

Analisi della qualità video per applicazioni webcasting

ing. Paola **Sunna Rai**
Centro Ricerche e
Innovazione Tecnologica
Torino

Presidente del gruppo
EBU/UER B/VIM

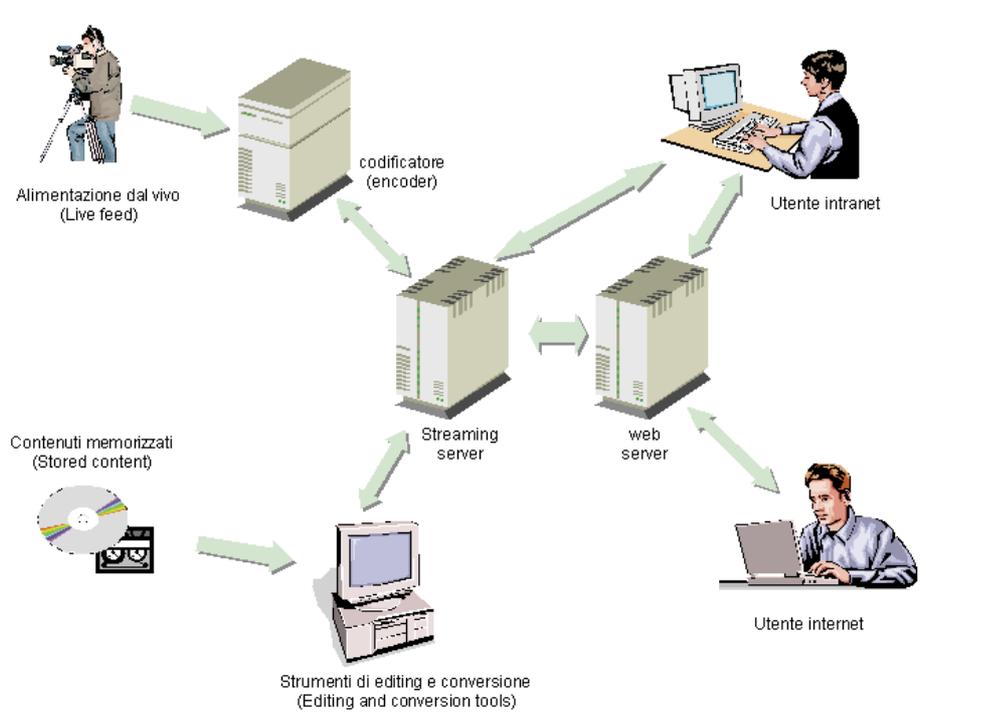
1. Introduzione

Il termine webcasting è utilizzato per indicare la produzione e trasmissione (streaming, download, video on demand) di contenuti audio, video, ... fruibili dagli utenti finali su terminali (PC, Palmari, PDA...) connessi ad Internet.

Una architettura tipica per il webcasting è riportata in figura 1.

I contenuti audio-video possono essere ripresi dal vivo ("live") oppure essere stati acquisiti e sottoposti ad editing per la creazione di presentazioni "ad hoc" per il web; la codifica è eseguita tramite opportune tecniche di compressione allo scopo di consentire la trasmissione nella banda a disposizione (modem, xDSL, fibra ottica) mentre la distribuzione verso gli utenti finali avviene utilizzando appositi server (web server o server di streaming).

Fig. 1 - Architettura Webcasting



Sommario

Nel 2001, il gruppo EBU B/VIM (Video in Multimedia) è stato creato allo scopo di studiare e definire una metodologia soggettiva per la valutazione della qualità video nell'ambito di applicazioni multimediali. Il contesto si riferisce all'utilizzo di codec per il web (Windows Media, Real Video, Mpeg4.) che, effettuando la compressione del segnale video per consentirne la trasmissione nelle bande tipicamente disponibili su Internet, introducono delle degradazioni sulla qualità fruita dall'utente finale. L'articolo descrive la metodologia proposta dal gruppo EBU, il tipo di test eseguiti ed i risultati preliminari ottenuti.

