

CG

**Computer
Gazette**



In questo articolo presenteremo un mare d'informazioni su quello che sta sotto la superficie dello standard MPEG. Vi aiuteremo a valutare i sistemi di codifica MPEG hardware e software oggi disponibili. L'MPEG sta diventando sempre più importante, specialmente per gli studi pieni di apparecchi video e strumenti di produzione. È il formato standard per il video su DVD e lo standard di trasmissione per la te-

codifica, l'output del processo e i componenti del decoder. E infatti i produttori di encoder possono usare molte tecniche per creare uno stream MPEG che soddisfi lo standard.

L'MPEG-1 è stato progettato per offrire audio e video compresso su CD-ROM e VideoCD. Inizialmente, era mirato a bassi data rate usando un quarto di risoluzione SIF a 352 x 240. Poi quel valore è stato abbassato per coprire la risoluzione video desktop di

maggiore, il formato MPEG-3 è stato inglobato nell'MPEG-2.

L'MPEG-4, che viene lanciato proprio in questo periodo, mira a offrire stream video e audio ad ampiezza di banda molto bassa su reti come Internet. Naturalmente, l'MPEG-4 ha molta concorrenza da parte dei codec proprietari, come quelli della Sorenson Vision e della RealNetworks. Solo il tempo e l'attività del mercato dirà se l'MPEG-4 verrà adottato dai principali

Tutti i segreti dell'MPEG

Una buona conoscenza della compressione MPEG potrà aiutarvi a scegliere il codec più adatto e a ottenere un video finale migliore

levisione digitale (DTV). Molti PC sono ora dotati di qualche applicazione di decodifica MPEG sia su scheda, sia come applicazione software in bundle. Anche i chioschi di presentazione e i briefing aziendali stanno usando l'MPEG per offrire contenuti video di alta qualità.

L'MPEG non è probabilmente il solo formato che gestite nel vostro ambiente di produzione, ma è un formato al quale dovrete prestare molta attenzione se volete entrare nella maggior parte dei mercati del futuro. Scegliere un encoder MPEG e usarlo al meglio può essere un compito difficile. La compressione video è sempre stata considerata un'arte magica, e l'MPEG non fa eccezione. Alcuni studi ingaggiano un tecnico dedicato con il solo compito di perfezionare l'arte della compressione MPEG.

L'approfondimento degli aspetti tecnici del formato video MPEG vi aiuterà anche a capire meglio il complesso processo di codifica e a offrire video di alta qualità ai vostri clienti.

Le molte versioni di MPEG

Prima di gettarci a testa bassa nell'argomento, facciamo una panoramica generale. MPEG sta per Motion Picture Expert Group, un sottocomitato dell'ISO, l'International Standards Organization. Il compito del gruppo è stato quello di definire un bitstream e un sistema di decodifica per offrire video e audio compresso su vari supporti. Il gruppo non ha definito il processo di

320 x 240.

L'MPEG-2 è stato originariamente sviluppato per il mercato broadcast, con data rate e qualità più alti rispetto all'MPEG-1. Questo standard è stato progettato per offrire video a pieno schermo e oggi viene usato per trasmettere segnali in definizione standard su DTV e via satellite. Sia l'MPEG-1, sia l'MPEG-2 sono compatibili con la specifica DVD-Video.

