ITALIAN PATENT OFFICE

Document No.

102012902101190A1

Publication Date

20140514

Applicant

STMICROELECTRONICS S.R.L.

Title

PROCEDIMENTO PER L'ESTRAZIONE DI INFORMAZIONI DISTINTIVE DA UN FLUSSO DI FRAME VIDEO DIGITALI, SISTEMA E PRODOTTO INFORMATICO RELATIVI

Front page automatically generated by WIPO

DESCRIZIONE dell'invenzione industriale dal titolo:
"Procedimento per l'estrazione di informazioni distintive da un flusso di frame video digitali, sistema e prodotto informatico relativi"
di: STMicroelectronics S.r.l., nazionalità italiana, Via C. Olivetti, 2 - 20864 Agrate Brianza (MB)
Inventore designato: Danilo Pietro PAU
Depositata il: 14 novembre 2012

****

TESTO DELLA DESCRIZIONE

Campo tecnico

La presente descrizione si riferisce alle tecniche per estrarre informazioni distintive (feature) da un flusso di immagini (frame) video digitali.

Varie forme di attuazione possano applicarsi al trattamento di immagini, ad esempio nell'ambito di processi di matching (ricerca degli abbinamenti o corrispondenza) e/o di retrieval (ricerca e recupero) di immagini.

Sfondo tecnologico

Documenti quali il documento ISO/IEC JTC1/SC29/WG11/W12929, July 2012, Stockholm, SE, Test Model 3: Compact Descriptors For Visual Search descrivono tecniche di elaborazione (ad esempio con architettura pipeline) basate sull'impiego di descrittori compatti utilizzabili per funzioni di matching e retrieval di immagini.

Il documento in questione non prende peraltro in considerazione il fatto che l'uso di tali descrittori, ad esempio per operazioni di abbinamento o recupero da parte di un server, può essere soggetto a vincoli in termini di

larghezza di banda o bitrate disponibile. Questa situazione può derivare, ad esempio, da limitazioni della larghezza banda disponibile su un collegamento, della variabilità della velocità di trasmissione dati, da fenomeni di congestione di rete o da considerazioni di varia natura legate alle prestazioni computazionali del server.

Scopo e sintesi

Varie forme di attuazione si prefiggono lo scopo di superare questa carenza.

Varie forme di attuazione permettono di conseguire tale scopo grazie a un procedimento avente le caratteristiche richiamate nelle rivendicazioni che seguono.

Varie forme di attuazione possono riferirsi ad un corrispondente sistema nonché ad un prodotto informatico caricabile nella memoria di almeno un elaboratore e comprendente parti di codice software suscettibili di realizzare le fasi del procedimento quando il prodotto è eseguito su almeno un elaboratore. Così come qui utilizzato, il riferimento ad un tale prodotto informatico è inteso essere equivalente al riferimento ad un mezzo leggibile da elaboratore contenente istruzioni per il controllo del sistema di elaborazione per coordinare l'attuazione del procedimento secondo l'invenzione. Il riferimento ad "almeno ad un elaboratore" è evidentemente inteso a mettere in luce la possibilità che la presente invenzione sia attuata in forma modulare e/o distribuita.

Le rivendicazioni formano parte integrante dell'insegnamento tecnico qui somministrato in relazione all'invenzione.

Varie forme di attuazione permettono di fissare il

numero di informazioni distintive o feature selezionato per ciascuna frame in funzione di un valore obiettivo (target) di bitrate con la conseguente possibilità di sfruttare in modo ottimale e dinamico le risorse disponibili in termini di bitrate per l'utilizzazione dei relativi descrittori.

Varie forme di attuazione permettono di codificare e inviare verso un server descrittori compatti così da poter migliorare le prestazioni in termini di precisione e di occupazione di canale (e di risorse, in generale) ad esempio da parte di un bitstream di richieste (o query).

Breve descrizione delle figure

Varie forme di attuazione saranno ora descritte, a puro titolo di esempio non limitativo, con riferimento alle figure annesse, in cui:

- le figure 1 e 2 sono schemi a blocchi funzionali che illustrano possibili funzioni di trattamento di immagini basate sulla estrazione di informazioni distintive (feature),

- la figura 3 è uno schema a blocchi funzionale che illustra possibili modalità di estrazione di feature,

- la figura 4 è uno schema a blocchi genericamente illustrativo di un possibile contesto di applicazione di forme di attuazione,

- la figura 5 è uno schema a blocchi rappresentativo della generale architettura di esempi di attuazione, e

- la figura 6 è uno schema a blocchi che illustra esempi di attuazione con diretto riferimento allo schema della figura 3.

Descrizione particolareggiata

Nella seguente descrizione sono illustrati vari

dettagli specifici finalizzati ad un'approfondita comprensione di vari esempi di forme di attuazione. Le forme di attuazione possono essere realizzate senza uno o più dei dettagli specifici, o con altri metodi componenti materiali, etc. In altri casi, strutture, materiali o operazioni noti non sono mostrati o descritti in dettaglio per evitare di rendere oscuri i vari aspetti delle forme di attuazione.

