



MINISTERO DELLO SVILUPPO ECONOMICO
DIREZIONE GENERALE PER LA LOTTA ALLA CONTRAFFAZIONE
UFFICIO ITALIANO BREVETTI E MARCHI

DOMANDA NUMERO	102006901436592
Data Deposito	27/07/2006
Data Pubblicazione	27/01/2008

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
G	08	B		

Titolo

METODO DI RILEVAZIONE EVENTI E SISTEMA DI VIDEO SORVEGLIANZA UTILIZZANTE IL METODO.

cui in diversi istanti di tempo si acquisiscono immagini di apprendimento di un'area sorvegliata in una situazione di assenza di eventi da rilevare, ed una fase operativa di rilevazione in cui vengono acquisite immagini correnti dell'area. Il metodo rileva un evento mediante il confronto di una immagine corrente con una immagine corrispondente ad una combinazione lineare di una pluralità di immagini di riferimento approssimanti, o coincidenti con, rispettive immagini di apprendimento.

DESCRIZIONE

Campo di applicazione dell'invenzione

La presente invenzione si riferisce ad un metodo di rilevazione eventi secondo il preambolo della rivendicazione 1 e ad un sistema di video sorveglianza che utilizza il metodo.

Descrizione dell'arte nota

Con sistema di video sorveglianza si intende nel presente brevetto far riferimento ad un sistema di sorveglianza che utilizza almeno un'unità di acquisizione immagini in grado di acquisire sequenze di immagini di un'area sorvegliata.

Nei sistemi di video sorveglianza noti, vengono solitamente previsti dei sistemi di rilevazione eventi in grado di generare allarmi una volta che si sia verificato un evento anomalo.

Alcuni dei sistemi noti si limitano a verificare una variazione (oltre una certa soglia impostata con la collaborazione dell'utente) della luminosità dei pixel in due immagini

temporalmente successive, quali due quadri di un segnale video o due immagini riprese in tempi diversi da una macchina fotografica.

Questi sistemi presentano il problema che improvvise variazioni di luce dovute ad esempio a passaggi di nuvole o a riflessi sull'acqua, così come movimenti naturali all'interno dell'area sorvegliata (ad esempio il movimento delle fronde di un albero o il passaggio autorizzato di autovetture su una strada) provocano numerosi falsi allarmi.

Per superare questi inconvenienti sono noti sistemi di video sorveglianza che prevedono una fase di apprendimento durante la quale il sistema costruisce un modello dell'area sorvegliata in una situazione di normalità, ossia una situazione in cui non devono essere generati allarmi. Durante la fase operativa, i pixel dell'immagine ripresa vengono confrontati con i pixel del modello. Se la differenza tra i pixel supera una certa soglia impostata dall'operatore, allora viene generato un allarme.

Nonostante la creazione di un modello relativo ad una situazione di normalità dell'area video sorvegliata, il confronto pixel per pixel dell'immagine acquisita con il modello risulta spesso poco efficiente perché lo scostamento di un singolo pixel dal suo modello è tale da generare un allarme. Questo porta alla generazione di moltissimi falsi allarmi.

Proprio per evitare tali falsi allarmi, alcune soluzioni note (brevetto americano US 5,892,856) di fatto sopprimono dalla

rilevazione quei pixel che durante la fase di apprendimento presentano variazioni di intensità (dovuti ad esempio a movimenti naturali della scena osservata).

Una tale soluzione, poco efficiente, risulta utilizzabile solo in pochi contesti come quello (oggetto del brevetto US 5,892,856) della rilevazione presenze ad una postazione di lavoro.

Altre soluzioni più sofisticate, ad esempio note dalla domanda di brevetto US 2004/0246336, prevedono in primis la creazione di un modello statistico di ogni pixel (con relativa media e varianza) e successivamente, in fase di rilevazione eventi il sistema prevede di estrapolare l'immagine dell'oggetto/persona rilevata per confrontarla con un set di modelli di oggetti autorizzati.

Questa soluzione presenta tuttavia lo svantaggio di richiedere un'elevata disponibilità di memoria per lo stoccaggio del modello della scena e degli oggetti autorizzati, ed una elevata capacità computazionale per analizzare l'intera immagine in tempo reale confrontando gli oggetti rilevati con quelli autorizzati.

Problema comune delle soluzioni note è poi legato al fatto che l'allarme viene segnalato quando lo scostamento di luminosità o di colore di un pixel in due quadri consecutivi supera una certa soglia impostata dall'operatore. Questo ha la conseguenza di legare l'efficienza del sistema alle capacità

dell'operatore, il che costituisce un problema là dove l'operatore non sia esperto.

