un valore preliminare, di filtro generico. Il suo scopo è quello di selezionare, tramite l'uso di indicazioni testuali, il tipo di archivio, la categoria o il genere di video, l'ambito temporale o la localizzazione dei soggetti, riducendo di gran lunga la mole dei documenti su cui operare.

Il secondo stadio, detto *searching*, costituisce il momento più importante della ricerca; il suo risultato è una lista di documenti che si candidano a soddisfare la maggior parte dei requisiti della query. In un primo passo della ricerca, può ancora avere un ruolo la modalità di recupero term-based, come mezzo per ridurre ulteriormente la quantità di documenti, almeno in relazione ai titoli, ai nomi, oppure a una descrizione generale di alcune sequenze.

Essenziali, comunque, per avanzare di un secondo passo in tale ricerca, sono ulteriori mezzi di descrizione di tipo content-based: quali il numero delle inquadrature, le immagini presenti, la dinamica e il grado di presenza delle figure, i dialoghi e la musica. I documenti della lista finale possono essere mostrati su un display. È assai improbabile però che la lista sia quella definitiva in questo stadio, perché queste modalità di ricerca non possono sviluppare la precisione fino a tenere conto degli elementi più riccamente dinamicoespressivi dei filmati.

Un terzo fondamentale stadio, necessario per il completamento della ricerca, è il *browsing*. Dalla lista dei documenti molto va ancora escluso; in questa fase si deve poter visionare una serie di scene che siano una buona anteprima dei filmati selezionati fino a questo momento, così l'utente potrà intuire velocemente il contenuto di ognuno, ed esaminare anche lunghe liste in breve tempo. Ciò permetterà di scegliere un insieme ristretto di documenti come risultato ultimo della ricerca, ovvero di stabilire un ulteriore raffinamento della query.

Per il browsing si pone, però, il problema della disponibilità di visual summaries veramente efficaci. Ottenere una sintesi del filmato significativa è un punto critico specifico del video data processing; il problema infatti non si pone nel browsing delle figure fisse, dove è sufficiente che siano visionabili in formato ridotto. La struttura di un visual summary invece deve essere prodotta in relazione al contenuto semantico del video, per consentirne una veloce interpretazione. A tale pratica si è usi da tempo – ad esempio, nel campo dei trailer pubblicitari dei film –, ma nel campo del Video Retrieval il problema è quello di agire secondo una prospettiva documentaristica, sapendo evidenziare gli elementi più rilevanti sotto tale punto di vista, e non dello spettacolo.

Quarto e ultimo stadio è il *viewing*. Esso, in sostanza, consiste nella visione di parti complete dei filmati selezionati come documenti finali, attraverso normali meccanismi di play, pause, fast-forward o reverse, allo scopo di utilizzare e valutare direttamente il video, eventualmente prima della scelta di recuperarlo per intero dal database.

A monte del processo della video query, perché sia aperto all'utente nella sua piena potenzialità, è il lavoro di preparazione dei documenti, consistente nella video analysis e nel video processing.

Dal punto di vista tecnico si pone lo sviluppo di potenti algoritmi per la *betweenshot analysis* come fondamento per un'efficace analisi dei video, la quale, nel rispetto della caratteristica che anima il senso di questi (ossia il montaggio), deve consentire il trattamento del filmato per l'archiviazione e la ricerca, comprendente anche la produzione degli indispensabili visual summaries.

Un'altra importante questione è quella dell'analisi e del processing della colonna sonora, fatta di voci, suoni e rumori: anche questi elementi possono essere usati come chiavi della video query nella sua omogeneità, e i sistemi di Video Retrieval devono tenerlo in forte considerazione.

2.3 Audio Retrieval

Nella pratica i sistemi di Audio Retrieval sembrano quelli che hanno dato i migliori risultati. Il miglioramento delle caratteristiche e della resa dei file audio, senza tralasciare la loro minore pesantezza (e di conseguenza facilità di manipolazione) legata al formato MP3, hanno contribuito ad aumentare l'interesse per il recupero di questo tipo di file.
La tecnica di ricerca migliore è il *browsing*, attuato attraverso una serie di informazioni sonore chiave relative a interi brani, ed è funzionale per il recupero di ogni genere di documenti audio: registrazioni musicali, vocali, o d'altro tipo di suoni e di rumori.
Le sintesi chiave di brani sonori (*audio key-information*), ascoltabili in carrellata, vengono trasmesse contemporaneamente ad indici chiave testuali, detti *text based indexes,* collegati al documento corrispondente alla chiave audio.
Il metodo di *query* tradizionalmente usato nei database audio, basato su parole chiave descrittive dei brani, non consente una ricerca veramente efficace, se non altro per la necessità che l'utente conosca un dato vocabolario d'interrogazione (senza sollevare nemmeno la questione della descrivibilità a parole di certi pezzi).
Viene, così, proposto un sistema di *browse search* che, più che integrare la ricerca terminologica, aiuti direttamente l'utente a scoprire e localizzare le informazioni audio che gli sono sconosciute.
Descrivendo lo schema del sistema proposto possiamo stabilire due fasi:
– la prima è quella della "synchronous provision of index and keyinformation", cioè *audio key-information,* brevi sintesi di interi brani, trasmesse dal server unitamente agli indici testuali;

– la seconda è la "audio browse search", nella quale l'utente cerca quasi inconsapevolmente, ascoltando in background tale flusso di brani chiave ma pronto a catturarli quando arriva qualcosa che può interessarlo.

Nella pratica, un ruolo fondamentale hanno i sistemi di trasmissione dell'informazione e la loro tecnologia. Infatti, nella prima fase, il server deve potere, anzitutto, realizzare degli opportuni *digest*[2] del documento audio, attraverso analisi e processing sonori che ne producano una sintesi rappresentativa. Successivamente, deve poterli inviare su un canale audio sincronicamente all'invio degli indici testuali su un canale parallelo.