Il riferimento ad "una forma di attuazione" nell'ambito di questa descrizione sta ad indicare che una particolare configurazione, struttura o caratteristica descritta in relazione alla forma di attuazione è compresa in almeno una forma di attuazione. Quindi, frasi come "in una forma di attuazione", eventualmente presenti in diversi luoghi di questa descrizione non sono necessariamente riferite alla stessa forma di attuazione. Inoltre, particolari conformazioni, strutture o caratteristiche possono essere combinate in ogni modo adeguato in una o più forme di attuazione.

I riferimenti qui utilizzati sono soltanto per comodità del lettore e non definiscono dunque l'ambito di tutela o la portata delle forme di attuazione.

Le figure 1 a 3 sono desunte dal documento ISO/IEC JTC1/SC29/WG11/W12929, già in precedenza citato, ed in particolare dal capitolo 1. Introduction e dal capitolo 2. Compact descriptor extraction del documento in questione, che, al successivo capitolo 5. Software Architecture and implementation details illustra esempi di implementazione di una architettura per la selezione delle informazioni distintive più significative da comprimere in ciascuna frame video digitale considerata.

In particolare, gli schemi delle figure 1 e 2

illustrano a livello di schema a blocchi esempi di architetture funzionali utilizzabili per svolgere una funzione di calcolo degli abbinamenti o corrispondenza (matching) M o di reperimento o ricerca (search o retrieval) S di frame video, ossia di immagini, digitali.

Nel primo caso (azione di matching M della figura 1) è possibile verificare in modo automatizzato (ad esempio tramite un computer) se due immagini rappresentano gli stessi oggetti o la stessa scena. Ciò avviene operando su su una immagine di interrogazione (Query Image) QI e su una immagine di riferimento (Reference Image) RI.

A tal fine, entrambe le immagini possono essere sottoposte ad un'operazione di estrazione di descrittori (genericamente indicata con 10) e ad un operazione di confronto condotta in M operando appunto sui descrittori ed informazioni distintive in 10 e diretta a rilevare l'eventuale adattamento o matching. Il risultato dell'elaborazione, indicato con R, indica se l'immagine di interrogazione o Query Image QI rappresenta oppure no gli stessi oggetti o la stessa scena, ossia si adatta o corrisponde a ("matches") l'immagine di riferimento RI.

La funzione di retrieval cui fa schematicamente riferimento la figura 2 opera fondamentalmente secondo gli stessi criteri con la differenza data dal fatto che, nel caso della figura 2, il confronto in S è fatto con più immagini di riferimento RI_i estratte da una base dati o database DB. In questo modo è possibile ricercare e reperire, nell'ambito del database DB, una o più immagini che riproducono gli stessi oggetti o la stessa scena dell'immagine di interrogazione QI.

Varie architetture e tecnologie di componenti suscettibili di realizzare le funzioni illustrate con

riferimento alle figure 1 e 2 sono definite nel cosiddetto Test Model 3.0 (TM3) for Compact Descriptors for Visual Search (noto anche con l'acronimo di CDVS): il modello TM3 implementa le funzionalità richieste per l'estrazione e il confronto di descrittori compatti limitati ad un insieme di lunghezze di descrittore (Descriptor Length) predeterminate.

Lo schema a blocchi della figura 3 illustra un possibile schema a blocchi di un sistema di elaborazione (ad esempio strutturata come pipeline) per l'estrazione di informazioni distintive o feature da frame video.

In varie forme di attuazione, un descrittore compatto di un'immagine I (che, in termini più generali, può essere vista sotto forma di trame o frame video digitali), può comprendere, oltre ad un certo numero di descrittori locali, anche un singolo descrittore globale suscettibile di rappresentare l'immagine nel suo insieme.

Lo schema della figura 3 illustra un esempio di architettura che permette di produrre un descrittore compatto di un'immagine/frame secondo il modello TM3 attraverso una sequenza di passi di elaborazione.

Nello schema della figura 3, il riferimento 12 illustra una funzione/modulo di identificazione di punti di interesse o punti chiave (interest points o keypoints) in un'immagine basata su una rappresentazione multiscala e suscettibile di essere realizzata, ad esempio, con una tecnica di differenze gaussiane (Differences-of-Gaussians o DoG).

Nello schema esemplificativo della figura 3, la funzione/modulo DoG 12 è seguita da una funzione/modulo 14 in cui i cosiddetti feature point sono descritti, ad esempio, con una trasformata invariante rispetto alla scala

(Scale Invariant Feature Transform o SIFT).

Nello schema esemplificativo della figura 3, i risultati dell'azione di trasformata realizzata in 14 vengono alimentati ad una funzione/modulo 16 di selezione dei punti di interesse o keypoint e descrittori locali associati, destinata a svolgere una funzione di selezione delle informazioni distintive o feature.

Ad esempio, in varie forme di attuazione, la funzione/modulo 16 può assegnare un valore positivo a ciascun tratto in funzione delle sue caratteristiche DOG, del relativo vettore SIFT e delle sue coordinate.

