



*Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni*

Utilizzo di uno schema MPEG-7 per la descrizione di percorsi di navigazione assistita di contenuti video

M. Tripiciano¹

RT-ICAR-PA-03-23

dicembre 2003



Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR)
– Sede di Cosenza, Via P. Bucci 41C, 87036 Rende, Italy, URL: www.icar.cnr.it
– Sezione di Napoli, Via P. Castellino 111, 80131 Napoli, URL: www.na.icar.cnr.it
– Sezione di Palermo, Viale delle Scienze, 90128 Palermo, URL: www.pa.icar.cnr.it



Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni

Utilizzo di uno schema MPEG-7 per la descrizione di percorsi di navigazione assistita di contenuti video

M. Tripiciano¹

Rapporto Tecnico N.:
RT-ICAR-PA-03-23

Data:
dicembre 2003

¹ Istituto di Calcolo e Reti ad Alte Prestazioni, ICAR-CNR, Sezione di Palermo .

I rapporti tecnici dell'ICAR-CNR sono pubblicati dall'Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche. Tali rapporti, approntati sotto l'esclusiva responsabilità scientifica degli autori, descrivono attività di ricerca del personale e dei collaboratori dell'ICAR, in alcuni casi in un formato preliminare prima della pubblicazione definitiva in altra sede.

Te.S.C.He.T.

A Technology System for Cultural Heritage in Tourism

**Utilizzo di uno schema MPEG-7
per la descrizione di percorsi di
navigazione assistita di contenuti
video**

Versione 1.0

Introduzione

Negli ultimi anni è nato un nuovo modo di considerare un filmato video, non più come una semplice sequenza temporale di fotogrammi indicizzata tramite una struttura logico-temporale (vedi fig.1), bensì come una combinazione strutturata di più oggetti video che ad esso appartengono; un Video Object può, quindi, essere una semplice sequenza temporale di frames (video object temporale), può identificare una o più regioni di uno o più frames (video object spaziale) o, infine, essere l'insieme di entrambi le accezioni individuando un oggetto in movimento (video object spazio-temporale).

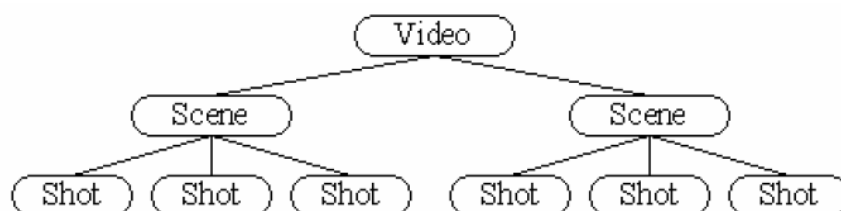


Figura 1: Esempio di struttura logica di un video

Questa interpretazione di un filmato è nata poiché una tale organizzazione permette una maggiore flessibilità; ad esempio si potrebbe visualizzare solo un oggetto, combinare oggetti provenienti da filmati diversi oppure inserire informazioni aggiuntive ed elementi non reali. Con una opportuna descrizione di tutti gli oggetti significativi che compongono il video è possibile organizzare dei databases di video object per consentire un facile reperimento e riutilizzo dell'informazione desiderata.

Lo scopo di questo rapporto è lo studio ed analisi dello standard MPEG-7 (*Moving Pictures Expert Group*) per individuare i Descrittori e gli schemi dedescrittivi idonei a descrivere una sequenza video in termini di metadati generali sulla stessa (nome, locazione dei dati, formato, ecc.), descriverne una o più indicizzazioni temporali (scene, shots, camera break) ed, inoltre, schemi descrittivi che consentano di annotare e gestire facilmente video objects di interesse individuati all'interno di una regione del frame (video objects spazio-temporali).

Una tale strutturazione descrittiva del filmato insieme a degli strumenti atti a gestirla e manipolarla consentirebbe lo sviluppo di applicazioni in diversi ambiti come, ad esempio, l'autoring video nella composizione di documenti multimediali di tipo didattico culturale (e-learning, videoconferenze...) o il monitoraggio ed il confronto dei risultati di una catena di processamento ed analisi di un ciclo di lavorazione (video restauro).

Nei domini applicativi sopra indicati potere annotare informazioni di carattere testuale riguardanti le strutture di indicizzazione del video, nonché dei contenuti semantici di basso livello dei video objects significativi, insieme con gli strumenti per editare, comporre, retronare e visualizzare tali informazioni, costituirebbe un valido ausilio alla realizzazione di un ambiente integrato che sfrutti al meglio la sempre più crescente mole di dati multimediali che vengono prodotti.