Nella seconda fase, entra in gioco il sistema dell'utente, una macchina che sia in grado di ricevere i due tipi di segnale, e di far ascoltare la carrellata di brani chiave.

Quindi, nel momento in cui l'utente rintraccia un'informazione che lo interessa, la macchina deve consentire, con una semplice operazione, di fissare e registrare l'indice testuale abbinato alla chiave sonora, che sarà utilizzato poi per inviare al server la richiesta di reperimento del brano intero.

Devono essere messi a punto macchine server e terminali utenti molto sviluppati, tanto che sia la *browsesearch* sia la registrazione e l'utilizzo degli indici dei brani risultino assolutamente semplici, realizzabili dall'utente in qualunque momento e senza doversi interessare di indicizzazione, preoccupandosi solo di registrare su una propria *memorycard* le chiavi di accesso ai documenti che sono stati scelti.

In tal modo, si ha dunque un sistema che consente di creare una sorta di grande repertorio personale di documenti sonori interessanti.

Il ricorso al metodo del *browsing,* attraverso una carrellata di *audio key-information*, consente di costruire velocemente e intuitivamente questo indice, sulla base dei contenuti sonori. Ciò dispensa dal conoscere e definire dati testuali per raggiungere l'informazione: essa si ascolta sintetizzata, si ferma semiautomaticamente il suo indirizzo, successivamente e sulla base delle necessità si può recuperare dal server.

[2] Il digest non è altro che una compressione estrema del documento.

In sostanza, questo può essere definito un vero e proprio sistema di ricerca *content-based*. Infatti, quando si desidera cercare un brano, ci si può collegare al server e ascoltare attentamente le chiavi audio inviate; così, esaminando in *browsing* i diversi brani, si può attuare la selezione sulla base dell'immediato contenuto sonoro dei documenti.
Sono dunque proprio le chiavi audio che consentono di riconoscere i documenti desiderati; l'indice testuale è, poi, solamente un sistema pratico per etichettarli, nel proprio archivio e in quello del server.
Per un sistema di Audio Retrieval avanzato, il principio di liberarsi dai sistemi tradizionali di archiviazione e recupero è già ben radicato in questo metodo, anche se si è ancora lontani da un sistema dove un utente generico possa, liberamente, impostare una *query* usando modelli o campioni di melodie, e recuperare tutto ciò che più gli si avvicina.

CAPITOLO 3
Applicazioni pratiche del MultiMedia Information Retrieval

3.1 Applicazione delle tecniche di Visual Retrieval

I campi di applicazione più importanti e dove sono più sfruttate le tecniche e i sistemi di ricerca visiva sono la ricerca biomedica, le scienze della Terra e dell'informazione geografica e la documentazione delle arti visive (gallerie d'arte, archivi di musei).
Altri campi di applicazione sono il disegno ingegneristico e architettonico, le banche di immagini scientifiche, gli archivi fotografici delle forze di polizia (utili per l'investigazione criminale) e il design di moda, oltre che l'utilizzo di immagini e foto nel web e nel mondo della comunicazione.
Tutti questi ambiti applicativi hanno reso questo campo di ricerca quello con la crescita più veloce nell'IT (Information Tecnology), richiedendo il contributo di diverse discipline di ricerca e ponendo nuove sfida a ingegneri e scienziati.

L'IBM si è sempre dimostrata interessata a questo campo. Lo dimostrano anche la messa a punto del più antico dei sistemi di Visual Retrieval, **QBIC** (**Query By Image Content**), nato alla fine degli anni Ottanta, e che con i necessari aggiornamenti rappresenta uno dei sistemi più all'avanguardia ancora oggi. QBIC si applica ai database di immagini e a quelli di video, ed è strutturato per il trattamento di documenti visivi prodotti in ogni campo specifico di applicazione come le immagini geografiche, ingegneristiche, d'arte o mediche ed è implementato nelle banche dati di ogni genere di istituto o azienda in molti paesi del mondo.

Le sue possibilità di indicizzazione e di ricerca content-based delle immagini sono le più ampie. Esso consente interrogazioni per forma, struttura, colore, relazioni spaziali, termini e combinazioni di queste modalità; è possibile proporre campioni dall'esterno come produrre modelli con gli strumenti messi a disposizione; strumenti di elaborazione delle immagini recuperate consentono, inoltre, modificazioni per rilanciare la query; in più, QBIC può riconoscere anche singole figure di un complesso e utilizzarle isolatamente per le interrogazioni, anche combinandole con quelle di altri complessi.

QBIC si basa su una combinazione di sistemi di indicizzazione automatici e semi-automatici, e utilizza funzionalità di *color histogram*.

L'utilizzo di color histogram consiste nel sovrapporre all'immagine una griglia di celle, all'interno di ognuna delle quali vengono applicati algoritmi atti al rilevamento del tipo e della percentuale di colore ivi presente. Il limite principale dei color histogram consiste nel mancare di informazioni spaziali relative al colore: se l'immagine viene per esempio ruotata o traslata le informazioni ottenute non sono più valide. I color histogram restano comunque un valido strumento al servizio di un CBIRS, a patto che le immagini restino posizionate e orientate così come esse si trovano al momento della loro indicizzazione.

Oltre alle tecniche di color histogram il QBIC utilizza funzioni di texture features, shape features e sketch features; le caratteristiche estrapolate dall'utilizzo di queste metodologie vengono salvate insieme alle immagini.

La caratteristica di QBIC è quella di consentire all'utente la composizione di una richiesta in base a diversi attributi visivi; per esempio l'utente può specificare una particolare composizione cromatica (x% del colore 1, y% del colore 2, etc.), una particolare texture, alcune caratteristiche relative alla forma e uno schizzo dei contorni dominanti nell'immagine target, insieme ai pesi relativi da dare ad ognuno di questi attributi.

QBIC viene utilizzato all'interno del sito web del museo dell'Hermitage per consentire agli utenti di reperire opere simili a quelle selezionate o per trovare opere che presentino colori e forme disegnate dall'utente stesso su una tavolozza.