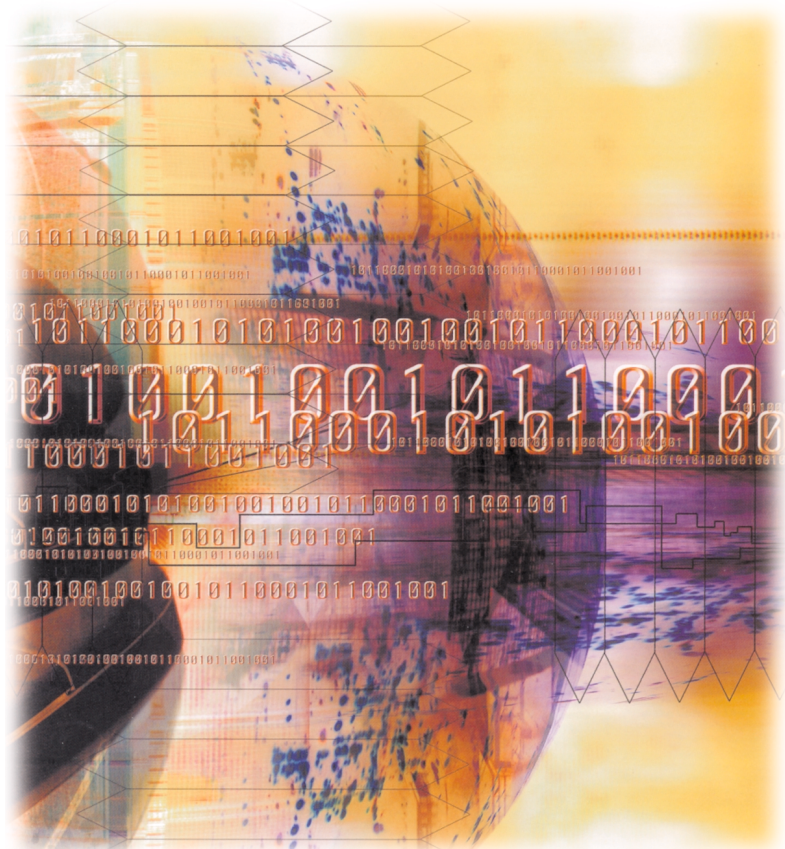
L'MPEG-3 era stato pensato per soddisfare le esigenze di video di qualità più alta della televisione ad alta definizione (HDTV). Dato che lo stream e le architetture di decoder dell'MPEG-2 potevano essere scalate per gestire questi segnali a larghezza di banda

protagonisti su Internet.

Dato che l'MPEG-2 è lo standard che si vede più spesso attualmente, mi concentrerò su quel formato. Guarderemo i componenti generali di uno stream video MPEG-2 e di un encoder MPEG-2. Tratteremo solo i formati video, non quelli audio.

Governato dal GOP

L'MPEG divide il video in tre tipi di frame. Gli Intraframe (I-frame) vengono codificati usando solo le informazioni all'interno di un singolo fotogramma. Sono necessari per mantenere la stabilità nel sistema di codifica, per



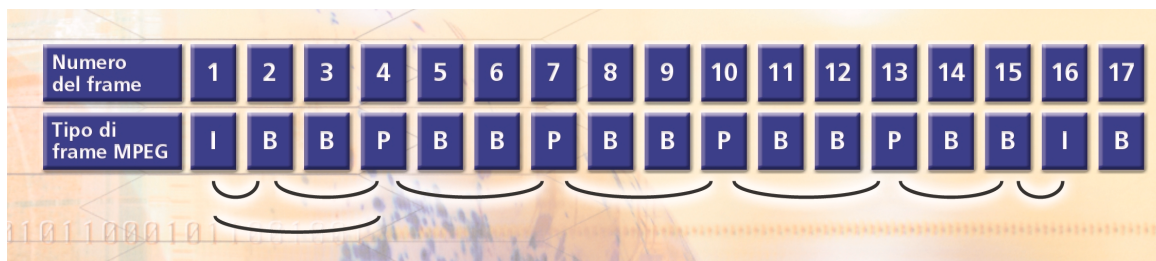
l'accesso casuale e per i tagli tra scene. Sono anche i frame che occupano la maggior parte dei dati. I Predictive frame (**P-frame**) vengono codificati stimando il movimento avvenuto rispetto al precedente I-frame o P-frame, e viene memorizzato solo quello che cambia nel fotogramma. I P-frame consentono un tasso di compressione medio. I Bidirectional frame (**B-frame**) analizzano il movimento rispetto al precedente I-frame o P-frame, e anche rispetto al successivo frame nella sequenza. I B-frame consentono un maggior rapporto di compressione a spese di un'elaborazione più complessa.