Scopo principale della presente invenzione è quello di superare gli inconvenienti dell'arte nota ed in particolare di presentare un sistema di video sorveglianza ed un metodo di rilevazione eventi che permettano una efficiente rilevazione di eventi contenendo il numero di falsi allarmi e preferibilmente senza richiedere elevate disponibilità di memoria e capacità computazionale.

Ulteriore scopo della presente invenzione è poi quello di presentare un sistema con un elevato grado di automazione in grado di calcolare automaticamente la soglia di errore oltre la quale discriminare tra una variazione normale di un pixel ed una situazione di allarme.

Formano poi scopo della presente invenzione un sistema di video sorveglianza ed un metodo di rilevazione eventi in grado di ottimizzare l'occupazione di memoria e di variare la complessità computazionale in funzione delle dinamiche presenti nell'area sorvegliata durante la fase di apprendimento.

Questi ed ulteriori scopi della presente descrizione sono raggiunti mediante un sistema ed un metodo di video sorveglianza secondo le rivendicazioni allegate, le quali formano parte integrante della presente descrizione.

Descrizione sintetica dell'invenzione

L'idea alla base della presente invenzione è quella di

tralasciare l'approccio tradizionale del confronto di ogni singolo pixel di una immagine corrente con un rispettivo pixel di una immagine di riferimento o con un modello di pixel.

L'invenzione si propone particolarmente di considerare delle regioni dell'immagine, ossia gruppi di pixel, così da tenere in considerazione anche la correlazione tra i pixel nella rilevazione di un evento e ridurre così il numero di falsi allarmi.

Il sistema di video sorveglianza secondo l'invenzione acquisisce delle immagini di un'area sorvegliata e ne confronta le singole regioni con relativi "modelli", che rappresentano una situazione di normalità, costituiti come spazio di immagini acquisite in una fase di apprendimento e relative alla scena osservata in una situazione di normalità.

L'immagine o la regione, viene trattata come un vettore immagine, il cui scostamento dalla situazione di normalità è misurato come errore di proiezione del vettore immagine su uno spazio delle immagini che costituisce il "modello" dell'area sorvegliata in una situazione di normalità.

Il "modello" viene costruito partendo da un insieme di immagini acquisite in fase di apprendimento riprendendo l'area in una situazione di normalità.

La fase di apprendimento può prevedere una fase di validazione del modello che consiste sostanzialmente nella simulazione di una fase operativa di rilevazione. La validazione utilizza

immagini acquisite in fase di apprendimento e relative alla scena in una situazione di normalità, e verifica se il modello costruito è buono.

Per la rilevazione degli eventi, il metodo secondo l'invenzione sfrutta particolarmente le proprietà dell'analisi delle componenti principali (PCA).

Vantaggiosamente, poi, il metodo prevede una opportuna riduzione del contenuto informativo delle immagini acquisite, così da trascurare fenomeni di minore importanza all'interno di una scena e ridurre il numero di falsi allarmi.

Descrizione sintetica dei disegni

Ulteriori scopi e vantaggi della presente invenzione risulteranno maggiormente chiari dalla descrizione che segue e dai disegni annessi, in cui:

- la fig. 1 mostra un sistema di video sorveglianza secondo una forma di realizzazione dell'invenzione;
- la fig. 2 mostra uno schema a blocchi del processo di elaborazione delle immagini acquisite da un sistema di video sorveglianza secondo l'invenzione;
- la fig. 3 mostra la scomposizione di una immagine acquisita in una pluralità di regioni.
- la fig. 4 mostra un esempio di rilevazione di un evento.

Descrizione dettagliata dell'invenzione

La figura 1 mostra un sistema di video sorveglianza 1. Attraverso un monitor 3 un operatore 2 controlla le immagini 6

acquisite da un'unità di acquisizione immagini 5.

In figura 1 l'unità di acquisizione immagini è una telecamera in grado di fornire in uscita un video, ossia una sequenza continua di immagini, ma va inteso che ai fini dell'invenzione potrebbe essere sostituita da qualsiasi altro mezzo equivalente, ad es. una camera digitale programmata che acquisisce immagini ad istanti di tempo regolari.

Un'immagine potrà quindi corrispondere ad un quadro o un semiquadro di un segnale video acquisito dalla telecamera, ad una immagine statica acquisita da una camera digitale, all'uscita di un sensore CCD, o più in generale ad una porzione di quanto sopra.

