

# Gestione Dati Multimediali: Modelli, Indici, Interfacce, Applicazioni

Ilaria Bartolini e Marco Patella

DEIS  
Università di Bologna

## Introduzione

La diffusione di nuovi strumenti tecnologici di acquisizione dati, quali smart-phone e fotocamere digitali, ha portato a una crescita senza precedenti della quantità di documenti multimediali a disposizione degli utenti, in particolare nell'ambito dei beni culturali. Esempi di tali raccolte possono includere archivi storici di fotografie digitalizzate, filmati di una cineteca, digitalizzazioni di opere all'interno di una collezione artistica. È divenuta quindi un'esigenza impellente definire strumenti di gestione efficace ed efficiente per l'esplorazione e la ricerca di tali dati. Le soluzioni presenti allo stato dell'arte offrono una varietà di tecniche eterogenee, che vanno dalla ricerca basata sul contenuto (ovvero su caratteristiche visuali, quali il colore di un'immagine), al recupero basato su parole chiave (ad esempio, i tag utilizzati per annotare una fotografia) e da gerarchie basate su argomenti specifici, a interfacce grafiche avanzate basate su nuove metafore di navigazione per supportare l'utente nell'esplorazione di grandi collezioni.

Nessuna delle soluzioni sopra descritte è però in grado da sola di raggiungere risultati soddisfacenti per l'utente. In questo contesto l'obiettivo dell'Unità UNIBO all'interno del progetto CoOPERARE è stato quello di fornire modelli, algoritmi e interfacce per la ricerca e l'esplorazione dei dati multimediali, allo scopo di facilitarne l'accesso per utenti generici. Tale soluzione rappresenta un efficace passo avanti nel fornire all'utente uno strumento integrato che permetta la ricerca e la navigazione dei dati sfruttando:

1. la natura oggettiva del dato
2. i suoi metadati
3. le annotazioni degli utenti.

Ad esempio, con riferimento all'ultimo degli archivi multimediali visualizzati in figura 1, possiamo notare come la descrizione del quadro ritraente la Gioconda appartenente ad una certa collezione artistica possa essere rappresentata da caratteristiche visuali (es., la distribuzione del colore all'interno del dipinto), da dati di natura archivistica (il periodo artistico d'appartenenza o l'autore) e da ulteriori dati inseriti da un utente «non esperto» (indicati con colore diverso in figura).

### Archivio Storico Fiat



- Trimotore Fiat G212
- Data: 1947
- Collezione: Tema di cultura industriale
- Tipologia: Immagine
- Aereo, Motore, Ali

### Cineteca



- Das Cabinet des Dr. Caligari
- Data: 1920
- Nazione: Germania
- Regista: Robert Wiene
- Genere: Horror
- Espressionismo, Ipnosi, Sonnambulismo

### Archivio Artistico



- La Gioconda
- Sito: Museo Louvre, Parigi
- Secolo: XVI
- Autore: Leonardo da Vinci
- Periodo: Rinascimento
- Data: 1503
- Dipinto, Ritratto, Sorriso

Figura 1: Esempi di archivi multimediali

Il modello che permette l'integrazione di tutte e tre le succitate caratteristiche descrive ogni documento multimediale attraverso un insieme di dimensioni, ognuna delle quali rappresenta un diverso punto di vista sul documento stesso. Ogni dimensione è rappresentata da una gerarchia di concetti semantici, ad ognuno dei quali possono essere associati diversi documenti: ciò permette di accedere ai dati secondo diversi livelli di astrazione. Inoltre, tali gerarchie possono essere personalizzate dall'utente ottenendo così percorsi diversificati in base all'utente e/o alle necessità informative del momento (si veda la figura 2 per un esempio di dimensione che può essere utile nella catalogazione di un'opera artistica).

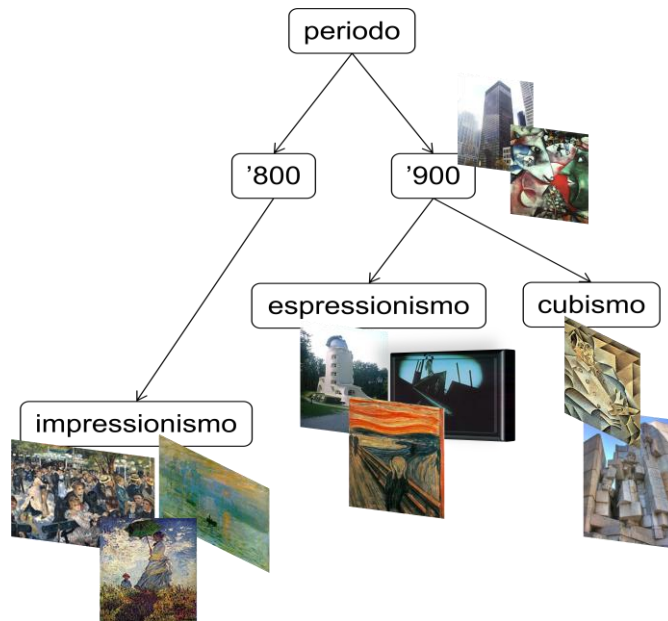


Figura 2: Dimensione per la caratterizzazione del periodo artistico di un'opera

La rappresentazione multidimensionale dei documenti multimediali si configura come strumento per l'accesso efficace ed efficiente ai documenti stessi, in quanto le gerarchie vengono sfruttate in fase di navigazione: selezionando un concetto all'interno di una dimensione, potranno essere

recuperati tutti i documenti associati a tale concetto e a concetti più specifici (ovvero ai discendenti di tale concetto nella gerarchia). È inoltre possibile selezionare concetti appartenenti a dimensioni diverse per raffinare la ricerca (si veda la figura 3).

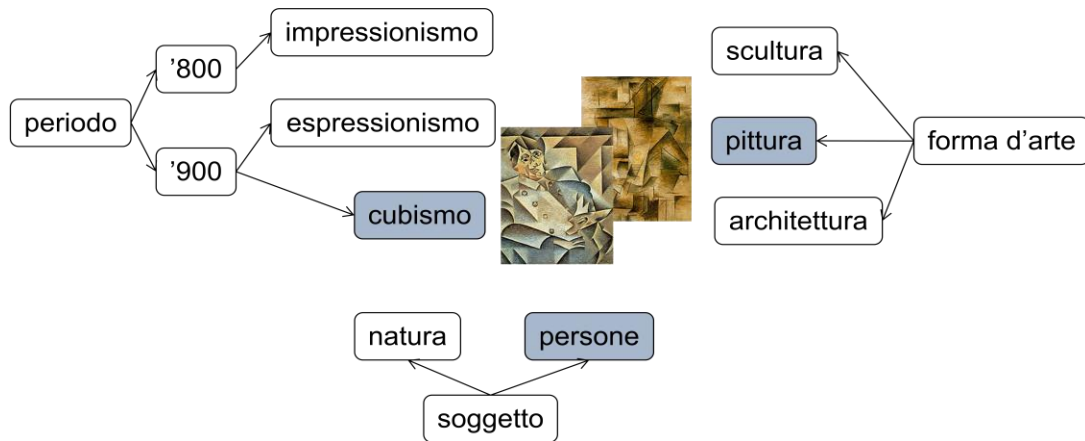


Figura 3: In fase di ricerca/navigazione, le dimensioni vengono sfruttate per filtrare i risultati che devono essere visualizzati

Tale modello si è concretizzato nella realizzazione di interfacce visuali per la ricerca e l'esplorazione di archivi di immagini e video che prevedono l'utilizzo di due tipi diversi di dimensioni. Il primo tipo consiste di tassonomie di concetti espressi tramite parole chiave: tali dimensioni testuali rappresentano, ad esempio, le annotazioni espresse dagli utenti in base alla propria visione sui dati e/o alla propria competenza. Tipologie diverse di utenti avranno infatti punti di vista diversi su uno stesso dato: ad esempio, uno studente, un archivista e uno storico dell'arte avranno visioni diverse su una stessa opera d'arte. Il secondo tipo di dimensioni, invece, si basa sulla rappresentazione del contenuto del dato, definito mediante la nozione di somiglianza. In questo caso, ogni concetto semantico della gerarchia coincide con un documento che è simile a tutti i suoi discendenti all'interno della gerarchia stessa. Nell'esempio di figura 4, tutti i dipinti inclusi nel sotto-albero di sinistra (rispettivamente, destra) sono «visivamente simili» al dipinto contenuto nel nodo radice di tale sotto-albero.