Ad esempio, nel caso in cui la n-esima feature di un'immagine sia indicata con s_n (con una connotazione che incorpora le caratteristiche DoG, il vettore SIFT e le relative coordinate), è possibile indicare con r una grandezza indicativa della rilevanza del punto di interesse (key point relevance); in questo modo, ad un determinato tratto o feature è possibile associare un valore $r(s_n)$, ed i suddetti valori di rilevanza possono essere ordinati in una scala decrescente facendo sì che, ad esempio, soltanto le prime L feature $n_1, \ldots, n_L$ siano mantenute, così da poter mantenere la lunghezza media della sequenza entro un certo limite di lunghezza.

Il risultato della elaborazione svolta in 16 (ossia la selezione di un numero limitato di punti chiave in grado di massimizzare una misura della qualità attesa in funzione dell'uso previsto - ad esempio per azioni di matching o search, così come esemplificate nelle figure 1 e 2) può essere sottoposto nell'ambito di una funzione/modulo 18 ad una compressione dei descrittori locali, ossia una quantizzazione scalare o vettoriale delle feature selezionate in 16.

Il risultato dell'elaborazione svolta in 18 è poi trasmesso ad una funzione/modulo 20 di compressione delle coordinate che comprime le coordinate dei punti di interesse selezionati così da generare in uscita i descrittori locali LD.

Il risultato della elaborazione svolta in 16 può poi essere inviato anche ad una funzione/modulo 22 che svolge una funzione di aggregazione delle feature (ad esempio SIFT) così da formare un descrittore globale GD.

Possibili criteri d'implementazione dello schema esemplificato nella figura 3 sono descritti in dettaglio nel documento ISO/IEC JTC1/SC29/WG11/W12929, già più volte citato, anche alla luce dell'articolo di D.G. Lowe: "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60, 91-110 (2004). Per motivi di brevità, la relativa descrizione di dettaglio non viene quindi ripetuta in questa sede, anche perché di per sé non essenziale ai fini della comprensione delle forme di attuazione.

Lo schema della figura 4 fa riferimento ad una possibile applicazione di elaborazione e trattamento di frame video digitali (ad esempio quali funzioni di matching o di search/retrieval, così come descritte in precedenza) attuata su un flusso di frame video in cui possono essere identificabili una prima frame F_i ed una seconda frame F_{i+1} qui considerate per semplicità separate da un intervallo di tempo 1/fps dove fps è il numero di frame per secondo nel flusso di frame video considerato: si apprezzerà peraltro che, in varie forme di attuazione, le frame F_i e F_{i+1} possono anche non essere l'una consecutiva all'altra nel flusso e cioè essere separate per multipli di 1/fps.

In varie forme di attuazione, una funzione di matching

come qui considerata può trovare applicazione anche in una tecnica di codifica di un flusso di frame video digitali così come descritta in una domanda di brevetto per invenzione industriale depositato in pari data dalla stessa richiedente.

Rifacendosi a quanto detto in precedenza con riferimento alle figure 1 a 3 (ed alla illustrazione di dettaglio fornita nei documenti ISO/IEC JTC1/SC29/WG11/W12929 e Lowe in precedenza citati), l'estrazione dei descrittori attuata nell'architettura descritta a titolo di esempio con riferimento alla figura 3 può essere ricondotta ad una generazione di rispettive sequenze (ad es. le sequenze utilizzate in M e S per confrontate le immagini di "query" QI con le immagini di riferimento RI e RIi) comprendenti un insieme di coppie [KP_{1i}, D_{1i}], [KP_{2i}, D_{2i},], ……, [K_{mi}, D_{mi}] in cui ai punti di interesse KP_{1i}, KP_{2i}, …… KP_{mi} sono associati i relativi descrittori D_{1i}, $D2_i$, ……, D_{mi}.

Una tale sequenza può essere compressa ad un numero limitato di bit limitando il numero delle coppie comprese nella sequenza. Ad esempio, il numero m delle coppie può essere limitato ad un valore di soglia T_i per ciascuna frame, ossia m = 1, …, T_i per la frame F_i con il numero di bit relativi limitato ad un valore num_bits_i, con la soglia T_i che identifica in pratica il numero di coppie punti di interesse/descrittori raggiunto il quale ogni ulteriore coppia, anche se rilevante, non contribuirebbe a generare bit in una sequenza corrispondente o, in alternativa, sarebbe compressa a 0 bit.

Così come illustrato nella figura 5, in varie forme di attuazione, il risultato dell'azione di estrazione, ossia la sequenza di coppie di punti di interesse e relativi

descrittori [KP_{1i}, D_{1i}], [KP_{2i}, $D_{2i,}$], ……, [KP_{mi}, D_{mi}] può essere inviato dal modulo estrattore 10 verso un server 24 destinato ad utilizzare le suddette sequenze per una funzione di elaborazione.