Come è ben noto, un'immagine (digitale o analogica) è scomponibile in pixel, ossia elementi fondamentali dell'immagine.

Ad ogni immagine possono quindi essere associate una o più matrici i cui elementi sono i valori di tensione del segnale video analogico o i valori di luminosità o di colore dei pixel. Nel caso di una telecamera a colori, all'immagine acquisita corrisponderà una matrice tridimensionale in cui ad ogni elemento della matrice (ossia ogni pixel) corrisponde una terna di valori corrispondenti ai valori dei segnali RGB.

Nel caso di immagine in scala di grigi, ad ogni elemento della matrice viene associato un valore corrispondente al valore di grigio del pixel corrispondente.

Nel corso della presente descrizione questa associazione immagine-matrice verrà sottintesa, così che nel proseguo si parlerà ad esempio di righe e colonne di una immagine.

Tornando alla figura 1, la telecamera 5 riprende un area, nel caso specifico un corridoio, e trasmette un segnale video che può essere visualizzato sul monitor 3.

Secondo l'invenzione, il sistema di sorveglianza 1 è provvisto di una unità di elaborazione immagini 4 in grado di rilevare eventi a partire dalle immagini acquisite dalla telecamera 5.

In figura 1 l'unità di elaborazione immagini 4 è rappresentata come un elaboratore elettronico operativamente connesso alla telecamera 5 e al monitor 3 per poter ricevere il segnale video dalla telecamera, elaborarlo e visualizzare delle immagini sul monitor.

In una forma di realizzazione preferita, l'unità di elaborazione immagini 4 è un videosever, ossia un elaboratore numerico che riceve un segnale video dall'unità di acquisizione immagini, lo elabora secondo il metodo della presente invenzione e trasferisce un segnale video ad uno o più terminali ad esso collegati, in particolare detto terminale essendo una postazione di lavoro dell'operatore.

Chiaramente sono possibili molte altre soluzioni, ad esempio l'unità di elaborazione immagini 4 può essere contenuta direttamente all'interno della telecamera 5 (che avrà quindi una unità di acquisizione immagini ed una unità di elaborazione

immagini) la quale viene collegata direttamente, o tramite uno switch video, al monitor 3.

L'unità di elaborazione immagini 4 è dotata di un software contenente porzioni di codice in grado di attuare il metodo di rilevazione eventi qui di seguito esposto.

Almeno una volta all'installazione, il metodo prevede una fase di apprendimento in cui il sistema costruisce un "modello" della scena osservata in una situazione di normalità.

Vantaggiosamente, tuttavia, il metodo prevede che la fase di apprendimento sia ripetuta più volte in situazioni ambientali (di luce, di traffico ecc...) differenti. Con ciò potendo costruire uno o più modelli.

Ad esempio, diversi modelli possono essere costruiti in differenti orari della giornata, il metodo prevedendo allora di confrontare un'immagine acquisita ad un certo orario con un modello valido per detto orario.

In una situazione di normalità, ossia in assenza di eventi da rilevare, l'operatore 3 avvia la fase di apprendimento del software in cui vengono acquisite immagini, che chiameremo "di apprendimento", dell'area sorvegliata.

Nella forma di realizzazione preferita, qui di seguito descritta, le immagini acquisite corrispondono ai quadri del segnale video generato dalla telecamera 5.

In questa fase è possibile che vengano ripresi oggetti in movimento quali foglie o autoveicoli che si muovono su una

strada posta sullo sfondo della scena ripresa, le immagini acquisite potranno quindi differire l'una dall'altra.

Conseguentemente il modello potrà rappresentare la scena osservata in una situazione dinamica, ma senza che vi siano eventi da rilevare.

Come prima operazione della fase di apprendimento, figura 2a, il metodo di rilevazione eventi prevede la selezione (202) di un insieme di N quadri $F_1, \dots, F_N$ a partire dal segnale video 201 acquisito dalla telecamera. Su questi quadri vengono preferibilmente eseguite alcune operazioni di elaborazione delle immagini quali, una conversione in scala di grigi (203), così da ridurre la dimensione dei dati da trattare, ed eventualmente anche, o in alternativa, un filtraggio passa-basso (204) con un kernel gaussiano per eliminare e attenuare tutte le variazioni ad alta frequenza che non interessa rilevare e ridurre così il contenuto informativo delle immagini da trattare, concentrando la rilevazione solo sul contenuto informativo interessante dell'immagine.

