

MPEG 4

Ing. Marco Bertini - Università degli Studi di Firenze
Via S. Marta 3 - 50139 - Firenze - Italy
Tel.: +39-055-4796540
Fax: +39-055-4796363
E-mail: bertini@dsi.unifi.it
Web: <http://viplab.dsi.unifi.it/~bertini>

Applicazioni di MPEG4

- Mpeg-4 si propone come standard per quattro diversi contesti applicativi:
 - real-time communication (videoconferenza);
 - televisione digitale;
 - applicazioni grafiche interattive (DVD, ITV);
 - World Wide Web.

■ Scopi di MPEG4

- Indipendenza delle applicazioni dai dettagli di basso livello
- Usabilità su un range di bitrate che vada da pochi kbit/s fino a qualche Mbit/s;
- Riutilizzabilità di tool di codifica e dei dati: l'informazione è rappresentata a livello di singole componenti chiamate "Audio-Visual objects (AVO)";
- Identificazione e gestione dell' IPR dei contenuti (Intellectual Property Right);
- Interattività non solo a livello di video (video A vs. video B) ma tra diversi AVO, stile pagina Web
- Possibilità di avere hyperlink per far interagire più sorgenti di informazione (sempre a livello di AVO), stile pagina web
- Capacità di gestire contemporaneamente informazione naturale/sintetica e real-time/non-real-time
- Capacità di comporre e rappresentare informazione secondo l'interazione con l'utente (paradigma VRML e computer grafica in generale)
 - MPEG4 estende il modello di MPEG2 aggiungendo la capacità di modificare la visualizzazione

Caratteristiche

- Lo standard MPEG-4 fornisce un insieme di tecnologie per soddisfare le necessità di:
 - autori;
 - service providers;
 - utenti finali.
- Lo standard è composto da 22 parti: es. la 2 per la compressione video, la 3 per l'audio, la 14 per il formato file, etc.
 - Alcune parti devono essere ancora finite

Caratteristiche

- Per *autori*, MPEG-4 consente la produzione di contenuto che ha maggiore riusabilità e flessibilità di quella oggi possibile con singole tecnologie come televisione digitale, grafica animata, pagine dal World Wide Web e loro estensioni. Rende anche possibile una migliore gestione e protezione dei diritti di autore sul contenuto.

Caratteristiche

- Per *fornitori di servizi di rete*, MPEG-4 fornisce informazione trasparente che può essere interpretata e tradotta negli appropriati segnali nativi di ciascuna rete.
 - Comunque, esclude considerazioni su Quality of Service, per cui MPEG-4 fornisce un generico insieme di parametri QoS per differenti media MPEG-4.
 - Il mapping esatto per queste traduzioni è al di fuori dello scopo di MPEG-4 ed è lasciato alla definizione dei network provider.

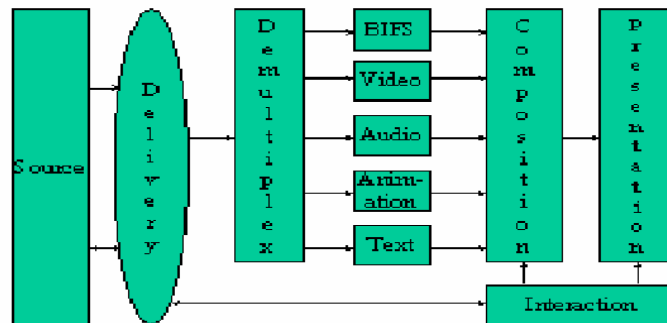
Caratteristiche

- Per *utenti finali*, MPEG-4 rende disponibili molte funzionalità a cui potenzialmente si potrebbe avere accesso da un singolo terminale e più alti livelli di interazione con il contenuto secondo i limiti imposti dall'autore. Tra le applicazioni, comunicazione real-time, sorveglianza e multimedia mobile.
- Per tutte le parti coinvolte, MPEG vuole evitare la proliferazione di una moltitudine di formati proprietari e formati non di rete.

Lo standard MPEG4

- Lo standard MPEG4 comprende le tecnologie per supportare:
 1. rappresentazione (codificata) per unità di contenuto aurale, visuale e audiovisuale (audio-visual objects" o AVO);
 2. il modo in cui gli AVO sono composti in una scena;
 3. il multiplexing e la sincronizzazione di AVO, per trasportarli su canali con QoS adeguato alla natura degli AVO (o QoS adeguato all'utente);
 4. Un'interfaccia generica tra applicazione e trasporto
 5. Il modo in cui un utente interagisce con la scena (es. punto di vista) e gli oggetti che la compongono (es. click su di un oggetto)
 6. Proiezione delle scene AV composte secondo il punto di vista/ascolto desiderato

Da MPEG2 a MPEG4

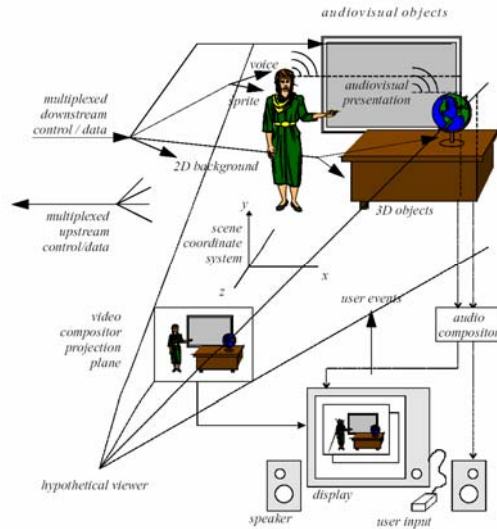


- Le scene audiovisive sono composte da diversi AVO, organizzati gerarchicamente, es.:
 - Sfondo fisso 2D
 - Immagine di una persona che parla (senza lo sfondo)
 - Voce associata alla persona
 - Oggetto sintetico (tavolo e mappamondo)
 - Suono sintetico (es. musica della sigla)
 - Etc.
- MPEG standardizza un insieme di primitive AVO per rappresentare oggetti naturali ed artificiali, 2D e 3D