1 Annotazione del video strutturato

Per descrivere sia le caratteristiche di carattere generale del video che quelle più specifiche riguardanti i video objects (VO), è stato scelto di utilizzare lo standard MPEG-7. Le motivazioni di tale scelta sono varie tra le quali quella di essere uno standard consolidato per la descrizione di documenti multimediali ed, inoltre, essendo basato sul formato di descrizione standard XML (*EXtensible Markup Language*), ne eredita tutte le peculiarità come la trasportabilità multiplatforma, l'estendibilità e l'utilizzo di schemi descrittivi che specificano il tipo dei dati descritti. Nel seguito darò un breve cenno sullo standard MPEG-7 per poi mostrare la struttura descrittiva proposta per annotare l'indicizzazione ed i video objects significativi delle sequenze video.

1.1 MPEG-7 standard

Lo standard MPEG-7 è stato sviluppato per fornire un modo standard per descrivere contenuti multimediali mediante l'utilizzo di metadati in modo tale da potere condividere tali informazioni tra le varie applicazioni.

MPEG-7 contiene strutture per descrivere il contenuto dei file multimediali (shots, oggetti in movimento...), aspetti di contenuto visuale (forma, contorno, colore...) come pure aspetti gestionali del tipo formato del file, dimensioni, frame rate ecc..

The standard MPEG7 è formalmente composto dalle seguenti parti:

1. Systems
 - Specifica strumenti per un'efficiente memorizzazione, trasporto e compressione dei documenti descrittivi.
2. Description Definition Language (DDL)
 - E' il linguaggio basato su XML usato per specificare Ds and DSs.
3. Visual
 - Specifica strumenti specifici per la descrizione del contenuto visuale (Dominante di colore, Texture ecc.)
4. Audio
 - Specifica strumenti specifici per la descrizione del contenuto audio
5. Multimedia Description Schemes (MDS)
 - Specifica strumenti per la descrizione generale audio-video di documenti multimediali
6. Reference Software
 - Fornisce una implementazione software dello standard
7. Conformance
8. Extraction and Use

Gli elementi base dello standard sono sostanzialmente tre:

1. Descriptor (D) usato per descrivere il singolo metadato
 2. Description Scheme (DS)
 - Un DS contiene Ds e altri DSs per definire strutture.
 3. Description Definition Language (DDL)
 - DDL è il linguaggio basato su XML usato per specificare Ds and DSs.
- Tramite il DDL l'utente ha la possibilità di definire Schemi Descrittivi propri non previsti dallo standard.

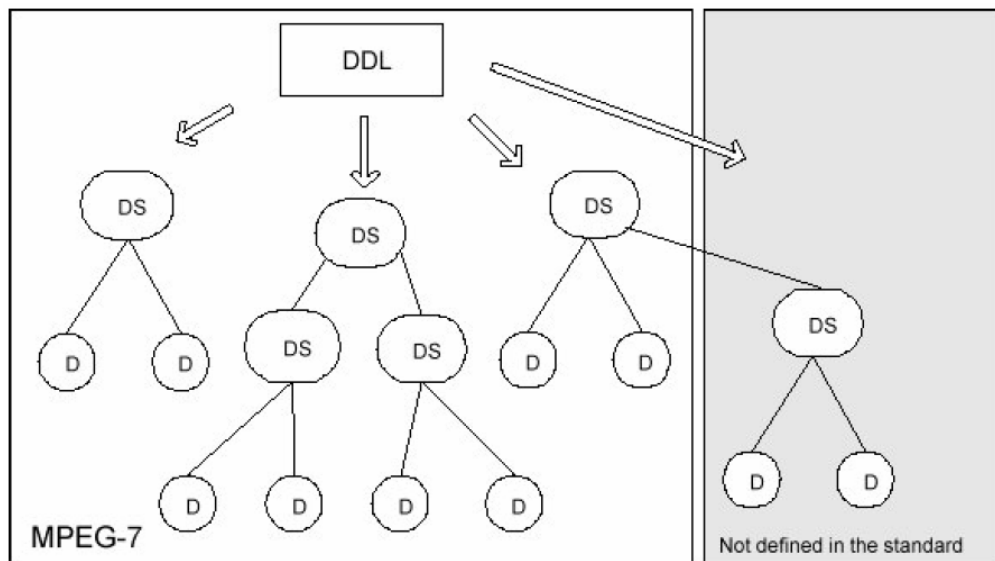


Figura 2: Relazione tra DDL DSs e D in MPEG-7

Nello standard MPEG-7 sono stati definiti un vasto numero di descrittori semplici (D) e di schemi descrittivi (DS) in modo tale da potere annotare un documento multimediale sotto vari aspetti, da quello di carattere generale (titolo, autore, locazione...) a quelli di segmentazione, sommarizzazione, contenuto, semantico e non, agli aspetti di fruibilità da parte di potenziali utenti (profili utente, scenari...); le parti 3 e 4 dello standard trattano gli aspetti più specifici audio e video e contengono schemi descrittivi per descrivere i contenuti di basso livello (forma, colore...).

