

## **INVENZIONE INDUSTRIALE**

**Titolo : Metodo di codifica video scalabile.**

**Inventore:**

### **Riassunto**

Un metodo di codifica video scalabile in cui il segnale video viene elaborato e codificato a diversi livelli di risoluzione spaziale dove tra coppie di livelli spaziali viene abilitato un meccanismo di predizione che consente un confronto a livelli di risoluzione spaziale omologhi senza impedire che vengano persi dati utili al fine di ottenere una buona efficienza di codifica.

**(Figura 1)**

**Titolare: Università degli Studi di Brescia**

**DESCRIZIONE**

L'invenzione riguarda i sistemi che necessitano di scambio di informazione di un segnale video numerico (videofonia (mobile), videocomunicazione (mobile), TV digitale terrestre, TV digitale satellitare, DVD, streaming video, video-on-demand, ...) e si riferisce in particolare ad un meccanismo di predizione incrociata tra rappresentazioni a diverse risoluzioni spaziali di uno stesso segnale video da utilizzare per un processo di codifica/decodifica dello stesso per garantire una piena scalabilità spaziale, temporale e di qualità di ricostruzione a partire di un unico flusso di codifica.

Sequenze temporali di immagini acquisite e rappresentate in formato numerico vengono denominate comunemente segnale video. Nell'attuale società dell'informazione, la compressione di un segnale video è un'operazione necessaria al fine di consentire e/o facilitare l'archiviazione di tali segnali e la comunicazione visiva a distanza. La compressione infatti è volta a ridurre l'occupazione in termini di unità d'archiviazione (comunemente bit) del segnale video, il che si traduce in una riduzione del ritmo di bit

necessario alla trasmissione dello stesso segnale su un canale di comunicazione digitale.

Il sistema di compressione o codifica riceve in ingresso il segnale video e restituisce un flusso di bit, di taglia inferiore rispetto alla taglia (espressa in bit) del segnale originale. E' possibile "contrattare" una minore taglia del flusso compresso a spese di una minore qualità del dato ricostruito dopo l'operazione inversa, detta di decompressione o meglio di decodifica.

La codifica viene effettuata, da chi possiede il segnale originale, mediante un sistema detto codificatore, mentre la decodifica è eseguita dal ricevente mediante un sistema detto decodificatore. Normalmente il segnale video originale ha caratteristiche di dimensione spaziale (coordinate x e y) e temporale (ritmo di immagini al secondo) che non vengono mutate dalle operazioni di codifica e decodifica che avvengono attorno ad un cosiddetto punto di lavoro caratterizzato dalla qualità (opportunosamente misurata) del dato decodificato. Secondo questo paradigma lavorano i più diffusi standard di compressione video, ad esempio quelli appartenenti alla famiglia MPEG e alla famiglia H.26x.

Tale paradigma viene però superato nel contesto della Codifica Video Scalabile (CVS). In tale contesto, il dato codificato rispetto ad un certo punto di lavoro

può essere decodificato secondo un certo numero, anche grande e non necessariamente definito a priori, di punti di lavoro cosiddetti inferiori, nel senso che sono ottenibili mediante la decodifica di una porzione del flusso codificato in origine. Tali punti di lavoro corrispondono alla possibilità di ricostruire (decodificare) un segnale video dove non solo la qualità ma anche la dimensione (o risoluzione) spaziale e il ritmo di immagini per unità di tempo risultino scalati (nel senso di ridotti) di certe quantità rispetto alle caratteristiche del segnale decodificato a partire dal flusso compresso intero.

Gli scenari applicativi che possono trarre beneficio da una CVS sono numerosi, ad esempio la produzione e distribuzione del video su canali a diversa capacità, e per dispositivi con diversa capacità di risoluzione spaziale o temporale (televisori, videotelefoni cellulari, palmari...); streaming del video su reti IP (Internet Protocol) eterogenee; sistemi evoluti di telesorveglianza; applicazioni di videoconferenza con banda non garantita; streaming video su reti mobili; consultazione veloce in archivi video; ed altri ancora.

Recentemente si è focalizzato un forte interesse attorno alla tipologia di codifica CVS grazie anche ad importanti soluzioni tecnologiche abilitanti la

scalabilità in termini spaziali e temporali, prima fra tutte la trasformata wavelet spaziale e la trasformata wavelet temporale nella sua versione moto compensata.

All'interno della comunità scientifica e soprattutto in seguito a lavori esplorativi nell'ambito di un gruppi di lavoro operante per l'organismo di standardizzazione ISO-MPEG, si è potuti arrivare recentemente ad una classificazione dei sistemi di CVS proprio secondo l'ordine in cui le suddette trasformazioni vengono applicate ai dati di partenza.

Scopo della presente invenzione è quello di proporre un metodo di codifica e di decodifica di segnali video che consenta di superare alcuni limiti delle architetture CVS note ed al contempo di fornire prestazioni di codifica del tutto competitive se non migliori rispetto a tutto lo stato dell'arte nota in materia di codifica video.

L'idea alla base della presente invenzione riguarda una nuova architettura per un sistema di CVS in cui il segnale video viene elaborato e codificato a diversi livelli di risoluzione spaziale dove tra coppie di livelli spaziali viene abilitato un meccanismo di predizione che consente un confronto a livelli di risoluzione spaziale omologhi senza impedire che vengano

persi dati utili al fine di ottenere una buona efficienza di codifica.