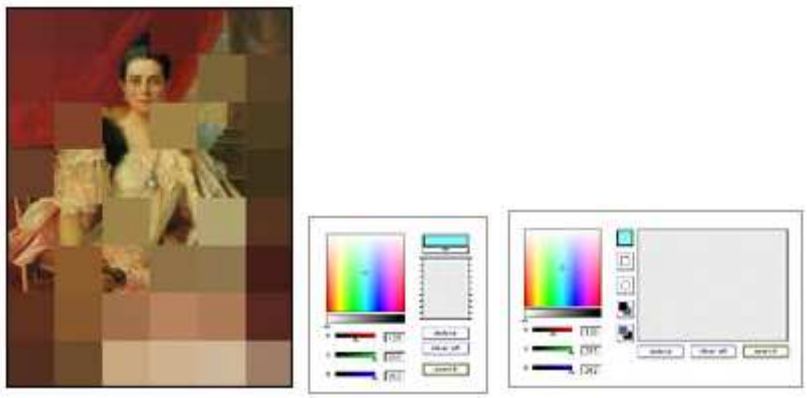

Figura 2: Esempio di QBIC Colour and Layout Search utilizzati dal museo Heritage

Questo è un buon esempio delle funzionalità dei color histogram: i dipinti saranno sempre presentati e usufruiti dall'utente in una determinata posizione, le informazioni spaziali possono quindi venire ignorate.

Un altro prodotto realizzato in collaborazione con IBM è **ImageMiner**, il quale è in grado di analizzare automaticamente le immagini producendo due ordini di indicizzazione del loro contenuto. Una funzione permette, interpretando le caratteristiche visive, di riferirle poi a dei termini, creando un vero e proprio tesauro terminologico; un'altra funzione estrae queste caratteristiche nella loro immediata concretezza figurativa, creando una sorta di tesauro visivo. Il sistema, dunque, è successivamente in grado di attuare ricerche sia sulla base dei termini, sia utilizzando dati relativi alle forme, le strutture e i colori.

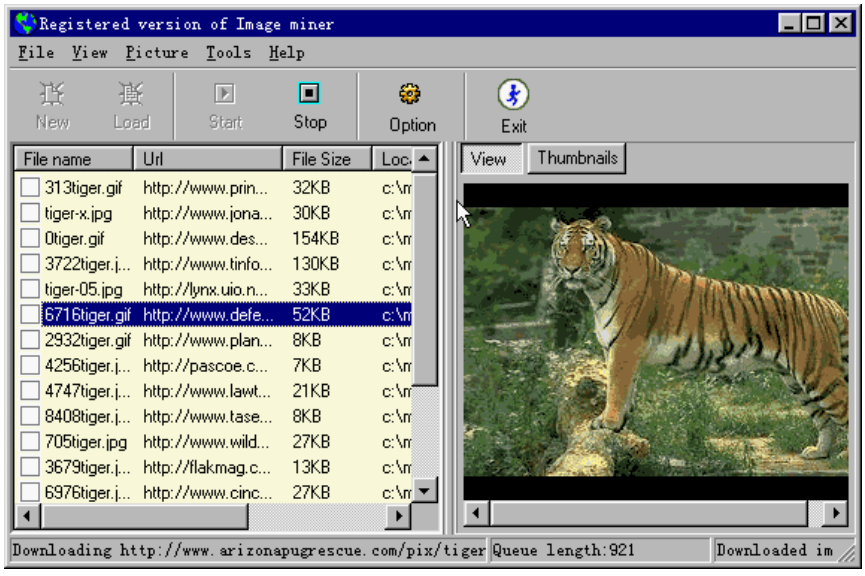

Figura 3: Uno screenshot di ImageMiner

Un importante sistema di indicizzazione e recupero di immagini è **VIPER** (**Visual Information Processing for Enhanced Retrieval**), realizzato dall'Università di Ginevra. Le ricerche possono essere impostate a partire dal browsing di una prima serie di immagini proposte dal sistema, ognuna delle quali è rappresentativa di una categoria e consente, se selezionata e inviata come dato di query, di continuare la ricerca con criteri content-based.

Un avanzato programma di Visual Retrieval è stato realizzato dall'Università di New York (**Columbia's Content-Bases Visual Query Project**) ed è diviso in diverse sezioni, ognuna delle quali è predisposta per rispondere a necessità di ricerca specifiche. Il modulo denominato VisualSEEk consente di impostare le queries in base al colore e al contorno delle figure, nonché l'uso di strumenti per la creazione dei modelli e degli esempi; il modulo WebSEEk è quello di impiego più semplice, basato sui testi e i colori, applicabile anche nel web; il modulo MetaSEEk è un'interfaccia utilizzabile per condurre ricerche su archivi differenti, un meta-motore di ricerca applicabile a motori di ricerca content-based compatibili.

Dalla comunità matematica, nell'ultima decade, è emerso un nuovo strumento, detto **wavelet transform**, per analizzare funzioni a diversi livelli di dettaglio. Le wavelet sono funzioni matematiche che dividono i dati ad esse sottoposti in diverse componenti di frequenza e successivamente studiano ogni componente con una risoluzione proporzionata alla sua scala. Nell'ambito delle immagini il wavelet transform è stato impiegato soprattutto per quanto concerne la compressione. Salvando i coefficienti più ampi di un'immagine sottoposta ad un algoritmo di wavelet transform (e buttando via tutti i coefficienti più piccoli) è possibile ottenere una rappresentazione dell'immagine in gran parte ancora accurata/corretta. Si può pensare ad un CBIRS che applichi ad ogni immagine un wavelet transform, si limiti a salvare un numero ridotto di coefficienti per ogni canale di colore (abbiamo visto come la

"descrizione" dell'immagine sarà comunque ancora corretta) e crei una piccola "firma" per ognuna delle immagini. Grazie alle dimensioni così ridotte di tale firma, la sua ricerca in un database potrà avvenire in tempi molto brevi. Rispetto ai tradizionali metodi (color histogram) le wavelet presentano numerosi vantaggi nell'ambito dell'analisi, primo fra tutti quello di poter formulare query ad ogni risoluzione, eventualmente diversa da quella dell'immagine target, nonché il fatto di preservare le informazioni spaziali relative all'immagine. Inoltre il tempo richiesto dalla creazione del database di immagini è indipendente dalla risoluzione delle immagini stesse. Tramite l'utilizzo di wavelet le informazioni di firma possono venire estrapolate da versioni *waveletcompressed* dell'immagine stessa consentendo al database di firma di venir creato in maniera veloce e conveniente direttamente a partire da un set di immagini compresse.