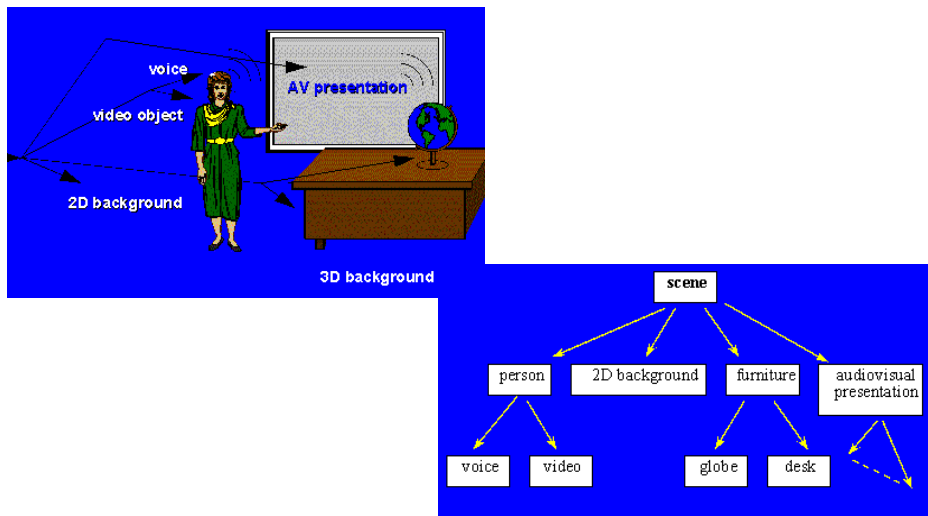
- La descrizione di una scena consiste delle seguenti informazioni:
 - Come gli oggetti sono raggruppati insieme (struttura gerarchica)
 - Come gli oggetti sono posti nello spazio e nel tempo: ogni oggetto ha coordinate locali che vengono rimappate sulle coordinate della scena, e l'apparizione degli AVO deve essere sincronizzata
 - Attributi degli AVO: alcuni elementi possono essere modificabili (es. contrasto o colore)

- BIFS: Binary Format for Scene description
 - Linguaggio binario (derivato da VRML) per la descrizione delle scene
 - La descrizione della scena è codificata in modo indipendente dagli stream dei media primitivi
 - Se il parametro può essere usato per la codifica dell'oggetto (es. parametri di moto) allora NON viene descritto nello stream BIFS

Esempio di scena MPEG4



Composizione della scena



Composizione della scena

- La composizione della scena consente di:
 - eliminare, spostare e posizionare a proprio piacere gli AVOs che compongono la scena stessa;
 - raggruppare diversi AVOs in modo da formare AVOs composti, così da consentire al fruitore di manipolare insieme consistenti di oggetti;
 - fornire dati di animazione agli AVOs in modo da personalizzarne e modificarne gli attributi (applicare una particolare tessitura ad un oggetto, mandare parametri di animazione ad un volto ecc.);
 - cambiare interattivamente il punto di vista nella scena.

Composizione della scena

- Il supporto di queste funzionalità è ottenuto attraverso una composizione e definizione della scena che adotta molte delle soluzioni ed impostazioni definite dallo standard VRML.
- Sono comunque eliminati molti vincoli che rendono troppo rigido lo standard VRML.
In particolare Mpeg-4 si propone (in più rispetto al VRML) di:
 - codificare informazione audio-video ad elevata qualità;
 - fare riferimento ad una scena dinamica che può essere cambiata in ogni momento.

Composizione della scena

- È previsto un livello di interazione con l'utente finale che consente a quest'ultimo di:
 - navigare nella scena;
 - eliminare o spostare gli oggetti all'interno della scena;
 - avviare l'esecuzione di alcune azioni (ad esempio selezionare un oggetto e provocare il play di un filmato);
 - selezionare uno tra più linguaggi nel caso siano presenti diverse tracce audio.

Raggruppamento degli oggetti

- Una scena MPEG-4 segue una struttura gerarchica che può essere rappresentata come un grafo diretto aciclico. Ogni nodo del grafo è un AVO.
 - La struttura non è necessariamente statica; gli attributi dei nodi (ad esempio parametri di posizione) possono essere cambiati mentre i nodi possono essere aggiunti, rimpiazzati o rimossi.

Posizionamento nello spazio e nel tempo

- Nel modello MPEG-4, oggetti audiovisivi hanno una estensione sia spaziale che temporale. Ciascun AVO ha un sistema di coordinate locale.
 - Un sistema *locale di coordinate*, per un oggetto è un sistema in cui l'oggetto ha una locazione spazio-temporale e una scala fissate.
 - Il sistema di coordinate locale serve per manipolare l'AVO nello spazio e nel tempo. AVO sono posizionati in una scena specificando una trasformazione di coordinate dal sistema di coordinate locale all'oggetto, in un sistema di coordinate globale definito da uno o più nodi di descrizione della scena nell'albero.

Selezione del valore degli attributi

- Singoli AVO e nodi descrittivi della scena espongono un insieme di parametri al layer di composizione attraverso i quali parte del loro comportamento può essere controllato.
 - Esempi includono la tonalità di un suono, il colore di un oggetto sintetico, l'attivazione o disattivazione di informazioni di accrescimento per codifica scalabile, etc.

Altre trasformazioni

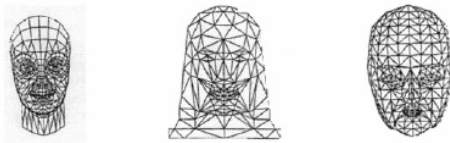
- La struttura di descrizione della scena e la semantica dei nodi sono fortemente influenzate dal VRML, incluso il suo modello degli eventi. Questo rende disponibili a MPEG-4 un ampio insieme di operatori di costruzione della scena, includendo primitive grafiche che possono essere usate per costruire scene sofisticate.