Come esempio, rifacendosi alle figure 1 e 2, si può pensare a contesti di impiego le sequenze [KP_{1i}, D_{1i}], [KP_{2i}, $D_{2i,}$], ……, [KP_{mi}, D_{mi}] estratte dalle immagini di query QI siano inviate verso un server 24 che svolge le funzioni di matching M o retrieval S operando su sequenze omologhe inerenti alle immagini di riferimento RI o Rii, ad esempio residenti presso il server 24 stesso.

La trasmissione delle sequenze di query [KP_{1i}, D_{1i}], [KP_{2i}, $D_{2i,}$], ……, [KP_{mi}, D_{mi}] può avvenire su un canale di trasmissione costituito, ad esempio, da un canale reale o virtuale nell'ambito di una rete N, con una banda o bitrate disponibile non costante, ma variabile nel tempo, ad esempio, per fenomeni di congestione della rete N e/o per il variare del rapporto segnale/rumore o segnale/interferenza sul canale di trasmissione (caso tipico delle reti wireless). Di conseguenza, il bitrate di cui il server 24 può disporre per l'elaborazione può variare e risultare limitato in modo diverso nel tempo.

Sempre a titolo di esempio, in aggiunta o in alternativa, possono determinarsi situazioni in cui il bitrate di cui il server 24 può disporre per l'elaborazione può variare per fenomeni legati al server stesso, ad esempio di eccessivo sovraccarico computazionale, perché il server 24 opera in multitasking ed il bitrate allocabile all'elaborazione delle query ricevute può variare nel tempo secondo il carico e la schedulazione dei task.

La figura 5 esemplifica il fatto che, in varie forme di attuazione, il valore corrente di bitrate disponibile al

server 24 (indicato con Target_bitrate) è trasmesso dal server 24 verso l'estrattore 10.

Così come schematicamente illustrato nella figura 6 (in cui parti e componenti identici o equivalenti a quelli già descritti con riferimento alla figura 3 sono indicati con gli stessi riferimenti che compaiono nella figura 3), in varie forme di attuazione il valore Target_bitrate può essere utilizzato in una funzione/modulo 26 per calcolare il valore di soglia di selezione Ti utilizzato dall'estrattore 10, ad esempio in 16, per selezionare i punti chiave e limitare il numero m = 1, …, T_i delle coppie [KP_{1i}, D_{1i}], [KP_{2i}, D_{2i},], ……, [KP_{mi}, D_{mi}] estratte per la generica frame F_i così da limitare in modo corrispondente al valore num_bits_i il numero di bit relativi

In varie forme di attuazione, il calcolo può essere quindi svolto in modo adattivo, limitando superiormente il numero di coppie punto di interesse/descrittore generate dall'estrattore 10 in funzione del valore Target_bitrate: questo valore identifica la capacità da parte del server 24 di ricevere ed elaborare tali sequenze. In varie forme di attuazione, la suddetta azione di limitazione permette di evitare di impegnare l'estrattore 10 nella estrazione di coppie punto di interesse/descrittore che al momento il server 24 non potrebbe ricevere o, anche ricevendole, non riuscirebbe ad elaborare.

In varie forme di attuazione, la funzione/modulo 26 può essere configurata per generare il suddetto valore di soglia Ti in funzione del valore di bitrate Target_bitrate indicato come disponibile dal server 24 (valore che si assume essere variabile nel tempo, idealmente da frame a frame, anche se le relative variazioni sono di solito più lente) tenendo in conto anche due altri parametri, ossia:

- il parametro fps, ossia il numero di frame per secondo nel flusso di frame in ingresso al blocco 12, e

- il numero di bit num_bits_i utilizzato per comprimere le coppie punto di interesse/descrittore, rilevato, ad esempio, da un contatore di bit 28 in grado di "sentire" i descrittori LD e GD.

In varie forme di attuazione, il modulo 26 può operare secondo i seguenti criteri.

Partendo dal numero bit num_bits_i utilizzato per "codificare" (ossia per esprimere) le coppie punto di interesse/descrittore per la frame F_i è possibile determinare un numero di bit medio $average_bit_couple_i$: il numero di bit utilizzato per ciascuna coppia può infatti essere diverso da coppia a coppia, ad esempio in funzione delle caratteristiche del singolo descrittore.

Per il suddetto numero medio $average_bit_couple_i$ può valere la seguente espressione:

$$average_bit_couple_i = num\ bits_i/T_i.$$

Partendo dalla grandezza Target_bitrate ricevuta dal server 24 è altresì possibile calcolare un valore di bit obiettivo (target) per ciascuna frame $Target_F_i$, esprimibile come:

$$Target_F_i = Target_bitrate/fps.$$

È poi possibile calcolare, con riferimento alla generica frame F_i, la differenza:

$$\delta_i = Target_F_i - num_bits_i$$

ossia la differenza (con segno) fra il valore di bit obiettivo (target) per la frame i-esima ed il numero effettivo di bit effettivamente utilizzati per "codificare" la frame in questione.