Le caratteristiche e i vantaggi del metodo di codifica e di decodifica secondo l'invenzione risulteranno comunque evidenti dalla descrizione di esempi preferiti di realizzazione, data a titolo indicativo e non limitativo con riferimento agli allegati disegni, in cui:

- la figura 1 rappresenta uno schema a blocchi che illustra il principio di funzionamento del codificatore secondo la presente invenzione;

- la figura 2 rappresenta uno schema a blocchi che illustra il principio di funzionamento del decodificatore associato al codificatore;

- la figura 3 rappresenta uno schema a blocchi di un esempio di realizzazione di un metodo di codifica e decodifica secondo l'invenzione;

- la figura 4 rappresenta uno schema di un esempio di realizzazione di una trasformata temporale;

- la figura 5 è una rappresentazione grafica di vettori di moto diretti e inversi per la descrizione del movimento di un oggetto tra due frames di una sequenza di immagini;

- la figura 6 è una rappresentazione grafica del concetto di MCTF (filtraggio temporale compensato

rispetto al moto), in cui la traiettoria del filtraggio nella direzione temporale anziché essere rettilinea segue la direzione dei campi di moto;

- la figura 7 rappresenta uno schema a blocchi di un ulteriore esempio di realizzazione di un metodo di codifica e decodifica secondo l'invenzione;

- la figura 8 mostra uno schema di dettaglio dei blocchi di codifica video e codifica video predittiva dello schema della figura 7;

- la figura 9 rappresenta uno schema a blocchi di un metodo di stima sincrona del campo di moto;

- la figura 10 è uno schema a blocchi di un codificatore progressivo;

- la figura 11 mostra un esempio di organizzazione di un flusso di bit codificato; e

- la figura 12 rappresenta uno schema di estrazione di un flusso di bit codificato come da figura 11.

I principi generali di funzionamento che governano il processo di codifica sono descritti dallo schema riportato nelle Figure 1 e 2.

Il metodo di codifica comprende in successione le fasi di:

- a) generare rappresentazioni a diverse risoluzioni spaziali del segnale video;

- b) trasformare temporalmente dette rappresentazioni;
- c) eseguire un'operazione di predizione tra tali trasformate temporali attraverso una messa in corrispondenza di informazioni inerenti a risoluzioni spaziali omologhe; e
- d) codificare entropicamente l'informazione risultante in modo progressivo, a partire da un primo livello di riferimento, detto base-layer.

Con riferimento allo schema della Figura 1, il segnale video originale, indicato con  $s_1$ , viene filtrato spazialmente (blocchi  $T_s$ ), prima colonna a sinistra dello schema, generando dei segnali  $s_2, s_2', \dots$  che rappresentano delle versioni delle immagini originali ad una risoluzione spaziale via via decrescente.

I segnali così ottenuti vengono filtrati, in modo indipendente, eseguendo una trasformazione temporale ( $T_t$ ) seguita da una ulteriore trasformazione spaziale ( $T_s$ ). Considerando ora quest'ultimo segnale appartenente al generico livello  $n$ , dove  $n$  coincide con il numero di trasformazioni spaziali eseguite nella fase iniziale del processo, si utilizza il segnale  $s_3, s_3', \dots$  proveniente dall'uscita del modulo temporale ( $T_t$ ) al livello  $n+1$  per predire la porzione di informazione ad esso compatibile. Tale operazione è rappresentata in Figura 1 dal nodo



differenza. I segnali così ottenuti subiscono una successiva trasformazione spaziale ( $T_s$ ) ed infine si applica una codifica entropica (EC) che genera i flussi di bit codificati finali (bitstream).

Il metodo di decodifica opera in modo duale rispetto al procedimento di codifica (figura 2). I flussi di bit codificati (bitstream) subiscono una *codifica entropica inversa* ( $EC^{-1}$ ) (*corretto?*) e quindi una trasformazione spaziale inversa ( $T_s^{-1}$ ). Considerando ora il segnale appartenente al generico livello  $n$ , dove  $n$  coincide con il numero di trasformazioni spaziali eseguite nella fase iniziale del processo, si utilizza il segnale proveniente dall'uscita del modulo spaziale inverso relativa al livello  $n+1$  per ricostruire l'informazione completa associata al segnale del livello  $n$ . Tale operazione è rappresentata in Figura 2 dal nodo somma. I segnali risultanti da questa operazione vengono quindi sottoposti ad una trasformazione spaziale inversa ( $T_s^{-1}$ ) e successivamente ad una trasformazione temporale inversa ( $T_t^{-1}$ ). In tal modo si ottengono segnali video  $Sv_0, Sv_1, \dots, Svn$  a varie risoluzioni spaziali.

La predizione può essere definita "incrociata" perché i segnali video a diverse risoluzioni spaziali subiscono una trasformazione prima di essere predetti da un livello di risoluzione spaziale all'altro. In altre

parole, e in accordo con un'implementazione preferita, al segnale ottenuto da una trasformazione temporale del video ad un certo livello di risoluzione spaziale (detto livello di riferimento) si applica un trasformazione spaziale che permetta di identificare (alcuni) dati ad una stessa risoluzione spaziale del segnale ottenuto per trasformazione temporale del video ad un diverso livello di risoluzione spaziale. Una predizione può così essere operata tra le rappresentazioni temporali così ottenute, riducendo pertanto l'entropia di informazione. Nel codificatore, l'informazione residua così ottenuta viene compressa entropicamente sfruttando una procedura di prioritizzazione del flusso di codifica che ne garantisca una completa scalabilità. Lo stesso principio di predizione incrociata può essere applicato in modo analogo a tutti i rimanenti livelli di risoluzione spaziale dei flussi video da codificare, sia partendo dal livello di riferimento iniziale, sia da un diverso livello di riferimento. Il decodificatore opera in modo duale rispetto al codificatore per quanto riguarda il principio di predizione incrociata. Può anche decodificare frazioni del flusso di codifica qualora si desideri ricostruire un unico flusso video ad un desiderato livello di qualità, e ad una particolare risoluzione temporale e spaziale.

In Figura 3 è mostrato uno schema di un esempio di realizzazione di un metodo di codifica e decodifica.