Audio Visual Objects (AVOs)

- MPEG4 consente di codificare figure di forme diverse (non solo rettangolari)
 - Il video diventa una composizione di oggetti 2D
 - Possono essere disposte in uno spazio 3D

Audio Visual Objects (AVOs)

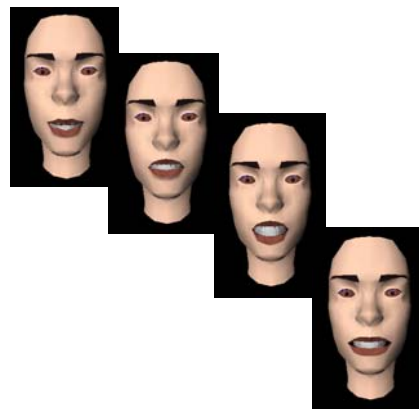
- Oltre alle primitive viste sopra (presenti nella figura) esistono AVO per:
 - Testo e grafica
 - talking heads e testo associato: usato dal ricevente per sintetizzare il parlato ed il movimento della testa
 - Animazione del corpo umano
 - Audio sintetico (es. TTS)



Audio Visual Objects (AVOs)

AVOs sintetici includono:

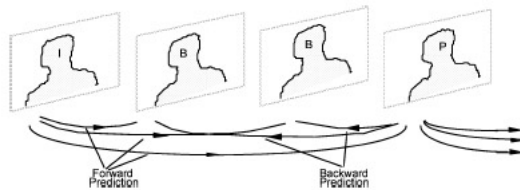
- volti animati;
- corpi animati;
- mesh 2D animate.



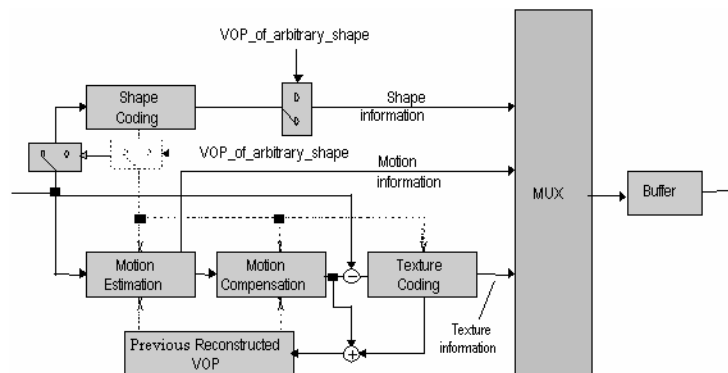
Rappresentazione codificata dei media objects

- Nella loro forma codificata, questi oggetti sono rappresentati nel modo più efficiente possibile.
 - La codifica di questi oggetti è il più efficiente possibile, anche tenendo conto che si supportano diverse funzionalità come la robustezza al rumore, estrazione ed editing di un oggetto o l'avere un oggetto disponibile in forma scalabile.
 - È importante notare che nella loro forma codificata, gli oggetti (audio o video) possono essere rappresentati indipendentemente da quelli circostanti e dalla sfondo.

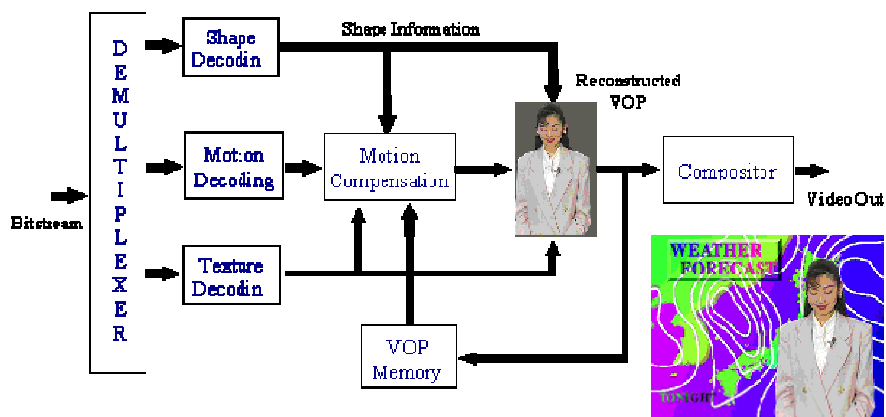
- MPEG-4 usa il concetto di Visual Objects (VOs)
- I Video Object Planes (VOPs) sono istanze di VOs in una sequenza video
- MPEG-4 usa I-VOP, P-VOP, B-VOP



- L'encoder è composto da due parti: lo shape coder e il coder tradizionale di motion e texture



Esempio di decodifica video



Interazione utente

- MPEG-4 consente l'interazione dell'utente con il contenuto presentato. Questa interazione può essere distinta in:
 - interazione lato server;
 - interazione lato client.
- L'interazione lato server, implica manipolazione del contenuto che occorre al lato trasmittente iniziata da una azione dell'utente.

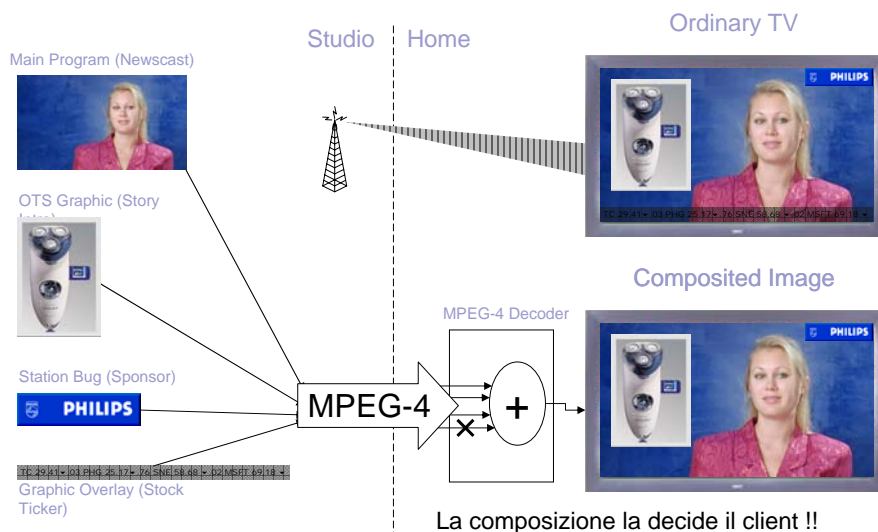
Interazione lato client

- L'interazione lato client implica manipolazione del contenuto localmente al terminale dell'utente finale e può assumere diverse forme:
 - modifica dell'attributo del nodo di descrizione di una scena (ad esempio, modifica della posizione di un oggetto, renderlo visibile o invisibile, modifica della dimensione del font per un nodo di testo sintetico) può essere implementata traducendo eventi dell'utente (click del mouse o comandi da tastiera) ad aggiornamenti della descrizione della scena.
 - I comandi possono essere elaborati dal terminale MPEG-4 esattamente nello stesso modo come se fossero stati originati dalla sorgente originale di contenuto.

