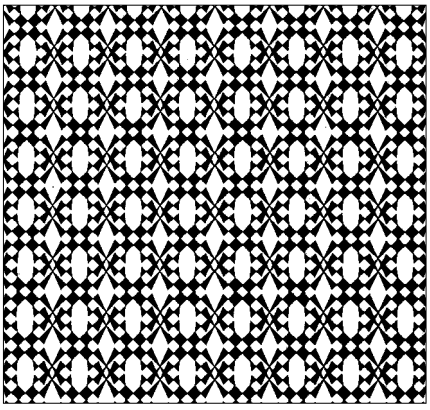
MultiMedia Information Retrieval

Un sistema avanzato per il reperimento dell'informazione multimediale

di Roberto Raieli

Si potrebbe interrogare un database bibliografico, seppure eccellente come Medline, tramite, ad esempio, l'immagine radiografica di un rene, se si stanno cercando articoli o altri scritti che discutano gli ultimi studi sui trapianti? Sarebbe impensabile secondo la logica del sistema, e di fatto impossibile.
Nei casi in cui si vogliano interrogare fonti di documenti testuali con mezzi non testuali tale confusione di "linguaggi" è sicuramente considerata paradossale, ma nei casi in cui s'interrogano tramite testo fonti documentarie visive o sonore lo scambio di linguaggi opposti viene invece considerato opportuno. Se non è possibile, però, ricercare e recuperare un documento scritto tramite mezzi di linguaggio visivi o sonori, allo stesso modo non dovrebbe essere considerato, con pochi dubbi, un metodo efficace recuperare documenti consistenti in suoni o figure attraverso l'uso di testi descrittivi. Piuttosto, ad esempio, dovrebbe apparire dispersivo cercare una fotografia di paesaggio, che rappresenti un tramonto con certe tinte, tramite una complicata descrizione "a parole" delle tonalità di colore desiderate, anziché sottoporre a un apposito sistema di ricerca un campione delle tinte stesse.
Lo stato attuale della tecnologia informatica, tra l'altro, fa da base e da spinta per lo sviluppo culturale di una società sempre più complessa ed esigente, la cui cultura estesa e onnicomprensiva è pienamente definibile "multimediale".
In base a tutto questo, allora, nell'odierna situazione culturale e tecnologica dovrebbero apparire evidenti i diversi limiti del continuare a operare nei termini di un generico Information Retrieval. Nella pratica tradizionale dell'Information Retrieval, infatti, ogni tipo di ricerca documentale è riportato alle condizioni di una ricerca tramite linguaggio testuale; è 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, o di metalinguaggio, più adatti alla sua natura propria.
Si potranno distinguere quindi, nella più generale metodologia del MultiMedia Information Retrieval:
– un sistema di Information Retrieval, basato su informazioni testuali per la ricerca di documenti testuali;
– un metodo di Visual Retrieval, secondo il quale i documenti visivi sono cercati e recuperati tramite dati visivi;
– un sistema di Audio Retrieval, nel quale l'informazione sonora è ricercata in misure di suoni;
– un Video Retrieval, dove per il recupero di documenti audiovisivi si utilizza il linguaggio audiovisivo.
In tale prospettiva, le diverse questioni sviluppate e risolte relativamente ai tradizionali sistemi e strumenti di Information Retrieval cambiano quasi completamente aspetto quando si parla di MultiMedia Information Retrieval.
La più importante differenza tra i sistemi di archiviazione e recupero di documenti testuali e quelli di documenti multimediali si evidenzia nel sistema di analisi ed estrazione degli elementi indicatori del contenuto del documento e dei descrittori specifici delle sue caratteristiche. A questo seguono le questioni che concernono l'impiego di tali dati di indicizzazione, sia per l'archiviazione dei documenti in una banca dati e per la creazione del relativo indice, sia, soprattutto, per la messa a punto dei metodi e dei sistemi di ricerca e recupero applicabili all'archivio.
Nei database multimediali risultano troppo riduttivi e poco efficaci 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 semplificativo e impreciso attribuire, dall'esterno, una descrizione testuale a contenuti che si fondano su un diverso regime di "senso".
Inoltre, se per i testi è adeguato anche il metodo dell'analizzarne il concetto, pur con i limiti posti dalla soggettività del documentalista, e dell'attribuire a questo un descrittore, non è invece egualmente efficace ripetere tale pratica per le immagini o gli audiovisivi, dacché, per prima cosa, i limiti soggettivi nel coglierne i concetti intimi sono maggiori. Non sempre, inoltre, il concetto interessa maggiormente del contenuto concreto di un documento multimediale, delle rappresentazioni in se stesse, dei tratti, delle forme, dei colori e dei suoni, cosa che, invece, avviene per gli scritti, dove il concetto espresso interessa spesso più della maniera di esprimerlo, della lingua e dello stile, a meno che non si tratti di testi letterari.