Il metodo di codifica comprende le seguenti fasi:

a) acquisire un primo segnale digitale ( $s_1$ ) rappresentativo di un segnale video, detto segnale digitale comprendendo una sequenza di immagini numeriche;

b) operare una trasformazione (TS) di detto primo segnale digitale ( $s_1$ ) per generare almeno un secondo segnale digitale video ( $s_2$ ) avente una risoluzione spaziale diversa da quella di detto primo segnale digitale ( $s_1$ );

c) effettuare una trasformazione (TT) dell' almeno un secondo segnale digitale video ( $S_2$ ) generando un terzo segnale digitale ( $S_3$ ), detta trasformazione lasciando sostanzialmente invariata detta risoluzione spaziale del secondo segnale digitale video ( $s_2$ ). Per risoluzione spaziale invariata si intende che il segnale ha subito una perdita di informazione spaziale inferiore, ad esempio, al 20%;

d) effettuare un'operazione di compressione (CE) di detto terzo segnale digitale ( $s_3$ ) per generare un primo segnale digitale compresso ( $S_c$ );

e) operare una trasformazione (CE-1) di detto primo segnale digitale compresso ( $S_c$ ) per generare un segnale ricostruito ( $S_r$ ) equivalente a quello ottenibile

mediante una decompressione di detto primo segnale digitale compresso (Sc);

f) effettuare un'elaborazione del primo segnale digitale video (s1) comprendente una trasformazione (TT) della tipologia della trasformazione della fase c) per generare un segnale digitale (St);

g) effettuare una trasformazione (TS) di detto segnale (St) della tipologia della trasformazione della fase b) per ottenere un segnale digitale (s4) comprendente una prima componente (10) e una seconda componente (11), dove dette prima e seconda componente rappresentano in combinazione lo stesso contenuto informativo del segnale digitale (St), detta trasformazione (TS) essendo adatta a generare detta prima componente (10) in modo che abbia sostanzialmente la stessa risoluzione spaziale del segnale digitale (s3) (v. Figure 1 e 8);

h) generare un segnale di predizione (Sp) rappresentativo della differenza fra detta prima componente (10) e il segnale ricostruito (Sr); e

i) effettuare un'operazione di compressione (CE) di detto segnale di predizione (Sp) e di detta seconda componente (11) per generare un secondo segnale digitale compresso (Sc2).

Il metodo di decodifica associato al metodo di codifica sopra descritto comprende le fasi di:

a) acquisire un segnale digitale comprendente almeno il primo ed il secondo segnale digitale compresso ( $S_c, S_{c2}$ );

b) effettuare un'operazione di decompressione ( $CE^{-1}$ ) di detto primo segnale digitale compresso ( $S_c$ ) per ottenere un segnale ( $S_{r'}$ ) rappresentativo del segnale ricostruito ( $S_r$ );

c) effettuare una trasformazione ( $TT^{-1}$ ) di detto segnale ( $S_{r'}$ ) per generare un segnale decodificato ( $S_d$ ) rappresentativo del secondo segnale digitale video ( $S_2$ );

d) effettuare un'operazione di decompressione ( $CE^{-1}$ ) di detto secondo segnale digitale compresso ( $S_{c2}$ ) per ricostruire un segnale ( $S_{p'}$ ) rappresentativo del segnale di predizione ( $S_p$ ) e un segnale  $11'$  rappresentativo della seconda componente  $11$  del segnale digitale ( $S_4$ );

e) effettuare una somma tra detto segnale ( $S_{p'}$ ) e detto segnale ( $S_{r'}$ ) per ricostruire un segnale  $10'$  rappresentativo della prima componente  $10$  del segnale digitale ( $S_4$ );

f) effettuare una trasformazione ( $TS^{-1}$ ) di detti segnali  $10', 11'$  per ricostruire un segnale ( $S_{t'}$ ) rappresentativo del segnale digitale ( $S_t$ ); e

g) effettuare una trasformazione ( $TT^{-1}$ ) di detto segnale ( $St'$ ) per generare un segnale decodificato ( $Sd1$ ) rappresentativo del primo segnale digitale video ( $S1$ ).

In altre parole, a partire dalla risoluzione originale, detta anche finale perché corrispondente all'ultimo stadio di raffinamento, si estraggono tramite trasformazione spaziale (TS), dei filmati a diverse risoluzioni. La risoluzione inferiore costituisce il livello "BASE" a cui viene associato il flusso codificato detto "base-layer". I livelli intermedi possono essere 0, 1 o più d'uno. Nello schema della Figura 3 è riportato l'esempio di un livello intermedio.

Una trasformata temporale (TT) è applicata per ogni livello di risoluzione spaziale al fine di decorrelare i dati lungo la direzione temporale.

A partire dal livello base viene istanziata un'architettura di relazioni inter-livello che prevede un meccanismo di predizione (P) di alcuni dati al livello di risoluzione spaziale corrente a partire dal livello di risoluzione spaziale inferiore.

Il blocco (CE) esegue la quantizzazione e codifica entropica dei coefficienti prodotti ai vari livelli di risoluzione spaziale.

Al fine di garantire un corretto aggiornamento (A) della predizione in fase di decodifica, si noti come nel

codificatore i dati utilizzati in predizione (P) e provenienti dal livello spaziale inferiore siano quelli che saranno presenti in decodifica (ovvero quelli decodificati tramite  $CE^{-1}$ ).

I flussi di bit codificati prodotti dal metodo di codifica e contenenti i segnali compressi ( $Sc$ ,  $Sc2$ ) vengono inviati ad un blocco denominato "Estrattore" che estrae e ricompone un nuovo flusso codificato nel quale i vari livelli spaziali possono, a seconda delle tecnologie utilizzate, subire scalature di risoluzione temporale e/o di qualità.