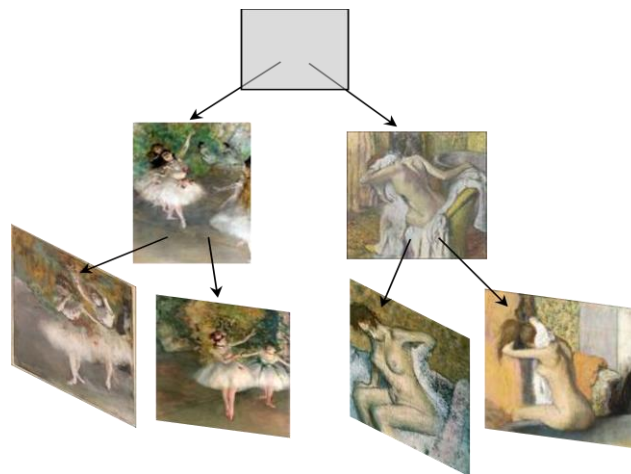


Figura 4: Un esempio di dimensione basata su somiglianza

Tali dimensioni sono in tutto e per tutto analoghe alle dimensioni testuali e, pertanto, possono essere sfruttate durante la fase di ricerca come visto in precedenza. Ad esempio, supponendo di essere giunti ad accedere al quadro in alto a sinistra in figura 4 («Petits Rats») ricercando quadri (dimensione forma d'arte) impressionisti (dimensione periodo) che raffigurano ballerine (dimensione soggetto), è possibile «navigare» la dimensione basata sulla somiglianza spostandosi verso opere d'arte visivamente simili, ma che possono non condividere altre caratteristiche con il dato originale.

La realizzazione efficiente di tale tipologia di dimensione richiede che il concetto di somiglianza si basi su caratteristiche estratte automaticamente dai documenti. In particolare, nell'ambito del progetto CoOPERARE, il contributo di UNIBO ha incluso anche la caratterizzazione in tal senso di documenti costituiti da immagini fisse e video, come verrà dettagliato nel seguito.

Infine, è da notare che l'utente può partecipare non solo alla «mappatura» dei contenuti sulle dimensioni esistenti, ma anche alla creazione/modifica delle dimensioni stesse: in tal modo, la visione dei dati degli archivi non è «imposta» all'utente ma «emerge» dall'interazione dell'utente con il dato (attraverso lo strumento) e con i (meta-)contenuti forniti da altri utenti della stessa categoria. Lo scopo è quello di creare, in maniera implicita, delle *folksonomies* nelle quali chi partecipa sono gli utilizzatori dei contenuti. In questo modo, si viene ad annullare il gap esistente tra le caratterizzazioni di un oggetto che vengono date da chi lo crea o lo archivia e quelle date invece da chi lo utilizza. L'utente diventa perciò «protagonista» del sistema, in quanto annota documenti multimediali in base alla propria visione sui dati e/o alla propria competenza (profilo utente) e personalizza le dimensioni. Nell'esempio di figura 5, si nota come diverse categorie di utenti (es., Studente, Archivista, Storico dell'arte) siano in grado di accedere allo stesso dato secondo percorsi diversificati (in base all'utente o alle necessità informative del momento).

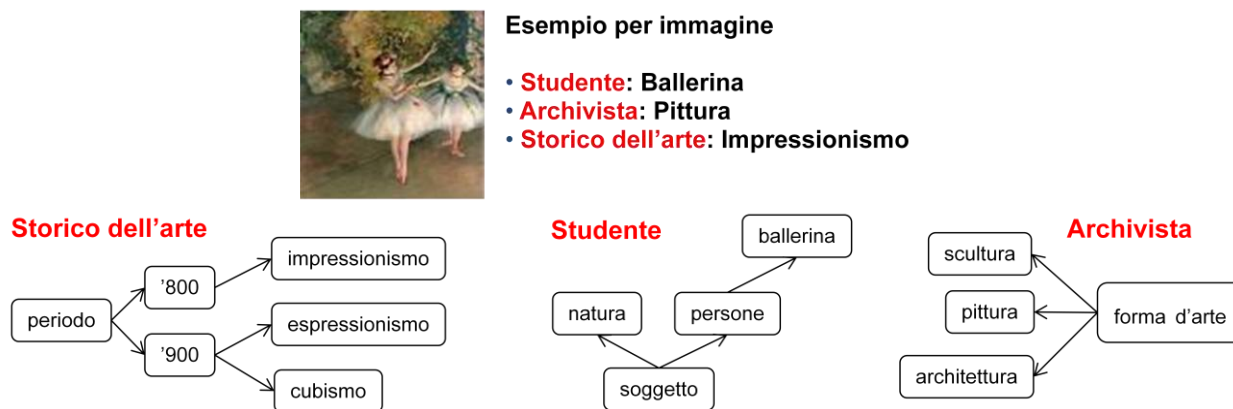


Figura 5: Categorie diverse di utenti possono annotare uno stesso documento secondo punti di vista differenti

Le dimensioni basate sulla somiglianza, invece, possono essere modificate dall'utente mediante «azioni» di personalizzazione del tutto *user friendly* (ad esempio, tramite *drag&drop* del mouse). L'effetto di tali azioni è anche quello di migliorare l'organizzazione dei dati, in modo che essa rispecchi più precisamente il concetto di somiglianza percepito dall'utente. Nell'esempio di figura 6 un intero sotto-albero è stato spostato all'interno della dimensione.