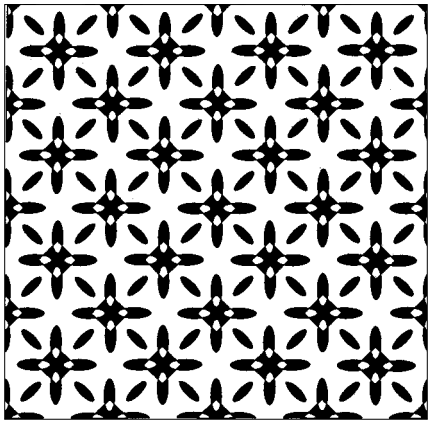
Term-Based e Content-Based Retrieval

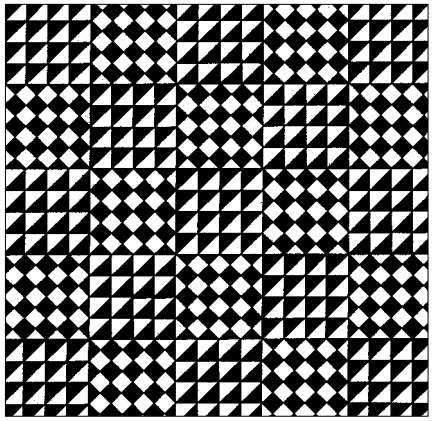
I sistemi di indicizzazione e ricerca tradizionali sono basati sui termini, sulla logica delle parole chiave, che ci si propone naturalmente, data la nostra abitudine a esprimere e rappresentare verbalmente e testualmente qualunque oggetto di conoscenza e informazione. I sistemi di recupero più innovativi richiedono, invece, una riflessione più avanzata, 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*.
In una logica di archiviazione e recupero *term-based* la *query* viene espressa testualmente, e va a incontrare i "surrogati testuali" dei documenti multimediali; surrogati che assumono la forma di termini di indicizzazione, di titoli o didascalie. A questi, nel database, è collegato il documento archiviato, dunque, recuperati i surrogati terminologici, essi trascinano automaticamente con loro la figura, il suono, o la relativa anteprima, come in una banca dati testuale le parole chiave fanno emergere il full-text o il suo abstract. Tutto viene allora tradotto "in parole": il documento, il suo contenuto, le chiavi d'accesso che lo identificano nella registrazione; quindi anche la formulazione e la struttura della *query* nella ricerca. Tale generalizzata 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 non debba essere costretta entro i limiti della lingua, ma 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.
Il sistema di archiviazione e recupero *content-based* si trova in una fase di sperimentazione piuttosto avanzata, benché 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 di surrogati linguistici, ma del contenuto "sostanziale" vero e proprio, composto di strutture e forme, suoni e colori.[1] 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 vincolante mediazione classificatoria.
Si deve pur sempre parlare di mediazione tra la fisicità originaria dell'oggetto – tela, marmo, pellicola, nastro magnetico – e la sua messa a disposizione informativa – ossia l'inevitabile trasformazione in documento digitale. Ma un primo grado di mediazione è il presupposto implicito di un sistema informativo in quanto tale, l'importante è che a questa mediazione di natura strutturale non se ne aggiungano altre, ad esempio quella di un sistema di classificazione. 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.

Il MMIR

MMIR è l'acronimo che indica l'insieme dei sistemi di MultiMedia Information Retrieval, non scindibile intreccio delle tecnologie del Visual Retrieval, dell'Audio Retrieval, del Video Retrieval e dei sistemi di Information Retrieval classici.
Tutto il complesso 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, nonché delle architetture necessarie nei nuovi grandi database multimediali. La necessità e l'efficacia di un complesso organico non escludono, però, che ognuno dei sistemi abbia un'individuabile specificità, sia dal punto di vista tecnico che metodologico.
Per ciò che riguarda l'intera compagine del MultiMedia Information Retrieval, un breve saggio di William Grosky, professore della Wayne State University di Detroit, condensa in un'efficace panoramica lo sviluppo delle banche dati e dei relativi principi e sistemi di gestione dell'informazione.[2] Lo studioso fa notare come, in particolare nell'ultimo decennio, il campo dei database non si sia arricchito solo di nuovi metodi e tecnologie per la gestione dei dati testuali, ma anche di importanti innovazioni che riguardano nuovi tipi di dati, come quelli visivi e audiovisivi.
Nei primi esperimenti sulla gestione di documenti multimediali nei database, gli studiosi svilupparono semplicemente una tecnica di descrizione basata sulla classica architettura relazionale, dove il documento e il suo contenuto erano rappresentati come una stringa di termini inserita all'interno di un sistema di faccette, classi e sottoclassi in relazione reciproca. Su tale architettura si definirono tutte le operazioni di indicizzazione, archiviazione e ricerca. L'esperienza di lavoro con sempre maggiori quantità di nuove tipologie di dati, però, mostrò presto come questo approccio avesse un'intrinseca debolezza: lo *scollamento* tra la vera natura dei documenti e la prospettiva solo terminologica in base alla quale, sia il sistema sia l'utente, venivano forzati a operare su di essi.
Non poteva essere efficiente la tecnica basata sulla messa a punto di surrogati descrittivi dei documenti multimediali, quindi i ricercatori cominciarono a interrogarsi su quale fosse effettivamente il tipo di informazione che poteva essere estratta da immagini, filmati o registrazioni sonore, e si chiesero come tali informazioni potessero essere rappresentate e organizzate per supportare richieste dimostratesi chiaramente orientate ai contenuti concreti. Si cominciò dunque a discutere sulle questioni relative a una indicizzazione multimediale, e ancora sulla natura delle interrogazioni multimediali e sulle tecniche di ottimizzazione delle ricerche. Si giunse così alla definizione del concetto di *multimedia query*, come qualcosa di ben differente e diversamente evoluto rispetto alla classica *query* terminologica, da impostare con riguardo alle caratteristiche specifiche dei materiali archiviati nel sistema interrogato.
Tuttavia lo stato dell'arte delude ancora Grosky. È stato compreso che l'architettura e la logica dei database multimediali devono essere *object-oriented*, basandosi sul contenuto concreto dei documenti, ma si è ancora lontani dal raggiungere un accordo sul modo migliore per realizzarle tecnicamente e con efficacia. Il processo di sviluppo dei nuovi database, in ogni caso, deve essere sempre guidato dall'attenta comprensione della natura dei documenti e dei dati multimediali, i quali rappresentano oggetti alfanumerici, testuali, grafici, figurativi,





animati, sonori e audiovisivi. È sterile tentare di ridurre a una serie di termini controllati tutta la complessità di oggetti direttamente ripresi dal mondo reale, tradotti in documenti che sono rappresentativi di cose e di situazioni, la cui natura propria è spesso posta come obiettivo specifico della *query*.