Esempi di composizione e object coding

- Quando sono utili object coding e composizione ?
 - La segmentazione è ancora difficile !
 - Due esempi di Philips...

MPEG-4 Object Coding



MPEG-4: Applicazione Pay-per-View

Comprate la palla !



Demo Philips : la palla è codificata come oggetto separato. Il resto della scena invece è un video rettangolare gratuito. Si paga per vedere la palla.

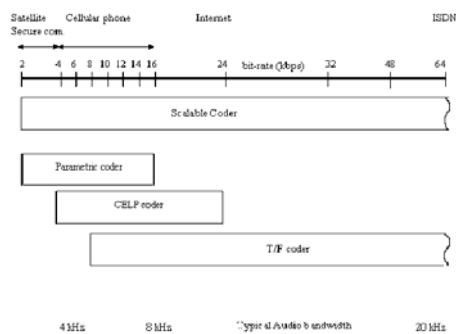


C'è già hardware dedicato per fare real-time tracking della palla, per il broadcast di video sportivi. E comunque si potrebbe fare anche in software. È un caso reale, già disponibile

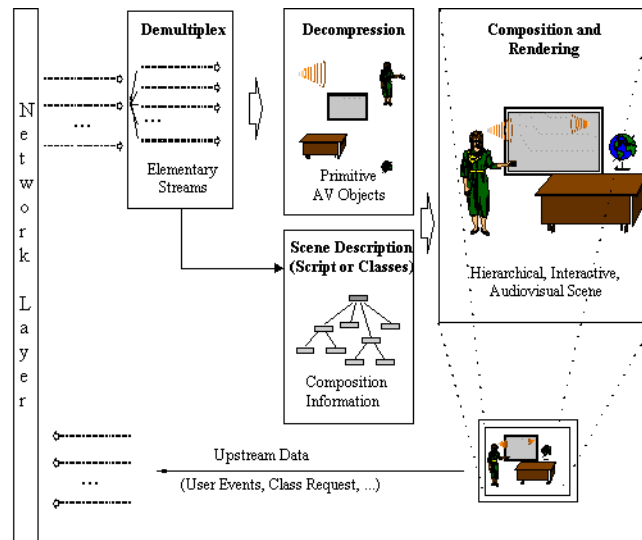
L'audio

- Sono definiti diversi algoritmi di compressione audio, secondo il bitrate che si vuole ottenere

- È definito anche un framework per rendere scalabile la compressione audio



La decodifica di MPEG4



Componenti principali

- Stream provenienti dalla rete (o da un dispositivo di memoria) sono demultiplexati per ottenere stream elementari.
- Gli stream elementari (ESs) sono esaminati e passati agli appropriati decodificatori.
 - La decodifica recupera il dato in un AVO dalla sua forma codificata ed esegue le operazioni necessarie per ricostruire l'AVO originale pronto per il rendering sul dispositivo appropriato.

Componenti principali

- L'AVO ricostruito è reso disponibile al layer di composizione per il suo uso potenziale durante il rendering della scena.
- Gli AVOs decodificati, insieme con informazione di descrizione della scena, sono usati per comporre la scena come descritto dall'autore.
- L'utente finale può interagire con la scena, secondo quanto stabilito dall'autore, che è eventualmente renderizzata e presentata.

Parte visual di MPEG-4

- Fornisce soluzioni nella forma di tools e algoritmi per:
 - compressione efficiente di immagini e video;
 - compressione efficiente di tessiture per il mapping di tessiture su mesh 2D e 3D;
 - compressione efficiente di stream geometrici variabili nel tempo che animano le mesh;
 - accesso casuale efficiente a tutti i tipi di oggetti visuali;
 - funzionalità di manipolazione estese per immagini e sequenze video;
 - codifica basata sul contenuto di immagini e video;
 - scalabilità basata sul contenuto di tessiture immagini e video
 - scalabilità spaziale, temporale e di qualità;
 - robustezza all'errore.

Oggetti sintetici

- Formano un sottoinsieme della classe più ampia di oggetti ottenuti con grafica computerizzata. Consideriamo i seguenti VOs sintetici:
 - rappresentazione parametrica di
 - descrizione sintetica di faccia e corpo umano
 - stream di animazione della faccia e del corpo
 - codifica di mesh statica e dinamica con mapping delle tessiture;
 - codifica della tessitura per applicazioni dipendenti dall'osservatore.

Animazione facciale

- Il volto è un AVO in grado di poter essere visualizzato ed animato.
 - Forma, tessitura ed espressione del volto sono controllate da uno stream di dati contenenti insiemi di *Facial Definition Parameters* (FDPs) e *Facial Animation Parameters* (FAPs).

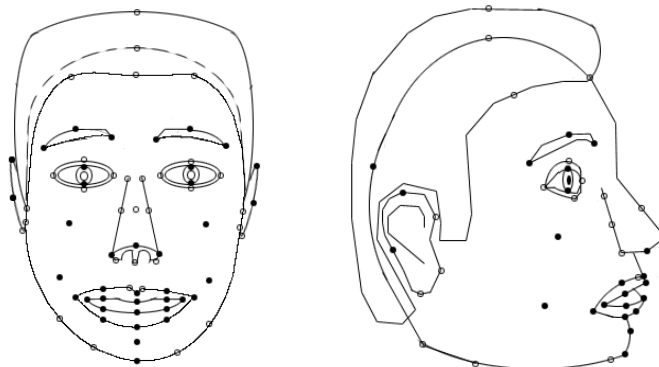


- Appena istanziato un volto rappresenta una generica faccia priva di espressioni particolari.