Per esempio, se si verifica un errore durante la codifica del frame 4 nella **Figura 1**, questo si può propagare fino al frame 15. Il B-frame numero 15 usa il P-frame numero 13 per stimare il movimento, ma il frame 13 ha sfruttato il frame 10, il frame 10 ha usato il frame 7, e il frame 7 è stato predetto in base al frame 4. Questo è un errore che rimane nell'immagine, e può anche peggiorare lungo gli 11 frame. Se l'I-frame non fosse posizionato ogni 15 fotogrammi, la propagazione dell'errore potrebbe durare anche più a lungo.

Gli I-frame consentono anche un accesso casuale alle sequenze video. Se

funzioni di calcolo per tutti i sistemi hardware, a parte quelli più sofisticati. Senza una buona stima del movimento, l'encoder deve codificare più dati per i P-frame e i B-frame, e l'equilibrio tra data rate e qualità sarà più difficile da raggiungere. Nuotare nelle acque oscure della stima del movimento vi farà conoscere meglio il vostro encoder MPEG e vi aiuterà a impostare parametri realistici per la vostra applicazione target.

Gli encoder MPEG dividono ogni frame in **macroblocchi**, normalmente da 16 x 16 pixel ciascuno. Per ogni macroblocco, lo stimatore di movi-



◀ **Figura 1:** una tipica struttura GOP di un video MPEG-2 è costituita da Intra frame, Predictive frame e Bidirectional frame, chiamati anche I-frame, P-frame e B-frame. Lo stimatore di movimento dell'encoder MPEG determina quale mix di tipi di frame è appropriato per un certo segmento

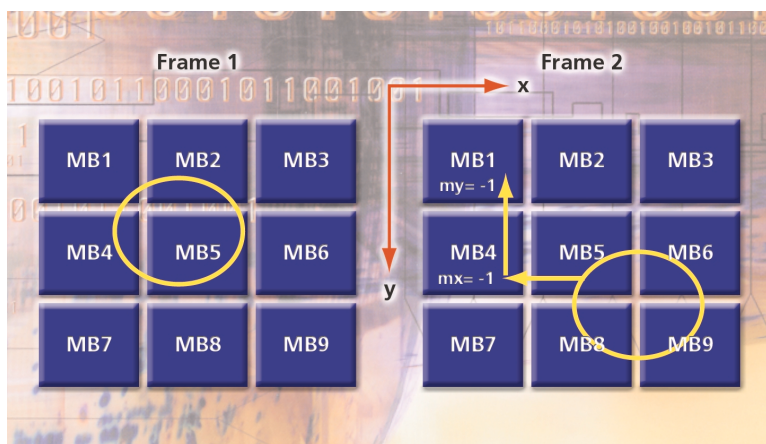
I-frame, P-frame e B-frame vengono riuniti in un Group Of Pictures (**GOP** o gruppo d'immagini). La struttura GOP di una sequenza video MPEG è il primo passo nell'impostazione di un bit-rate target per l'encoder, a causa dei differenti requisiti di data rate per i tre tipi di frame. Più I-frame ci sono, più alto sarà il bit-rate. Più sono distanziati fra loro gli I-frame, più basso sarà il bit-rate raggiungibile dall'encoder. Potete vedere una tipica struttura GOP nella **Figura 1**. Gli I-frame sono spazati di 15 fotogrammi. Ci sono il doppio di B-frame rispetto agli I-frame e