Rappresentazione di documenti multimediali

Riferendoci ancora allo scritto di Grosky, possiamo fare il punto anche su questioni di natura più tecnica.[3]
Negli ultimi anni, si è detto, ricercatori e tecnici hanno ben acquisito la coscienza della natura dei dati e dei documenti multimediali. Essi consistono di "oggetti" di ogni tipo, che per via della loro complessa struttura non sono efficacemente rappresentabili con la logica dei sistemi *term-based*. Questi oggetti, definiti da Grosky "oggetti multimediali", sono ripresi dal mondo reale, estratti anche con strumenti di registrazione diretta – fotografica o sonora, magnetica o digitale – e tradotti in rappresentazioni dell'"oggetto reale". Con le tecniche attuali – ad esempio nel campo dell'*image processing* o della *speech recognition* – i sistemi elettronici riescono a identificare propriamente gli oggetti reali, estraendo, magari semiautomaticamente, determinate informazioni dal corrispondente oggetto multimediale. Tali informazioni sono dette *features* e sono contenute in quello che è definito il *data model* dell'oggetto multimediale, insieme di dati meno complesso e voluminoso di quello dell'oggetto che rappresenta.
Dagli oggetti reali, dunque, appartenenti al mondo dell'esperienza quotidiana, vengono ripresi gli oggetti multimediali, documenti di vario genere di questa esperienza trattabili dai sistemi informativi, e da questi ultimi sono poi estratti i *data model*, insiemi di dati digitali identificativi del documento che possono essere trattati dai sistemi informativi più avanzati. L'intero processo, che dalle cose della realtà quotidiana porta al loro trattamento documentario tramite i *data model*, è chiamato da Grosky "multimedia data modelling", e si propone come il nucleo teorico e tecnico del MultiMedia Information Retrieval.
Il *data model* rappresenta quindi le proprietà delle cose, le loro relazioni e le operazioni definite su di esse, e tali concetti astratti, nondimeno, sono in esso tradotti in dati digitali, fisicamente situabili nel sistema del database.[4] Così, attraverso la mediazione del *data model*, le *queries* e altre operazioni sugli oggetti reali possono essere trasformate in operazioni sulle rappresentazioni astratte di tali oggetti, che sono a loro volta trasformate in operazioni sui dati digitali che traducono le rappresentazioni astratte nel linguaggio del sistema elettronico.
Grosky propone un elenco dei diversi generi di informazioni che possono essere catturate in un *data model* multimediale:

– The detailed structure of the various multimedia objects
– Structure-dependent operations on multimedia objects
– Properties of multimedia objects
– Relationships between multimedia objects and real-world objects
– Portions of multimedia objects with representation relationships with real-world objects, the representation relationships themselves, and the methods used to determine them
– Properties, relationships, and operations on real-world objects.[5]

Per le immagini, ad esempio, i *data model* possono contenere dati come il formato, la risoluzione, il numero di pixel, i valori dei colori, i valori caratteristici della struttura, eccetera. È essenziale allora sviluppare tali modelli rappresentativi, in quanto praticamente tutte le operazioni relative agli oggetti del database multimediale si basano su di essi. Non mancano ricerche ed esperimenti, ricorda lo studioso, per la messa a punto di sempre più perfezionati strumenti che siano in grado di analizzare l'oggetto multimediale, per ricercare ed estrarre le sue caratteristiche rilevanti.

La query *multimediale*

Alla trattazione del *data modelling*, e dei processi di creazione dell'archivio digitale, Grosky fa seguire la trattazione del "query processing".[6]
La *query* in un database multimediale, ribadisce lo studioso, è una procedura ben differente da quella operante in una banca dati testuale. Il *browsing*, intanto, assume una particolare importanza in questo ambito; le stringhe di interrogazione, poi, possono contenere direttamente valori del *data model* o oggetti multimediali di esempio. Inoltre, i risultati di tali *query* non sono basati su un esatto *matching* tra i dati inviati e quelli ritrovati, ma solo su certi gradi di similitudine.
Nel procedimento più tipico, l'utente avvia la ricerca con la preliminare selezione di un archivio, e di una sua parte, tramite l'uso di stringhe di testi appropriati. In questo caso, con valore propedeutico alla ricerca più propriamente multimediale, i termini, i titoli e i nomi possono essere il metodo più adatto e veloce per una prima riduzione della mole di tutto il materiale potenzialmente interrogabile.
Si procede, quindi, con un *browsing* d'esplorazione, tramite cui si possono inviare al sistema delle richieste semplicemente selezionando con il mouse un oggetto, o una sua parte, sul display. Così, muovendosi tra tanti oggetti che assomigliano a quello cercato, o che con esso sono in relazione, si possono inviare all'elaboratore diversi *data model*, contenenti i dati caratteristici ➤

che dovranno essere rintracciati negli oggetti dell'archivio per estrarli come risultati della *query*.
Questo è il metodo più pratico per stringere la ricerca attorno al documento che si desidera. Però, nei database più avanzati, oltre a inviare i parametri di *query* tramite la selezione di un oggetto già posseduto dal sistema, si possono compilare delle griglie specificando liberamente tali parametri; oppure il sistema può dare la possibilità di inserire dall'esterno dei modelli di esempio, che verranno analizzati derivandone il *data model*.
Con tale metodologia il database potrà, ad esempio, essere in grado di soddisfare richieste del tipo:
– con l'immagine di un particolare di un'opera di Van Gogh, trovare l'opera intera e altre opere simili;
– attraverso un identikit, trovare le fotografie di un criminale;
– tramite la fotografia di una ballerina, rintracciare diversi video dove danza;
– con le note principali di un motivo, trovare diverse composizioni che lo hanno sviluppato.
I risultati di questo genere di *query*, si è detto, non sono basati su esatte corrispondenze tra i parametri della stringa di richiesta e quelli dell'oggetto restituito come risultato, ma solo su un certo grado di similitudine e di vicinanza tra i valori dei rispettivi dati. È piuttosto raro che due *data model* rappresentanti lo stesso oggetto corrispondano perfettamente; è più facile, cioè, che anche due fotografie con la stessa inquadratura della stessa persona, ad esempio, abbiano valori di colori e forme solamente vicini.
La misura della similitudine tra due oggetti multimediali è solitamente indicata al sistema come variabile tra 0 (completamente differente) e 1 (esattamente corrispondente). Teoricamente il risultato di una *query* è sempre costituito dall'intero insieme degli oggetti archiviati nel database, in rank da 1 a 0 rispetto al modello della richiesta. Di fatto, ovviamente, è sempre definito un valore minimo, vicino a 1, per cui tutti gli oggetti il cui valore di similitudine al campione stimato dal sistema non superi tale soglia non vengono estratti.
Alla base del *query processing* deve esserci, ovviamente, un metodo di indicizzazione appropriato alle caratteristiche del materiale multimediale.
Il concetto di indicizzazione va inteso però, in questo contesto, in senso più largo rispetto alla sua accezione comune. Esso, nell'ambito dei documenti multimediali, va riferito a una tecnica di creazione dell'indice del database tramite l'estrazione da documenti non testuali di elementi che non sono termini né sono traducibili in termini. L'indice viene creato, quindi, impostando come collegamenti di accesso ai documenti i dati costitutivi del loro stesso contenuto multimediale: forme, colori, suoni, e altri simili elementi.
La tecnica generale indicata da Grosky prevede l'estrazione dall'oggetto multimediale di *n* caratteristiche, valutate e definite numericamente, e rappresentate in una struttura a *n* dimensioni. L'analisi ed estrazione di tali dati numerici può essere effettuata manualmente *computer-assisted,* oppure dal sistema in modo interamente automatico. Le *n* caratteristiche possono, poi, essere valutate e registrate indipendentemente l'una dall'altra, oppure come parti di un inscindibile complesso nella struttura del *data model*. L'indice sarà infine utilizzato per valutare la prossimità dei valori numerici, e quindi stimare e assegnare un grado di similitudine tra queste *n* caratteristiche e le *n* caratteristiche della stringa di ricerca.[7]
Il processo della *query* multimediale è dunque vario e complesso, e una continua interazione con l'utente è ancora più indispensabile che nelle *query* classiche per il suo successo; ma i sistemi più all'avanguardia dovranno consentire di svolgere le diverse operazioni di costruzione, gestione e consultazione del database con la maggiore facilità possibile dinanzi al risultato più completo.
Dall'insieme del MultiMedia Information Retrieval si enucleano, comunque, alcuni aspetti di Retrieval particolari, poiché nell'attuale circuito dell'informazione non circolano solo documenti multimediali *in toto*, ma fanno parte di esso anche documenti specificamente visivi, audiovisivi o sonori. Conviene accennare quindi separatamente al Visual Retrieval, al Video *retrieval* e all'Audio Retrieval.