Il Multimedia Description Scheme contiene un gran numero di Schemi Descrittivi atti a descrivere innumerevoli aspetti di un documento multimediale oltre a una serie di Descrittori e strumenti di base; la figura 3 mostra le parti che compongono l'MDS.

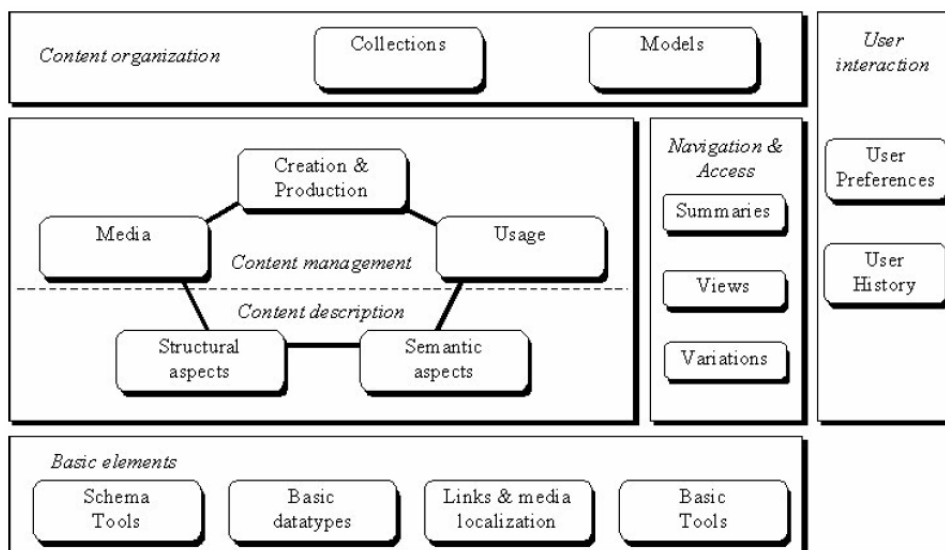


Figura 3: Componenti del Multimedia Description Scheme (MDS)

La struttura descrittiva proposta utilizza un sottoinsieme degli schemi descrittivi definiti nello standard ed, in particolare, utilizzeremo alcune features di basso livello definite nella parte 3 dello standard (Visual) ed un sottoinsieme dei DSs definiti nel Multimedia Description Schemes (MDS) per quanto riguarda le informazioni generali (Content management) e la struttura indicizzata in shots e video objects del filmato (Content description).

1.2 La struttura video proposta

La descrizione del video da noi proposta allo stato attuale è suddivisa in 4 strutture di base e precisamente:

1. CreationInformation
 - Contiene le informazioni generali sul filmato (titolo, formato, dimensioni ecc.)
2. MediaTime
 - Contiene la durata
3. TemporalDecomposition
 - Contiene le informazioni riguardanti la segmentazione in shots
4. SpatialDecomposition
 - Contiene le informazioni riguardanti i video objects

Nella figura 4 è possibile vedere il primo livello strutturale da noi proposto per descrivere una sequenza video.

Mpeg7	
xmlns	urn:mpeg:mpeg7:schema:2001
xmlns:xm1	http://www.w3.org/XML/1998/namespace
xsi:schemaLocation	urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd
xmlns:mpeg7	urn:mpeg:mpeg7:schema:2001
xmlns:xsi	http://www.w3.org/2001/XMLSchema-instance
xsi:schemaLocation	urn:mpeg:mpeg7:schema:2001 C:\DOCUME~1\tripiciano\Documents\Lavoro\MPEG\VIDEO~1\Mpeg7-2001.xsd
DescriptionMetadata	id=ICAR-MT
DescriptionUnit	
xsi:type	VideoType
id	DU-001
Video	
id	FILM-001
xsi:type	AnalyticEditedVideoSegmentType
mediaTimeBase	/Mpeg7/DescriptionUnit/Video/CreationInformation/RelatedMaterial/MediaInformation/MediaProfile/MediaInstance/MediaLocator
mediaTimeUnit	PT1N30F
CreationInformation	
MediaTime	
TemporalDecomposition	id=TD-1 overlap=false gap=false criteria=ShotDecomposition
SpatialDecomposition	id=SD-2 criteria=VideoObjectsDecomposition

Figura 4: Struttura generale della descrizione MPEG-7 proposta

1.2.1 La struttura degli shots

L'indicizzazione temporale in shots è anch'essa suddivisa in 4 strutture di base e precisamente:

1. MediaLocator
 - Contiene le informazioni riguardanti il keyframe dello shot
2. TextAnnotation
 - Contiene informazione testuale libera
3. Semantic
 - Contiene una label associata allo shot
4. MediaTime
 - Contiene le informazioni temporali riguardanti lo shot (inizio, durata)

Nella figura 5 è possibile vedere il primo livello strutturale da noi proposto per descrivere uno shot del video.