Un esempio di CBIRS basato su *wavelet transform* è **ImgSeek**, sviluppato da Ricardo Niederberger Cabral e rilasciato sotto licenza GNU GPL. Per utilizzare ImgSeek l'utente formula una richiesta ad un database di immagini utilizzando l'interfaccia utente per "dipingere" uno schizzo di ciò che vuole trovare o fornendo al sistema un'immagine-esempio. La ricerca dell'immagine maggiormente simile può venire ostacolata da diversi fattori. L'immagine-esempio è di norma diversa dall'immagine-target, quindi il metodo di recupero deve consentire alcune distorsioni. Se l'immagine proviene da uno scanner potrebbe presentare artefatti quali micro spostamenti di colore, scarsa risoluzione, effetti di dithering, etc. Se invece l'immagine fosse dipinta dall'utente, presenterebbe tutti gli eventuali limiti artistici dell'utente stesso. Ciò nonostante, l'obbiettivo è quello di recuperare le immagini in maniera veloce e interattiva anche con database contenente migliaia di immagini.
Durante la fase di preprocessing viene effettuato un wavelet transform su ogni immagine presente nel database. Considerando solo i coefficienti più alti restituiti dalla decomposizione si genera una piccola "firma" per ogni immagine. Le firme vengono salvate e organizzate in maniera tale da rendere facile e veloce la comparazione di ognuna di esse con una nuova

firma. Quando un utente sottopone una richiesta viene effettuata una wavelet transform che genera una firma per l'immagine sottoposta. Questa nuova firma viene confrontata con quelle relative alle immagini del database e quella che corrisponde meglio viene recuperata e consegnata all'utente.

Le firme generate da ogni wavelet transform sono molto piccole, la qual cosa rende ImgSeek veramente molto veloce nell'indicizzazione e nella ricerca e decisamente efficace nel trovare il cosiddetto "best match".

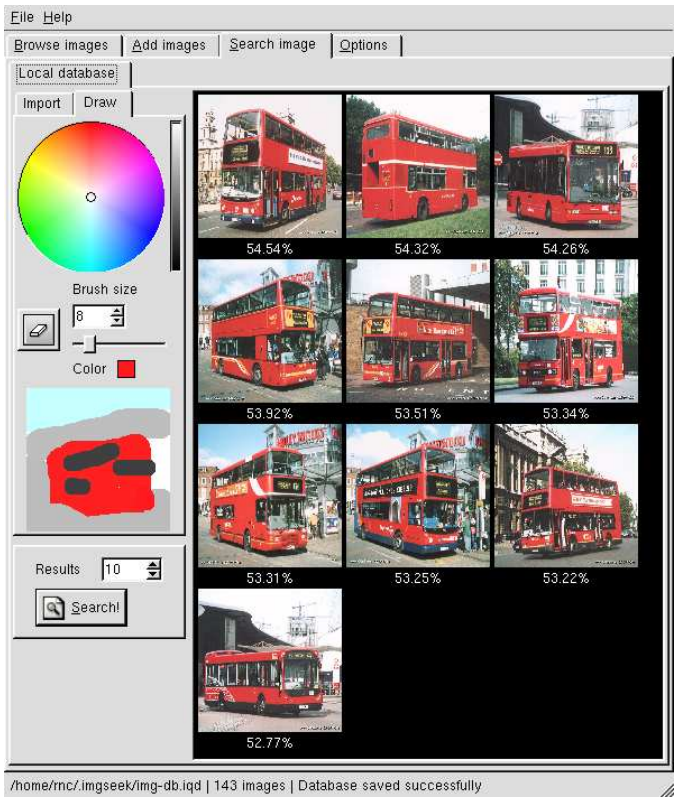

Figura 4: Disegnando un rettangolo rosso con dei segni neri, immaginando un cielo azzurro e l'asfalto grigio (vedi lo schizzo a sinistra nella figura) ImgSeek ha ottenuto questi 10 bus su una collezione di 143 immagini.