Il presente articolo riporta i risultati preliminari ottenuti durante l'esecuzione di una serie di test mirati ad ottenere indicazioni sulla qualità fornita dai principali codec diffusi su web.

Generalmente la valutazione della qualità degli algoritmi di codifica si effettua tramite prove soggettive formali oppure attraverso prove expert viewing; per esempio, la Raccomandazione ITU-R BT.500 [1] riporta le metodologie adottate per la valutazione soggettiva della qualità video in ambito di applicazioni broadcasting mentre, tuttora, non esistono Raccomandazioni che indichino i criteri con cui valutare la qualità dei codec usualmente utilizzati per applicazioni multimediali.

2. B/VIM (Video in Multimedia)

Nel 2001, un nuovo gruppo EBU, il B/VIM ^{Nota 1}, è stato creato allo scopo di studiare e definire una metodologia soggettiva per la valutazione della qualità video nell'ambito di applicazioni di webcasting.

I codec introducono dei degradamenti più o

meno visibili in funzione del bit-rate di codifica e della complessità del materiale sorgente; i degradamenti possono consistere in perdite di frame con conseguente percezione a scatti del movimento, presenza di blocchettizzazione, perdita di risoluzione e così via. Metodologie oggettive basate sulla valutazione del rapporto segnale rumore non forniscono informazioni relative al degradamento sufficientemente correlate con la qualità percepita dall'osservatore, per cui il metodo più attendibile per ottenere un'indicazione delle prestazioni di nuovi algoritmi di compressione o di nuovi codec resta quello delle prove soggettive.

La necessità di formulare una nuova metodologia, differente rispetto a quelle riportate nella [1], nasce dal fatto che il terminale di fruizione non è più un CRT (Cathode Ray Tube), ma diventa un PC e la risoluzione dei formati non è quella televisiva (Full Format 720x576), ma si riduce (CIF oppure QCIF) per far fronte alla ridotta banda disponibile per le connessioni Internet (Modem, ISDN, ADSL...).

La metodologia soggettiva è stata proposta

Nota 1 - al progetto B/VIM (Video in Multimedia) dell'UER/EBU hanno aderito oltre alla Rai (il Centro Ricerche e Innovazione Tecnologica), l'IRT (Institut für Rundfunktechnik) che è il centro ricerca di televisioni tedesche (ARD, ZDF, DLR), austriache (ORF) e svizzere (SRG/SSR), la televisione norvegese NRK (Norsk Rikskringkasting AS) e FTRD (France Telecom R&D).

I codec introducono dei degradamenti più o La metodologia soggettiva è stata proposta