Visual Retrieval

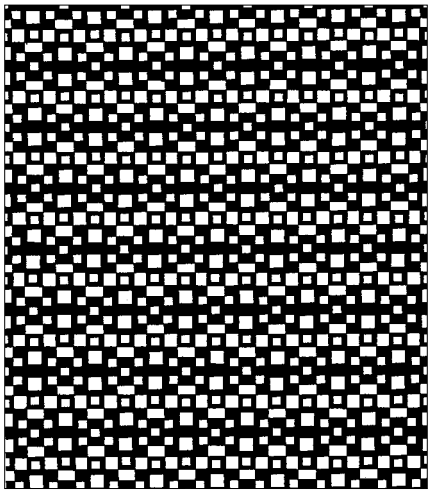
Discussi gli orientamenti complessivi del Content-Based Retrieval e dell'indicizzazione del materiale multimediale, può essere facilmente intuito il loro rapporto con le problematiche dei sistemi applicati ai documenti visivi.
Per attuare l'indicizzazione e le successive operazioni di recupero dei documenti visivi, e in particolare per il trattamento automatico dei valori relativi ai loro elementi e proprietà, è fondamentale che l'operatore umano, così come il sistema elettronico, siano in grado di produrre un certo livello di astrazione del documento. Un'astrazione che, ovviamente, non ha nulla a che vedere con l'astrazione concettuale; un'astrazione che significa, piuttosto, estrazione di alcuni elementi del contenuto concreto dell'immagine. Le varie operazioni, infatti, potranno meglio avvenire trattando tale forma astratta, distanziata dal documento immediatamente e indistintamente preso, base del *data model* come lo ha indicato William Grosky, la quale sarà il risultato di una serie di operazioni di identificazione, isolamento, valuta-

zione ed estrazione dei valori rappresentativi delle immagini digitali.
Nei sistemi specializzati di Visual Retrieval sono possibili, in genere, cinque differenti 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.
Tali modi di trattamento dei dati, inoltre, possono costituire la struttura del sistema, presi singolarmente o in combinazione tra loro.
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.[8]

Il processo della query *visiva*

Nel fare il punto sullo stato generale delle tecnologie e delle possibilità del Visual retrieval, può venirci incontro uno scritto di Alberto Del Bimbo, professore dell'Università di Firenze, relativo al "visual browsing, querying and Retrieval" nei database di immagini e video.[9]
Lo studioso scrive che le applicazioni di questo sistema sono studiate e messe in opera sia in campi di utilizzo specifico, sia nel caso di attività generiche di trattamento delle immagini. Tra le aree più importanti, è quella della ricerca biomedica ad aver coinvolto la maggior parte degli interessi e degli studi; molto ampie sono, poi, le applicazioni nel campo delle scienze della Terra e dell'informazione geografica; un settore in considerevole sviluppo, inoltre, è quello della documentazione delle arti visive. Altri importanti ambiti di applicazione sono, successivamente, il disegno ingegneristico e architettonico, le banche di immagini scientifiche, gli archivi fotografici delle forze di polizia e il design di moda. Infine, anche per ambiti molto generici sono stati messi a punto sistemi di ricerca visiva, magari con supporto terminologico, che consentono di rintracciare qualunque tipo di documento figurativo; seppure senza troppa raffinatezza, alcuni sistemi sono utilizzabili, inoltre, sullo spazio del web.
Del Bimbo riepiloga poi, e tratta in generale, le diverse questioni preminenti dello specifico del Visual Retrieval.[10]
Anzitutto si accenna alla particolare struttura dei documenti visivi, e viene ribadita la differenza – che oramai è ben chiara – tra i contenuti figurativi e quelli testuali, e quindi la necessità di un appropriato trattamento *content-based*, che consideri le proprietà prettamente visive delle immagini.
Sono discussi, quindi, i metodi di descrizione del contenuto e di indicizzazione delle immagini. In specifico, lo studioso divide i livelli di astrazione delle caratteristiche visive in proprietà "globali" e proprietà "locali": le proprietà globali vengono indicate nel "colore" e nella "struttura"; le proprietà locali riguardano la "forma" e le "relazioni spaziali".[11]
Passando alla rassegna delle diverse soluzioni per l'interrogazione della banca dati, il tradizionale sistema *term-based* si può considerare inadatto per esprimere e centrare propriamente il contenuto dell'immagine. Esso però mantiene comunque un valore propedeutico, cioè una precisa utilità nello sfoltire il database di ampie porzioni dell'archivio che non si rivelerebbero di alcun interesse in base all'obiettivo della ricerca.
La forma di interrogazione visiva più adeguata è ovviamente quella *content-based*, che esplora direttamente la natura figurativa dell'immagine, tanto attraverso l'impiego di modelli e campioni quanto ➤

