

ITALIAN PATENT OFFICE

Document No.

102010901852378A1

Publication Date

20111228

Applicant

SISVEL TECHNOLOGY S.R.L.

Title

METODO PER LA DECODIFICA 2D-COMPATIBILE DI FLUSSI VIDEO
STEREOSCOPICI

Descrizione dell'Invenzione Industriale dal titolo: **-SVT014-
"METODO PER LA DECODIFICA 2D-COMPATIBILE DI FLUSSI VIDEO
STEREOSCOPICI"**

di Sisvel Technology, di nazionalità italiana, con sede in Via Castagnole 59, 10060 None (TO), ed elettivamente domiciliata presso i Mandatari Ing. Roberto Dini (No. Iscr. Albo 270 BM), Ing. Marco Camolese (No. Iscr. Albo 882 BM), Ing. Antonio Di Bernardo (No. Iscr. Albo 1163 BM), Dott. Giancarlo Reposio (No. Iscr. Albo 1168 BM) ed Ing. Andrea Grimaldo (No. Iscr. Albo 1060 BM) c/o Metroconsult S.r.l., Via Sestriere 100, 10060 None (TO).

Inventori designati:

- BALLOCCA Giovanni, Via Lavazza 56/8, 10131 Torino
- D'AMATO Paolo, Via Costabella 21, 00195 Roma
- CELIA Saverio, Via Torino 45, 10040 Rivalta di Torino (TO)

Depositata il _____ No. _____

DESCRIZIONE

La presente invenzione si riferisce ad un metodo per la decodifica di un flusso video stereoscopico digitale, ossia di un flusso video che, elaborato in modo opportuno in un dispositivo di visualizzazione, produce sequenze di immagini del tipo che vengono percepite come tridimensionali da un osservatore.

Tale metodo consente all'utente che dispone di un decoder e/o di un televisore convenzionali (non stereoscopici) la visualizzazione delle immagini stereoscopiche in 2D, nonché la visualizzazione 2D su decoder e/o televisore stereoscopici (3D).

Nel corso degli ultimi anni, il mondo della produzione cinematografica, stimolato dalla disponibilità di nuovi strumenti di produzione resi possibili dall'avvento delle tecnologie digitali, ha rivolto grande attenzione e ingenti

risorse alla produzione di contenuti 3D di tipo stereoscopico.

L'interesse per il 3D si sta estendendo al possibile uso domestico, cioè con visualizzazione delle immagini su di un televisore. Ad esempio, alcuni operatori di pay-TV stanno per iniziare a trasmettere programmi in 3D.

L'approccio più comune nella presentazione di contenuti video stereoscopici è quello basato sulla visualizzazione di due flussi video indipendenti, destinati rispettivamente all'occhio destro e all'occhio sinistro, che vengono ricomposti dal cervello nella rappresentazione di un oggetto tridimensionale.

I contenuti stereoscopici per uso domestico sono generalmente contenuti video in alta definizione e possono essere distribuiti su un supporto di memorizzazione di massa (DVD, Blu-ray disc, supporti magneto-ottici o a stato solido e così via) oppure attraverso canali di broadcasting (cablati o wireless) o via rete telematica (IP).

In ambiente di produzione, però, può risultare impossibile, con le strutture esistenti, trasferire e gestire in modo separato i due flussi prodotti dalle telecamere stereoscopiche che riprendono la scena da due punti di vista separati.

Inoltre, le reti di distribuzione che permettono di raggiungere l'utente finale hanno un dimensionamento tale da rendere antieconomico l'uso di due flussi ad alta definizione indipendenti per l'erogazione di un singolo servizio. Di conseguenza, è necessario adottare alcuni accorgimenti in fase di produzione che consentano di ridurre il bit-rate necessario al trasferimento ed alla fruizione dei contenuti.

Poiché i risultati degli studi sulle differenze nella percezione dei dettagli nelle immagini bidimensionali e

tridimensionali sembrano indicare che, anche abbassando la risoluzione dei contenuti stereoscopici rispetto a quella dei contenuti bidimensionali, la qualità percepita dall'utente rimane accettabile, sono state sviluppate tecniche differenti per l'impacchettamento (anche denominato con l'equivalente espressione inglese "frame packing") delle due immagini componenti la vista stereoscopica in un singolo fotogramma.