volete fermarvi al B-frame numero 6, il decoder dovrebbe prendere le informazioni dei P-frame numeri 4 e 7. La decodifica del frame 4, però, richiede anche l'informazione del frame 1. Di conseguenza, il sistema deve avere la capacità di memorizzare tre frame simultaneamente e la potenza di calcolo per decodificare quei tre frame di video per ricostruire il B-frame desiderato. Per fermarsi su un I-frame, invece, il decoder ha bisogno solo dei dati di quel fotogramma. Ciò spiega perché attualmente è più diffuso l'editing con la precisione dell'I-frame. Un buon en-

mento cerca di determinare dov'era nel frame precedente quel blocco d'informazioni, o uno molto simile. La stima viene effettuata confrontando le informazioni del macroblocco corrente con le informazioni dei macroblocchi del frame precedente. In una ricerca completa, vengono valutati tutti i macroblocchi del fotogramma precedente. È possibile utilizzare molti algoritmi di ricerca limitata per prendere decisioni intelligenti sul movimento e quindi ridurre i requisiti di elaborazione.

Una volta trovata una corrispondenza, il sistema definisce un **vettore di movimento** per il macroblocco e codifica la differenza tra i due macroblocchi. Considerate i set di macroblocchi nei due diversi fotogrammi mostrati nella **Figura 2**. Le informazioni nel macroblocco 6 del frame 2 sono molto simili a quelle nel macroblocco 2 del frame 1. Un vettore di movimento con una componente X di -1 e una componente Y di -1 punterà alla posizione appropriata per recuperare le informazioni per il macroblocco 6. Tuttavia, ci sono alcune differenze tra i due macroblocchi. Quelle differenze verranno codificate insieme al vettore di movimento, in modo da poter inviare un'immagine completa del macroblocco 6 nello stream MPEG. Quando il decoder vuole ricostruire il macroblocco 6 nel frame 2, recupererà le informazioni dal macroblocco 2 del frame 1 e poi aggiungerà le informazioni residue che sono state codificate per il macroblocco 6.

◀ **Figura 2:** mentre un oggetto si muove attraverso l'inquadratura di fotogramma in fotogramma, l'encoder MPEG definisce un vettore di movimento con componenti X e Y per descrivere come si è spostato l'oggetto rispetto alla posizione precedente



P-frame combinati. Potete anche farvi un'idea di come funziona la predizione per i P-frame e i B-frame. Il primo P-frame è una predizione tra se stesso e il primo I-frame. Il primo B-frame analizza il movimento tra se stesso, il precedente I-frame e il successivo P-frame, poi prende una decisione intelligente sul modo migliore per codificare il movimento che è stato rilevato.

Nell'MPEG, c'è sempre un compromesso tra bit-rate e qualità. Senza I-frame, l'encoder può perdere traccia del movimento in una sequenza, e gli errori possono iniziare ad accumularsi.

Il decoder MPEG permette d'impostare liberamente la struttura GOP e d'inserire I-frame dove necessario per assicurare una migliore codifica e l'accesso casuale alla sequenza.

La codifica del movimento

Il compito più complesso dell'encoder MPEG è la stima del movimento. Nelle applicazioni software, le fasi di ricerca e di decisione in questa parte dell'algoritmo risultano lunghe. Sono anche onerose in termini di memoria e

Se l'algoritmo di stima del movimento non riesce a trovare una buona corrispondenza tra macroblocchi, i blocchi subiranno una codifica intraframe, come spiegherò nel paragrafo seguente. Questa codifica richiede più bit, quindi meglio l'encoder MPEG riesce a gestire la stima del movimento, più alta sarà la qualità finale del video MPEG a un certo bit-rate target.



Le difficoltà di codifica

Vediamo un paio d'esempi di come tutte queste manipolazioni matematiche influenzino il modo in cui un encoder MPEG vede il vostro video. Le transizioni rappresentano un problema speciale per gli encoder MPEG. Tenete conto che in una dissolvenza incrociata, ogni pixel di ogni frame può cambiare. A seconda della soglia di sensibilità dello stimatore di movimento dell'encoder MPEG, potrebbe non essere possibile trovare un vettore di movimento valido. Quindi, i macroblocchi in tutti i fotogrammi devono essere codificati come grandi I-frame, invece che come P-frame e B-frame più compatti. I requisiti di data rate di questi frame aumenteranno drasticamente.