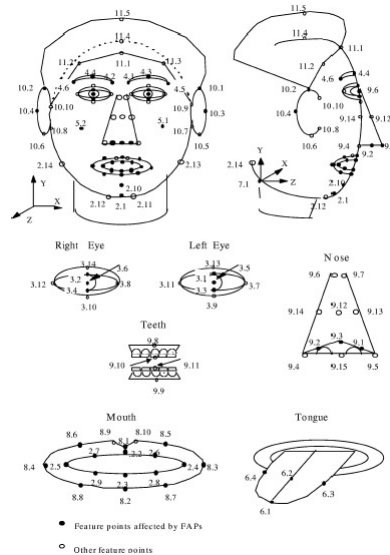
Animazione facciale

- Il volto può essere animato attraverso i parametri di animazione (FAP).
- I FDP possono essere utilizzati per personalizzare l'aspetto del volto:
 - è possibile passare da una generica faccia ad una con particolari caratteristiche in termini di forma e tessitura.
- Localmente è consentito all'utente di interagire con il modello del volto:
 - ad esempio è possibile amplificare i movimenti della bocca e rendere così più facile l'interpretazione labiale del parlato.

Face Animation



Face Animation - 2



Face Animation - 3

- Visemes e Expressions
- I Visemes sono combinazioni pre-impostate di FAP
- I Visemes sono usati per il parlato
- Le Expressions sono usate per esprimere emozioni

1	2	3	4	5	6	7	8	9	10	11	12	13	14
(p,b,m)	(f,v)	(T,D)	(t,d)	(k,g)	(tS,dZ,S)	(s,z)	(n,l)	(r)	(A:)	(e)	(l)	(Q)	(U)
put	far	think	tip	call	chair	sir	lot	red	car	bed	tip	top	book
bed	voice	that	doll	gas	join	zeal	not						
mill					she								

Corpi animati

- Il corpo è un AVO in grado di poter essere visualizzato ed animato.
- Forma, tessitura e postura del corpo sono controllate da uno stream di dati contenenti insiemi di *Body Definition Parameters* (BDPs) e *Body Animation Parameters* (BAPs).
- Appena istanziato un corpo si presenta in modo eretto con le braccia allineate ai fianchi.
 - Il corpo può essere animato attraverso parametri di animazione (BAP).
 - I BDP possono essere utilizzati per personalizzare le caratteristiche del corpo (altezza, larghezza spalle etc.).

Mesh 2D animate

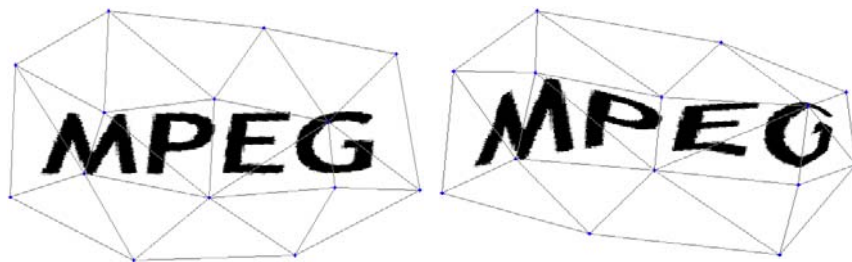
- Una mesh 2D è una tassellazione di una regione planare in elementi poligonali.
 - I vertici dei poligoni sono chiamati *node points* della mesh.
- Lo standard Mpeg-4 prevede l'impiego di soli elementi triangolari per la definizione delle mesh.
- Una mesh 2D può essere resa dinamica attraverso la specifica della mesh iniziale e dei vettori di moto dei *node points* all'interno di un intervallo temporale.

Mesh 2D animate

- Le mesh dinamiche sono di particolare utilità per modellare tessiture animate.
- Lo standard prevede infatti di poter applicare una tessitura ad una mesh:
 - quando i node points della mesh vengono spostati dai vettori di moto, alla tessitura di ogni elemento triangolare viene applicato un mapping parametrico che porta ad un warping della tessitura in modo da rimanere ancorata alle nuove coordinate dei node points.

Object Mesh Coding

- Si fa il tracking solo dei nodi
- Warping per creare nuove immagini



Indicizzazione video basata su contenuto

- La rappresentazione con mesh consente la creazione di snapshot chiave animati per la sintesi visuale del movimento di oggetti.
- La rappresentazione con mesh fornisce informazione accurata sulla traiettoria di un oggetto che può essere usata per ritrovare VOs con moto specifico.
- La rappresentazione con mesh fornisce rappresentazione di forma di un oggetto basata su vertici che è più efficiente della rappresentazione con bitmap per la ricerca di oggetti basata sulla forma.

Manipolazione di VOs

- Realtà aumentata:
 - fonde insieme immagini virtuali (generate da computer) con immagini reali in movimento (video) per creare informazione di visualizzazione migliorata.
- Trasformazione/animazione di oggetti sintetici:
 - sostituzione di un oggetto video naturale in un video clip con un altro oggetto video. L'oggetto video di sostituzione può essere estratto da un altro video clip naturale o può essere trasformato da una immagine singola usando l'informazione di moto dell'oggetto da rimpiazzare (richiede una rappresentazione temporalmente continua del moto).
- Interpolazione spazio-temporale:
 - modellazione del moto delle mesh fornisce una interpolazione temporale compensata dal moto più robusta.

Compressione di VOs

- La modellazione di mesh 2D può essere usata per la compressione se è scelto di trasmettere mappe di tessitura solo a key-frame selezionati e animare queste mappe di tessitura per i frames intermedi.

Mesh 3D generiche

- Mpeg-4 supporta anche mesh generiche per rappresentare oggetti 3D sintetici.
 - Queste mesh supportano proprietà come colore, normali per effetti di ombreggiatura, coordinate di tessitura per il mapping di tessiture naturali, immagini e video sulle mesh.