Ad esempio, nel caso di un singolo fotogramma C ad alta definizione (1920 x 1080 pixel), vengono acquisite le due immagini che compongono il canale sinistro e quello destro (nel seguito L e R) con una risoluzione orizzontale pari alla metà della risoluzione di un fotogramma ad alta definizione ed in seguito disposte fianco a fianco in un singolo fotogramma (formato side-by-side), come rappresentato in Figura 1a.

In questo modo, è possibile utilizzare un singolo flusso ad alta definizione per il trasporto dei due canali video indipendenti; al momento della decodifica i due semifotogrammi vengono separati e riportati al formato 16/9 applicando opportune tecniche di interpolazione.

Allo stesso modo, è possibile applicare un procedimento analogo dimezzando la risoluzione verticale, mantenendo intatta quella orizzontale, e disponendo i due fotogrammi L ed R uno sopra l'altro (formato top-bottom), come rappresentato in Figura 1b.

Il flusso video stereoscopico costituito dai fotogrammi compositi viene quindi compresso, al fine di ridurre il bit-rate di trasporto per la distribuzione su rete di broadcasting, rete IP o supporto di memorizzazione di massa.

Uno dei requisiti sui quali si concentra l'attenzione dei fornitori di servizio, in particolare i broadcaster

responsabili del servizio pubblico, è quello della 2D-compatibilità dei segnali stereoscopici.

Infatti, per consentire agli utenti che già possiedono un decoder ad alta definizione di fruire dei servizi trasmessi, è desiderabile che i programmi 3D possano essere visualizzati anche come programmi 2D. Allo stesso modo è desiderabile che un contenuto 3D su DVD, Blu-ray disc 3D o sito Internet, possa essere visualizzato dai televisori, o monitor, sia 2D che 3D.

Questo risultato può essere raggiunto in due modi: trasmettendo simultaneamente la versione 2D e quella 3D di uno stesso programma, oppure adottando una opportuna tecnica di codifica del flusso stereoscopico.

La prima opzione comporta ovviamente uno spreco di banda che gli operatori del servizio desiderano evitare per quanto possibile.

Per quanto riguarda la seconda opzione, sono note nell'arte tecniche per la generazione di flussi stereoscopici 2D-compatibili.

Una di queste tecniche è legata all'applicazione delle cosiddette "mappe di profondità", descritte ad esempio nelle domande di brevetto statunitensi no. US 2002/0048395 e no. US 2004/0101043.

In pratica, al video a colori bidimensionale viene associato un segnale, in forma di video supplementare in bianco e nero, che trasporta le mappe di profondità. Un opportuno decodificatore può ricostruire un video stereoscopico a partire dai dati ricevuti. Tuttavia, questa tecnica soffre degli stessi problemi della succitata trasmissione in 2D ed in 3D dello stesso programma: è infatti necessario trasferire due segnali video in parallelo, con elevato dispendio in termini di bit-rate di trasporto.

Un'altra tecnica di codifica del flusso stereoscopico di tipo 2D-compatibile è ad esempio quella denominata "multiview".

Poiché le coppie di immagini destra e sinistra che compongono il flusso video stereoscopico sono caratterizzate da un elevato grado di somiglianza, si utilizzano le stesse tecniche di soppressione della ridondanza spazio-temporale impiegate nella codifica dei flussi bidimensionali. Infatti, sottratto un certo offset dovuto alla distanza geometrica dei punti di ripresa (la distanza interoculare), le differenze tra l'immagine destra e quella sinistra sono piccole.

Lo standard MPEG2 è stato esteso con una specifica supplementare chiamata Multi View Profile (MVP); allo stesso modo, il successivo standard H.264/AVC è stato esteso con l'inclusione della specifica Multi View Coding (MVC).

Carattere comune a queste due specifiche è l'uso della codifica video scalabile: il flusso video stereoscopico viene compresso in un layer di base (il flusso base 2D) più un enhancement layer, che trasporta la seconda vista. La sintassi del flusso codificato garantisce la decodificabilità del video 2D anche da parte di decoder di vecchia generazione, purché conformi agli standard MPEG2 o H.264/AVC.

Tuttavia, il bit-rate necessario alla codifica dei flussi stereoscopici in uno dei formati sopra descritti risulta, ad oggi, ancora troppo elevato per permetterne l'impiego in ambiente di broadcasting e, di conseguenza, i formati di frame packing rimangono l'unica soluzione praticabile nel breve termine per l'avvio dei servizi 3D.