Ci sono due modi per aggirare il problema della codifica delle transizioni. Il più semplice è quello di alzare il bit-rate target. I buoni encoder MPEG-2 offrono la codifica a Variable BitRate (VBR o a bit-rate variabile), che consente di regolare il bit-rate scena per scena. Senza VBR, è necessario alzare

di movimento per quei nuovi macroblocchi. Ecco perché i B-frame sono così importanti per l'efficienza dell'MPEG. Questo tipo di frame guarda in avanti oltre che indietro per trovare il movimento. Le nuove aree svelate dello sfondo non esistono nel frame precedente, ma esisteranno nel frame successivo o in quello dopo ancora, quindi sarà possibile assegnare un vettore di movimento a quei macroblocchi. I vettori di differenza risultanti possono essere molto piccoli, e l'efficienza di codifica viene mantenuta.

I concetti chiave della compressione

Come ho detto nel paragrafo precedente, ogni immagine viene divisa in macroblocchi. Una volta determinato il vettore di movimento per un macroblocco in un P-frame o B-frame, viene stabilita la differenza tra il macroblocco corrente e quello da cui deriva. Se lo stimatore di movimento dell'encoder MPEG ha fatto bene il suo lavoro, quella differenza sarà molto piccola o

nel quantizzatore, dove riduce il numero di bit usati per rappresentare quei coefficienti che non sono vitali per il dettaglio complessivo dell'immagine. Troppa quantizzazione elimina i dettagli necessari dall'immagine e genera una "blocchettizzazione" visibile intorno ai contorni degli oggetti in movimento. Una quantizzazione troppo piccola mantiene l'encoder al di sopra del suo bit-rate target. Gli encoder MPEG hanno quindi un loop di feedback sul controllo del rate che permette al quantizzatore di regolarsi in base al bit-rate target che l'encoder deve ottenere.

La codifica di entropia organizza i coefficienti quantizzati in ordine di occorrenza, in modo che l'algoritmo possa essere ancora più efficiente. Poi viene usata una tabella di codifica a lunghezza variabile per assegnare codici a ciascun elemento, usando codici brevi per gli elementi più comuni e codici più lunghi per gli elementi meno comuni in una sequenza. Quindi, da dove deriva il risparmio di data rate offerto dall'MPEG? La parte con perdita d'informazione dell'algoritmo MPEG deriva dal processo di quantizzazione,



▲ **Figura 3:** dopo la stima di movimento per i P-frame e i B-frame, i macroblocchi di ogni frame vengono portati dalla DCT nel dominio delle frequenze. Il quantizzatore riduce poi il numero di bit usati per descrivere i dati di frequenza. La codifica di entropia ordina i dati quantizzati per consentire una compressione ulteriore. Il buffer mantiene i parametri di quantizzazione, in modo da poterli inserire nelle informazioni di header del blocco MPEG finale

il livello di un encoder a bit-rate costante per l'intero video. Anche se la codifica a bit-rate costante può essere perfettamente adeguata per la vostra applicazione, è utile avere a disposizione l'opzione VBR.

L'altro metodo per gestire le transizioni implica il cambiamento del vostro contenuto. Se lo spazio o l'ampiezza di banda a vostra disposizione sono limitati, come succede quando si cerca di far stare più di due ore di video su un DVD a singolo lato, dovete raggiungere un compromesso tra qualità e quantità. Usate transizioni il più possibile corte, e mantenete semplice il contenuto dietro la transizione. Se tutto il resto fallisce, provate un semplice taglio o una tendina al posto di una dissolvenza. Quando effettuate una zoomata sul vostro soggetto, la dimensione apparente degli oggetti cambia in modo tale che lo stimatore di movimento possa rilevare grandi differenze, ma può anche non riuscire a individuare un vettore di movimento valido. Se il vostro video MPEG evidenzia dei difetti mentre la videocamera zooma, dovrete provare a comprimere di nuovo. Alzate il data rate, abbreviate la scena, o usate al suo posto una scena statica se ne avete una.