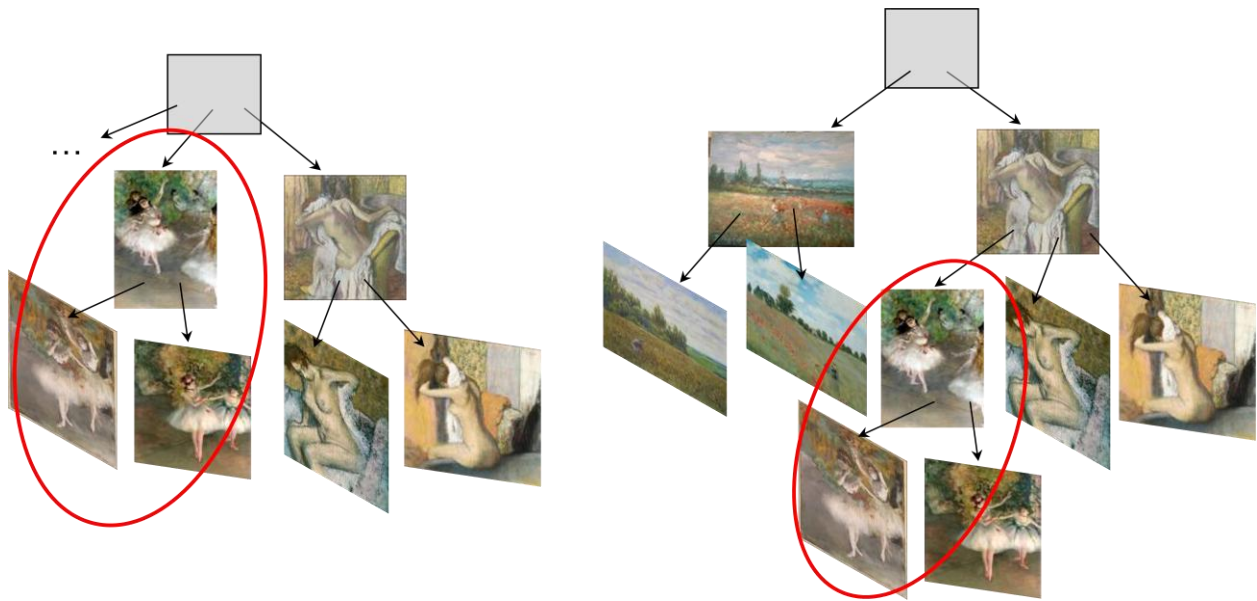


Figura 6: Personalizzazione della dimensione basata sulla somiglianza

### Applicazione al caso immagini

Uno dei contributi dell'Unità UNIBO all'interno del progetto CoOPERARE ha riguardato la realizzazione di un prototipo che permettesse la navigazione di archivi di immagini secondo i principi enunciati nell'introduzione. In tale contesto è da sottolineare come la gestione delle dimensioni di tipo testuale non ponga requisiti di particolare complessità dal punto di vista sia dell'efficacia che dell'efficienza. Viceversa, il trattamento delle caratteristiche di basso livello necessita di particolare attenzione poiché con esse si vuole rappresentare il contenuto informativo del dato. A tal fine, risulta fondamentale il concetto di similarità, con il presupposto che dati simili avranno lo stesso contenuto informativo (o quantomeno contenuti simili). In questo modo è possibile raggruppare assieme immagini «simili», creando così gruppi di immagini, permettendo all'utente la navigazione gerarchica dei dati e la modifica delle relazioni di similarità degli stessi attraverso le summenzionate azioni di personalizzazione.

Al fine di dimostrare le funzionalità più rilevanti di tale prototipo, nel seguito si descriverà un esempio d'uso dello stesso nel caso di esplorazione/navigazione di un archivio fotografico personale. Come mostrato dalla figura 7, l'utente ha caratterizzato le proprie foto secondo 6 dimensioni:

- animal: per le fotografie che raffigurano animali, viene indicato il tipo di animale;
- vegetation: indica il tipo di vegetazione eventualmente raffigurato;
- landscape: caratterizza il tipo di paesaggio rappresentato nella foto;
- geo: contiene i riferimenti geografici (metadati) automaticamente estratte dalla fotocamera tramite rilevatore GPS;
- device: contiene i parametri utilizzati dalla fotocamera durante lo scatto (es., tempo di esposizione, distanza focale);
- feature: rappresenta la dimensione basata sulla somiglianza del contenuto, in questo caso basata sulla distribuzione del colore all'interno della fotografia.

Si supponga che l'utente sia alla ricerca di immagini che soddisfino un ben definito requisito semantico, come fotografie di tigri del Bengala in mezzo all'erba. A tal scopo, nell'esempio di figura 7, l'utente ha già selezionato la coordinata «animal», contenente due concetti, «orso» (*bear*) e «felino» (*feline*). Poiché ambedue i concetti presentano concetti figli, il prototipo visualizza per entrambi un «cubo», che simboleggia un contenitore di immagini ulteriormente esplorabile, ad esempio tramite doppio clic.

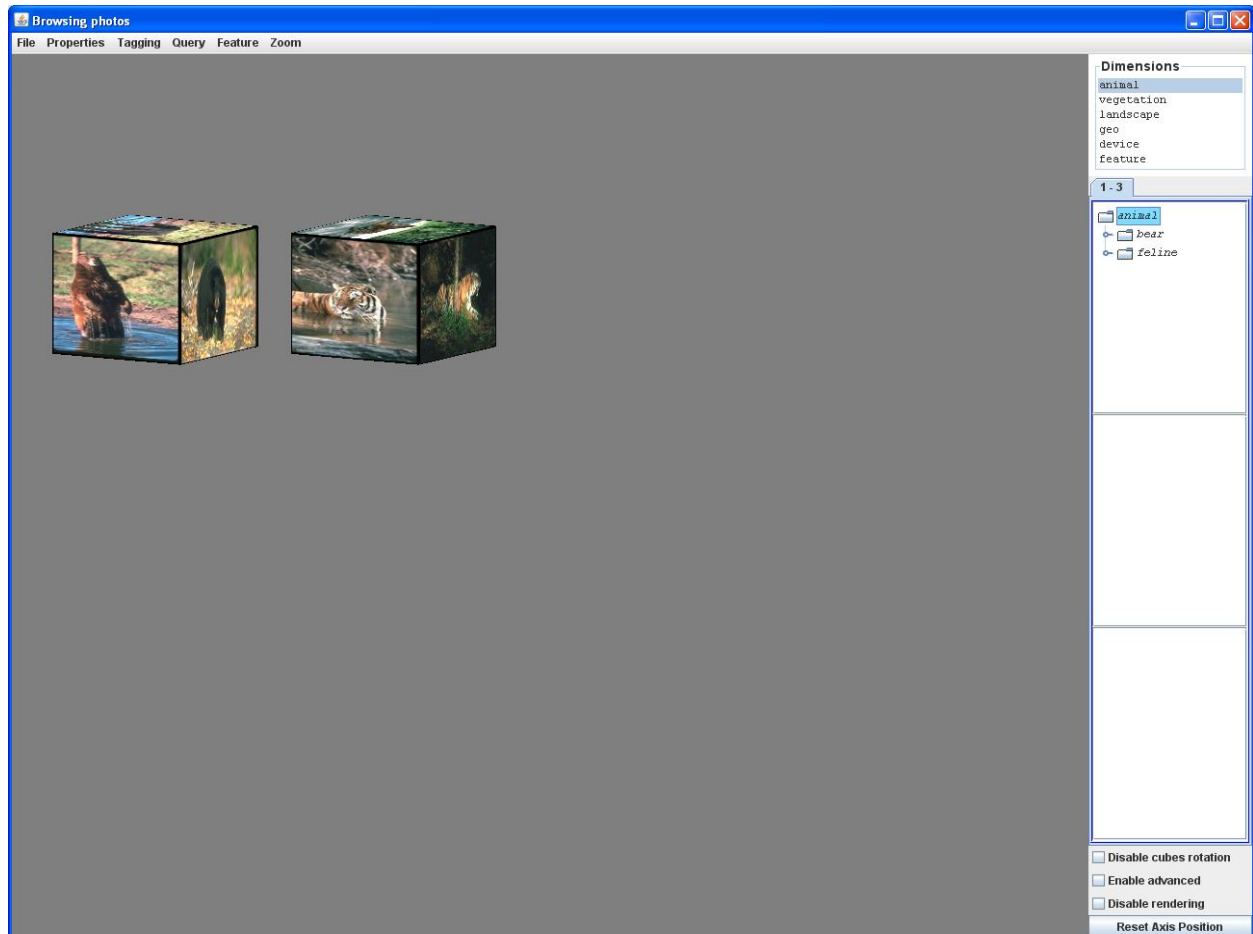


Figura 7: Esplorazione di un archivio fotografico (1 di 6)

L'utente, quindi, decide di selezionare anche la dimensione «vegetation», in modo da esplorare fotografie che raffigurino animali tra la vegetazione (si veda la figura 8). Infatti, la selezione di due o più dimensioni fa sì che vengano visualizzati solamente quei documenti che siano stati annotati usando concetti provenienti da tutte le dimensioni selezionate. Essendo state selezionate due dimensioni, il prototipo è in grado di visualizzare le immagini usando una metafora bidimensionale, in cui all'interno di ogni coordinata sono visualizzati i concetti di una singola dimensione. In figura 8, ad esempio, la dimensione «vegetation» contiene i concetti «foresta» (*forest*), «erba» (*grass*), «piante» (*plants*) e «albero» (*tree*). Cliccando sul «cubo» evidenziato dal cerchio in figura 8, l'utente indica di essere interessato alle fotografie che ritraggono felini nell'erba.