Scopo della presente invenzione è pertanto quello di indicare un metodo di decodifica che permetta di estrarre

da un flusso video stereoscopico digitale un segnale video di tipo 2D-compatibile (2D) e, in particolare, un metodo di decodifica di un flusso video stereoscopico digitale di tipo 2D-compatibile basato sull'uso di fotogrammi compositi, che sia applicabile qualunque sia il modo in cui vengono impacchettate le immagini destra e sinistra all'interno di detti fotogrammi compositi.

Questo ed altri scopi della presente invenzione sono ottenuti mediante un metodo di decodifica di un flusso video stereoscopico incorporante le caratteristiche delle rivendicazioni allegate, le quali formano parte integrante della presente descrizione.

Ulteriori scopi e vantaggi della presente invenzione appariranno maggiormente chiari dalla descrizione che segue di alcuni esempi di realizzazione forniti a titolo esplicativo e non limitativo.

Tali esempi di realizzazione vengono descritti con riferimento ai disegni allegati, in cui:

- le Figure 1a, 1b ed 1c, già menzionate in precedenza, rappresentano un fotogramma composito di un flusso video stereoscopico, rispettivamente in formato side-by-side, in formato top-bottom ed in un formato alternativo;
- le Figure 2a, 2b e 2c rappresentano rispettivamente i fotogrammi compositi delle Figure 1a, 1b ed 1c, in cui una regione di fotogramma è opportunamente evidenziata per illustrare uno specifico passo di realizzazione del metodo secondo la presente invenzione;
- la Figura 3 rappresenta un sistema di produzione e fruizione di un flusso video di immagini comprendente un decodificatore ed un apparato di riproduzione video secondo l'invenzione.

Con riferimento alla Figura 3, viene illustrato un sistema di produzione e fruizione di un flusso video di immagini

comprendente uno stereoplexer 200 (dispositivo responsabile della produzione del fotogramma composito, in breve stereoplexing) che è configurato per ricevere coppie di immagini stereoscopiche sinistra e destra contraddistinte rispettivamente dalle lettere L e R. Nel flusso video destinato alla distribuzione, le immagini stereoscopiche della stessa coppia sono combinate in un singolo fotogramma composito FC secondo uno specifico formato di impacchettamento. Nell'esempio di realizzazione le immagini sono combinate utilizzando il formato di impacchettamento descritto precedentemente con riferimento alla Figura 1a.

Il segnale in uscita dallo stereoplexer 200 può essere compresso da un encoder 202, eventualmente multiplexato con altri segnali compressi e utilizzato per la diffusione su reti broadcast, oppure può essere distribuito in forma di file via rete telematica o su supporto di memorizzazione. Si noti che lo stereoplexer 200 può essere un dispositivo a sé stante oppure essere inglobato nell'encoder 202.

Le elaborazioni successive alla produzione del fotogramma composito FC vengono realizzate utilizzando apparati e metodi noti, che non sono oggetto dell'invenzione e non verranno, pertanto, ulteriormente descritti.

Si noti che durante la fase di stereoplexing, compressione o multiplexing, è prassi comune introdurre nel flusso dati risultante opportuni metadati di segnalazione, così da descrivere quale metodo di frame packing è stato utilizzato per la codifica del video stereoscopico. Tali metadati possono essere inseriti, ad esempio, nei messaggi SEI (Supplemental Enhancement Information) della codifica H.264 oppure nella Service Information che descrive la composizione del transport stream nel quale è stato multiplexato il contenuto video. Alternativamente, i metadati possono essere inseriti nello stesso fotogramma

composito FC, secondo il metodo descritto nella domanda di brevetto IT2008M0000267. In realtà, i suddetti metadati possono essere inseriti in qualsiasi forma e modalità (ad esempio, potrebbero assumere la forma di descrittori XML, o di apposite strutture dati binarie, o codici derivati da una tabella di look-up nota al dispositivo decodificatore), possono essere trasportati come segnalazione in banda o fuori banda e possono essere associati al flusso video in corrispondenza di qualsiasi punto della catena di produzione e distribuzione del contenuto. Tali modalità non costituiscono parte dell'invenzione e non verranno qui ulteriormente trattate.