Quando un oggetto si muove attraverso l'immagine, vengono rivelate nuove informazioni dello sfondo che non erano visibili nei frame precedenti. Lo stimatore di movimento nei P-frame non sarà in grado di trovare vettori

nulla. Se non è così, viene impostato un flag e i macroblocchi subiscono una codifica intraframe.

Terminata questa elaborazione del macroblocco, i P-frame e i B-frame vengono ulteriormente divisi in blocchi da 8 x 8 pixel. Anche gli I-frame vengono suddivisi nello stesso modo per effettuare l'elaborazione intraframe. Nella **Figura 3** è visibile un diagramma di questo algoritmo blocco per blocco. La base dell'elaborazione dei blocchi negli encoder MPEG è la Discrete Cosine Transform (DCT o trasformata discreta del coseno), che trasforma le informazioni spaziali degli elementi dell'immagine nel dominio delle frequenze. Ciò permette all'encoder di valutare il modo in cui cambiano le informazioni dell'immagine, invece di valutare come appaiono ai nostri occhi. A ogni elemento d'immagine del blocco viene assegnato un coefficiente numerico che rappresenta diversi livelli di cambiamento nell'immagine. Se lo stimatore di movimento ha fatto il suo lavoro, i coefficienti dei P-frame e dei B-frame rappresentano solo cambiamenti tra i due frame. I nostri occhi sono meno sensibili ai dettagli nelle aree in movimento dell'immagine, quindi i coefficienti a frequenza più alta possono essere cambiati, o persino eliminati, senza che lo spettatore se ne accorga. È in questo modo che l'MPEG ottiene data rate bassi conservando un'alta qualità.

L'encoder altera i coefficienti DCT

che può cambiare all'interno del frame. I coefficienti DCT non scartano i dati. Come mostrato nella **Figura 3**, i parametri di quantizzazione arrivano in un buffer, in modo da poter essere inviati come informazioni di header dei blocchi. I dati vengono ulteriormente compressi usando vari algoritmi di codifica di entropia, ma anche in questo caso non c'è perdita d'informazioni in questa fase.

Anche gli elementi senza perdita d'informazioni del processo di codifica influenzano la quantità di lavoro che l'encoder MPEG deve svolgere. La profondità di bit usata per calcolare la DCT, il controllo del rate e la flessibilità del quantizzatore, e la complessità della codifica di entropia, sono tutti fattori che influenzano l'efficienza di un encoder MPEG.

MPEG, dissolvenze e tendine

Guardiamo come tutte queste manipolazioni matematiche possono influenzare il vostro video. Considerate una dissolvenza incrociata applicata a due immagini fisse. Ci saranno differenze in ogni blocco dell'immagine a causa della dissolvenza. Ma dato che entrambe le immagini sono fisse, lo stimatore di movimento dovrebbe definire un vettore di movimento di (0, 0). D'altra parte, i cambiamenti durante la transizione porteranno alla generazione di coefficienti DCT per ogni pixel in

un blocco e la quantizzazione potrebbe diventare piuttosto drastica per conservare il data rate. Potreste iniziare a vedere un po' di blocchetti durante la dissolvenza, specialmente se ci sono molti dettagli ad alto contrasto nell'immagine sottostante.