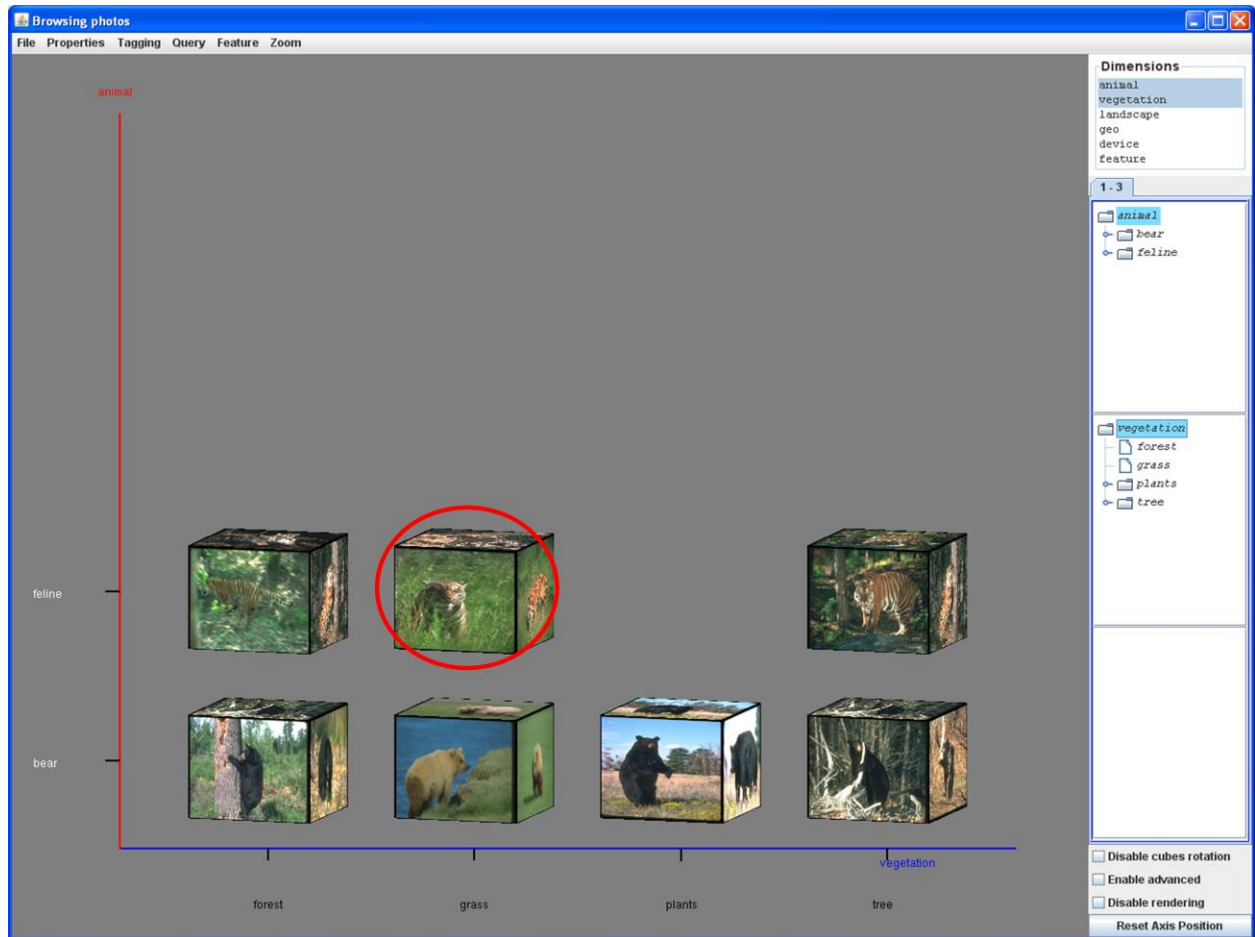


Figura 8: Esplorazione di un archivio fotografico (2 di 6)

Il sistema, quindi, recupera le immagini richieste e le visualizza sullo schermo (si veda la figura 9). Dopo alcuni passi di esplorazione all'interno della dimensione «animal», del tutto analoghi al precedente, l'utente giunge ad una situazione in cui il concetto «erba» della dimensione «vegetation» non presenta ulteriori specificazioni, ovvero non ha alcun «figlio»; pertanto si ha una diversificazione tra le immagini solamente per quello che riguarda la dimensione «animal». Siccome l'utente è interessato alle tigri del Bengala, seleziona il cubo corrispondente, evidenziato dal cerchio in figura 9.