Un decodificatore 212 (qui denominato anche con l'equivalente termine inglese decoder), ai fini della riproduzione del flusso video, partendo dalla descrizione del formato di frame packing trasportata nei metadati sopra descritti, riesce a determinare la regione occupata da una delle due immagini stereoscopiche (rappresentata nelle figure 2a, 2b e 2c) e ritagliare tale regione per inviarla all'apparato di visualizzazione 2D.

Questa determinazione può essere ottenuta in modo esplicito o implicito a partire dai metadati.

La modalità esplicita si ha quando la geometria del formato di frame packing viene descritta in modo esaustivo nei metadati (ad esempio indicando esplicitamente le coordinate dei vertici delle aree occupate da ciascuna delle immagini stereoscopiche).

La modalità implicita, invece, si ha quando nei metadati è indicato solamente il tipo di frame packing (ad esempio side-by-side), in cui tale indicazione è sufficiente al decodificatore per determinare tali aree.

In una prima forma di attuazione, il segnale stereoscopico codificato, compresso e diffuso in broadcast, oppure letto

da un supporto di memorizzazione, giunge ad un decoder digitale 212 collegabile ad un apparato di visualizzazione 222 (ad esempio, un televisore di tipo convenzionale).

Il decoder 212 è dotato degli opportuni componenti software che gli permettono di acquisire e analizzare la segnalazione che descrive il segnale stereoscopico. Ad esempio, questi componenti software potrebbero essere installati durante la fabbricazione, o successivamente per aggiornamento, tramite uno qualsiasi di diversi meccanismi disponibili: ad esempio, ricevendo una nuova versione del firmware over-the-air, scaricandolo via rete telematica, oppure leggendolo da una periferica connessa su interfaccia USB. Tali componenti software comprendono anche l'implementazione degli algoritmi necessari a ricavare una delle due immagini componenti il video stereoscopico in funzione del formato utilizzato per il frame packing.

Ad esempio, nel caso della modalità implicita ed indicazione del formato side-by-side (figura 1a), il metodo secondo l'invenzione prevede la determinazione dell'area occupata da una delle immagini stereoscopiche per divisione del fotogramma in due parti aventi area uguale lungo l'asse verticale del fotogramma (figura 2a). Invece, nel caso del top-bottom (figura 1b), la divisione viene fatta lungo l'asse orizzontale (figura 2b).

Una delle parti in cui è stato suddiviso il fotogramma viene selezionata come immagine bidimensionale e presentata a video dopo le opportune elaborazioni (ridimensionamento). Nel caso in cui la segnalazione contenga una descrizione esplicita della geometria del formato di frame packing (ad esempio il formato di figura 1c), il decoder 212 seleziona le informazioni necessarie a ritagliare dai fotogrammi del flusso video stereoscopico una delle due immagini

componenti. Ad esempio, nella segnalazione potrebbero essere indicate le coordinate dei vertici dei poligoni che racchiudono una delle immagini costituenti la coppia stereoscopica (ad esempio, i punti A, B, C, D di figura 2c): in questo caso, il decoder 212 dovrebbe solamente selezionare i dati relativi ad una delle due immagini ed applicare l'algoritmo di ritaglio corrispondente al formato di frame packing.

Il successivo ridimensionamento della sezione selezionata può essere delegato allo scaler del decoder 212 o, in modo equivalente, a quello del dispositivo di visualizzazione 222. Questo componente applica opportuni algoritmi di interpolazione per riportare l'immagine al suo formato selezionato per la visualizzazione, ricostruendo i pixel mancanti o modificando il fattore di forma utilizzato per la rappresentazione dei pixel.

In una forma di attuazione alternativa, il decoder 212 è integrato nel dispositivo di visualizzazione 222.

Si noti che le parti di fotogramma composito che non sono contenute all'interno della finestra di visualizzazione 2D (parti non scurite delle fig. 2a,b,c) non vengono utilizzate dall'apparato di visualizzazione e quindi, in linea di principio, potrebbero anche non essere decodificate. Pertanto, se il decodificatore legge le informazioni sul formato di impacchettamento, potrebbe anche evitare di decodificare le parti di fotogramma composito inutili alla visualizzazione 2D. Ad esempio, un decoder in grado di elaborare soltanto fotogrammi 1080i e non 1080p 50 o 60 Hz potrebbe riuscire a decodificare l'immagine 720p rappresentativa di una delle due viste inserita nel fotogramma 1080p di fig. 1c.