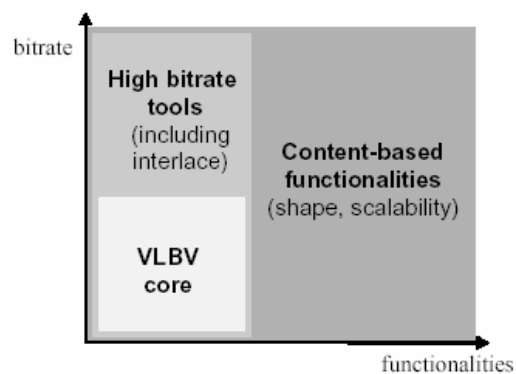
Mesh 3D generiche

- Sono forniti algoritmi per:
 - compressione efficiente di mesh generiche;
 - (*livello di dettaglio, LOD*) *scalabilità di 3D meshes* - consente al decoder di decodificare un sottoinsieme dello stream di bit totale per ricostruire una versione semplificata della mesh contenente meno vertici dell'originale. Tali rappresentazioni semplificate sono utili per ridurre il tempo di rendering di oggetti che sono distanti dall'osservatore (gestione del LOD), e consente anche una minor potenza dei motori di rendering per renderizzare l'oggetto ad una qualità ridotta.
 - *Scalabilità spaziale* - consente al decoder di decodificare un sottoinsieme dello stream totale di bit generato dal codificatore per ricostruire la mesh ad una ridotta risoluzione spaziale. Questa caratteristica è più utile quando usata combinata con la scalabilità LOD.

La compressione

- La base della compressione MPEG4 è quella di MPEG1 e 2:
 - Sequenze di frame rettangolari a vari:
 - Livelli di bitrate
 - Livelli di frame rate
 - Formati di input
 - Scalabilità di qualità
 - Scalabilità spaziale
 - Scalabilità temporale

- Progressive & interlaced
- SQCIF/QCIF/CIF/4*CIF/CCIR 601, fino a 2048*2048
- Y/Cb/Cr/Alpha
- Inizialmente 4:2:0, in futuro 4:2:2 e 4:4:4 per qualità studio
- Continuous variable frame rate



La figura fornisce una classificazione di base per i bit rates e le funzionalità attualmente fornite dallo standard Visual di Mpeg-4 per immagini naturali e video.

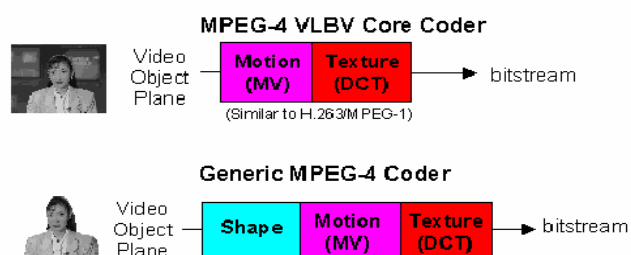
- VLBV: Very Low Bit-rate Video: algoritmi e tool per applicazioni che operano a bitrate tra 5 e 64 kbits/s:
 - sequenze di immagini a bassa risoluzione spaziale (max. CIF)
 - basso frame rate (fino a 15 fps).

VLBV Core

- Codifica di sequenze di immagini rettangolari
- Accesso diretto, fast forward e reverse
- Queste funzioni sono previste anche per bitrate più elevati, con input fino a ITU-Rec.601

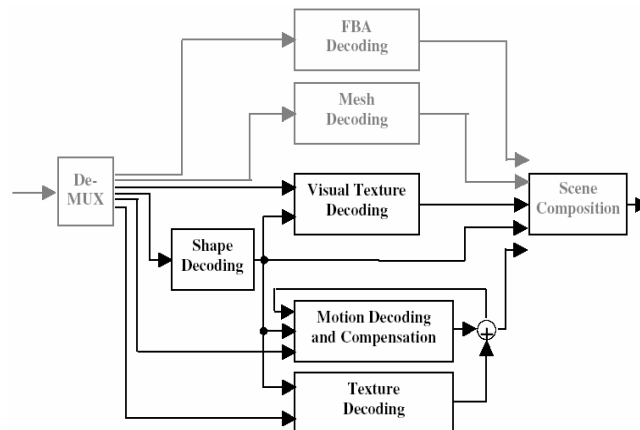
Funzioni Content-based

- (de)codifica di contenuti
- Interattività
- rappresentazione flessibile
- manipolazione di VO nel dominio compresso
- Complemento della compressione VLBV e HBV



- La shape può essere codificata come canale alfa 8 bit) o bitmask

Coding



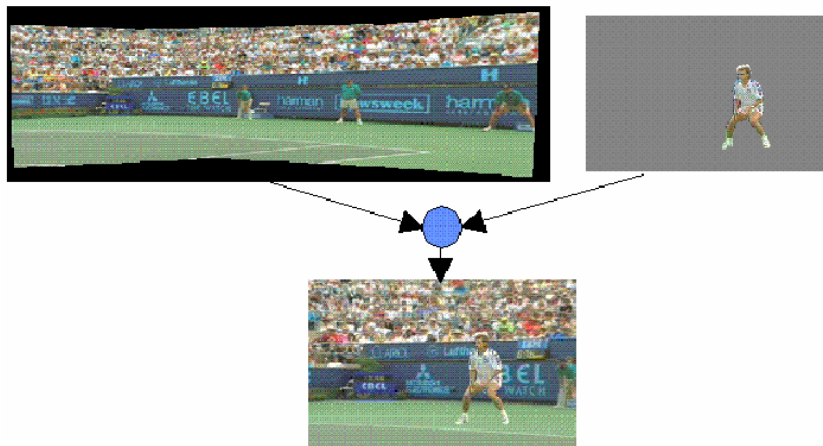
- Schema generale della codifica video MPEG4

- Si possono codificare immagini o sequenze video usando gli stessi tool:
 - Visual texture decoding: immagini
- Se non si tiene conto della shape si ha l'encoder basato su Motion compensation e Texture decoding:
 - Come MPEG1 e MPEG2
 - Anche in MPEG4 si usano i blocchi!

- La struttura base del codificatore include codifica di forma (per VOs di forma arbitraria), codifica di tessitura basata su compensazione di moto e DCT (utilizzando DCT 8x8 standard o DCT adattativa alla forma).
- Un importante vantaggio dell'approccio di codifica basato sul contenuto è che l'efficienza di compressione può essere migliorata in modo significativo per alcune sequenze video utilizzando tecniche di predizione del moto appropriate e dedicate basate su oggetti, per ciascun oggetto nella scena.