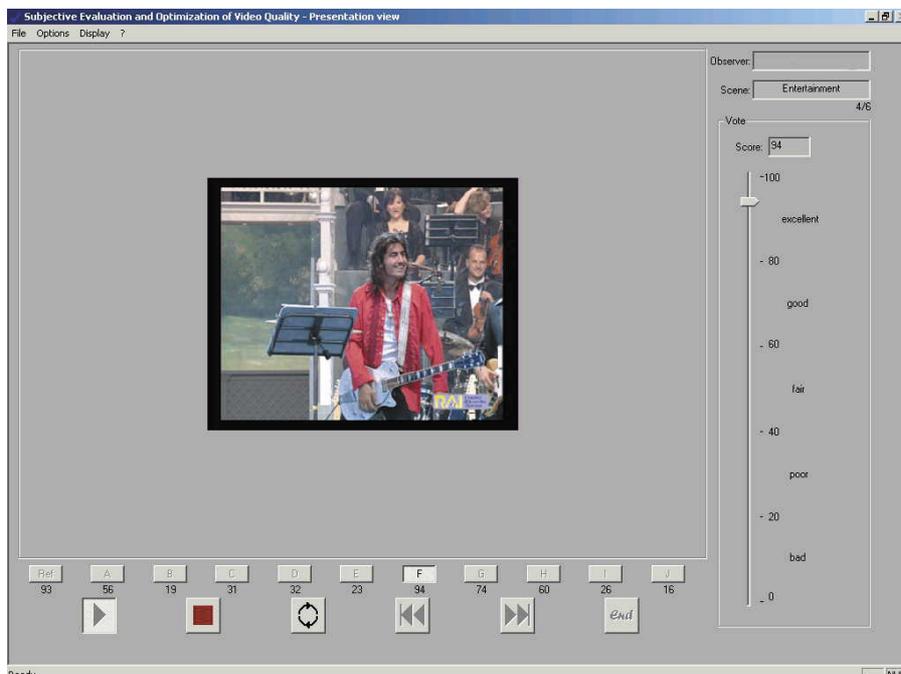


Fig. 2 - Interfaccia grafica per la somministrazione dei test

Connessione	Bit-rate	Bit-rate Effettivo (kbps)	Audio (kbps)	Video (kbps)	QCIF Frame Rate (Hz)	CIF Frame Rate (Hz)
Modem	56	40(+/- 10%)	8 mono	32(+/- 10%)	6.5	
Isdn	128	100(+/- 10%)	20 mono	80(+/- 10%)	12.5	6.5
DSL	256	200(+/- 10%)	32 stereo	168(+/- 10%)	25	12.5
DSL	500	400(+/- 10%)	48 stereo	352(+/- 10%)		25
DSL	700	560(+/- 10%)	64 stereo	500(+/- 10%)		25

Tab. 1 - Parametri utilizzati per le odifiche

da FTRD e si basa su interfaccia grafica su PC (figura 2) che consente all'osservatore di esprimere il suo giudizio di qualità utilizzando una scala continua compresa tra 0 e 100.

Come si può osservare dalla figura 2, l'interfaccia è caratterizzata dalla presenza di una serie di bottoni indicati con Ref ed una successione di lettere A, B, C..

2.1 Descrizione del funzionamento dell'interfaccia

All'osservatore viene chiesto di sedersi e di scegliere la distanza di visione che preferisce. Viene informato che la sequenza associata al tasto Ref (Reference) è quella di riferimento,

caratterizzata dalla qualità migliore in quanto rappresenta la sequenza (clip) in formato non compresso. I bottoni (A, B, C,..) corrispondono alla stessa sequenza codificata con codec e bit-rate differenti.

Tra le sequenze sotto test viene inserito nuovamente il Reference senza che l'osservatore ne sia al corrente allo scopo di verificare la correttezza dell'esecuzione del test, legata al fatto che l'osservatore sia in grado di riconoscere il Reference e di votarlo all'incirca come il Reference esplicito

L'osservatore può rivedere la stessa clip quante volte preferisce e modificare, eventualmente, il suo giudizio.

Fig. 3 - Basket



Fig. 4 - Entertainment



Nota 2 - I metodi di codifica disponibili sulla maggior parte dei codec si distinguono in "single-pass" e "two-pass": quest'ultimo differisce dal primo in quanto il codificatore processa la sequenza due volte: durante il primo passo, il codec acquisisce informazioni sulla complessità del contenuto mentre durante il secondo passo esegue la codifica ottimizzando il processo in base alle informazioni raccolte durante il primo. La codifica "two pass" non può essere utilizzata per codifiche live.

Il tasto >> consente di passare alla sequenza successiva.

Al termine della sessione i dati espressi dai singoli osservatori vengono salvati su un file per essere successivamente elaborati.

2.2 Test condotti dal gruppo B/VIM

Allo scopo di verificare la validità della metodologia proposta da France Telecom, il gruppo ha deciso di selezionare cinque sequenze (Basket, Entertainment, Kayak, News e Flower & Garden) che fossero rappresentative della normale programmazione televisiva, dato che lo scopo non era quello di mettere in crisi i codec sotto test. Le figure 3-4-5-6-7-8 corrispondono ad un fotogramma rappresentativo delle sequenze sotto test.