In un'ulteriore forma di realizzazione dell'invenzione, il metodo descritto in precedenza può essere applicato da un

decoder 212 o visualizzatore 3D nel caso in cui l'utente decida di passare temporaneamente dalla visualizzazione 3D a quella 2D (ad esempio, nel caso di eccessivo affaticamento del sistema visivo dovuto alla visualizzazione stereoscopica). Lo spettatore, in tale situazione, può commutare tra la visualizzazione 3D e quella 2D utilizzando un apposito comando (ad esempio, la pressione di uno specifico tasto sul telecomando).

In conclusione, il metodo e il decodificatore secondo l'invenzione forniscono 2D-compatibilità per flussi video tridimensionali verso apparati di riproduzione convenzionali 2D ed apparati 3D.

Il decoder 212 può essere un qualsiasi dispositivo in grado di ricevere e decodificare secondo l'invenzione un flusso video stereoscopico, per esempio un ricevitore di televisione digitale, un lettore di supporti ottici o magnetici digitali (DVD, Blu-ray player, personal video recorder). Il decoder 212 può anche essere parte integrante di un apparato comprendente il dispositivo di visualizzazione.

La presente invenzione non è limitata ad un metodo per la decodifica 2D-compatibile di un flusso video stereoscopico e relativi dispositivi, ma è passibile di varie modificazioni, perfezionamenti, sostituzioni di parti ed elementi equivalenti senza però allontanarsi dall'idea dell'invenzione, così come è precisato meglio nelle seguenti rivendicazioni.

RIVENDICAZIONI

1. Metodo per elaborare un flusso video di immagini digitali, il metodo comprendente i passi di:

- ricevere il flusso video che comprende almeno un fotogramma composito (FC) contenente una coppia di immagini digitali stereoscopiche (L,R) secondo un predeterminato formato di impacchettamento;

- ricevere metadati che descrivono il formato del fotogramma composito (FC);

- generare un flusso video di uscita riproducibile su un apparato di visualizzazione (222),

caratterizzato dal fatto che il metodo comprende i passi di

- determinare l'area nel fotogramma composito (FC) occupata da una sola immagine della coppia stereoscopica all'interno del fotogramma composito in base a detti metadati;

- estrarre dal fotogramma composito (FC) l'immagine contenuta in tale area, e

- generare un fotogramma di uscita contenente detta immagine estratta.

2. Metodo secondo la rivendicazione 1, in cui i metadati comprendono un indicatore di detto formato di impacchettamento.

3. Metodo secondo la rivendicazione 1, in cui i metadati sono posizionati nei messaggi SEI dello standard H.264.

4. Metodo secondo la rivendicazione 1, in cui i metadati sono posizionati nella Service Information del transport stream.

5. Metodo secondo la rivendicazione 1, in cui i metadati sono posizionati all'interno del fotogramma composito (FC).

6. Metodo secondo la rivendicazione 2, in cui i metadati comprendono una descrizione esplicita del formato di impacchettamento.

7. Metodo secondo una qualsiasi delle rivendicazioni da 1 a 6, in cui viene decodificata soltanto la parte di fotogramma composito (FC) contenente l'immagine da estrarre.

8. Dispositivo di decodifica di flussi video di immagini digitali, caratterizzato dal fatto di implementare un metodo secondo una qualsiasi delle rivendicazioni da 1 a 7.

9. Dispositivo di decodifica secondo la rivendicazione 8, in cui il metodo di decodifica viene attivato su comando dell'utente.

10. Dispositivo di decodifica secondo la rivendicazione 9, in cui il metodo di decodifica viene attivato dall'utente per mezzo della pressione di uno specifico tasto del telecomando.

11. Apparato di riproduzione video comprendente un dispositivo di decodifica secondo una qualsiasi delle rivendicazioni da 8 a 10.

CLAIMS

1. Method for processing a video stream of digital images, the method comprising the steps of:

- receiving the video stream which comprises at least one composite frame (FC) containing a pair of stereoscopic digital images (L,R) according to a predetermined packing format;

- receiving metadata which describes the format of the composite frame (FC);

- generating an output video stream which can be reproduced on a visualization apparatus (222),

characterized in that the method comprises the steps of:

- determining the area in the composite frame (FC) which is occupied by one image of the stereoscopic pair within the composite frame based on said metadata;

- extracting the image contained in said area from the composite frame (FC), and

- generating an output frame containing said extracted image.