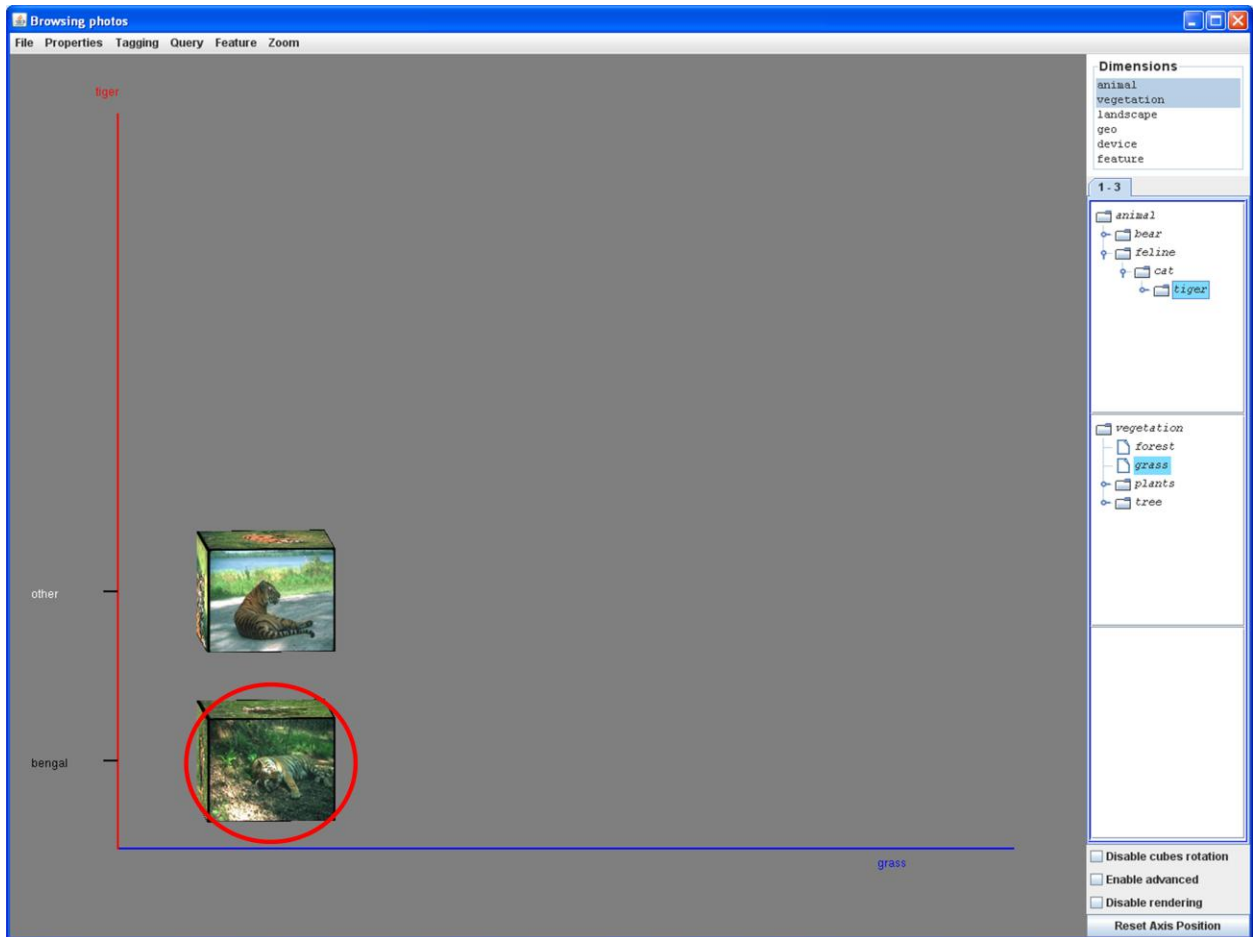


Figura 9: Esplorazione di un archivio fotografico (3 di 6)

Dopo questo passo, l'utente raggiunge un concetto senza figli anche all'interno della dimensione «animal», quindi il prototipo è in grado di visualizzare tutte le fotografie che sono state annotate con i concetti selezionati, ovvero «animal/feline/cat/tiger/bengal» e «vegetation/grass». Tali fotografie, come mostrato dalla figura 10, sono visualizzate sullo schermo mediante una rappresentazione spaziale che evidenzia la somiglianza tra tali immagini, secondo il principio per cui immagini simili sono visualizzate vicine sullo schermo.



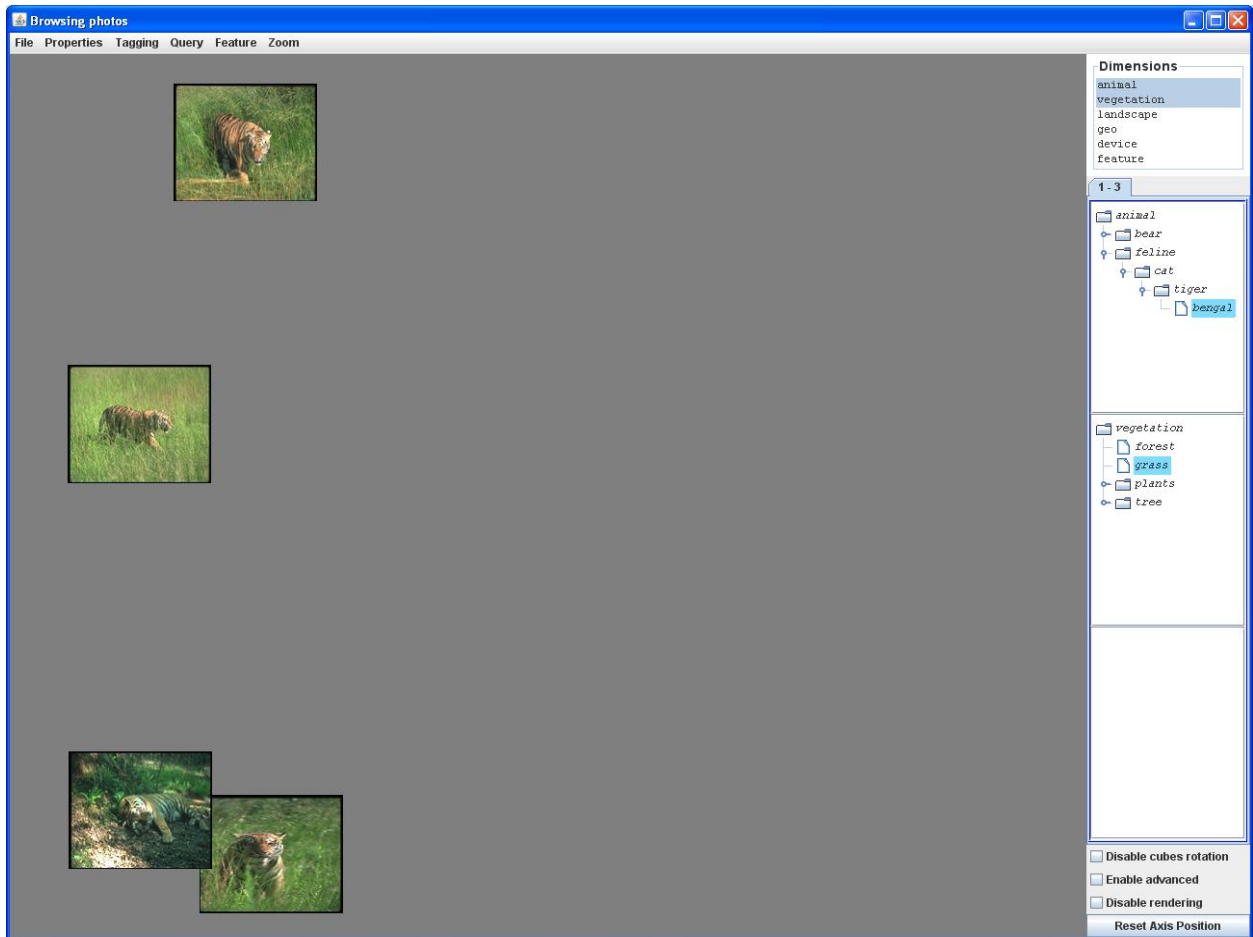


Figura 10: Esplorazione di un archivio fotografico (4 di 6)

Si supponga ora che l'utente, soddisfatto dalla ricerca precedente, desideri ora ritornare alla situazione di figura 8, ovvero di richiedere fotografie annotate sia secondo la dimensione «animal» che secondo la dimensione «vegetation», ma che voglia anche navigare la dimensione basata sulla somiglianza. Come mostrato dalla figura 11, la selezione di una terza dimensione fa sì che il prototipo passi ad una visualizzazione tridimensionale (ovviamente simulata tramite visuale assonometrica). Si noti che la dimensione basata sulla somiglianza presenta concetti che rappresentano «gruppi» di immagini simili tra loro. Questo viene riflesso anche dalla visualizzazione tridimensionale, in quanto la dimensione «feature» distingue proprio tali gruppi (si noti come, in figura 11, la distribuzione di colore vari solamente lungo tale coordinata). Per ciascun gruppo, l'immagine visualizzata dal prototipo è semplicemente una fotografia rappresentativa di tale gruppo.