I codec considerati sono riportati di seguito:

- 1) Windows Media Encoder 8 (WM8)
- 2) Real Video 8 (RV8)
- 3) Sorenson 3
- 4) Quick Time (MPEG-4)
- 5) Dicas (MPEG-4)

I dettagli tecnici relativi ai singoli codec sono riportati nelle schede allegate

La tabella 1 indica il bit-rate (audio e video),

il frame-rate e le risoluzioni adottate per lo svolgimento dei test.

3. Caratteristiche generali dei codec sotto test

Nell'ambito dei test eseguiti le piattaforme sono state utilizzate per codificare file non compressi in formato AVI; i codec sono stati utilizzati in modalità CBR, un passo ^{Nota 2}.

I codec consentono di specificare una serie di parametri che si ripercuotono sulla qualità finale del filmato e sulla dimensione del file ottenuto.

Tali parametri riguardano:

- audio bit-rate
- video bit-rate
- dimensione del buffer (s): questo parametro deve essere specificato per la codifica CBR in singolo e doppio passo mentre per quella VBR è l'encoder che determina la dimensione ottimale del buffer
- tipo di codifica: CBR, VBR, due passi CBR, due passi VBR
- formato dell'immagine: altezza e larghezza
- frame-rate massimo (Hz)
- distanza tra i key frame (s): rappresenta la distanza temporale tra due fotogrammi chia-

Fig. 5 - Horse Riding

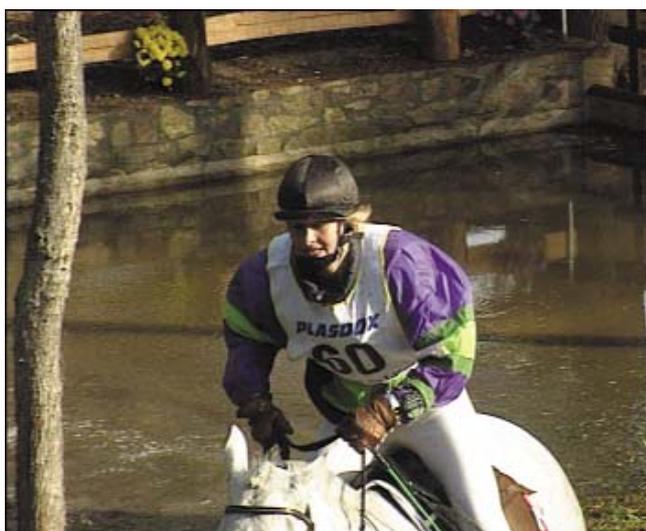


Fig. 6 - Kayak



ve, ossia editabili; i fotogrammi compresi tra due key frame contengono esclusivamente informazioni differenziali rispetto al key frame precedente. Il valore dovrebbe essere pari a pochi secondi per sequenze caratterizzate da elevato movimento e maggiore per sequenze più statiche

- qualità: a seconda del valore assunto da questo parametro, l'encoder privilegia la nitidezza del contenuto o la fluidità del movimento

In generale valgono le seguenti considerazioni:

- il valore associato alla qualità può avere ripercussioni sul frame-rate effettivo di codifica in quanto l'encoder, nel caso in cui i contenuti siano particolarmente critici in termini di dettaglio spaziale e di presenza di movimento, può eliminare alcuni frame (drop) per rispettare le impostazioni relative alla qualità
- il bit-rate (kbps) ottenuto in riproduzione potrebbe non rispecchiare fedelmente quello impostato durante la codifica, ma essere superiore di qualche kbps a seconda delle dimensioni degli header inseriti dal codec
- la dimensione finale del file non coincide con quella ottenibile moltiplicando il bit-rate impostato per la lunghezza della sequenza, ma è superiore di qualche decina di kbyte in funzione della dimensione del buffer e

della distanza dei key frame; buffer lunghi e distanze tra key frame brevi provocano un aumento delle dimensioni del file finale rispetto a quelle attese.