Il valore assoluto della differenza δ_i indica di quanto l'azione di estrazione svolta dall'estrattore 10 si

discosta dalla capacità di ricezione/elaborazione del server 24 ed il segno identifica il verso dello scarto:

- positivo, se il server 24 può ricevere/elaborare più di quanto l'estrattore 10 gli manda;

- negativo, se il server 24 non è in grado di ricevere/elaborare quanto l'estrattore 10 gli manda.

In funzione della differenza δ_i è allora possibile stabilire un valore obiettivo $Target_F_{i+1}$ definito come:

$Target_F_{i+1} = Target_F_i + \delta i$

ossia aggiungendo o sottraendo dal processo di codifica così come al momento attuato dall'estrattore 10 un numero di bit pari a δ_i, ad esempio calcolando un nuovo valore di soglia T_{i+1} per il numero di coppie punto di interesse/descrittore da utilizzare per la frame successiva F_{i+1}, ad esempio sulla base della relazione:

$T_{i+1}=target_F_{i+1}/average_bit_couple_i$.

In varie forme di attuazione, il parametro average_bit_couple potrebbe essere stimato per la frame i+1 con un primo tentativo di codifica per poi essere applicato in un secondo passo reale di codifica.

Naturalmente, fermo restando il principio dell'invenzione, i particolari di realizzazione e le forme di attuazione potranno variare, anche in modo significativo, rispetto a quanto qui illustrato a puro titolo di esempio non limitativo senza per questo uscire dall'ambito di protezione. Tale ambito di protezione è definito dalle rivendicazioni annesse.

RIVENDICAZIONI

1. Procedimento per estrarre informazioni distintive o feature da frame video digitali (F_i, F_{i+1}) in un flusso, il procedimento comprendendo:

- estrarre con un estrattore (10) da dette frame video (F_i, F_{i+1}) rispettive sequenze di coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) limitando (16) ad un valore di soglia (T_i) il numero (m) di coppie estratte per ciascuna frame,

- inviare le sequenze di coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) estratte con l'estrattore (10) verso un server (24) per l'elaborazione con un bitrate avente un valore che varia nel tempo,

- ricevere detto valore di bitrate variabile nel tempo all'estrattore (10) come bitrate obiettivo (Target_bitrate) per l'estrazione,

- limitare (16) ad un valore di soglia (T_i) variabile nel tempo in funzione di detto bitrate obiettivo (Target_bitrate) il numero (m) di coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) estratte dall'estrattore (10).

2. Procedimento secondo la rivendicazione 1, in cui detta elaborazione al server (24) comprende un'elaborazione compresa fra un abbinamento o corrispondenza (matching - M) ed un reperimento o ricerca (search o retrieval - S).

3. Procedimento secondo la rivendicazione 1 o la rivendicazione 2, comprendente:

- determinare, in funzione di detto bitrate obiettivo

(Target_bitrate), un numero di bit obiettivo (Target_F_i) disponibile per le coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) estratte per una certa frame (F_i),

- determinare, per detta frame (F_i), una differenza con segno (δ_i) fra il suddetto numero di bit obiettivo (Target_F_i) disponibile per le coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) estratte per una certa frame (F_i) ed il numero di bit (num_bits_i) compreso nelle coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) estratte per la stessa frame (F_i), e

- determinare un numero di bit obiettivo (Target_F_{i+1}) disponibile per le coppie di punti di interesse/descrittori da estrarre per una frame (F_{i+1}) successiva nel flusso aggiungendo detta differenza con segno (δ_i) al suddetto numero di bit obiettivo (Target_F_i) disponibile per le coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) estratte per detta certa frame (F_i).

4. Procedimento secondo la rivendicazione 3, comprendente:

- determinare un numero medio di bit ($average_bit_couple_i$) compreso nelle coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) estratte per detta certa frame (F_i), e

- determinare detto valore di soglia (T_{i+1}) per le coppie di punti di interesse/descrittori da estrarre per una frame (F_{i+1}) successiva nel flusso quale rapporto fra il numero di bit obiettivo (Target_F_{i+1}) disponibile per le coppie di punti di interesse/descrittori da estrarre per detta frame (F_{i+1}) successiva nel flusso e detto numero

medio di bit (average_bit_couple$_i$) compreso nelle coppie di punti di interesse/descrittori ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], …, [KP_{mi}, D_{mi}]) estratte per detta certa frame (F_i).

5. Sistema per estrarre informazioni distintive o feature da un flusso di video frame digitali, il sistema comprendendo un estrattore (10) configurato per operare con il procedimento secondo una qualsiasi delle rivendicazioni 1 a 4.

6. Prodotto informatico caricabile nella memoria di almeno un elaboratore e comprendente porzioni di codice software per attuare le fasi del procedimento secondo una qualsiasi delle rivendicazioni 1 a 4.