I quadri così modificati vengono inseriti in un buffer di apprendimento (205).

Alternativamente, le operazioni sopra indicate di elaborazione delle immagini possono essere ripetute ciclicamente su ogni quadro acquisito come mostrato in fig.2b. Qui all'avvio il parametro n viene posto uguale ad 1, si acquisisce un'immagine (202b), la si converte in scala di grigi (203b), la si filtra

con un filtro passa-basso (204c) e la si memorizza nel buffer (205c). Successivamente si incrementa il valore di n e si ripetono operazioni fino allo stoccaggio nel buffer di N immagini.

Una volta creato, il contenuto del buffer di apprendimento viene suddiviso in due parti: un primo gruppo di quadri, detti quadri "di addestramento", sui quali viene eseguita un'analisi delle componenti principali (PCA) e un secondo gruppo di quadri, detti quadri "di validazione", utilizzati per convalidare i risultati ottenuti dalla PCA.

La fase di apprendimento prevede quindi, corrispondentemente, una fase di addestramento ed una di validazione.

Secondo una forma di realizzazione preferita, nella fase di addestramento ciascuno dei quadri F_n di addestramento (che nell'esempio si prendono in numero pari a S), e preferibilmente ciascuno dei quadri memorizzati nel buffer di apprendimento, viene suddiviso, mediante una griglia predefinita, in regioni (ossia piccole immagini preferibilmente quadrate o rettangolari) $R_{i,j}$ di massimo $M \times M = m$ pixel.

L'effetto, mostrato in figura 3, è quello di ricavare una pluralità di porzioni di immagini da ogni quadro.

La dimensione della griglia dipende dalle dimensioni tipiche del target che si vuole individuare nella scena osservata.

Tale griglia potrà allora essere impostata in fase di installazione da un installatore in funzione dell'inquadratura

della scena e delle esigenze dell'operatore, o essere predefinita in fase di fabbricazione.

Per ciascuna regione $R_{i,j}$ di un quadro F_n si ricava un corrispondente vettore colonna $IR_{i,j}(F_n)$.

Questo vettore $IR_{i,j}(F_n)$ è ottenuto sostanzialmente inserendo progressivamente gli elementi della matrice $R_{i,j}$ (ossia i valori dei pixel della regione) che si incontrano scorrendo le colonne dall'alto verso il basso e da sinistra a destra. Così l'elemento $IR_{i,j}(F_n)(2)$ corrisponde al pixel posto sulla seconda riga della prima colonna dell'immagine $R_{i,j}$.

Affiancando i vettori colonna di una medesima regione $R_{i,j}$ si crea una corrispondente matrice della normalità $Y_{i,j} = (IR_{i,j}(F_1), IR_{i,j}(F_2), \dots, IR_{i,j}(F_s))$.

Le colonne della matrice della normalità generano uno spazio vettoriale delle immagini.

Le colonne portano l'informazione relativa ad una regione della scena osservata in diversi istanti di tempo ed in una situazione di normalità, mentre gli autovettori della rispettiva matrice di covarianza ne costituiscono le componenti principali, ossia le direzioni lungo le quali è maggiore la varianza delle colonne di $Y_{i,j}$, ossia delle immagini raccolte.

Una volta ottenuta la matrice $Y_{i,j}$, viene eseguita una scomposizione in valori singolari (SVD) al fine di ottenere tre matrici $U_{i,j}$, $V_{i,j}$, $\Sigma_{i,j}$ tali che

$$Y_{i,j} = U_{i,j} \cdot \Sigma_{i,j} \cdot V_{i,j}^T$$

con

$$U_{i,j} = [u_1 \dots u_s]$$

$$\Sigma_{i,j} = \text{diag}(\sigma_1, \dots, \sigma_s)$$

$$V_{i,j} = [v_1 \dots v_s]$$

Dove u_i (vettori $\in \mathbb{R}^m$) sono gli autovettori della matrice di covarianza di $Y_{i,j}$, e $\sigma_1, \dots, \sigma_s$ sono i valori singolari di $Y_{i,j}$.