4. Analisi della qualità video

L'analisi statistica dei dati forniti dai singoli laboratori, eseguita da FTRD, ha dimostrato la correttezza e la riproducibilità dei risultati ottenuti applicando la metodologia proposta dal gruppo B/VIM.

Di seguito sono riportate:

- le considerazioni derivanti dall'esecuzione di prove expert-viewing, ossia prove soggettive che coinvolgono un numero limitato di osservatori (4-5 persone) esperte nell'ambito della valutazione della qualità video
- i risultati delle prove soggettive formali.

Le expert-viewing sono state eseguite riproducendo localmente i filmati su un PC (900 MHz, 1024x768) e chiedendo agli osservatori di esprimere un giudizio globale, su una scala a cinque livelli (Pessimo, Cattivo, Discreto, Buono ed Eccellente) in funzione della definizione del dettaglio spaziale e della fluidità del movimento.

I risultati emersi dalle expert-viewing sono

Fig. 7 - Teaching



Fig. 8 - Flower & Garden



riportati di seguito:

- WM8 e RV8 forniscono una qualità confrontabile anche se i difetti introdotti dai due codec sono diversi: le sequenze codificate con WM8, soprattutto a bassi bit-rate, presentano una forte blocchettizzazione mentre le sequenze trattate con RV8 sono caratterizzate da una forte perdita di risoluzione nei

dettagli

- Sorenson 3 fornisce una qualità nettamente inferiore a quella di WM8 e RV8, soprattutto a bassi bit-rate
- Dicas e QuickTime forniscono qualità confrontabili ed inferiore, rispetto a WM8 e a RV8

Per quanto riguarda la qualità video fornita dagli encoder valgono le seguenti considerazioni:

- per il formato QCIF, almeno 128-256 kbps sono necessari per ottenere un indice di qualità compreso tra discreto e buono
- per il formato CIF, almeno 500-700 kbps occorrono per ottenere un'indicazione di qualità compresa tra discreto e buono

ovviamente tali considerazioni sono strettamente legate al tipo di contenuto che viene codificato per cui a parità di bit-rate la qualità soggettiva percepita può essere fortemente diversa a seconda che il contenuto sia semplice (es. News, Teaching) o più complesso (es. Kayak, Basket).

Le considerazioni scaturite dalle expert-viewing hanno trovato conferma nei risultati delle prove soggettive formali per i formati CIF e QCIF riportati rispettivamente nelle figure 9 e 10.

Fig. 9 - Risultati delle prove soggettive per il formato CIF

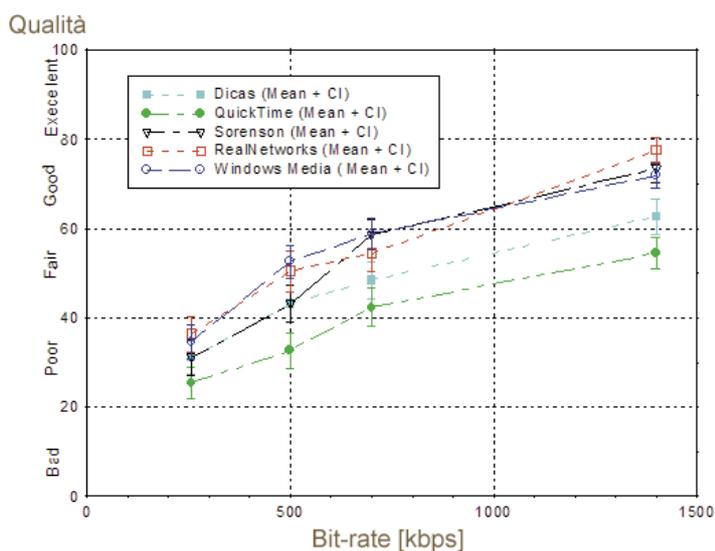
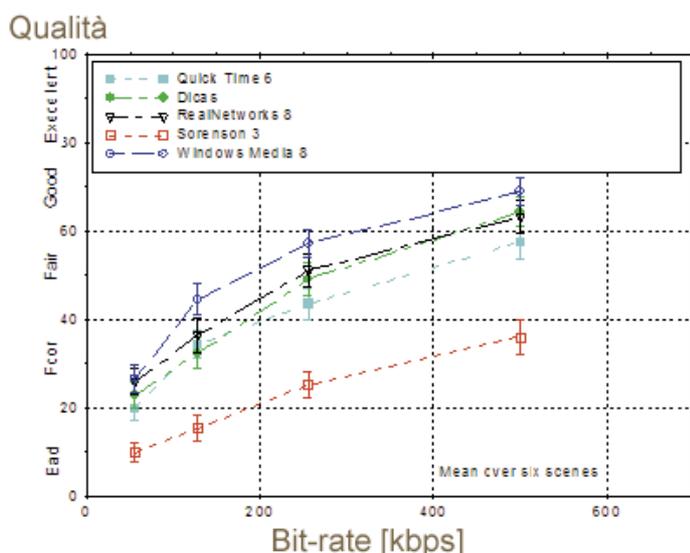