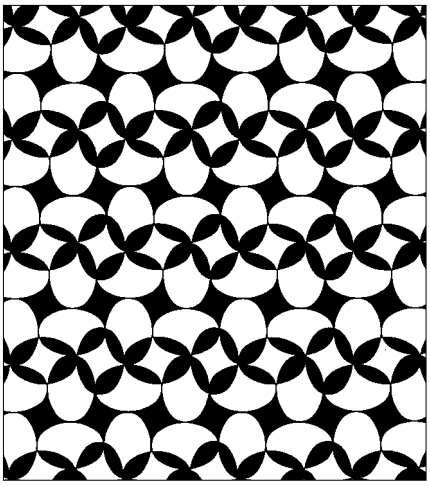
di griglie di valori, in base ai quali il sistema cercherà documenti di simile contenuto.
In tal modo le stringhe della *query* visiva useranno elementi realmente appropriati, ad esempio:
– immagini intere (per *query* relative alla struttura o alla forma);
– dettagli e icone (nelle *query* basate sulle relazioni spaziali);
– bozze esemplificative (per *query* sulle forme);
– griglie di valori (nelle *query* basate sui colori).
Come per la *query* multimediale in generale, diventa rilevante il fatto che le espressioni di ricerca possono essere solo approssimazioni, formulazioni parziali, o individuali, dell'effettivo contenuto visivo che l'utente desidera recuperare. Questo regime di imprecisione della *query* viene però incontro al fattore tecnico dell'approssimazione della modalità di confronto dei dati, in quanto non vi sarà mai un *matching* esatto tra il *data model* inviato per la ricerca e quello archiviato. Dunque, dovendo il recupero avvenire comunque per similitudine, da un punto di vista tecnico non ci sarà necessità di essere troppo rigorosi nel comporre la stringa di *query*.
Un problema critico, piuttosto, è rappresentato dalla differente valutazione e riconoscimento di forme, colori e strutture, che possono dare elaboratore ed essere umano. Nella ricerca per forme, ad esempio, possono entrare in considerazione 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. Nel programmare il sistema di recupero, allora, bisognerà accordare il metodo di calcolo geometrico



e matematico usato dall'elaboratore per stimare forme, strutture e colori, con la ben più sfumata sensibilità umana nel vedere le immagini.
Il punto più originale della breve trattazione di Del Bimbo è quello relativo alle fasi del *browsing*, della visualizzazione e, soprattutto, alla generale interazione dell'utente con l'intera ricerca. Il processo di ricerca e di reperimento delle immagini, per sua natura, deve basarsi sull'interattività più di quello di una tradizionale ricerca di documenti testuali; il giudizio dell'utente tratto dal *browsing* e dalla visione dei campioni è ancora più determinante per la correzione e il perfezionamento della *query*. Infatti:

> Results of an early query are used to issue a more fitting query, and editing facilities are used to modify image content to closely fit the desidered image appearance.[12]

Si può iniziare il processo di ricerca con una selezione terminologica di massima di una parte dell'archivio, e visualizzare una certa quantità di immagini rappresentative di un insieme. Si può quindi selezionare una figura d'esempio e inviarla come stringa di *query*. Oppure, nei sistemi più avanzati, tale prima stringa può essere costituita dal disegno di un modello sulla base della forma o del colore, come anche dall'immissione di un'immagine campione dall'esterno.
Successivamente, valutando le immagini ottenute come primi risultati, si possono cambiare i valori dei parametri di ricerca e rilanciare la *query*. Oppure, ad esempio, si può selezionare un'immagine tra quelle recuperate, estrarne e cambiarne la forma, il colore o la struttura, e rilanciare in questo modo una nuova *query* diversamente definita; così come, in alcuni sistemi, si possono combinare le figure di due immagini distinte. Tutte queste fasi, ovviamente, possono essere ripetute, in ordini e modi diversi, fino a ottenere un risultato soddisfacente.
Deve essere considerata anche la possibilità di continuare a usare parametri testuali per perfezionare la ricerca, senza limitarsi alla fase preliminare, anche dopo aver effettuato le prime operazioni di *query* per colore o forma. 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.



Applicazioni pratiche

Uno dei più avanzati gradi di progresso delle attività europee in direzione del Visual Retrieval, è rappresentato dalle sperimentazioni dell'Università di Brema.[13] Nei laboratori dell'università è stato messo a punto, in collaborazione con l'IBM, il sistema ImageMiner, il quale è in grado di analizzare automaticamente le immagini producendo due ordini di indicizzazione del loro contenuto. Un modulo, interpretando le caratteristiche visive, le riferisce poi a dei termini, creando un vero e proprio tesauro terminologico; un altro modulo, invece, 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.[14]
Un altro importante sistema di indicizzazione e recupero di immagini è VIPER (Visual Information Processing for Enhanced Retrieval), realizzato dal Computer Vision Group dell'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*.[15]
Molto più ricco e vario di quello europeo è il panorama delle esperienze e delle realizzazioni negli Stati Uniti. Il più antico dei sistemi di Visual Retrieval è QBIC (Query By Image Content), prodotto dall'IBM, messo a punto già dalla fine degli anni Ottanta; ovviamente si è evoluto con i tempi, e a tutt'oggi rappresenta uno dei sistemi più all'avanguardia. QBIC si applica ai database di immagini e a quelli di video, 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, come combinarle con quelle di altri complessi.[16]
Altro avanzato programma di Visual Retrieval è quello realizzato dalla Columbia University di New York, denominato Columbia's Content-Based Visual Query Project. Il programma è diviso in diverse sezioni, ognuna delle quali è predisposta per rispondere a necessità di ricerca specifiche. Il modulo denominato VisualSEEk è quello principale, e 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.[17]

Video Retrieval

I database di audiovisivi sono oggi sempre più utili e diffusi, in ogni campo di possibile applicazione: televisione e cinema come studi ingegneristici o aerospaziali; quindi anche la questione di una metodologia di ricerca ottimale per i documenti video cresce di importanza. Ma la natura del video, consistente in un flusso continuo di immagini, rende piuttosto complessa la problematica della rappresentazione e della ricerca del suo contenuto.
Per presentare lo stato dell'arte del Video Retrieval, possiamo fare riferimento a un articolo scritto da Ruud Bolle, Boon-Lock Yeo e Minerva Yeung, ricercatori dell'IBM e dell'INTEL.[18] Tutto parte, secondo gli autori, dalla sempre maggiore tendenza alla realizzazione dei documenti video in formato digitale – sviluppata anche dalla recente diffusione dei dvd – che rende possibile il *video data processing*, quindi l'estrazione di informazioni digitali per la strutturazione del database e 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".[19] 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.
Nella *video data abstraction* l'ideale sarebbe utilizzare un sistema di riconoscimento ed estrazione automatico dei dati caratteristici del video, ma la tecnica attuale consente solo sistemi impostati manualmente e *computer-assisted*. Fondamentale però per i tre studiosi è che, al di là del metodo, 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 *data model*, 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*.[20]