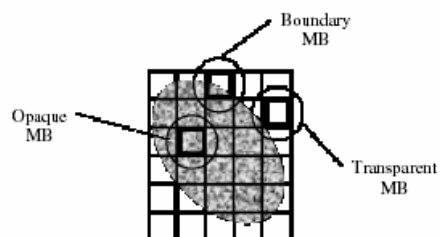
- È l'aggiunta dello shape decoding che consente la creazione della codifica content-based
 - Non solo forme rettangolari...
 - ... e comunque anche quando si usa codifica object-based l'analisi dei frame è basata su blocchi!
 - La codifica delle forme rettangolari è comunque migliorata: bitrate più bassi a parità di qualità

- Uno dei vantaggi della codifica content-based è che si può adattare la compressione secondo l'oggetto:
 - Es. background più compresso del foreground
 - Trasmissione di un'immagine panoramica (e parametri di moto della camera) + oggetto foreground come altro VO



- L'algoritmo da usare per determinare la forma dell'oggetto da codificare non è definito, solo il bitstream per rappresentarlo
 - Ne esistono diversi completamente automatici, o assistiti

- Viene usata la bounding box dell'oggetto da codificare, dividendola in MB di 16x16. Eventualmente si ingrandisce la b.box per renderla divisibile in MB

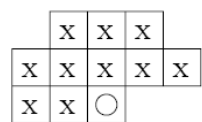


Shape Coding

- Binary alpha maps
- Bounding box che copre il VOP
- BB divisa in 16X16 pixel (binary alpha blocks: BAB)
- Content based arithmetic encoding (CAE) è usata per il coding

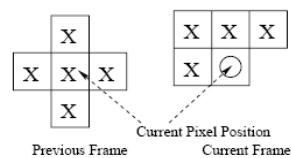
Shape Coding - 2

- Si considera un alto grado di correlazione tra pixel. Ogni pixel è codificato secondo una probabilità condizionale dipendente dal suo contesto



○ Current Pixel

(a) Template for Intra Mode



(b) Template for Inter Mode

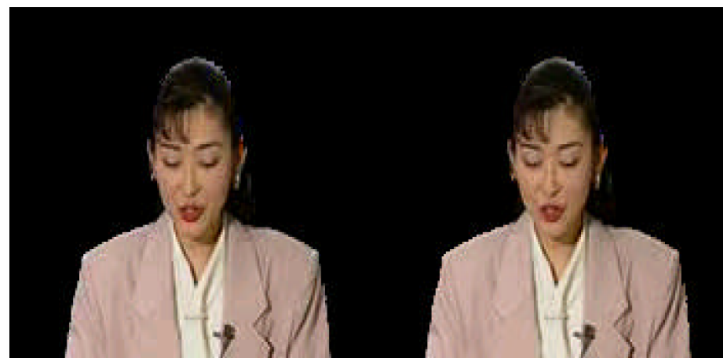
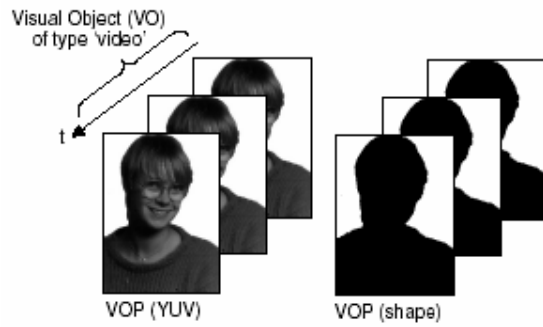
Shape Coding - 3

- Grayscale alpha maps
- Rappresentano il grado di trasparenza dei pixel
- 0 è trasparente, 255 è opaco
- Può essere usato in video ibrido (naturale/sintetico)

Motion Coding

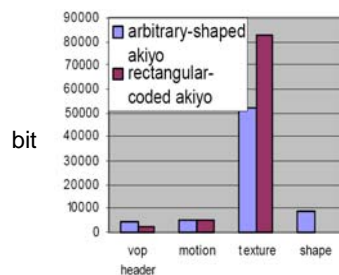
- P-VOPs stimati usando il VOP corrente e il VOP precedente
- B-VOPs stimati usando il VOP corrente ed i VOP circostanti
- Motion Block Predictor
- Si creano motion vector (codificandone le differenze tra I circostanti) per blocchi di 8X8 or 16X16 pixel

- Esempio di VO video, composto da Video Object Planes
- Nota: se siamo nel caso in cui l'unico VO è il frame rettangolare allora VOP = frame di MPEG2
 - È quello che fanno in realtà la maggior parte di codec in commercio



Arbitrary shaped

Rectangular coding



- La maggiore compressione di MPEG4 deriva da:
 - Motion compensation più precisa, si usa una risoluzione di $\frac{1}{4}$ di pixel (interpolando, ovviamente...)
 - Global motion compensation: si usa un solo set di parametri per l'intero VOP; tiene conto del moto della telecamera, usando quindi pochi parametri. Funziona bene nel caso di immagini senza oggetti in movimento.
 - Un miglioramento della predizione del moto nei B-VOP

- Anche la codifica DCT è stata cambiata:
 - Due sistemi di quantizzazione diversi, uno stile MPEG2 ed uno che non usa tabella di quantizzazione (stile H.263)
 - Predizione di coefficienti AC/DC in intraMB
 - Due nuove scansioni della matrice risultante

H.264/MPEG-4 AVC/MPEG-4 Part 10

- La parte 10 dello standard MPEG-4 (ISO/IEC 14496-10) è uno standard di codec video che mira alla massima compressione possibile. La parte è nota anche come H.264
 - L'idea è mantenere la stessa qualità della compressione MPEG-4 parte 2 a metà del bitrate
 - È comunque abbastanza flessibile da essere usato anche in media come DVD:
 - Gli standard futuri come HD-DVD e Blu-ray Disc lo prevedono nei player
 - È previsto il suo uso anche in ambito broadcast