Il decodificatore inverte tutta la catena di elaborazione della codifica restituendo i filmati decodificati a varie risoluzioni spaziali.

La scalabilità spaziale è ottenuta semplicemente scartando, a partire da un certo layer di interesse (che diventa il layer finale), tutti i flussi codificati dei layer a risoluzione spaziale superiore. Si noti che al fine di eseguire un'operazione di predizione è necessario confrontare i dati allo stesso livello di risoluzione spaziale. Questo può avvenire ad esempio in due modi:

- 1) attraverso un'interpolazione dei dati a livello spaziale superiore,
- 2) attraverso una decimazione dei dati a più alta risoluzione attivando al contempo un meccanismo di

conservazione dell'informazione scartata in fase di decimazione.

In accordo con una forma preferita di realizzazione illustrata nelle Figure 4-6, il blocco di trasformata temporale (TT) viene realizzato tramite una tecnica di filtraggio temporale moto-compensata (MCTF). L'operazione di stima del moto produce per ogni coppia di immagini definite come "frame di riferimento A- frame attuale B" un campo di moto, rappresentabile tramite vettori di moto (in Figura 5 vengono rappresentati i vettori di moto che descrivono lo spostamento di un oggetto in movimento). Il campo di moto stimato viene codificato al fine della rappresentazione compressa. Il campo di moto viene utilizzato dal filtraggio wavelet MCTF che con l'ausilio di tale informazione effettua una decorrelazione temporale molto più efficace rispetto alla sola applicazione del normale filtraggio wavelet lungo la direzione temporale. I dati (frames) creati da una trasformata wavelet MCTF, detti sottobande temporali, possono essere organizzati in livelli di risoluzione temporale ed è questa organizzazione che abilita la codifica con scalabilità temporale. Il principio di un filtraggio MCTF è rappresentato in Figura 6, dove la curva (c) rappresenta la traiettoria seguita dal filtraggio. Matematicamente, due pixel ad istanti di



tempo successivi (frame A e B) verranno visitati dal filtro nella sequenza di coordinate moto-compensate  $(x,y,tA) \rightarrow (x+k, y+l,tB)$ , anziché  $(x,y,tA) \rightarrow (x,y,tB)$ . Dettagli sulle problematiche di occlusione e scopertura di aree da parte di oggetti in movimento possono essere trovati ad esempio in [Ohm94, Cho99]. Nello schema proposto il filtraggio MCTF viene effettuato su ciascun livello di risoluzione per cui campi di moto a diverse risoluzioni spaziali vengono generati codificati.

In accordo con una forma di realizzazione vantaggiosa del metodo di codifica e di decodifica, illustrato nelle Figure 7 e 8, l'informazione ad un determinato livello di risoluzione spaziale nel dominio trasformato viene ottenuta attraverso una elaborazione spaziale dell'informazione trasformata di un diverso livello di risoluzione spaziale, omologa all'elaborazione utilizzata per la generazione del video allo stesso diverso livello di risoluzione spaziale.

Più in dettaglio, i filmati di riferimento ai diversi livelli di risoluzione spaziale sono ottenuti con l'utilizzo di una trasformata wavelet discreta (TWD) spaziale (Figura 7). Nello schema è ancora evidente la struttura predittiva dove le relazioni inter-livello vengono intercettate da un codificatore video predittivo (CVP), dettagliato (assieme al codificatore video di

base-layer CV) in Figura 8. Si noti come la stessa TWD utilizzata per la riduzione di risoluzione spaziale viene utilizzata all'interno del CVP al fine di creare sottobande spazio-temporali di bassa risoluzione (in alto a destra) direttamente confrontabili con le sottobande temporali decodificate al livello di risoluzione inferiore. In questo modo si può realizzare in modalità DPCM (ovvero codificando la sola differenza) una predizione tra sottobande omologhe. Le sottobande tra le quali si esegue la differenza si ritengono omologhe in quanto:

- 1) hanno lo stesso livello di risoluzione spaziale,
- 2) hanno subito la stessa tipologia di trasformazione spazio-temporale, anche se trasformata spaziale e temporale sono state eseguite nei due casi in ordine inverso.

In accordo con una forma preferita di realizzazione, l'informazione di moto utilizzata per determinare il dominio trasformato della rappresentazione del video è ricavata in modo sincrono e per raffinamento tra almeno due livelli di risoluzione spaziali considerati.

Al fine di ottenere una buona decorrelazione dei coefficienti temporali è essenziale che le funzioni di wavelet utilizzate siano applicate lungo sequenze di pixel omologhi, cioè appartenenti ad uno stesso oggetto.

L'operazione che permette identificare tali sequenze di pixel è la stima del campo di moto (blocco 20 in Figura 9), ossia la stima dei valori di roto traslazione che descrivono l'evoluzione temporale di un blocco di pixel omologhi nella sequenza S di immagini che deve essere codificata.

Lo stimatore di moto da utilizzarsi nella fase di codifica del base layer può essere scelto tra le soluzioni tecniche disponibili più perforanti, quali, ad esempio, H264. Nella fase di codifica di un livello a più alta risoluzione spaziale, rispetto al base layer, si dovrà utilizzare uno stimatore in grado di supportare la codifica del moto multi livello, base layer più enhancement layer, opportunamente modificato. In particolare, il blocco di retroazione che modifica i parametri dello stimatore deve tenere conto anche della distorsione introdotta sull'immagine differenza nel dominio MCTF e non solo della distorsione introdotta sulla immagine originale, come evidenziato dalla freccia (F) in Figura 9.

Al fine di rispettare i vincoli appena citati il campo di moto (partizionamento dei macro blocchi e valori dei vettori di moto) del livello spaziale predetto dovrà quindi essere coerente con il campo di moto corrispondente stimato sul base layer.