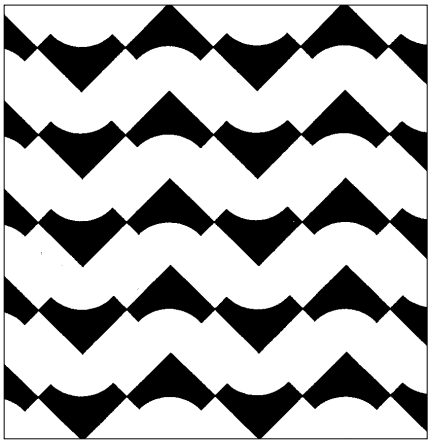
Modalità di video query

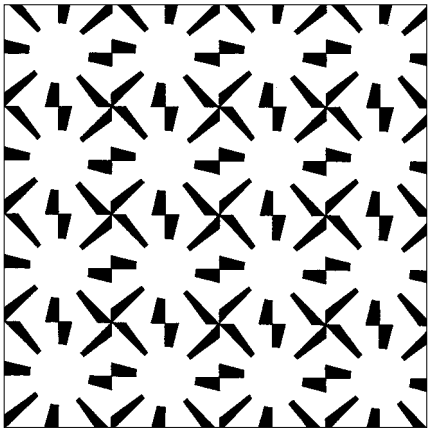
Bolle, Yeo e Yeung descrivono il processo di ricerca come una serie di stadi concatenati, i primi dei quali agiscono da filtri per i successivi, riducendo sempre più la mole dei documenti fino a individuare quelli che soddisfano le necessità informative dell'utente. L'utente potrà dunque rivolgersi al 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.[21]
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 grandemente 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, breve e centrata, 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 dinamico-espressivi dei filmati.
Un terzo fondamentale stadio necessita, allora, per il completamento della ricerca, e può essere individuato nel *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 documentalistica, 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*. Bolle, Yeo e Yeung differenziano tra la problematica dell'analisi della singola inquadratura, e quella dell'analisi di più inquadrature inserite in una sequenza di montaggio. Vi è una grande diversità tra l'analisi dell'inquadratura che, a parte il fattore del movimento, si riporta a quella delle immagini fisse che sta alla 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 particolari.[22]
Dal punto di vista tecnico, i tre ricercatori pongono lo sviluppo di potenti algoritmi per la *between-shot 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.*[23]
Un'ulteriore 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 organicità, e i sistemi di Video Retrieval non devono trascurarli. Oltre ciò che riguarda il contesto della ricerca nei database di video, però, la problematica del trattamento dell'informazione sonora rappresenta una questione autonoma, cui accenneremo in specifico.

Applicazioni pratiche

Un sistema per il trattamento e la ricerca di documenti video abbastanza diffuso in commercio è quello prodotto dall'azienda statunitense Virage. Questo programma consente di indicizzare i video, e quindi di recuperarli, in base tanto alle caratteristiche visive quanto a quelle audio.
Non manca 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. Un modulo di analisi del montaggio consente, inoltre, di valutare i contenuti relativi alle immagini in movimento delle se-





quenze. 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.[24]

Audio Retrieval

Odiernamente, anche i database di documenti sonori stanno conoscendo un ampio sviluppo in vari campi, e in particolare in quello della musica, dove il miglioramento delle caratteristiche e della resa dei file audio – a cui non è indifferente la messa a punto del formato MP3 –, contribuisce a rilanciare l'interesse per le questioni legate al trattamento dei documenti sonori.
Il metodo di Audio Retrieval che nella pratica sembra aver dato i risultati migliori è basato sulla ricerca tramite *browsing*, 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. Una discussione esemplificativa può essere tratta da un articolo di un gruppo di ricercatori della Nippon Telegraph and Telephone Corporation – Ryuji Kubozono, Kazuhiro Gomi, Seishi Kinohara, Michihiro Inagaki e Yumiko Matsuura –, i quali presentano un progetto nel campo dei sistemi di fornitura di prodotti on-demand.[25]
Gli autori analizzano un sistema nel quale sintesi chiave di brani sonori, definite *audio key-information*, ascoltabili in carrellata, sono trasmesse contemporaneamente a 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 – per non 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.[26]
Descrivendo lo schema del sistema proposto, i cinque tecnici stabiliscono due fasi:
– la prima è quella della "synchronous provision of index and key-information", 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 cat- ➤

turarli 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* 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.[27]
Il funzionamento del sistema viene chiarito dai tecnici discutendo la sperimentazione di un prototipo. Devono essere messi a punto macchine server e terminali utenti molto sviluppati, tanto che sia la *browse-search* 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 *memory-card* 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.[28]
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.
In direzione di 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.

Applicazioni pratiche

Altri generi di sistemi sono in corso di sperimentazione. Tra quelli che vogliono essere al servizio dell'utenza più generale, rilevante è l'esempio di SoundFischer, prodotto dalla statunitense Muscle Fish LLC, sistema per il trattamento dei documenti audio basato sulle loro caratteristiche sonore.
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; ma il modulo specifico per tale recupero non è ancora uscito dalla fase sperimentale.[29]

I motori di ricerca

Al di là delle applicazioni del MultiMedia Information Retrieval studiate per il ristretto ma sicuro ambito delle banche dati, anche i motori di ricerca di rete hanno ritenuto di doversi meglio adeguare alla logica dell'informazione multimediale. Difatti, accanto alla loro tradizionale finestra di ricerca basata sulle informazioni testuali, sono state inserite nuove finestre dedicate alla ricerca di immagini, suoni o video, nel tentativo di almeno avvicinarsi ad una prospettiva di reperimento *content-based*.
La struttura di *query* è, in genere, realizzata secondo la classica architettura relazionale dei motori di ricerca e delle guide di rete. Accanto alla riga per l'inserimento di parole chiave, le schermate mostrano elenchi gerarchici i cui singoli punti portano a successivi elenchi gerarchici, fino all'ultima schermata con i link ai siti della sottoclasse finale. La novità sta però nel fatto che le strutture di ricerca delle risorse specifiche – musicali, video, fotografiche – sono state realizzate *ad hoc*, dedicandone una a ogni singola categoria di materiale, non confondendole con quelle dei siti più generali, che interessano principalmente in base all'informazione testuale contenuta. Inoltre, sono in fase di sperimentazione nuove modalità, basate sulla ricerca intuitiva in relazione alle caratteristiche proprie del contenuto dei diversi media.
Tra i sistemi più evoluti è quello sperimentato da Yahoo Image Surfer. Image Surfer è specializzato nella ricerca di immagini, che, pur