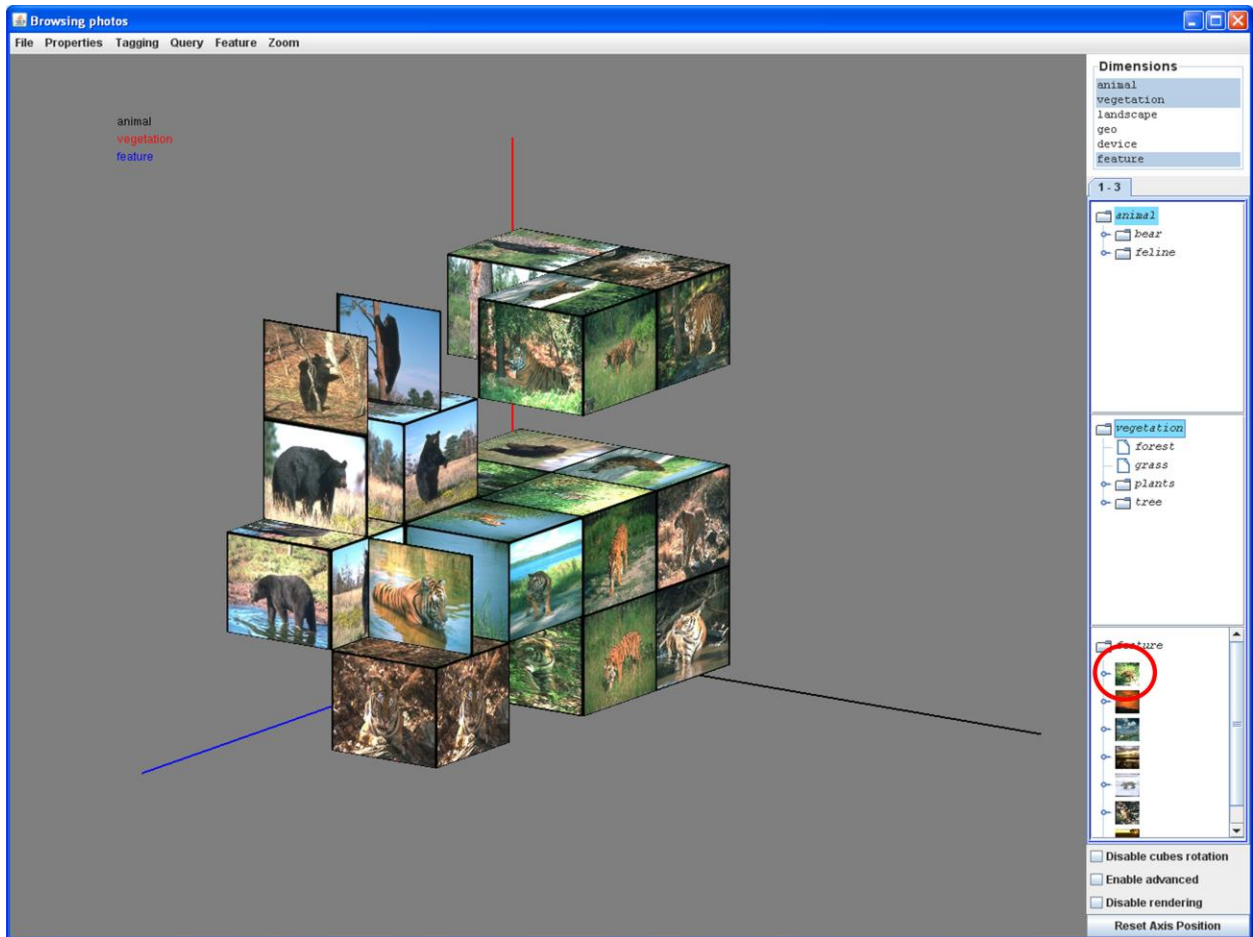


Figura 11: Esplorazione di un archivio fotografico (5 di 6)

Infine, quando l'utente seleziona il gruppo evidenziato da un cerchio in figura 11, il prototipo visualizza solamente le immagini contenute all'interno di tale gruppo. Come mostrato dalla figura 12, tale gruppo contiene cinque sotto-gruppi, che possono essere ulteriormente esplorati dall'utente, nel caso egli desideri raffinare ulteriormente la sua richiesta.

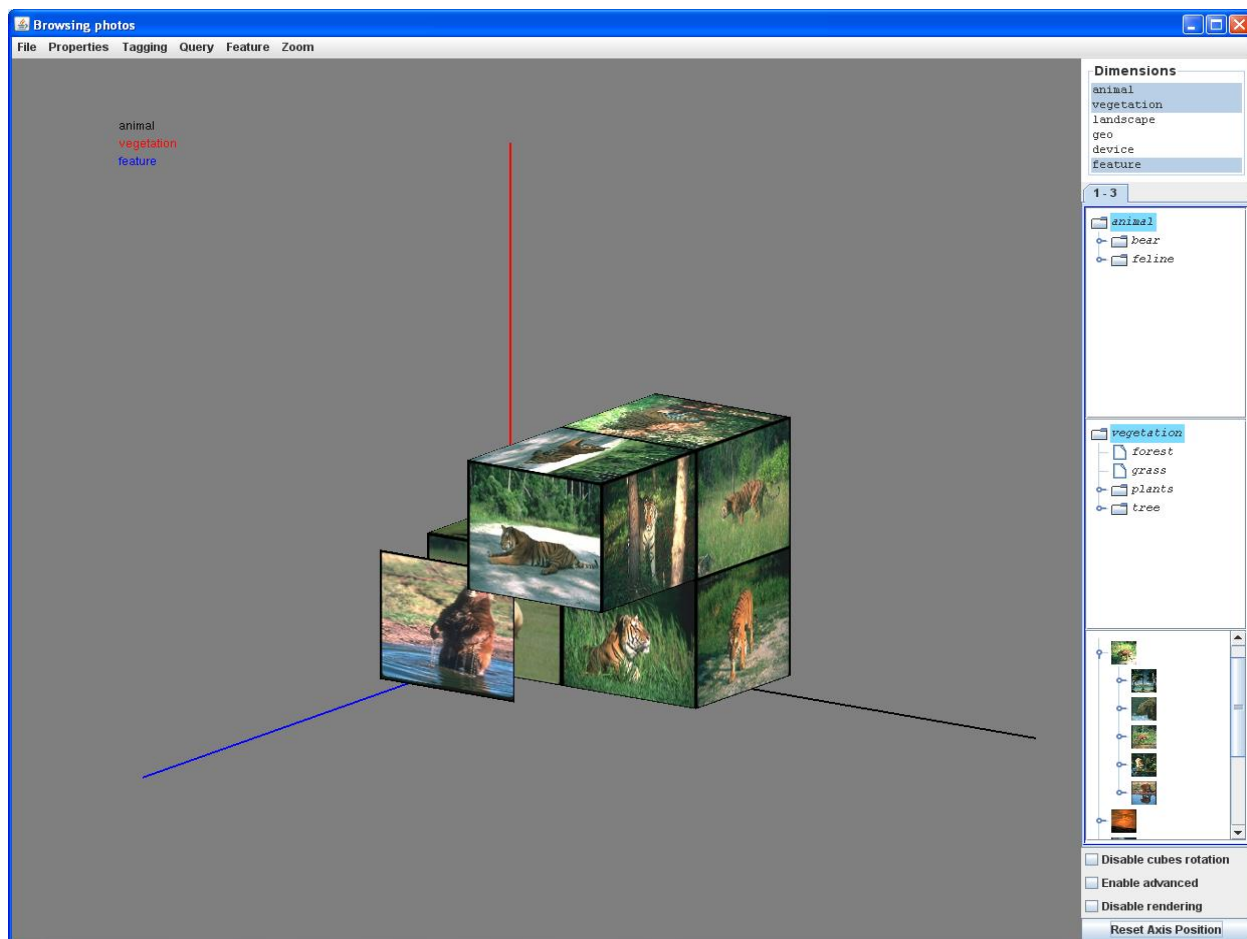


Figura 12: Esplorazione di un archivio fotografico (6 di 6)

### Applicazione al caso video

Il secondo contributo principale dell'Unità UNIBO all'interno del progetto CoOPERARE ha riguardato la definizione di soluzioni efficaci ed efficienti per la caratterizzazione e la rappresentazione del contenuto di archivi video.