CLAIMS

1. A method of extracting features from digital video frames (F_i, F_{i+1}) in a flow, the method including:

- extracting at an extractor (10) from said video frames (F_i, F_{i+1}) respective sequences of couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) by limiting (16) to a threshold value (T_i) the number (m) of couples extracted for each frame,

- sending the sequences of couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) extracted by the extractor (10) towards a server (24) for processing with a bitrate having a value that varies over time,

- receiving said bitrate value variable over time at the extractor (10) as a target bitrate (Target_bitrate) for extraction,

- limiting (16) to a threshold value (T_i) which varies over time as a function of said target bitrate (Target_bitrate) the number (m) of couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) extracted by the extractor (10).

2. The method of claim 1, wherein said processing at the server (24) includes processing selected between matching (M) and search or retrieval (S).

3. The method of claim 1 or claim 2, including:

- determining as a function of said target bitrate (Target_bitrate), a target number of bits (Target_F_i) available for the couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) extracted for a

certain frame (F_i),

- determining, for said certain frame (F_i), a difference with sign (δ_i) between the target number of bits (Target_F_i) available for the couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) extracted for said certain frame (F_i) and the number of bits (num_bits$_i$) included in the couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) extracted for the same frame (Fi), and

- determining a target number of bits (Target_F_{i+1}) available for the couples to be extracted for one subsequent frame (F_{i+1}) in the flow by adding said difference with sign (δ_i) to said target number of bits (Target_F_i) available for couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) extracted for said certain frame (F_i).

4. The method of claim 3, including:

- determining an average number of bits (average_bit_couple$_i$) included in the couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) extracted for said certain frame (Fi), and

- determining said threshold value (T_{i+1}) for the couples of points of interest/descriptors to be extracted for a subsequent frame (F_{i+1}) in the flow as the ratio of the number of target bits (Target_F_{i+1}) available for the couples of points of interest/descriptors to be extracted for said subsequent frame (F_{i+1}) in the flow to said average number of bits (average_bit_couple$_i$) included in the couples of points of interest/descriptors ([KP_{1i}, D_{1i}], [KP_{2i}, D_{2i}], ..., [KP_{mi}, D_{mi}]) extracted for said certain frame (Fi).

5. A system for extracting features from a flow of digital video frames, the system including an extractor (10) configured for operating with the method according to any one of claims 1 to 4.

6. A computer program product loadable into the memory of at least one computer and including software code portions for implementing the steps of the method according to any one of claims 1 to 4.