inquadrata in una logica di relazioni gerarchiche, procede solo in parte in base ai nomi delle classi, a cui vengono affiancate, piuttosto, immagini esempio relative a ciò che le classi comprendono; è possibile procedere fino in fondo nella ricerca anche solo visionando e seguendo intuitivamente queste ultime.[30]
Altri importanti motori di ricerca, che hanno sviluppato apposite sezioni per il recupero sia di immagini, sia di suoni sia di video, sono ad esempio i noti AltaVista e Infoseek.[31]

Conclusioni

Il continuo miglioramento, insieme alla sempre maggiore "amichevolezza" nei confronti dell'utente e al progressivo abbassamento dei costi, sta facendo delle nuove tecnologie di trattamento dei documenti multimediali, non solo in campi di utilizzo tecnico e professionale, il mezzo di espansione e sviluppo della comunicazione multimediale in tutta la comunità mondiale.
Siamo di fronte a sistemi veramente innovativi, sempre più specializzati in un efficiente trattamento dell'informazione digitale, basati su una tecnologia di archiviazione e recupero che tratta direttamente il contenuto dei documenti, definita per questo *content-based*, in opposizione ai tradizionali sistemi di indicizzazione e ricerca basati su termini descrittori di tale contenuto, detti *term-based*. 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.
Nel concludere, comunque, si deve chiarire che un buon livello di precisione nel reperimento dei documenti multimediali si può raggiungere solo utilizzando in combinazione, e mutua integrazione, tecniche e tecnologie di ricerca basate tanto sulla definizione dei concetti, tramite termini controllati, quanto sulla rappresentazione del contenuto, attraverso elementi multimediali.
I due sistemi possono infatti essere armonizzati, non essendoci motivo, nella prospettiva dell'innovazione, di non considerare la validità che il sistema tradizionale continua a mantenere in molte occasioni. L'interrogazione *term-based* può, per prima cosa, essere un ottimo metodo preliminare per selezionare una parte della grande quantità di documenti di un archivio, e per centrare la ricerca in base a dati quali gli ambiti propri delle informazioni, le tipologie, le classi, i titoli, gli autori. Successivamente, essa può costituire un sistema finale di ripulitura dall'inevitabile "rumore" specifico di un'interrogazione *content-based*, precisando quello che il sistema non è in grado di avvertire nell'analisi diretta delle caratteristiche rappresentative del documento.
Soprattutto, però, i due procedimenti possono operare in costante interazione, in un unico sistema e con un'unica schermata di ricerca, nella composizione di una formula di *query* che combinando immagini e testi, suoni e testi, o ancora tutti questi insieme, possa servire per la ricerca di documenti molto complessi, il cui contenuto informativo si estende a tutti i livelli di senso e significato, dove anche le definizioni concettuali hanno un'importanza, soprattutto per soddisfare esigenze di ricerca molto sofisticate. ■

Bibliografia e risorse di rete

BEARMAN, DAVID – TRANT, JENNIFER (eds.), *Museum Interactive MultiMedia 1997: cultural heritage systems design and interfaces. Selected papers from ICHIM '97*, Pittsburgh, Archives and Museum Informatics, 1997.

BESSER, HOWARD, *Image database resources*, 1996, <http://sunsite.berkeley.edu/Imaging/Databases>.

BOLLE, RUUD M. et al., *Video query: research directions,* "IBM Journal of Research and Development", 42 (1995), 2, p. 233-252.

CASTELLI, VITTORIO et al., *Progressive search and retrieval in large image archives*, "IBM Journal of Research and Development", 42 (1998), 2, p. 253-268.

COMPUTER VISION GROUP, *Links to other image database systems*, 2000, <http://cui.unige.ch/~viper/other_systems.html>.

VIPER, *Visual information processing for enhanced retrieval*, 2000, <http://cui.unige.ch/~viper>.

DEL BIMBO, ALBERTO (ed.), *Image and video databases: visual browsing, querying and retrieval*, "Journal of Visual Languages and Computing", 7 (1996), 4 (speciale).

ENSER, PETER G. B., *Pictorial information retrieval. Progress in documentation*, "Journal of Documentation", 51 (1995), 2, p. 126-170.

GROSKY, WILLIAM I., *Managing multimedia information in database systems*, "Communications of the ACM", 40 (1997), 12, p. 73-80.

HEIDORN, BRYAN P. – SANDORE, BETH (eds.), *Digital image access & retrieval. Papers presented at the 1996 clinic on library applications of data processing*, Urbana-Champaign IL, University of Illinois, 1997.

IBM, *QBIC. Query by image content*, <http://wwwqbic.almaden.ibm.com>.

KUBOZONO, RYUJI et al., *Browse search using audio key-information for multimedia on-demand systems*, "IEEE Transactions on Consumer Electronics", 42 (1996), 4, p. 900-906.

MAYBURY, MARK T. (ed.), *Intelligent multimedia information retrieval*, Cambridge, MIT Press, 1997.

NOACK, MANFRED, *Image mining. Stand der Entwicklung auf dem Gebiet von Image-Retrieval-Systemen*, "NFD Information Wissenschaft und Praxis", 49 (1998), 2, p. 73-76.

QI, HAIRONG – SNYDER, WESLEY E., *Content-Based Image Retrieval in picture archiving and communications systems*, "Journal of Digital Imaging", 12 (1999), 2 suppl., p. 81-83.

RORVIG, MARK E., *A method for automatically abstracting visual documents*, "Journal of the American ➤

Society for Information Science", 44 (1993), 1, p. 40-56.
SOUNDFISHER, 1999, <http://www.soundfisher.com>.
SVENONIUS, ELAINE, *Access to nonbook materials: the limits of subject indexing for visual and aural languages*, "Journal of the American Society for Information Science", 45 (1994), 8, p. 600-606.
UNIVERSITÄT BREMEN, *ImageMiner*, 1998, <http://www.tzi.uni-bremen.de/BV/ImageMiner>.
UNIVERSITY OF NEW YORK, COLUMBIA, *Columbia's content-based visual query project*, <http://www.ee.columbia.edu/~sfchang/vis-project>.
VIRAGE, *Virage products demo*, 1999, <http://www.virage.com/products/demo/index.html>.
YAHOO IMAGE SURFER, 1999, <http://isurf.yahoo.com>.