Secondo un altro aspetto dell'invenzione, la codifica progressiva viene ottenuta attraverso un meccanismo di codifica entropica dei simboli ottenuti dall'approssimazione di frame trasformati relativi a medesimi livelli di risoluzione temporali, in modo da garantire un'adeguata scalabilità temporale e in termini di qualità del flusso di codifica.

Dopo la trasformata temporale, i coefficienti sono decorrelati temporalmente. Si può osservare che, complessivamente, ovvero a seguito della trasformazione temporale e della trasformazione spaziale (con o senza il meccanismo di predizione), vengono generati insiemi di coefficienti decorrelati i quali possono essere classificati secondo una loro collocazione di risoluzione spazio-temporale. Per la realizzazione di un codificatore video scalabile oltre che in senso di risoluzione spaziale anche a livello di risoluzione temporale e qualitativa è necessario che la descrizione codificata dei coefficienti sia organizzata secondo una prioritizzazione definita sui dati di partenza. Tipicamente, per quanto riguarda la scalabilità temporale, il codificatore dovrà garantire un accesso esclusivo ai dati relativi ad una certa risoluzione temporale, mentre per quanto riguarda la scalabilità in qualità dovrà essere garantita una quantizzazione dei

coefficienti ridefinibile e raffinabile a seconda della qualità di riproduzione desiderata. E' inoltre desiderabile che questo raffinamento di qualità possa essere effettuato progressivamente e finemente.

In quest'ottica, la realizzazione preferenziale della codifica entropica di seguito proposta comprende un sistema di codifica che agisce congiuntamente su dati di medesimo livello temporale e in grado di realizzare un raffinamento progressivo della qualità dei suddetti dati.

A tal fine è possibile utilizzare un qualsiasi codificatore entropico per immagini fisse purché adattato a gestire gruppi di frame. Questo adattamento può riguardare la semplice gestione di più frame oppure delle modifiche apportate ai meccanismi di scansione dei coefficienti e della loro codifica che beneficino della coerenza strutturale e statistica dei frame raggruppati.

In accordo con una forma vantaggiosa di realizzazione, detta codifica entropica viene ottenuta attraverso un meccanismo di codifica dei simboli che operi per bit-plane di frame trasformati relativi a medesimi livelli di risoluzione temporali.

Più in dettaglio, i coefficienti trasformati (spazialmente, temporalmente o che subiscono entrambe le trasformazioni) presentano tipicamente porzioni di maggiore attività (in senso energetico) laddove le

suddette operazioni di correlazione lineare non riescono ad agire efficacemente. Sui dati video questo si verifica tipicamente in corrispondenza di discontinuità spaziali o temporali le quali tipicamente individuano gli oggetti e le relative proprietà di moto. La coerenza morfologica degli oggetti rappresentati nel video e la relativa regolarità locale del moto fanno sì che le aree di maggiore attività dei coefficienti trasformati si concentrino tipicamente a formare degli agglomerati (cluster) spazio-temporali. La descrizione della posizione e del valore dei coefficienti che è necessario fare (implicitamente o esplicitamente) attraverso un algoritmo di codifica entropica potrà quindi sfruttare utilmente questa informazione a-priori (ipotesi di clusterizzazione) per aumentare le prestazioni di codifica. Si decide inoltre di sfruttare l'ipotesi di clusterizzazione dei coefficienti attivi secondo diverse soglie di attività, tipicamente individuabili attraverso i piani di bit di una rappresentazione binaria dei dati stessi. Tale codificatore prende il nome di Codificatore Progressivo basato su Dilatazione Morfologica Multilivello su Gruppi di Frames a Risoluzione Temporale Omogenea: CPDMM-GFRTO ed è una estensione al caso di filmati della tecnica di codifica per immagini statiche EMDC (Embedded Morphological Dilatino Coding). Questa

tecnica di codifica è stata finora concepita solo per dati fissi. Al contrario, CPDMM-GFRTO:

- 1) gestisce problematiche di codifica congiunta di gruppi di un numero qualsiasi di frame,
- 2) può adattare la sua struttura algoritmica alle esigenze di complessità ed eventualmente a requisiti di scalabilità spaziale.

Nello schema di riferimento della Figura 10, i due blocchi tratteggiati intervengono tipicamente nei casi di coefficienti in ingresso organizzati in una struttura multirisoluzione (ad esempio nel caso si utilizzi una trasformata wavelet per la decorrelazione spaziale, come indicato nell'implementazione preferita). Per rappresentazioni non-multirisoluzione spaziale, per le quali si possa far valere l'ipotesi di clusterizzazione dei coefficienti trasformati (es. generati da una trasformazione temporale) è possibile semplicemente togliere i blocchi tratteggiati.

La codifica CPDMM-GFRTO ha una struttura che si itera sui piani di bit in modo da realizzare al contempo la scansione progressiva della mappa dei coefficienti significativi e la loro quantizzazione e codifica entropica. A tal fine vengono gestite due liste. La principale è la lista dei coefficienti trovati significativi "LCS" rispetto alla soglia determinata dal

di piano di bit corrente. Una seconda lista tiene memoria dei coefficienti non significativi nelle zone di confine "LCC" dei cluster di coefficienti significativi (v. dettaglio dei blocchi). I dati in ingresso sono tipicamente coefficienti appartenenti a dei gruppi di frames di medesimo livello temporale (blocco "GdF"). La ragione di questo sta nella possibilità di abilitare la scalabilità temporale in modo efficace semplicemente scartando i flussi di bit relativi ai livelli temporali non utili. E' altresì importante raggruppare più frames assieme in modo da poter sfruttare le similarità statistiche dei frames di pari livello temporale per poter costruire tabelle di probabilità più affidabili ai fini di una eventuale codifica entropica adattativa dei flussi di informazione. La codifica entropica adattativa fa parte di ciascun blocco di elaborazione ed ha il compito di eliminare le ridondanze statistiche delle catene di simboli generate da ciascun blocco. Generalmente, data la strutturazione spaziale dei dati, si ricorre a tecniche di codifica aritmetica adattativa e basata su contesti (selezione delle tabelle di probabilità a seconda dello stato di un intorno di coefficienti rispetto al coefficiente in esame).