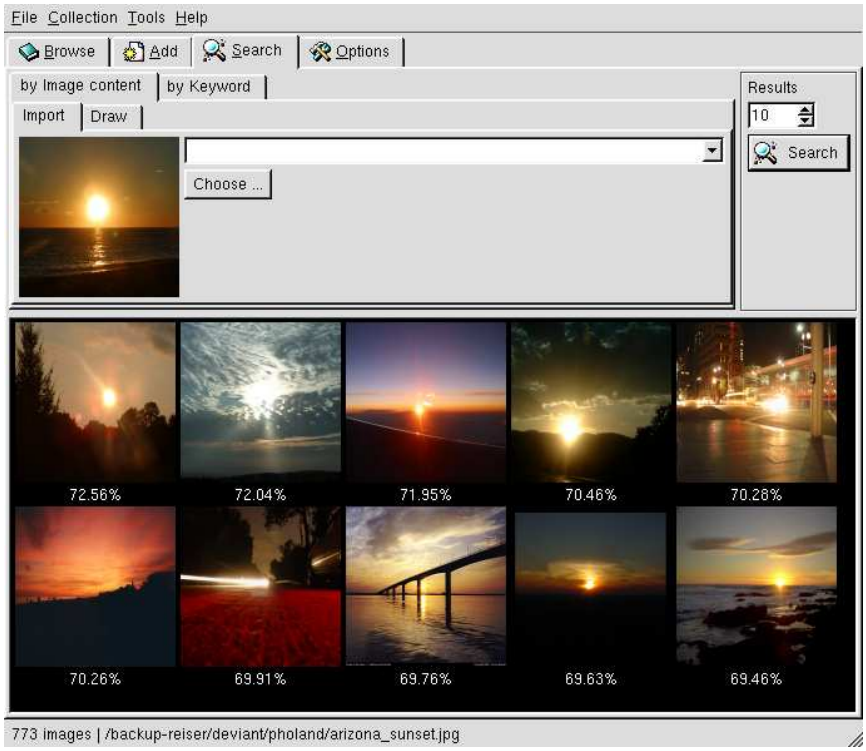

Figura 5: Impostando un campione (in questo caso un tramonto) come query ImgSeek mostra tutte le immagini che abbiano caratteristiche simili al mio modello.

Per quanto riguarda l'applicazione nel campo web gli esempi piu' importanti sono la famosa funzione di ricerca immagini di **Google** (www.google.com) e il motore di ricerca **AltaVista** (www.altavista.com), che permette di ricercare immagini, mp3 e file audio e anche video.
Le ricerche delle immagini su internet avvengono ancora grazie soprattutto a sistemi term-based, analizzando l'url dell'immagine (il suo nome e il percorso "virtuale" dell'immagine nel web), scomponendo l'intero codice HTML della pagina dove è contenuta l'immagine, esaminando il testo nel campo ALT del tag img[3] e controllando i link riferiti a quel file, ma è possibile recuperare importanti informazioni direttamente dall'immagine, come il suo tipo (gif, jpeg, ecc…), la data di creazione e la sua dimensione.

[3] Per visualizzare un'immagine in un documento html bisogna utilizzare il seguente comando: <img src="percorso/nomefile.estensione">
Nel tag si può anche specificare un "alternativa" all'immagine, nome che verrà utilizzato dai motori di ricerca per recuperare l'informazione, e che verrà visualizzato da chi naviga con una versione testuale del proprio browser internet.
es. <img src="percorso/nomefile.estensione" alt="descrizione dell'immagine">

Figura 6: Uno screenshot tratto dall'homepage di AltaVista

Altavista inoltre, ha dei record davvero invidiabili:

• Viene creato il primo indice Web di Internet (1995)

• La prima possibilità di ricerche multilingue in Internet

• Il primo motore di ricerca in Internet a lanciare servizi di ricerca di immagini, audio e video

• I servizi e le funzioni di ricerca in Internet più all'avanguardia: ricerca di contenuti multimediali, traduzione e riconoscimento della lingua, ricerca di specialità

• Riceve 61 brevetti relativi alla ricerca, molti di più di qualsiasi altra impresa del settore della ricerca in Internet

3.2 Applicazione delle tecniche di Video Retrieval

Un sistema per il trattamento e la ricerca di documenti video abbastanza diffuso in commercio è quello prodotto dall'azienda statunitense **Virage**, *Video Logger*, un potente software che permette la codifica e l'indicizzazione dei video, trasformandoli in beni accessibili e flessibili. Questo programma permette infatti di recuperare i dati ricercati, in base tanto alle caratteristiche visive quanto a quelle audio. Inoltre permette la creazione di diverse e differenti versioni codificate in

base al bit rate nello stesso momento in cui sta indicizzando il video e consente inoltre una ricerca immediata ed esatta. Non manca poi la possibilità di riferire ai video, o a loro frammenti, etichette testuali che contengano titoli, autori, date, generi o descrizioni, per garantire possibilità di ricerca terminologiche. Un dispositivo di analisi visiva può, invece, estrarre le caratteristiche propriamente figurative dalle immagini dei singoli fotogrammi. Nello stesso momento in cui il video viene codificato, le tecnologie avanzate di cattura e analisi lavorano in tempo reale per creare un indice strutturato circa il contenuto. VideoLogger permette tre livelli di indicizzazione: in aggiunta alla potente analisi automatica del video, effettua l'indicizzazione automatica da fonti esterne - quali trascrizioni, dati GPS, analisi dei log da database – e permette annotazioni manuali che possono essere aggiunte in tempo reale o dopo il processo.

Un modulo di analisi del montaggio consente, inoltre, di valutare i contenuti relativi alle immagini in movimento delle sequenze. Infine, un componente dedicato alla speech recognition permette, tramite l'analisi della colonna sonora vocale, di rappresentare il video in base a ciò che in esso è espresso verbalmente.

Analisi temporale

La dimensione temporale di un video è un dato caratteristico di questo tipo di documento. L'analisi temporale di un video richiede la sua suddivisione in elementi base. Questa suddivisione può essere effettuata in quattro differenti livelli temporali:

- Frame level: ogni frame è trattato separatamente. Non ci sono analisi temporali in questo livello.
- Shot-level: è il livello chiave, un insieme di frame contigui acquisiti attraverso una registrazione continuata. La suddivisione del video in "shots" generalmente non si riferisce ad alcun'analisi semantica e viene utilizzata solo l'informazione temporale.

- Scene-level: Una scena è un insieme di scatti contigui aventi un significato semantico comune.
- Video-level: Il video viene trattato nella sua interezza.