Ora considerate una tendina orizzontale. Senza movimento, ci saranno grandi porzioni dell'immagine che rimangono costanti tra i fotogrammi. Dato che non ci sono cambiamenti da fotogramma a fotogramma nell'immagine, i coefficienti DCT sono zero e l'efficienza di codifica è molto alta. Nel punto dell'immagine in cui avviene la tendina, appariranno i coefficienti DCT ma solo per la direzione orizzontale. Non ci sono componenti verticali per la DCT e quindi ci saranno meno dati da codificare rispetto alla DCT di una dissolvenza incrociata.

Un semplice taglio richiederà l'inserimento da parte dell'encoder di un nuovo I-frame nella sequenza. Quindi, a seconda della lunghezza della transizione e del contenuto sottostante, ci sono casi in cui un taglio sarà la scelta migliore. Tenete presente, però, che molti tagli veloci tra scene complesse richiederanno l'inserimento di più I-frame, che potrebbero risultare altrettanto onerosi per l'encoder di una dissolvenza. Se i limiti di data rate non sono troppo stringenti, sono perfettamente accettabili transizioni e contenuti più complessi. Nelle scene con molto movimento, la DCT viene applicata solo alle piccole differenze dovute alla sfocatura della videocamera, all'illuminazione, o a nuovi oggetti che vengono rivelati dietro l'oggetto in movimento. Se lo stimatore di movimento sta facendo il suo lavoro, la cosa finisce qui. Le sottili differenze da fotogramma a fotogramma appaiono nella tabella DCT come coefficienti a bassa frequenza, ed è possibile applicare un alto livello di quantizzazione, ottenendo una codifica molto efficiente.

La cosa importante da ricordare è che i dati vengono persi solo durante la fase di quantizzazione. Ma l'invisibilità della quantizzazione verrà determinata da quanto bene l'encoder gestisce tutti i passi precedenti e dalla qualità del meccanismo di controllo del rate. Infine, il vostro bit-rate target determina quanto duramente dovrà lavorare il quantizzatore.

Bit-rate variabile

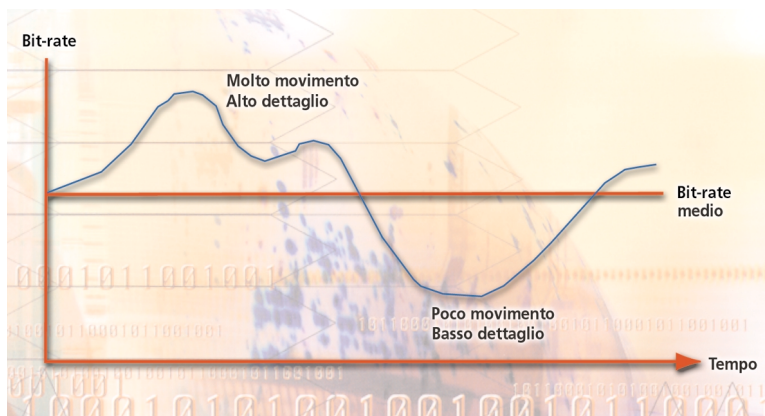
Fortunatamente, alcuni encoder MPEG offrono altre funzionalità per aiutare a raggiungere il vostro bit-rate target. Due tra le più importanti sono il primo menzionato **VBR** e la **codifica in più passate**.

L'MPEG-1 è stato progettato per essere uno stream di dati a bit-rate costante, nel senso che a ogni GOP viene assegnato lo stesso numero di bit da suddividere tra i tre tipi di frame. Anche l'MPEG-2 può essere creato usando un bit-rate costante, ma lo standard

consente anche di utilizzare il più avanzato VBR. Con il VBR, porzioni differenti di una sequenza possono essere codificate con data rate differenti (**Figura 4**). Il VBR permette alle porzioni ad alto movimento del vostro video di usare più bit e alle sequenze con meno movimenti di usarne meno. La qualità complessiva sarà più alta, perché i bit vengono usati dove servono di più. Quindi un sistema di codifica MPEG che gestisce il VBR può raggiungere una qualità d'immagine più alta rispetto a un sistema che non lo gestisce.