La difficoltà principale nell'affrontare grandi collezioni di video risulta il recupero (semi-)automatico di informazioni, quali ad esempio il contenuto visuale e/o il genere. Le soluzioni proposte dai più popolari motori di ricerca per la gestione di video (come Google o YouTube) consistono nell'analizzare il contesto «attorno» al video per inferire informazioni sul video stesso; ad esempio, si utilizza il titolo della pagina web contenente il video. Poiché non sempre è possibile analizzare il contesto all'interno del quale si trova il video (in quanto scorrelato dalle informazioni contenute nel video o perché semplicemente mancante), diversi studi si sono recentemente concentrati sul recupero di video basato sul contenuto visuale. La letteratura esistente sull'argomento generalmente suddivide il problema in due attività principali: il rilevamento dei cambi di inquadratura e l'indicizzazione semantica del video.

La prima problematica è stata uno degli argomenti di ricerca preminenti negli ultimi vent'anni nel recupero di video, con il fine di segmentare un video in scene, all'interno delle quali i fotogrammi

condividessero le medesime caratteristiche visuali, in modo da facilitare l'analisi del contenuto delle scene stesse. Ciò è generalmente ottenuto rilevando tagli e transizioni graduali (come dissolvenze, ecc.), attraverso il confronto delle caratteristiche visuali tra fotogrammi consecutivi e la successiva applicazione di criteri di separazione basati sull'utilizzo di soglie.

D'altro canto, l'indicizzazione semantica ha lo scopo di estrarre concetti semantici dai filmati, in modo da fornire all'utente una descrizione compatta, utile ad esempio per la categorizzazione del contenuto dei video. Per questo, l'approccio più comune è quello di applicare tecniche ideate per il contesto immagini ai fotogrammi caratteristici delle scene; questo permette di rilevare il contenuto delle scene video e generare, quindi, una descrizione opportuna. Attualmente, il metodo più comunemente utilizzato per rappresentare tale descrizione risulta l'utilizzo di tecniche di annotazione (*tagging*) dei filmati, che raffigurano il contenuto video tramite l'uso di etichette testuali. In questo modo, il concetto semantico riprodotto da ciascuna etichetta testuale può essere trasferito alla scena video. Ovviamente, nel nostro caso, tali annotazioni verranno rappresentate dai concetti presenti all'interno delle varie dimensioni testuali, analogamente a quanto visto per il caso degli archivi di immagini. Secondo questo paradigma, quindi, la caratterizzazione e la rappresentazione del contenuto di archivi video richiede di risolvere il problema dell'annotazione automatica di filmati secondo le diverse dimensioni presenti.

In questo contesto, è importante notare che la segmentazione di un filmato per scopi di annotazione è un processo che presenta vistose differenze rispetto ad altri tipi di segmentazione video. Ad esempio, nel caso in esame non è necessaria un'esagerata precisione nel determinare l'esatto momento di transizione tra due scene, in quanto si desidera solamente identificare i fotogrammi che riportano lo stesso contenuto visuale della sequenza.

L'attività di ricerca svolta dall'Unità UNIBO durante il progetto ha portato alla definizione ed alla realizzazione del prototipo SHIATSU (Semantic Hlerarchical Automatic Tagging of videos by Segmentation Using cuts), un sistema completo di elaborazione video in grado di offrire una descrizione accurata del contenuto di filmati. Dato un filmato, questo viene per prima cosa diviso in scene (sequenze di fotogrammi); quindi, per ogni scena viene identificato un insieme rappresentativo di fotogrammi; tale insieme viene fornito al modulo di etichettatura, generando un insieme di etichette candidate che descrivano la scena (tali etichette possono essere verificate dall'utente prima dell'associazione con la scena); infine, il filmato completo viene etichettato in maniera gerarchica utilizzando una propagazione pesata delle etichette delle scene componenti. Questo permette la realizzazione di una piattaforma di *browsing* gerarchica a due livelli: il livello scena ed il livello filmato. In questo modo, l'utente del sistema SHIATSU è in grado di recuperare non solo le scene, ma anche i filmati caratterizzati da una specifica etichetta.

Un'enfasi particolare all'interno di SHIATSU è stata posta sull'efficienza del processo di etichettatura e di recupero/*browsing*. Se il secondo compito è svolto grazie all'utilizzo di un database relazionale (in grado, quindi, di recuperare in tempo reale le scene ed i filmati che includono una data etichetta testuale specificata dall'utente), per la prima attività è stata necessaria un'attenzione particolare per far sì che i processi di segmentazione e di analisi delle scene potesse essere svolto in tempi ragionevoli. A tal scopo, si sono utilizzate tecniche all'avanguardia per la caratterizzazione a basso livello dei fotogrammi e per la risoluzione efficiente di interrogazioni di similarità. Quest'ultima è necessaria in quanto il modulo di annotazione funziona in base al principio secondo il quale «risorse simili tendono ad essere annotate con gli stessi termini». In questo modo, le etichette suggerite per una scena vengono selezionate tra quelle associate ad immagini simili ai fotogrammi della scena stessa. Per ogni scena da annotare, è quindi necessario

risolvere diverse interrogazioni di similarità, ciascuna delle quali utilizzi le caratteristiche di basso livello delle immagini/fotogrammi per stabilirne la similarità. Siccome quest'ultimo risulta essere un compito alquanto gravoso dal punto di vista computazionale, il confronto esaustivo di ogni fotogramma con l'insieme di immagini annotate rappresenta una soluzione chiaramente non percorribile. Per questo, ci si è avvalsi dell'esperienza del gruppo di ricerca UNIBO nel campo dell'indicizzazione di spazi metrici per realizzare un metodo di accesso che permettesse di risolvere in maniera efficiente ogni singola interrogazione di similarità, ottenendo così un processo di annotazione che non risultasse particolarmente oneroso, con tempi di risposta dell'ordine del secondo per il suggerimento delle etichette relative ad una singola scena.

Il sistema SHIATSU è stato valutato sperimentalmente utilizzando alcune collezioni *benchmark* di video. Per quel che riguarda l'efficienza del processo di segmentazione, SHIATSU si è rivelato estremamente efficiente, ottenendo una capacità di elaborazione in tempo reale dei filmati su un sistema «low-end». Per quel che riguarda la precisione del processo di segmentazione, i risultati ottenuti sono paragonabili a quelli di altri sistemi allo stato dell'arte, che risultano però estremamente più lenti nell'elaborazione dei fotogrammi. Infine, l'accuratezza del sistema di etichettatura risulta ottima, in particolare considerando che il sistema è in grado di effettuare tale operazione in tempo reale, con tempi di esecuzione di qualche secondo per ogni sequenza.

È infine da rimarcare che, rispetto ad altri strumenti esistenti per l'etichettatura automatica di filmati, SHIATSU considera solamente il contenuto visuale di un filmato per estrarre le etichette descrittive: tale approccio può però essere facilmente complimentato con altre tecniche (eventualmente più complesse) per l'analisi di audio e parlato, in modo da fornire una caratterizzazione più precisa dei filmati. Per il momento, è importante sottolineare che il sistema SHIATSU, nonostante la sua semplicità ed efficienza, è in grado di ottenere ottimi livelli di precisione nell'etichettatura.