Le liste sono inizializzate come vuote. Nell'eventualità che non vi siano coefficienti da



analizzare tra quelli precedentemente individuati dai blocchi di elaborazione, si procede in qualsiasi momento alla codifica esplicita di un nuovo coefficiente.

Il processo termina una volta raggiunta una condizione di STOP. Questa può essere determinata dal raggiungimento di una particolare piano di bit (arresto orientato alla precisione) o al raggiungimento della quantità di bit assegnata (arresto orientato alla quantità di bit a disposizione). Anche se non riportato nello schema è altresì possibile monitorare costantemente la qualità del dato decodificato (o una stima di essa) ed arrestare il processo di codifica con un criterio orientato alla qualità del dato decodificato.

Viene ora descritto più in dettaglio il metodo di codifica entropica schematizzato a blocchi nella Figura 10.

- Blocco Dilatazione morfologica ("Dm"): vengono analizzati i coefficienti della LCS e viene realizzata un'operazione di dilatazione morfologica attorno ad essi con l'obiettivo di trovare coefficienti clusterizzati. Ogni nuovo coefficiente trovato significativo viene aggiunto alla LCS e l'operazione continua sino all'esaurimento della lista. I coefficienti non significativi si trovano giocoforza ai bordi dei cluster analizzati e vengono posti nella LCC. Viene codificata la

significatività dei coefficienti analizzati e l'informazione di segno. Si genera quindi un flusso binario che può, come già detto, essere ulteriormente codificato entropicamente.

- Blocco di Predizione multirisoluzione ("Pm"): nelle rappresentazioni multirisoluzione sono solitamente presenti dipendenze inter-banda (inter-scala) che si possono sfruttare per una più efficace compressione costruendo relazioni di filiazione da parte di coefficienti "padre" presenti ai livelli di risoluzione inferiore. In questo caso, esaurito il compito del blocco precedente si ricercano nuovi coefficienti significativi non caso ma tra i possibili figli di padri già trovati significativi. Questo rende più efficace la ricerca di nuovi coefficienti significativi (questi sono infatti distribuiti in modo clusterizzato ma sparso nelle rappresentazioni multirisoluzione). Come si vede dallo schema, una volta trovato un nuovo coefficiente significativo viene immediatamente richiamata la procedura di dilatazione morfologica, in quanto è quella che risulta più efficace in termini di codifica.

- Bollo di Dilatazione morfologica dei confini ("Dmc"): sempre in riferimento a rappresentazioni multirisoluzione ed in particolare a rappresentazioni wavelet, si è notato che i cluster di coefficienti significativi spesso hanno

una struttura ad arcipelago, ovvero sono concentrazioni di coefficienti connessi (cluster) con alcune appendici non necessariamente "strettamente connesse" ma comunque posizionate nelle immediate vicinanze del un cluster principale. Questo rilievo suggerisce una seconda strategia per la ricerca di nuovi coefficienti significativi. Quindi, esaurite le azioni dei blocchi precedenti, non si procede a caso nella ricerca di nuovi coefficienti significativi ma si considerano dapprima i bordi dei cluster trovati (le posizioni dei coefficienti di tali bordi sono elencate nella LCC) e si esegue una dilatazione morfologica su di essi. Nuovamente quando viene trovato un coefficiente significativo si esegue immediatamente la procedura di dilatazione morfologica.

- Blocco ("R") di Raffinamento sul piano di bit: una volta individuati tutti i coefficienti significativi per il piano di bit corrente, vengono raffinati tutti i coefficienti trovati significativi al piano di bit precedente per riportarli alla precisione del piano di bit corrente.

Ogni simbolo emesso dall'algoritmo descritto viene codificato attraverso un codificatore aritmetico adattativo basato su contesti statistici.

Un ulteriore aspetto dell'invenzione riguarda un sistema di rappresentazione del partizionamento del

flusso di codifica adatto ad essere sottoposto al metodo di decodifica della presente invenzione.

Detto sistema comprende meccanismi di indicizzazione che permettano un frazionamento del flusso di codifica per raggiungere operativamente il flusso da associare ad un certo livello di qualità per un dato punto di funzionamento in termine di risoluzione spaziale e temporale, e che possano consentire inoltre di rispettare requisiti di massimo ritardo codifica-decodifica.

In un sistema di codifica scalabile parte del flusso codificato al punto di lavoro più elevato deve essere "estratta" e trasmessa al decodificatore. L'operazione di estrazione deve essere semplice (non deve richiedere operazioni di decodifica o transcodifica) in quanto le proprietà di scalabilità del flusso codificato devono consentire un'estrazione immediata col solo ausilio di semplice informazione di intestazione per la navigazione e la scansione strutturale del flusso video codificato.

Detto questo vanno considerate almeno due tipologie di estrazione: 1) estrazione singola; 2) estrazione multipla.

1) Estrazione singola: per estrazione singola si intende l'operazione descritta sopra dove il flusso estratto è decodificabile al punto di lavoro desiderato.

Dal flusso estratto non è normalmente possibile eseguire un'estrazione ottima di punti di lavoro inferiori.

2) Estrazione multipla: in questo caso ci si riferisce ad un'estrazione conservativa dell'informazione necessaria per estrarre il più ampio sottoinsieme consentito di punti di lavoro inferiori a partire dal flusso estratto.