Fig. 10 - Risultati delle prove soggettive per il formato QCIF



5. Altri codec per applicazioni Webcasting

Esiste una notevole proliferazione di codec per applicazioni di Webcasting e l'evoluzione del mercato è estremamente rapida. Parecchi produttori offrono soluzioni complete end-to-end che comprendono oltre alla piattaforma di codifica, quella di authoring, streaming e di play-out. La maggior parte delle soluzioni sono "proprietarie" e la tendenza dei costruttori è quella di fornire dei plug-in che consentano di utilizzare Real Video oppure Windows Media come player del materiale da loro codificato.

La tabella 2 riporta un elenco di altri prodotti disponibili sul mercato, alcuni dei quali sono freeware, mentre altri sono a pagamento.

Codec	Società	Sito Web	Architettura
Vp5	On2	www.on2.com	Plug-in per RealVideo
ZyGoVideo	ZyGoVideo	www.ZyGoVideo.com	Plug-in QuickTime
3ivx	Happy Machine	www.3ivx.com	Plug-in per QuickTime
Divx	Divx	www.divx.com	Proprietaria
H.263	Apple	www.apple.com	QuickTime
MPEG-4		www.m4if.org/products	Alcune case produttrici forniscono dei plug-in per Real e/o per QuickTime mentre altre forniscono soluzioni end-to-end proprietarie

Tab. 2 - Codec per il video su web

H.264

Un commento a parte merita l'H.264 (noto anche come MPEG-4/Parte 10), un codec di ultima generazione sviluppato inizialmente dal gruppo ITU-T VCEG (Video Codec Expert Group), successivamente ITU e ISO hanno coordinato i loro sforzi creando il JVT (Joint Video Team) allo scopo di finalizzare uno standard che consentisse di ottenere un guadagno di almeno il 50% rispetto agli standard precedenti. L'approvazione da parte delle due organizzazioni internazionali ed il rilascio della versione definitiva del software è previsto per Maggio 2003. Le prestazioni di questo codec sembrano essere davvero promettenti sia in termini di rapporto qualità/bit-rate che di robustezza su canali in presenza di errori. Informazioni su H.264 sono reperibili in [2] e [3].

La tabella 3 riporta i risultati di alcuni test condotti dall'HHI su sequenze in formato CIF allo scopo di misurare il guadagno dell'H.264

rispetto ad altri algoritmi di codifica (H.263 High level profile, MPEG-4 Advanced Single Profile e MPEG-2).

E' necessario tenere presente che gli elevati guadagni riportati in tabella 3 sono ottenuti a discapito della notevole complessità computazionale (ved. Scheda allegata per dettagli tecnici sul funzionamento dell'algoritmo) che caratterizza l'algoritmo.