Come esempio d'utilizzo del prototipo SHIATSU, la figura 13 mostra il processo di annotazione di un video. Nella parte superiore sinistra dell'interfaccia si possono notare i pulsanti che permettono di caricare un nuovo video, visualizzarlo ed annotarlo, eventualmente salvando le annotazioni suggerite dal sistema. La suddivisione del video in sequenze è mostrata nella parte superiore destra dell'interfaccia. Selezionando una sequenza in tale riquadro, è possibile visualizzare la sequenza (parte centrale sinistra dell'interfaccia) e scegliere secondo quale dimensione, tra quelle presenti nella parte centrale destra della finestra, si vuole caratterizzare tale sequenza. Nella parte inferiore dell'interfaccia è possibile navigare la tassonomia di concetti per selezionare manualmente alcuni di essi come annotazioni (parte sinistra) o richiedere al sistema di suggerire automaticamente i concetti più «appropriati» per la descrizione di tale sequenza. Una volta annotate le varie sequenze, è poi possibile propagare opportunamente tali annotazioni all'intero video.

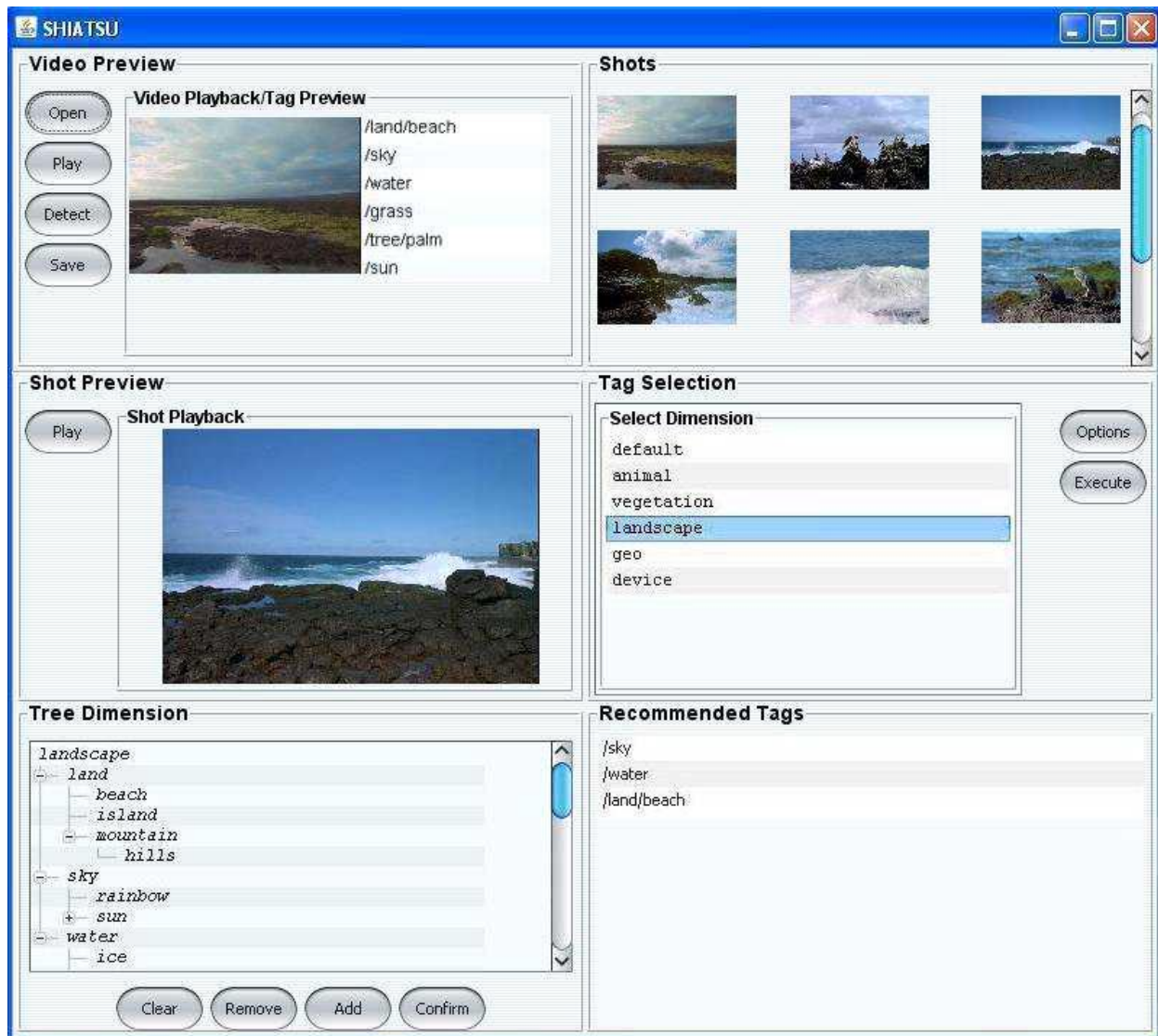


Figura 13: L'interfaccia del prototipo SHIATSU durante la fase di annotazione

## Bibliografia

- Bartolini, I. (2009). A Multi-faceted Browsing Interface for Digital Photo Collections. *Proceedings of the 7th International Workshop on Content-Based Multimedia Indexing (CBMI 2009)*, Chania, Grecia, Giugno 2009.
- Bartolini, I. & Ciaccia, P. (2008). Scenique: A Multimodal Image Retrieval Interface. *Proceedings of the 2008 International Working Conference on Advanced Visual Interfaces (AVI 2008)*, Napoli, Italia, Maggio 2008.
- Bartolini, I., Ciaccia, P. & Patella, M. (2006). Adaptively Browsing Image Databases with PIBE. *Multimedia Tools and Applications Journal (MTAP)*, 31(3):269-286, Dicembre 2006.

Bartolini, I., Ciaccia, P. & Patella, M. (2007). PIBE: Manage Your Images the Way You Want! *Proceedings of the 2007 IEEE ICDE International Conference on Data Engineering (ICDE 2007)*, Istanbul, Turchia, Aprile 2007.

Bartolini, I., Ciaccia, P. & Patella, M. (2010). Query Processing Issues in Region-Based Image Databases. *Knowledge and Information Systems (KAIS)*, 25(2):389-420, Novembre 2010.

Bartolini, I., Patella, M. & Romani, C. (2010). SHIATSU: Semantic-Based Hierarchical Automatic Tagging of Videos by Segmentation using Cuts. *Proceedings of the 3rd International ACM MM Workshop on Automated Information Extraction in Media Production (AIEMPro10)*, Firenze, Italia, Ottobre 2010.

Bartolini, I., Patella, M. & Stromei, G. (2011). The WINDSURF Library for the Efficient Retrieval of Multimedia Hierarchical Data. *Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP 2011)*, Siviglia, Spagna, Luglio 2011.

Bartolini, I. & Romani, C. (2010). SHIATSU: Annotating Your Videos the Easy Way!. *Proceedings of the 3rd International Conference on Similarity Search and Applications (SISAP 2010)*, Istanbul, Turchia, Settembre 2010.

Bartolini, I. & Romani, C. (2010). Efficient and Effective Similarity-based Video Retrieval. *Proceedings of the 3rd International Conference on Similarity Search and Applications (SISAP 2010)*, Istanbul, Turchia, Settembre 2010.