Verrà ora descritto un esempio di un meccanismo di composizione ed estrazione di un flusso di bit codificato e scalabile. In particolare, nello schema della Figura 11 si propone un esempio di organizzazione di un flusso codificato (blocco "GOP") relativo ad un intero filmato secondo una suddivisione a GOP (Group of Pictures). La prima suddivisione caratterizzante (a livello di GOP o a livello dell'intera sequenza video) è quella corrispondente alla risoluzione spaziale (blocco "Rs" nell'esempio in figura, in cui vengono evidenziati i livelli QCIF, CIF e 4CIF). Ciascun sottoflusso di livello spaziale è suddiviso in livelli temporali (blocco "Lt") ad ognuno dei quali è associata l'informazione codificata relativa ai campi di moto (nell'esempio si riporta la suddivisione del livello CIF). Ciascun sotto-flusso temporale può essere troncato secondo un criterio volto ad ottenere una certa qualità globale o un certo ritmo di bit complessivo (situazione più comune). Mentre le

intestazioni per l'accesso ai GOP ed ai livelli spaziali contengono unicamente le posizioni di inizio dei frammenti di flusso, nel caso dei livelli temporali, oltre all'informazione di accesso, si può rendere necessario trasmettere un'informazione aggiuntiva per l'estrattore su come tagliare i vari flussi di livello temporale al fine di ottimizzare la qualità di ricostruzione per un la quantità di bit assegnata (bit-budget).

E' inoltre possibile inserire un livello di suddivisione ulteriore in gruppi di frame (blocco "GF<sub>n</sub>") di medesimo livello temporale. Tale suddivisione, lievemente svantaggiosa in termini di efficienza di codifica, potrebbe rivelarsi molto utile in applicazioni che richiedono un basso ritardo tra le temporizzazioni assolute di codifica e decodifica (ritardo temporale, annullando i tempi di trasmissione e di calcolo, tra l'istante di codifica di un frame ed istante in cui lo stesso frame è disponibile dopo la decodifica). Il numero di suddivisioni per ciascun livello temporale dipende dalle caratteristiche di accesso desiderate e dalla tipologia di MCTF adottata.

In Figura 12 viene riportata un'esemplificazione dell'operato dell'estrattore. L'esempio riguarda un'estrazione dalla risoluzione spaziale CIF e al livello

temporale di 15 frame per secondo e ritmo di bit assegnato, partendo da un flusso complessivo di una codifica 4CIF a 60 fps. Si notino le porzioni di flusso completamente eliminate e quelle troncate relative ai livelli da mantenere. Si noti inoltre il tratteggio sull'eliminazione dell'intestazione che contiene le istruzioni per il troncamento. Tale eliminazione avviene infatti solo nel caso di modalità d'estrazione singola.

In accordo con una forma preferita di realizzazione, nel caso della codifica entropica CPDMM-GFRT0 descritta in precedenza è utile inserire nell'intestazione dei livelli temporali le lunghezze parziali del flusso codificato per ogni passo di codifica e per ogni piano di bit. Con queste informazioni, infatti, l'estrattore è in grado di effettuare un riempimento dei sottoflussi il più omogeneo possibile secondo un criterio teso alla minimizzazione della distorsione.

## RIVENDICAZIONI

1. Metodo di codifica di un segnale video comprendente le fasi di:

a) acquisire un primo segnale digitale (s1) rappresentativo di un segnale video, detto segnale digitale comprendendo una sequenza di immagini numeriche;

j) operare una trasformazione (TS) di detto primo segnale digitale (s1) per generare almeno un secondo segnale digitale video (s2) avente una risoluzione spaziale diversa da quella di detto primo segnale digitale (s1);

k) effettuare una trasformazione (TT) dell' almeno un secondo segnale digitale video (S2) generando un terzo segnale digitale (S3), detta trasformazione lasciando sostanzialmente invariata detta risoluzione spaziale del secondo segnale digitale video (s2) (*esempio meno del 20%*);

l) effettuare un'operazione di compressione (CE) di detto terzo segnale digitale (s3) per generare un primo segnale digitale compresso (Sc);

m) operare una trasformazione (CE-1) di detto primo segnale digitale compresso (Sc) per generare un segnale ricostruito (Sr) equivalente a quello ottenibile mediante una decompressione di detto primo segnale digitale compresso (Sc);



n) effettuare un'elaborazione del primo segnale digitale video ( $s_1$ ) comprendente una trasformazione (TT) della tipologia della trasformazione della fase c) per generare un segnale digitale ( $St$ ),

il metodo essendo caratterizzato dal fatto di comprendere inoltre le fasi di:

o) effettuare una trasformazione (TS) di detto segnale ( $St$ ) della tipologia della trasformazione della fase b) per ottenere un segnale digitale ( $s_4$ ) comprendente una prima componente ( $10$ ) e una seconda componente ( $11$ ), dove dette prima e seconda componente rappresentano in combinazione lo stesso contenuto informativo del segnale digitale ( $St$ ), detta trasformazione (TS) essendo adatta a generare detta prima componente ( $10$ ) in modo che abbia sostanzialmente la stessa risoluzione spaziale del segnale digitale ( $s_3$ );

p) generare un segnale di predizione ( $Sp$ ) rappresentativo della differenza fra detta prima componente ( $10$ ) e il segnale ricostruito ( $Sr$ ); e

q) effettuare un'operazione di compressione (CE) di detto segnale di predizione ( $Sp$ ) e di detta seconda componente ( $11$ ) per generare un secondo segnale digitale compresso ( $Sc_2$ ).