6. Conclusioni

Il mercato dei codec per il web è in continua evoluzione; al momento della redazione dell'articolo sono già disponibili nuovi prodotti software da parte di Microsoft (Windows Media 9), di Real (Real Video 9) e Sorenson (Sorenson Squeeze). Questi nuovi prodotti saranno presi in considerazione dal gruppo B/VIM nella seconda fase del progetto che partirà nel mese di giugno 2003. Alcuni risultati relativi al confronto delle prestazioni tra Windows Media 9 e Real

Coder	Guadagni di bit-rate (Formato CIF)		
	MPEG-4 ASP	H.263	MPEG-2
JVT/H.264	38,62%	48,80%	64,46%
MPEG-4 ASP	-	16,65%	42,95%
H.263 HLP	-	-	30,61%

Tab. 3 - Guadagni dell'H.264. Risultati di test condotti da HHI (Heinrich Hertz Institute) che è fortemente impegnato nello sviluppo del software.

Video 9 sono riportati in [4]: i test sono stati eseguiti da un laboratorio indipendente Keylabs incaricato da Real di condurre la survey.

Per tutti i codec considerati, la qualità video fornita è strettamente dipendente dal contenuto e dai parametri di codifica; tali parametri sono a loro volta legati alla banda a disposizione e consentono ampi margini di libertà.

Un'altra caratteristica comune ai codec nati per applicazioni web è la loro graduale migrazione verso applicazioni di tipo broadcasting (SDTV – Standard Definition Television, HDTV – High Definition Television); Windows Media 9 e Real Video 9, così come l'H.264, si stanno ponendo come possibili candidati per un'eventuale sostituzione di MPEG-2; i guadagni dichiarati dai produttori sono dell'ordine di circa 2 – 3 volte rispetto a MPEG-2 a parità di bit-rate. Al momento della redazione dell'articolo, il Centro ricerche ed Innovazione Tecnologica

sta conducendo dei test allo scopo di verificare il guadagno di questi nuovi algoritmi di codifica rispetto a MPEG-2 nel caso di codifiche su materiale SDTV; i risultati saranno oggetto di un prossimo articolo

Bibliografia

1. Recommendation ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures"
2. ITU-T H.26L Standardisation (ITU-T Q6/16, VCEG), www.tnt.uni-hannover.de/plain/project/vceg
3. R. Schaefer, T. Wiegand, H. Schwarz: "The emerging H.264/AVC standard", EBU Technical Review, January 2003, www.ebu.ch/trev_293-schaefer.pdf
4. www.keylabs.com/results/realnetworks/vidperf9.shtml

Acronimi e sigle

<i>AVI</i>	Audio Video Interleave: formato file
<i>CBR</i>	Constant Bit Rate
<i>CIF</i>	Common Intermediate Format: formato immagine, 352 pixel e 288 righe
<i>EBU/UER</i>	European Broadcasting Union/ Union Européenne de Radio-Télévision (www.ebu.ch)
<i>Frame</i>	Quadro televisivo
<i>Frame rate</i>	Frequenza di quadro: numero di quadri al secondo [Hz]
<i>Key-frame</i>	fotogramma di tipo Intra utilizzato per predire i quadri successivi
<i>ISDN</i>	Integrated Services Digital Network: standard internazionale di comunicazione a 64 kbps
<i>ISO</i>	International Organisation for Standardisation (www.iso.org)
<i>ITU</i>	International Telecommunication Union (www.itu.org) Union Internationale de Télécommunication Unión Internacional de Telecomunicaciones
<i>MPEG</i>	Motion Picture Expert Group: gruppo di lavoro congiunto ISO/IEC
<i>PDA</i>	Personal Digital Assistant: dispositivo palmare
<i>Q-CIF</i>	Quarter Common Intermediate Format: formato d'immagine, 176 pixel per 144 righe
<i>VBR</i>	Variable Bit Rate
<i>xDSL</i>	tecnologie Digital Subscriber Lines: modem e protocolli, per trasferire dati su doppino