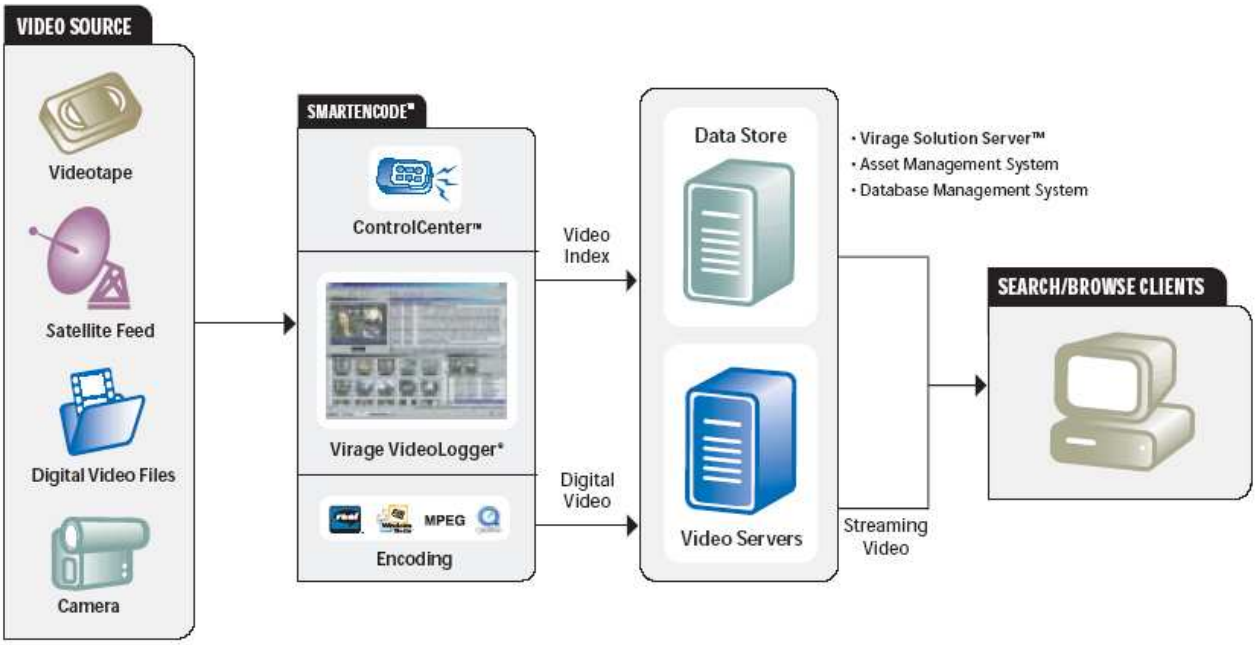

Figura 7: Dalla fonte video alla ricerca da parte dell'utente finale

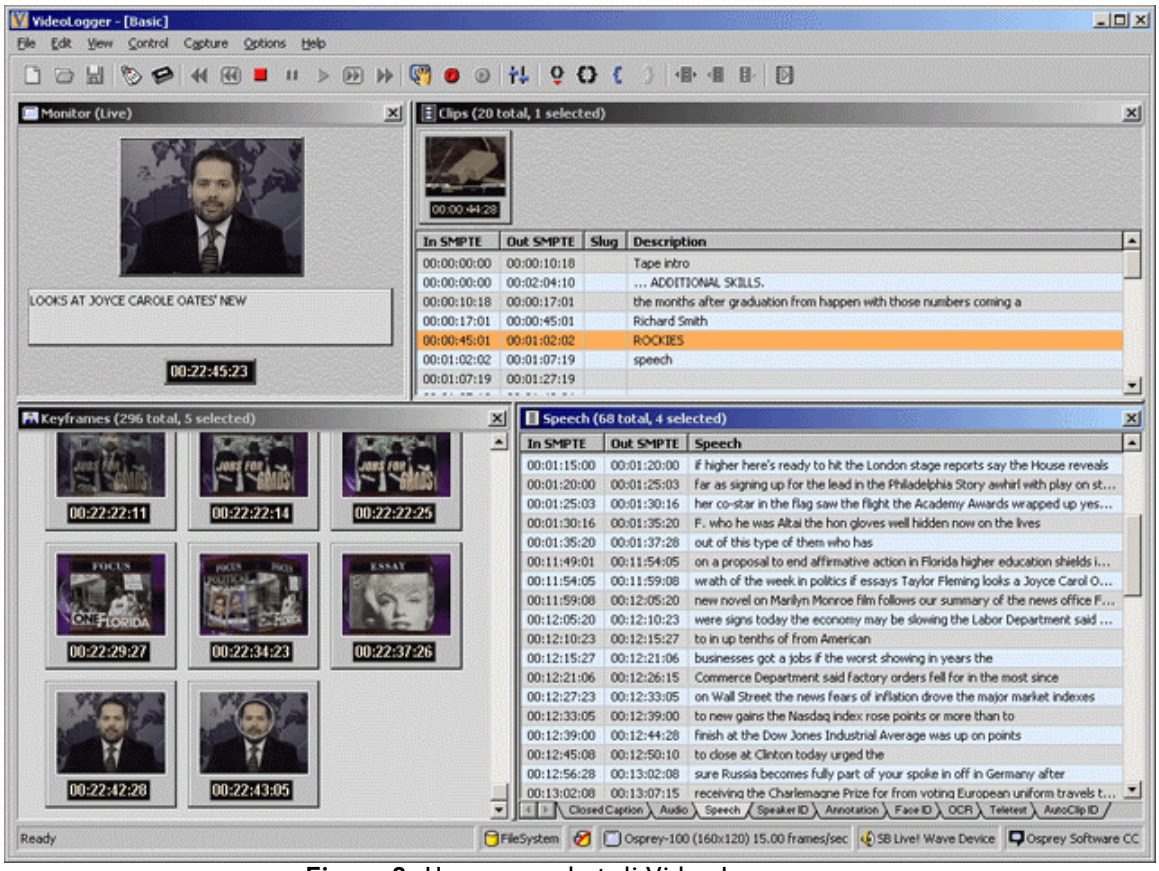

Figura 8: Uno screenshot di Video Logger

3.3 Applicazioni delle tecniche di Audio Retrieval

Tra i servizi di recupero dei file audio, quello che vuole essere al servizio dell'utenza più generale, è sicuramente **SoundFisher**, sistema per il trattamento dei documenti audio basato sulle loro caratteristiche sonore.