FIG. 1

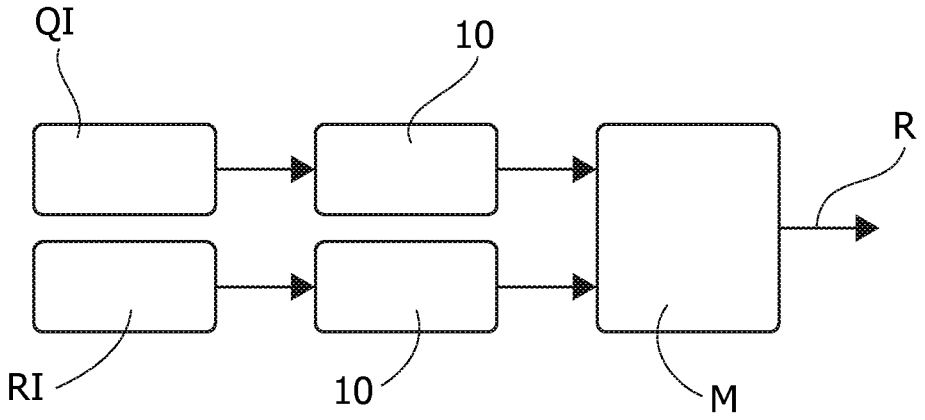


FIG. 2

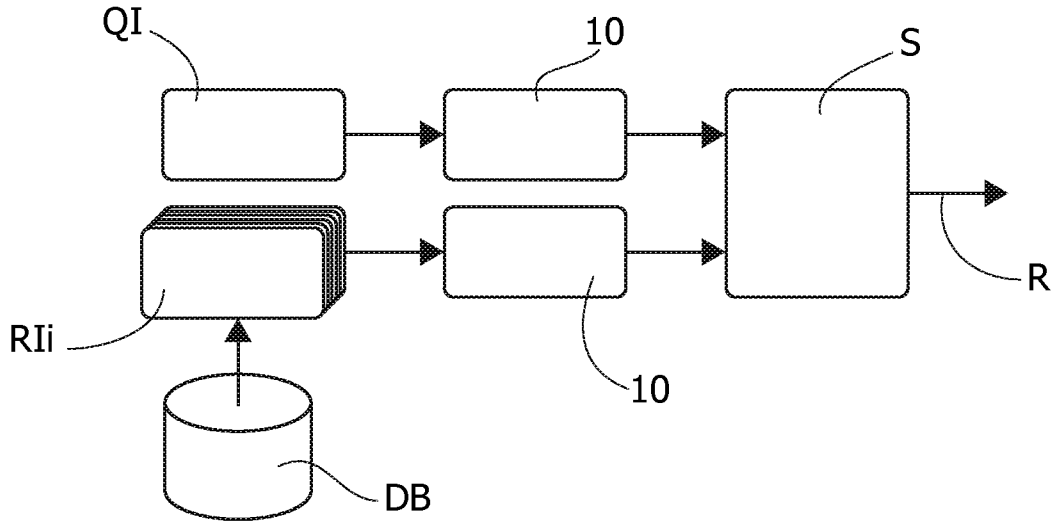


FIG. 3

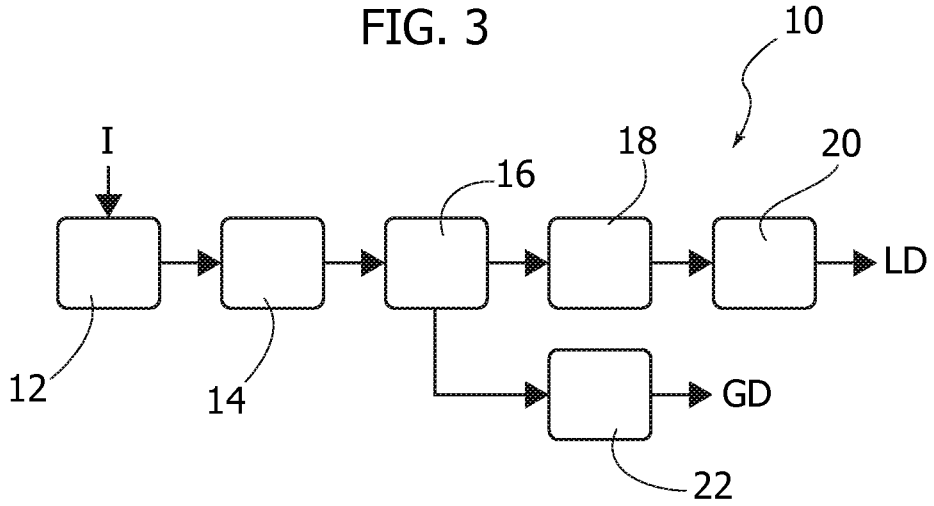


FIG. 4

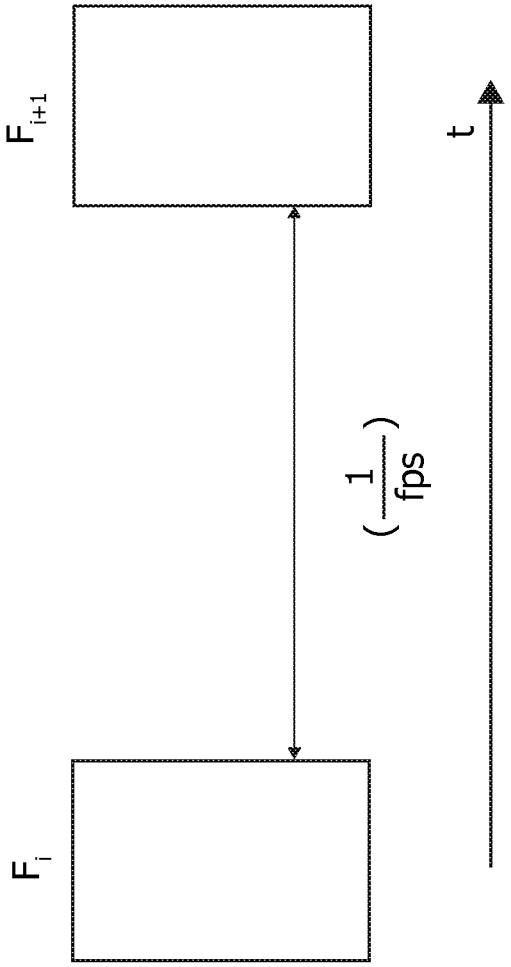
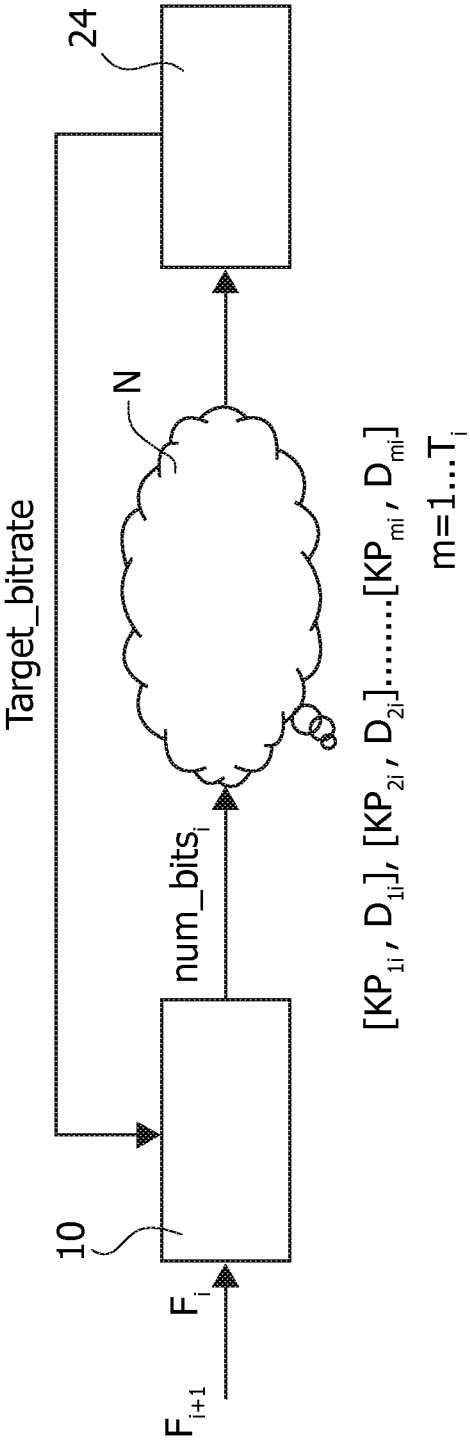