Gli encoder MPEG possono usare due modi per implementare il VBR. Uno è quello di tenere traccia delle prestazioni precedenti e allocare bit per i frame e i GOP mentre vengono elaborati. Il secondo metodo, più sofisticato, è quello di passare il materiale due volte.

Durante la prima passata, l'encoder analizza l'intera sequenza video e poi prende le decisioni iniziali su come allocare al meglio i bit per le differenti parti del video. La seconda passata usa quelle decisioni per codificare i GOP.



Anche se i sistemi di codifica MPEG a due passate sono più costosi di quelli a passata singola e richiedono più tempo di codifica, la loro strategia permette di raggiungere una qualità dell'immagine superiore.

Capire gli artefatti

Se avete lavorato con la grafica 3D, potete essere dei potenziali esperti di compressione. L'elemento chiave per capire la matematica che sta dietro il complesso algoritmo dell'MPEG sta nel visualizzare come elabora i cambiamenti di un'immagine nel tempo. Uno degli artefatti più fastidiosi che appaiono nel video MPEG è "l'effetto zanzara" (in inglese "mosquito noise"), simpatico nome dato alle piccole e sporadiche modifiche dei pixel intorno ai contorni di oggetti in movimento, come la testa di una persona che parla. Dato che soggetti di questo tipo non si muovono molto, un buon stimatore di movimenti può predire con precisione il movimento fotogramma per fotogramma. Tuttavia, le sottili differenze nell'illuminazione e nel dettaglio dei

capelli possono comunque venire codificate nei differenziali dei vettori di movimento.

Queste piccole differenze vengono trasformate in coefficienti DCT che normalmente sono principalmente ad alta frequenza. Un codificatore può alterare la quantizzazione di questi coefficienti blocco per blocco e frame per frame. Le modifiche frame per frame si manifestano con gli artefatti di rumore che noi chiamiamo "zanzare".

Ricordatevi il vecchio adagio: "spazzatura in entrata, spazzatura in uscita". Tutti i trucchi che avete imparato comprimendo video per i CD-ROM vi saranno d'aiuto nella compressione MPEG. L'illuminazione, le tecniche di pre-elaborazione e le apparecchiature utilizzate influenzeranno la qualità del vostro prodotto finale.

Seguire la corrente

La cosa positiva dell'MPEG è che offre una buona probabilità di generare video compresso di ottima qualità. Anche se gli encoder MPEG utilizzano ti-



◀ **Figura 4:** la compressione a bit-rate costante (la linea retta orizzontale) usa la stessa quantità di dati per descrivere immagini complesse e immagini semplici. Ciò può portare a una qualità d'immagine variabile. La codifica VBR (la linea curva) usa quantità diverse di dati per tipi differenti di contenuto video, permettendo di ottenere una migliore qualità d'immagine complessiva

picamente uno dei pochi chip di codifica MPEG esistenti sul mercato, ci sono comunque molti parametri che possono influenzare la qualità del loro output. La codifica software può offrire una qualità elevata e magari algoritmi superiori di stima del movimento, ma, con le CPU attuali, la codifica software richiede tempo.

Usate le informazioni di questo articolo su come sono strutturati gli stream MPEG e come un encoder influenza la qualità visiva del vostro video per trovare un encoder che offra le migliori caratteristiche e la qualità che potete ottenere con il vostro budget.

Infine, ricordate che diventare esperti di compressione video MPEG richiede tempo. Potrete ottenere video MPEG di qualità più elevata sfruttando l'esperienza di un tecnico dedicato che sa realmente come ottimizzare gli strumenti. E se starete attenti ad applicare la giusta illuminazione e le tecniche appropriate di postproduzione al vostro video sorgente, chi effettua la compressione non avrà la sensazione di nuotare controcorrente durante le sessioni di codifica.

(© DV)

CG