Figura 9: Una schermata di SoundFisher

Il sistema analizza automaticamente tutti i tipi di file audio, estrae da essi varie informazioni tecniche, e consente di collegare ai documenti delle parole chiave. Si possono, così, impostare ricerche complesse basate su diversi generi di dati.

Esso promette, inoltre, di poter ricercare brani sulla base delle loro qualità e dei loro contenuti propriamente sonori, nonché per similitudine a dati modelli, anche se questa caratteristica è ancora in una fase sperimentale.

CAPITOLO 4
Stato attuale delle metodologie di gestione dell'informazione multimediale

4.1 Le differenze tra CBIRS e CBVRS

Il sistema content-based di ricerca immagini (CBIRS - Content-based Image Retrieval Systems) sta iniziando ad espandersi sempre di più. I risultati stanno migliorando continuamente e i principi alla base della ricerca vantano una vasta gamma di diversità. I sistemi CBVRS (Content-based Video Retrieval Systems) sono, invece, meno comuni e, ad una prima occhiata, sembrano la naturale estensione del CBIRS. In questo capitolo ripercorriamo l'evoluzione fatta nello sviluppo del CBVRS e analizziamo la sua relazione con il CBIRS. Nel nostro percorso dimostreremo che attualmente il CBVRS non è un'evidente estensione del CBIRS.

I sistemi di Content-based Video Retrieval (CBVR) appaiono come un evoluzione naturale (o fusione) di Content-based Image Retrieval (CBIR) e Content-Based Audio Retrieval systems.

Tuttavia, ci sono diversi fattori che vengono ignorati quando ci si occupa delle immagini e che invece vanno presi in considerazione analizzando i video. Questi fattori sono collegati alle informazioni e alla dimensione temporali disponibili in un documento video. Se da un lato possono complicare il sistema d'interrogazione, allo stesso tempo possono aiutare nel caratterizzare le informazioni utili per l'interrogazione.

Le problematiche principali, trattando dati audiovisivi, sono quelle dell'analisi della singola inquadratura e quelle dell'analisi di più inquadrature inserire in una sequenza di montaggio. Vi è una grande diversità tra l'analisi dell'inquadratura che, a parte il fatto del movimento, si riporta a quella delle immagini fisse che sta a base del Visual Retrieval, e

l'analisi degli attacchi di montaggio che, per via del loro meccanismo di connessione e differenziazione del discorso visivo, pone problemi assolutamente rilevanti.

Il passo da sistemi di indicizzazione che operano su immagini a sistemi che operano su segmenti video è breve: se nel caso delle immagini applico algoritmi su due dimensioni spaziali il video introduce il vettore temporale, ovverosia dovrò operare su n fotogrammi di x per y dimensioni.

I Content Based Video Retrieval Systems (CBVRS) permettono di trovare sequenze video in un database. Operano di norma dividendo il video in segmenti (video segmentation), indicizzando questi segmenti e fornendo all'utente funzionalità di interrogazione del database generato (video querying).

La segmentazione di uno stream video, ovverosia rilevare i cambiamenti bruschi tra gruppi vicini di fotogrammi, è molto simile alla ricerca di immagini così come viene intesa da un CBIRS: si tratta di confrontare immagini. La ricerca di una sequenza in un database di video è analoga alla ricerca di un'immagine in un database di immagini.

Anche nell'ambito dei CBVRS l'utilizzo dei color histogram è stato fino a poco tempo fa l'approccio maggiormente seguito. Operando nel dominio dei pixel, cercando di rilevare i tagli (cut) presenti in uno stream video in base alla similitudine nella distribuzione dei colori nei fotogrammi, si sono raggiunti buoni risultati, specie abbinando le caratteristiche rilevate sotto questo profilo con quelle relative al movimento, alla forma e al contorno. Il problema principale è sempre stato quello relativo alle notevoli risorse computazionali necessarie a eseguire tali operazioni per ogni fotogramma disponibile.

Applicate alla segmentazione e indicizzazione di stream video le wavelet si sono rivelate essere ancora una volta uno strumento estremamente efficace, dato che consentono un'approssimazione molto buona al contenuto dell'immagine anche con solo pochi coefficienti: questa proprietà è di fondamentale importanza quando si ha a che fare con

compressione di tipo lossy[4], il che è quanto accade nella maggioranza dei casi nell'ambito della compressione video. I coefficienti risultato di una decomposizione tramite wavelet forniscono informazioni indipendenti dalla risoluzione originale dell'immagine, permettendo così confronti tra immagini, quindi anche video, di diversa risoluzione.

L'autore di ImgSeek, Ricardo Niederberger Cabral, aveva cominciato a sviluppare un CBVRS basato su wavelet prima di decidere di dedicarsi esclusivamente a un CBIRS, con l'intenzione di aggiungere funzionalità di recupero video in futuro.

Il CBVRS di Cabral, VideoQuery utilizza i wavelet transform nel seguente modo:

– in una prima fase i video vengono segmentati, ovverosia divisi in scene, in base ai cambiamenti bruschi rilevati tra fotogrammi vicini

– le scene vengono indicizzate in un database; per ogni scena vengono salvate informazioni relative al file di appartenenza, il fotogramma chiave che identifica questa scena, la wavelet signature relativa al fotogramma stesso con i relativi coefficienti e una miniatura dell'immagine

– il motore di interrogazione del database prevede che l'utente fornisca un'immagine al sistema, che questa subisca i medesimi processi di wavelet transform applicati in precedenza ai video indicizzati ed infine che i coefficienti propri della sua wavelet signature vengano confrontati con quelli delle altre nel database.

– il fotogramma la cui firma risulta essere più simile a quella della query viene restituito all'utente, insieme a informazioni relative al file cui appartiene, il numero di fotogramma all'interno del video e il coefficiente di similitudine in base alla metrica del sistema.