TemporalDecomposition	
id	TD-1
overlap	false
gap	false
criteria	ShotDecomposition
VideoSegment	
xsi:type	ShotType
id	VS-SH-1
MediaLocator	xsi:type=ImageLocatorType
TextAnnotation	type=Shot
Semantic	
MediaTime	

Figura 5: Indicizzazione temporale – la struttura proposta per gli shots

1.2.2 La struttura dei video objects

La descrizione dei video objects è suddivisa in 5 strutture di base e precisamente:

1. CreationInformation
 - Contiene informazioni generali sull'oggetto (titolo, classificazione, documenti correlati ecc.)
2. Semantic
 - Contiene tutte le principali informazioni legate alle caratteristiche visuali del video object (labels, keyframe, maschere, colore dominante, texture, ecc.)
3. MatchingHint
 - Contiene i descrittori che specificano il grado di confidenza delle caratteristiche visuali calcolate
4. PointOfView (non implementato)
 - Contiene le informazioni semantiche supplementari del video object (chi, quale oggetto, quale azione, quando, perché, come)
5. SpatioTemporalLocator
 - Contiene le informazioni temporali riguardanti il video object (inizio, durata)

Nella figura 6 è possibile vedere il primo livello strutturale da noi proposto per descrivere un video object.

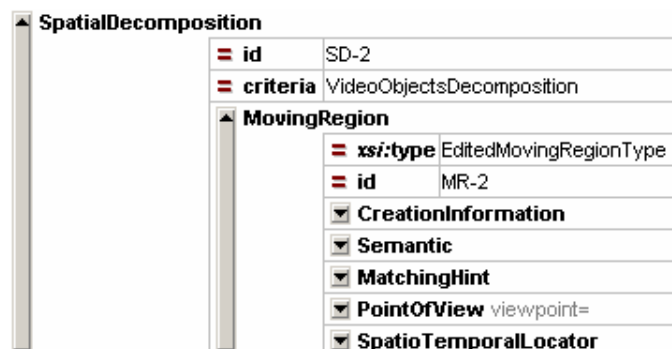


Figura 6: La struttura proposta per i Video Objects

I video object considerati hanno una loro evoluzione spazio-temporale e per la loro descrizione abbiamo scelto di utilizzare uno degli schemi descrittivi base di MPEG-7; il tipo Regione in Movimento (*MovingRegion*). L'oggetto individuato si muove all'interno di una regione per un intervallo di tempo T; lo schema descrittivo scelto permette di definire all'interno della sua struttura per ogni oggetto individuato, sia le informazioni relative alla posizione di tale regione, sia i descrittori visuali dell'oggetto (*DominantColorType* DS ...).

1.3 Considerazioni finali

La struttura gerarchica individuata consente di annotare le informazioni che sono utili e significative per lo sviluppo di applicazioni nell'ambito dei domini applicativi individuati all'inizio di questo rapporto; inoltre grazie alla semplicità e compattezza strutturale dovrebbe essere meno difficoltoso sviluppare un insieme di strumenti software che consentano di gestire, editare, comporre e visualizzare graficamente le strutture descritte.

In questi mesi sono in corso di sviluppo una serie di strumenti che permettano di estrarre da un filmato in modo semi-automatico i video objects ritenuti significativi, calcolarne alcune features visuali ed annotare nella struttura sopra descritta le informazioni ottenute; sono in fase di sviluppo delle API C++ che permettano di accedere ed editare tali documenti descrittivi ed, inoltre, sarà implementato un primo prototipo che permetta di visualizzare graficamente e di navigare i filmati video secondo le strutture annotate nel relativo documento MPEG-7.

1.4 Bibliografia

- [1] World Wide Web Consortium's (W3C) XML web site <http://www.w3.org/XML>.
- [2] Information Technology-Multimedia Content Description Interface – Part 3: Visual (ISO/IEC 15938-3:2002/FPDAmd 1)
- [3] Information Technology-Multimedia Content Description Interface – Part 5: Multimedia Description Schemes (ISO/IEC 15938-5:2003/FPDAmd 1)