F_i
F_i+1
(1/fps)
t

FIG. 5


F_i+1
F_i
10
num_bits_i
N
24
Target_bitrate
[KP_1i, D_1i], [KP_2i, D_2i].....[KP_mi, D_mi]
m=1...T_i

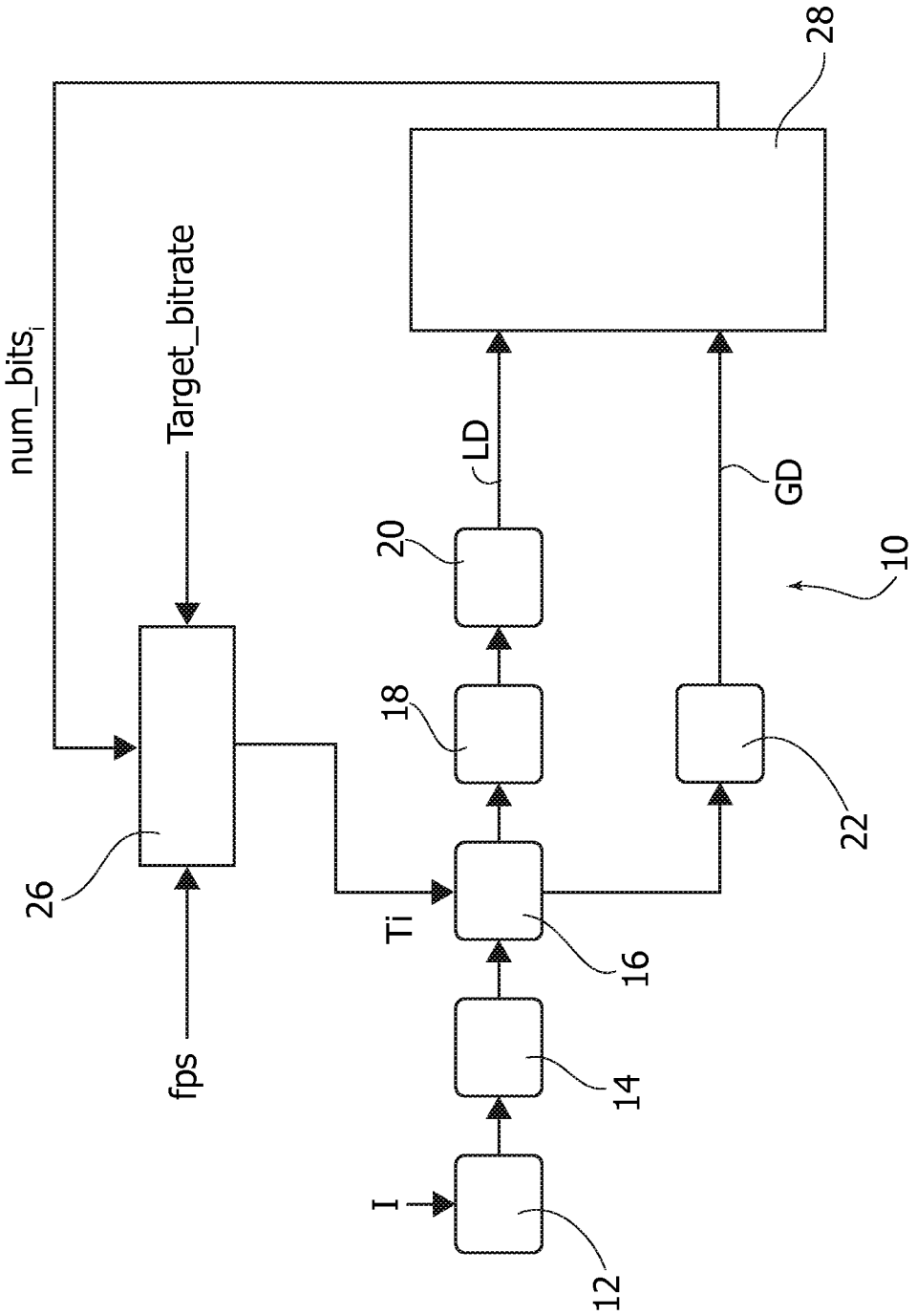



FIG. 6