2. Method according to claim 1, wherein the metadata comprises an indicator of said packing format.

3. Method according to claim 1, wherein the metadata is positioned in the SEI messages of the H.264 standard.

4. Method according to claim 1, wherein the metadata is positioned in the Service Information of the transport stream.

5. Method according to claim 1, wherein the metadata is positioned within the composite frame (FC).

6. Method according to claim 2, wherein the metadata comprises an explicit description of the packing format.

7. Method according to any of claims 1 to 6, wherein only that part of the composite frame (FC) which contains the image to be extracted is decoded.

8. Device for decoding video streams of digital images, characterized by implementing a method according to any of claims 1 to 7.

9. Decoding device according to claim 8, wherein the decoding method is activated upon the user's command.

10. Decoding device according to claim 9, wherein the decoding method is activated by the user by pressing a specific remote control button.

11. Video playback device comprising a decoding device according to any of claims 8 to 10.

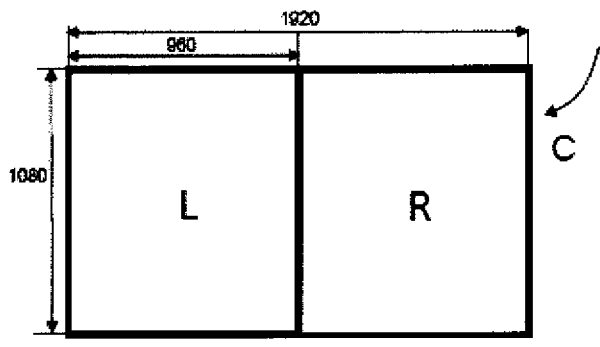


Fig. 1a

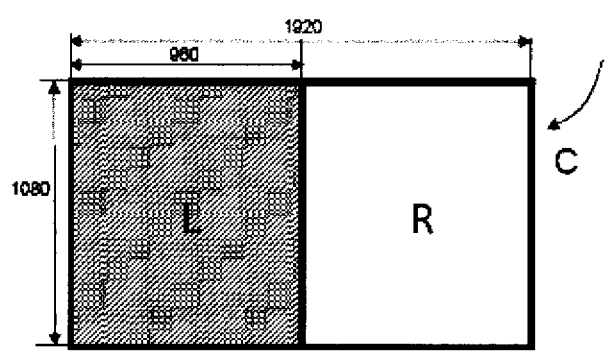


Fig. 2a

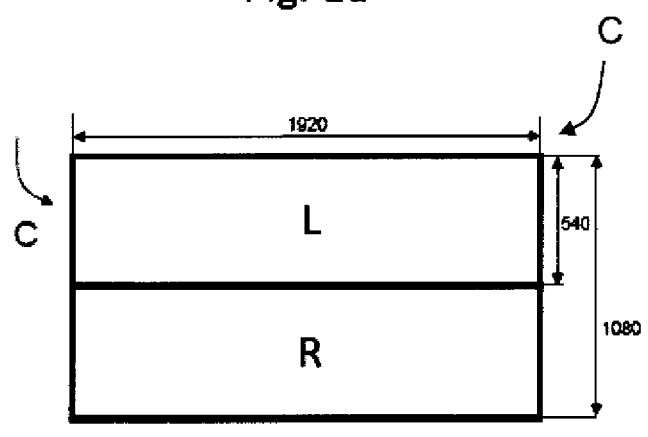


Fig. 1b

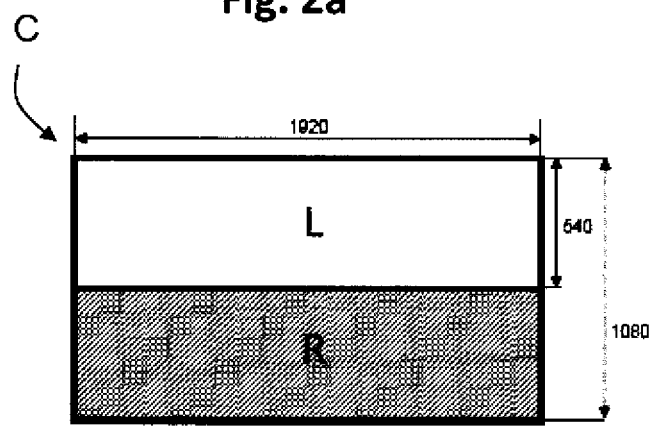


Fig. 2b

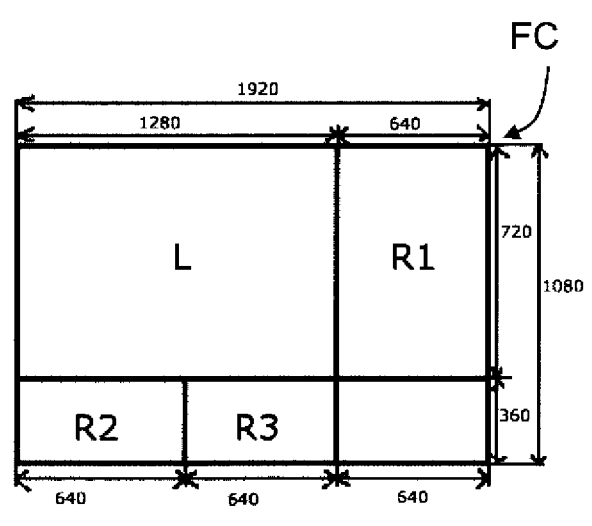


Fig. 1c

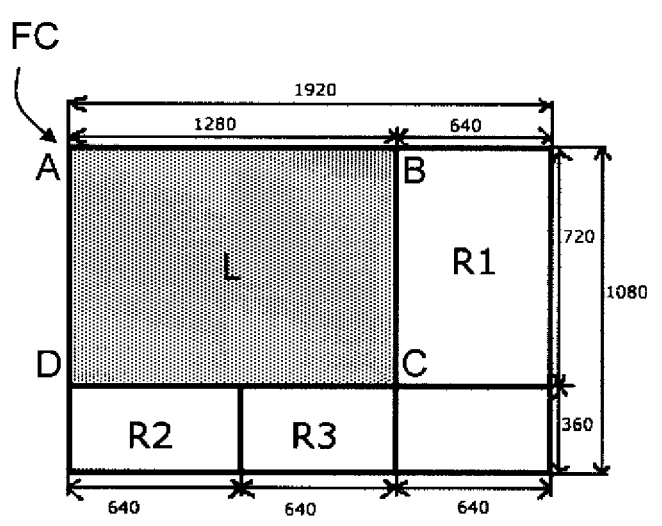


Fig. 2c

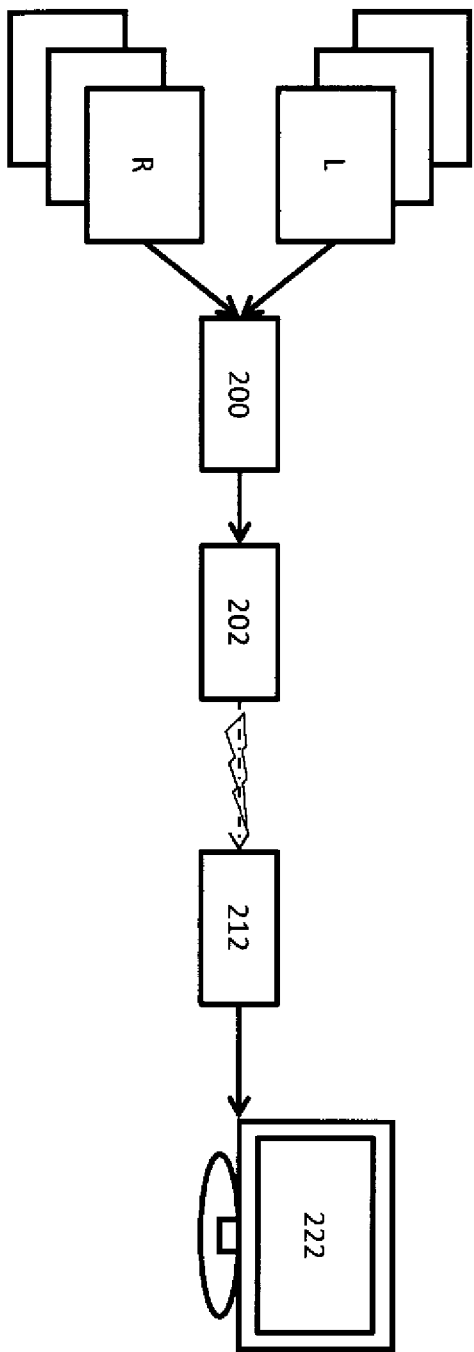


Fig. 3