**2.** Metodo di codifica secondo la rivendicazione 1, in cui l'elaborazione del primo segnale digitale video

(s1) alla fase f) comprende inoltre una trasformazione (TS) di detto primo segnale digitale (s1) per generare almeno un segnale digitale video (s2) avente una risoluzione spaziale diversa da quella di detto primo segnale digitale (s1).

**3.** Metodo di codifica secondo la rivendicazione 1 o 2, in cui la trasformazione (TS) genera in uscita un segnale digitale video avente una risoluzione spaziale inferiore rispetto a quella del segnale in ingresso.

**4.** Metodo di codifica secondo una qualsiasi delle rivendicazioni precedenti, in cui la trasformata (TS) alla fase b) è una trasformata spaziale.

**5.** Metodo di codifica secondo la rivendicazione 4, in cui detta trasformata spaziale è una trasformata wavelet spaziale.

**6.** Metodo di codifica secondo una qualsiasi delle rivendicazioni precedenti, in cui la trasformata (TT) alla fase c) è una trasformata temporale.

**7.** Metodo di codifica secondo la rivendicazione 6, in cui detta trasformata temporale è una trasformata wavelet temporale.

**8.** Metodo di codifica secondo la rivendicazione 7, in cui detta trasformata wavelet temporale è una trasformata wavelet temporale moto-compensata.

**9.** Metodo di codifica secondo una qualsiasi delle

rivendicazioni precedenti, in cui le fasi dalla c) alla i) sono ripetute per ulteriori secondi segnali (S2).

**10.** Metodo di decodifica di un segnale video codificato secondo una qualsiasi delle rivendicazioni precedenti, comprendente i passi di:

h) acquisire un segnale digitale comprendente almeno il primo ed il secondo segnale digitale compresso (Sc, Sc2);

i) effettuare un'operazione di decompressione (CE-1) di detto primo segnale digitale compresso (Sc) per ottenere il segnale ricostruito (Sr);

j) effettuare una trasformazione (TT-1) di detto segnale ricostruito (Sr) per generare un segnale decodificato (Sd) rappresentativo del secondo segnale digitale video (S2);

k) effettuare un'operazione di decompressione (CE-1) di detto secondo segnale digitale compresso (Sc2) per ricostruire il segnale di predizione (Sp) e la seconda componente (11) del segnale digitale (S4);

l) effettuare una somma tra detto segnale di predizione (Sp) e detto segnale ricostruito (Sr) per ricostruire la prima componente (10) del segnale digitale (S4);

m) effettuare una trasformazione (TS-1) di dette prima e seconda componente (10, 11) per ricostruire il

segnale digitale (s4); e

n) effettuare una trasformazione (TT-1) di detto segnale digitale (s4) per generare un segnale decodificato (Sd1) rappresentativo del primo segnale digitale video (S1).

**12.** Un metodo di decodifica di un segnale video numerico in grado di decodificare una frazione del flusso di codifica per ricostruire una rappresentazione in un dominio trasformato del video corrispondente ad un certo livello di riferimento (in termini di qualità, risoluzione spaziale e risoluzione temporale), la quale rappresentazione viene utilizzata per predire la stessa tipologia di rappresentazione in un dominio trasformato del video ad un diverso livello di risoluzione spaziale.

**13.** Metodo di codifica di un segnale video, comprendente in successione le fasi di:

- a) generare rappresentazioni a diverse risoluzioni spaziali del segnale video;
- b) trasformare temporalmente dette rappresentazioni;
- c) eseguire un'operazione di predizione tra tali trasformate temporali attraverso una messa in corrispondenza di informazioni inerenti a risoluzioni spaziali omologhe; e

d) codificare entropicamente l'informazione risultante in modo progressivo, a partire da un primo livello di riferimento, detto base-layer.

**14.** Metodo di codifica secondo una qualsiasi delle rivendicazioni precedenti, dove il tipo di rappresentazione del video in un dominio trasformato è rappresentata da una trasformata wavelet temporale (realizzata anche con moto compensazione, i.e. MCTF) del segnale video per ogni livello di risoluzione spaziale desiderato.

**15.** Metodo di codifica secondo una qualsiasi delle rivendicazioni precedenti, dove l'informazione ad un determinato livello di risoluzione spaziale nel dominio trasformato viene ottenuta attraverso una elaborazione spaziale dell'informazione trasformata di un diverso livello di risoluzione spaziale, omologa all'elaborazione utilizzata per la generazione del video allo stesso diverso livello di risoluzione spaziale.

**16.** Metodo di codifica secondo una qualsiasi delle rivendicazioni precedenti, dove l'informazione di moto utilizzata per determinare il dominio trasformato della rappresentazione del video è ricavata in modo sincrono e per raffinamento tra almeno due livelli di risoluzione spaziali considerati.

**17.** Metodo di codifica secondo una qualsiasi delle rivendicazioni precedenti, dove la codifica progressiva viene ottenuta attraverso un meccanismo di codifica entropica dei simboli ottenuti dall'approssimazione di frame trasformati relativi a medesimi livelli di risoluzione temporali, in modo da garantire un'adeguata scalabilità temporale e in termini di qualità del flusso di codifica.

**18.** Metodo di codifica secondo la rivendicazione precedente, dove detta codifica entropica viene ottenuta attraverso un meccanismo di codifica dei simboli che operi per bit-plane di frame trasformati relativi a medesimi livelli di risoluzione temporali.

**19.** Un sistema di rappresentazione del partizionamento del flusso di codifica adatto ad essere sottoposto al metodo di decodifica secondo una qualsiasi delle rivendicazioni precedenti, comprendente meccanismi di indicizzazione che permettano un frazionamento del flusso di codifica per raggiungere operativamente il flusso da associare ad un certo livello di qualità per un dato punto di funzionamento in termini di risoluzione spaziale e temporale, e che possa consentire inoltre di rispettare requisiti di massimo ritardo codifica-decodifica.