La traccia utilizzata per la ricerca in VideoQuery è quello del query-by-example, nel caso specifico limitato alle sole immagini statiche come esempio da fornire al sistema, ma è già pronto il codice per permettere all'utente di fornire come query al sistema uno stream MPEG2. Il sistema restituisce all'utente un'immagine, o meglio una sua anteprima, e

[4] La compressione di tipo lossy è una compressione con perdita di informazioni

successivamente è possibile effettuare il play del file video cui appartiene la signature che corrisponde maggiormente, a partire dal fotogramma relativo alla firma stessa.

4.2 MMIR nelle biblioteche digitali

Le Biblioteche Digitali, inizialmente nate per la gestione di documenti testuali, vengono sempre più utilizzate per l'archiviazione e la gestione di oggetti digitali multimediali. Le funzionalità offerte variano a seconda dell'ambito applicativo e possono spaziare dalla possibilità di gestire media diversi (immagini, audio, video, testi, ecc.), all'utilizzo di modelli di metadati specifici, alla possibilità di effettuare ricerche basate sul contenuto degli oggetti digitali o sulla loro struttura.
Esiste una nuova generazione di strumenti di information retrieval, strumenti che si pongono al servizio di una utenza dalle esigenze avanzate, che non può accontentarsi dei tradizionali tools di ricerca.
Che sia esclusivamente bibliografica, per monografie e/o articoli di periodici, cataloghi (anche nella forma di OPAC e MetaOPAC) e banche dati, oppure allargata a fonti ipertestuali (motori di ricerca), attualmente la ricerca riguarda, in massima parte, oggetti comunque testuali.
La novità dei Multimedia Information Retrieval systems sta nella possibilità di ricercare anche documenti sonori, filmati, immagini digitali, a partire non tanto dalla loro descrizione quanto dal contenuto, evidenziato anche dalla costruzione di metadati appositamente concepiti: uno spostamento non indifferente dalla forma esterna alla struttura dell'oggetto, che l'ambiente digitale ha reso indispensabile.
L'argomento conosce negli ultimi tempi un rinnovato interesse, probabilmente legato alle più recenti innovazioni tecnologiche dei programmi di elaborazione e gestione dei dati digitali multimediali, dopo una fioritura sorprendente ma precoce nei primi anni '90.

Attuali limiti

Il superamento della tradizionale mentalità descrittiva, connaturata in larga misura alla natura analogica dei supporti dei documenti sinora trattati dalla comunità bibliotecario-documentalista, deve avvenire nell'ottica più globale del superamento della "rappresentazione linguistica" del documento, nonché della mediazione "classificatoria" in qualche modo intrinseca alla professione stessa. Alle classiche interfacce dei database testuali, che consentono la ricerca in un indice composto esclusivamente di termini estratti dai documenti o inseriti in metadati testuali, devono succedere interfacce che permettano di formulare la query in diverse dimensioni, non solo tramite i termini ma anche attraverso le immagini e i suoni, rendendo in tal modo possibile la ricerca in indici composti da testi tratti dalle didascalie o dal parlato, da immagini chiave di una sequenza, da semplici figure, da melodie, da forme, colori e suoni. Nuove prospettive nel panorama della letteratura professionale, prospettiva che non mancherà di influenzare, negli anni futuri, anche le biblioteche italiane, uscendo dalla fase teorico-sperimentale attuale per approdare ad applicativi diffusi.

Conclusioni

Nel corso di queste pagine si è compreso come l'architettura e la logica dei database multimediali debbano essere *object-oriented*, basati quindi sul contenuto concreto dei documenti.
Tuttavia si è ancora lontani dal raggiungere un accordo sul modo migliore per realizzare tecnicamente e con efficacia questo sistema. Il processo di sviluppo dei nuovi database deve essere quindi sempre guidato dall'attenta comprensione della natura dei documenti e dei dati multimediali nonostante sia difficile ridurre ad una serie di termini controllati tutta la complessità di oggetti direttamente ripresi dal mondo reale.

Attualmente un buon livello di precisione e di richiamo nella ricerca è stato raggiunto tramite l'impiego complementare di tecniche e tecnologie basate sia sulla definizione dei soggetti, tramite termini che descrivono l'oggetto, sia sulla rappresentazione del contenuto, attraverso elementi multimediali.

Se già adesso, con tutti i limiti imposti da una tecnologia così recente, abbiamo la possibilità di recuperare un file audio semplicemente dialogando con il nostro computer, o recuperare un quadro disegnando il contorno o la forma di un oggetto presente in quell'opera d'arte, possiamo immaginare, dunque, in futuro, il recupero dei dati da noi desiderati solamente pensando ad un'immagine fissa nella nostra testa?

Bibliografia

MULTIMEDIA INFORMATION RETRIEVAL: Content-based information retrieval from large text and audio databases
di Peter Schauble
1997 – Boston, Kluwer Academic

MULTIMEDIA INFORMATION RETRIEVAL: Metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione
a cura di Roberto Raieli e Perla Innocenti
2004 – Roma, AIDA

Readings in INFORMATION RETRIEVAL
di Karen Sparck Jones and Peter Willet
1997 – San Francisco (Californai) - Morgan Kaufmann

HYPERTEXT – INFORMATION RETRIEVAL – MULTIMEDIA 97
di Norbert Fuhr / Gisbert Dittrich
1997

INFORMATION RETRIEVAL SYSTEMS (Theory and Implementation)
di Gerald Kowalski
1997 - Boston, Kluwer Academic

PRACTICAL DIGITAL LIBRARIES: Books, bytes and bucks
di M. Lesk
1997 - San Francisco (California), Morgan Kaufmann

DIGITAL LIBRARIES
di William Y. Arms
2000 - Cambridge, London, The MIT pres

VISUAL INFORMATION RETRIEVAL
di Alberto del Bimbo
1999 - San Francisco (California), Morgan Kaufmann