Note

[1] Un contenuto "concreto", quindi, rappresentativo, non inteso nel senso tradizionale e letterario di "concetto".
[2] Cfr. WILLIAM I. GROSKY, *Managing multimedia information in database systems*, "Communications of the ACM", 40 (1997), 12, p. 73-80, p. 73-74.
[3] Cfr. WILLIAM I. GROSKY, *Managing multimedia information ...*, cit., p. 74-76.
[4] Esattamente un *data model*, per Grosky, è un insieme di dati, che però sono relativi, in ultima analisi, a una sorta di raccolta di *abstract concepts*, a una griglia di categorie che può essere usata per rappresentare le caratteristiche degli oggetti reali. I valori elettronici concreti, riferiti ai concetti astratti – colorazione, linearità, sonorità, strutturazione –, si definiscono attraverso l'inevitabile passaggio dell'oggetto reale per l'oggetto multimediale, ossia il documento trattato dal sistema informativo secondo un primo grado di mediazione, che solo può essere digitalizzato e ulteriormente trattato nel processo di *data modelling* (cfr. WILLIAM I. GROSKY, *Managing multimedia information ...*, cit., p. 74-75).
[5] Cfr. WILLIAM I. GROSKY, *Managing multimedia information ...*, cit., p. 75.
[6] Cfr. p. 76-77.
[7] Cfr. WILLIAM I GROSKY, *Managing multimedia information ...*, cit., p. 77.
[8] Un'ulteriore discussione delle modalità di "astrazione", insieme con un'ampia analisi della problematica del Visual Retrieval, si può trovare in PETER G. B. ENSER, *Pictorial information retrieval. Progress in documentation*, "Journal of Documentation", 51 (1995), 2, p. 126-170.
[9] ALBERTO DEL BIMBO, *Image and video databases: visual browsing, querying and retrieval,* "Journal of Visual Languages and Computing", 7 (1996), 4, p. 353-359.
[10] Cfr. ALBERTO DEL BIMBO, *Image and video databases,* cit., p. 353-356.
[11] *Ibidem,* p. 353-354.
[12] *Ibidem,* p. 355.
[13] In proposito si può vedere MANFRED NOACK, *Image mining. Stand der Entwicklung auf dem Gebiet von Image-Retrieval-Systemen*, "NFD Information Wissenschaft und Praxis", 49 (1998), 2, p. 73-76.
[14] L'Università di Brema fornisce una presentazione di ImageMiner all'indirizzo web, <http://www.tzi.uni-bremen.de/BV/ImageMiner>.
[15] Il Computer Vision Group mette a disposizione una demo di VIPER, e una ricca serie di materiali bibliografici, all'indirizzo <http://cui.unige.ch/~viper>.
[16] L'IBM fornisce una presentazione di QBIC, così come versioni demo e materiale informativo e bibliografico, all'indirizzo <http://wwwqbic.almaden.ibm.com>.
[17] Tutti i moduli del progetto sono pienamente disponibili in rete: VisualSEEk all'indirizzo <http://www.ctr.columbia.edu/VisualSEEk>; WebSEEk all'indirizzo <http://disney.ctr.columbia.edu/WebSEEk>; e MetaSEEk allo <http://www.ctr.columbia.edu/MetaSEEk>. L'intero programma è disponibile, insieme a diversi materiali bibliografici, all'indirizzo <http://www.ee.columbia.edu/~sfchang/vis-project>.
[18] RUUD M. BOLLE et al., *Video query: research directions*, "IBM Journal of Research and Development", 42 (1998), 2, p. 233-252.
[19] Estrarre dei frammenti che siano chiavi rappresentative di un filmato intero, così come un abstract o un descrittore rappresentano un intero documento testuale, implica che si possa opportunamente sintetizzare il video. Il sistema di sintesi proposto da Bolle, Yeo e Yeung, comporta che il video intero sia diviso in parti definite, unità con proprio contenuto semantico, organizzate tra loro secondo una "gerarchia strutturale" (cfr. RUUD M. BOLLE et al., *Video query: research directions*, cit., p. 238-239).
[20] Cfr. RUUD M. BOLLE et al., *Video query: research directions*, cit., p. 233-235.
[21] *Ibidem,* p. 235-241.
[22] *Ibidem,* p. 241-244.
[23] *Ibidem,* p. 244-249. Per un ulteriore approfondimento si può vedere MARK E. RORVIG, *A method for automatically abstracting visual documents*, "Journal of the American Society for Information Science", 44 (1993), 1, p. 40-56.
[24] La Virage mette a disposizione in rete una versione demo del proprio prodotto all'indirizzo <http://www.virage.com/products/demo/index.html>.
[25] RYUJI KUBOZONO et al., *'Browse search using audio key-information' for multimedia on-demand systems*, "IEEE Transactions on Consumer Electronics", 42 (1996), 4, p. 900-906.
[26] Il principio è quello di sfruttare le attività quotidiane di ascolto degli utenti, per ancorarvi il sistema di comunicazione delle chiavi audio che poi serviranno nel recupero dei documenti sonori. In qualunque momento quindi, tramite la radio o la televisione, può essere inviato all'utente un flusso di *audio key-information* con i corrispondenti *text-index*, lasciandogli acquisire naturalmente un nuovo sistema di indicizzazione e recupero semplice e intuitivo (cfr. RYUJI KUBOZONO et al., *Browse search using audio..., cit.* p. 900-901).
[27] Cfr. RYUJI KUBOZONO et al., *Browse search using audio..., cit.* p. 900-903.
[28] Cfr. RYUJI KUBOZONO et al., *Browse search using audio..., cit.* p. 903-904.
[29] Una presentazione di SoundFisher e del nuovo modulo *content-based*, insieme a dei più tradizionali motori di ricerca, sono disponibili allo <http://www.soundfisher.com>.
[30] Yahoo Image Surfer è disponibile in rete all'indirizzo <http://isurf.yahoo.com>.
[31] AltaVista è presente all'indirizzo <http://www.altavista.com>; Infoseek è in rete allo <http://infoseek.go.com>.