MPEG-4 AVC: caratteristiche

- Multi-picture motion compensation: fino a 32 immagini di riferimento (anziché le due dei B frame); utile in casi particolari come flash e scene ripetute.
- Weighted prediction dei vettori di moto (vedi lucidi MPEG1)
- Variable Block-size motion compensation: i blocchi vanno da 16x16 a 4x4 pixel, consentendo segmentazione precisa delle regioni dell'immagine
- Context-adaptive binary arithmetic coding (CABAC): tecnica lossless per la compressione di elementi sintattici dello stream
- Context-adaptive variable-length coding (CAVLC): tecnica più semplice di CABAC per la compressione dei coefficienti quantizzati delle trasformate, più efficienti dei meccanismi standard MPEG
- Nuova tecnica per il VLC: Exponential-Golomb (Exp-Golomb) coding, da usare quando non si usa CABAC o CAVLC
- Miglioramenti vari nelle tecniche utili per correzione errori e streaming
- In media la stessa qualità di MPEG2 viene ottenuta a metà della banda necessaria

H.264/MPEG-4 AVC/MPEG-4 Part 10

		Standards		
		MPEG-2	MPEG-4 ASP	MPEG-4 H.264
Features	I, P, B-frames	✓	✓	✓
	Interlace	✓	✓	✓
	Coding	Huffman	Huffman	Huffman or Arithmetic
	Block size	fixed 16x16	fixed 16x16	variable down to 4x4
	¼ pixel		✓	✓
	GMC		✓	
	Loop Filter (aka deblocking filter)			✓
	Slice-based motion prediction			✓
	Multiple reference frames			✓
	MB AFF (improved interlaced management)			✓
	RDO (Rate Distortion Optimisation)			✓
	WP (Weighted Prediction)			✓
	Switching pictures (for fast change channel)			✓

Codifica tessiture e immagini

- Le immagini e le texture da applicare su mesh sono codificate con wavelet:
 - Ottima compressione
 - Scalabilità
 - Codifica shape di forma arbitraria

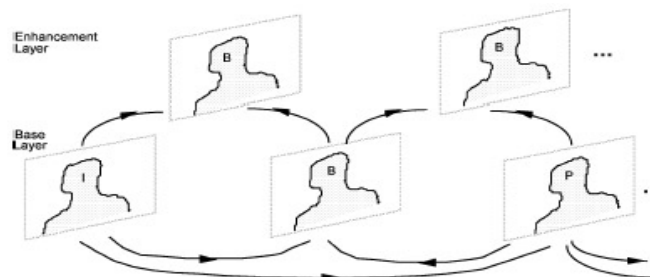
Altre funzionalità

- La scalabilità spaziale, temporale e di qualità è a livello di VO
- Correzione e mascheramento degli errori



Scalable Video Coding

- Spatial Scalability
- Temporal Scalability



Robust Video Coding

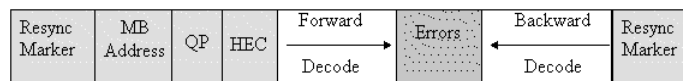
- Resynchronization
 - metto marker nello stream ed eventualmente salto i bit corrotti fino al prossimo
- Resynchronization markers tra le righe dei macroblocchi (GOBs)
- La posizione dei marker è arbitraria, ma usualmente a intervalli uniformi (così sappiamo dove cercarli)
- Non c'è dipendenza dei dati tra i pacchetti

Robust Video Coding - 2

- Data Partitioning
- Localizza l' errore dentro un video packet
- I dati di moto e DCT sono separati
- All'interno dei dati di moto e DCT, i dati sono partizionati

Robust Video Coding -3

- Reversible Variable Length Codes
- VLC: se trovo un errore il resto del pacchetto video deve essere scartato
- RLC: se trovo un errore il decoder legge all'indietro a partire dal prossimo marker fino a ritrovare l'errore
 - I codici sono meno efficienti ma più sicuri



Implementazione

- L'implementazione di sistemi MPEG4 è piuttosto complicata: svariate migliaia di pagine di standard...
 - Ricerca attiva su alcuni algoritmi: es. tracking di oggetti all'interno di un video per estrarre VO

- Il formato dei file (basato su Quicktime) è *streamable*:

- È adatto sia per uso locale che per lo streaming in rete
 - Fornisce informazioni al server di streaming mediante metadati contenuti nel file, è compito del server di streaming usarli all'interno di un protocollo di streaming

- Profili: subset delle funzioni di MPEG4, usati per applicazioni specifiche

- Ad ogni profilo si può associare un livello che limita la complessità computazionale
- Come in MPEG2: non importa implementare tutte le funzioni dello standard, mi posso limitare ad un profilo

Es.: profili visuali

- Es.: simple profile, codifica di oggetti visuali rettangolari con correzione di errori, adatto per applicazioni su terminali mobili
- Simple scalable profile: aggiunge il supporto per la scalabilità temporale e spaziale al simple profile. Adatto per servizi su Internet
- Core profile: aggiunge il supporto per oggetti di qualsiasi forma, scalabili temporalmente, al simple profile

■ Esistono profili anche per:

- animazioni facciali;
- audio;
- mesh;
- grafica;
- scene;
- oggetti (IPR, object descriptor, etc.)

MPEG4 - Profiles e Levels

Visual Tools	Simple	Core	Main	Simple Scalable
Basic	X	X	X	X
• I-VOP				
• P-VOP				
• AD/DC Prediction				
• 4-MV				
Unrestricted MV				
Error resilience	X	X	X	X
Short Header	X	X	X	
B-VOP		X	X	X
P-VOP				
Method 1/Method 2 Quantisation		X	X	
P-VOP based temporal scalability		X	X	
• Rectangular				
• Arbitrary Shape				
Binary Shape		X	X	
Grey Shape			X	
Interface			X	
Sprite			X	
Temporal Scalability (Rectangular)				X
Spatial Scalability (Rectangular)				X
N-Bit				
Scalable Still Texture				

Codec basati su MPEG4

- Divx:
 - Nato da reverse engineering di implementazione Microsoft, per attivare funzioni altrimenti disattivate
- Opendivx / XVid:
 - Implementazione open-source
 - Simple profile e ASP