Come è noto, utilizzando una scomposizione SVD gli elementi della matrice diagonale $\Sigma_{i,j}$ sono legati dalla relazione

$$\sigma_1 \geq \dots \geq \sigma_r \geq \sigma_{r+1} \geq \dots \geq \sigma_s \geq 0$$

Secondo l'invenzione, per ottimizzare la rilevazione degli eventi e concentrarsi sul contenuto informativo rilevante dell'immagine, la matrice $Y_{i,j}$ viene approssimata dalla matrice

$$Y_{i,j}^r = U_{i,j}^r \cdot \Sigma_{i,j}^r \cdot (V_{i,j}^r)^T$$

in cui

$$U_{i,j}^r = [u_1 \dots u_r]$$

$$\Sigma_{i,j}^r = \text{diag}(\sigma_1 \dots \sigma_r)$$

$$V_{i,j}^r = [v_1 \dots v_r]$$

ed in cui il numero di valori singolari r considerati per costruire $Y_{i,j}^r$ è ottenuto trascurando tutti i valori singolari inferiori ad una soglia.

La matrice $Y_{i,j}^r$ ha le medesime dimensioni della matrice $Y_{i,j}$, ma porta con sé solo l'informazione legata alle prime r componenti principali della matrice $Y_{i,j}$.

Le colonne $u_1 \dots u_r$ della matrice $U_{i,j}^r$ sono le componenti principali della matrice $Y_{i,j}$.

Per il fissaggio della soglia sono stati condotti diversi esperimenti i quali hanno mostrato che si riesce ad avere una buona rilevazione di eventi nel caso in cui si approssimi il contenuto informativo di $Y_{i,j}$ rinunciando al 20%-30%, preferibilmente il 25%, dell'energia di $Y_{i,j}$ (ossia delle porzioni di immagine $R_{i,j}$ utilizzate per costruire questa matrice).

Per una nota proprietà legata ai valori singolari, la percentuale dell'energia $\%E(r)$ legata alle prime r componenti principali della matrice $Y_{i,j}$ (con S valori singolari), è :

$$\%E(r) = \frac{\sum_{k=1}^r \sigma_k}{\sum_{k=1}^S \sigma_k}$$

A partire da queste considerazioni, per ogni regione $R_{i,j}$ si determina un rispettivo valore di r e si costruisce la matrice $Y_{i,j}^r$ che rappresenta l'essenza di quanto appreso durante la fase di apprendimento.

A questo punto segue la fase di validazione, il cui compito è di verificare che l'insieme di apprendimento costituito dai quadri di addestramento sia sufficientemente rappresentativo della scena ripresa in una situazione di normalità.

La verifica passa attraverso la simulazione di una fase operativa di rilevazione in cui al posto di immagini correnti

dell'area sorvegliata si utilizza almeno un quadro F^{VAL} di validazione, ossia una immagine di apprendimento non appartenente all'insieme di apprendimento, quindi non utilizzata per costruire la matrice della normalità $Y_{i,j}$.

Operativamente, per la validazione si prende almeno un quadro di validazione F^{VAL} e lo si suddivide in una pluralità di regioni $R_{i,j}(F^{VAL})$ attraverso la medesima griglia utilizzata per suddividere i quadri di addestramento.

Per ogni regione $R_{i,j}(F^{VAL})$, si crea il corrispondente vettore $IR_{i,j}(F^{VAL})$ come fatto per i quadri di addestramento.

Successivamente, il vettore $IR_{i,j}(F^{VAL})$ viene proiettato su uno spazio delle immagini di apprendimento per determinare la "distanza" tra l'immagine di validazione e la situazione di normalità sintetizzata nella matrice $Y_{i,j}^r$.

Come è noto, data la matrice $Y_{i,j}^r = (y_1^r, y_2^r \dots y_s^r)$ con elementi in $\mathfrak{R}^m$ si può definire lo spazio $\text{Range}(Y_{i,j}^r)$, sottospazio di $\mathfrak{R}^m$, costituito da tutte le combinazioni lineari delle colonne $y_1^r, y_2^r \dots y_s^r$. Per un noto teorema di algebra lineare, il sottospazio $\text{Range}(Y_{i,j}^r)$ coincide con il sottospazio $\text{Range}(U_{i,j}^r)$, costituito da tutte le combinazioni lineari delle colonne di $U_{i,j}^r$, ossia delle prime r componenti principali di $Y_{i,j}$.

La proiezione di una porzione di immagine $R_{i,j}$ del quadro F^{VAL} , ossia del vettore $IR_{i,j} \in \mathfrak{R}^m$ sulle componenti principali di $Y_{i,j}$ si ottiene quindi usando l'operatore proiettore definito come

$$P_{\text{Range}(Y_{i,j}^r)}$$

$$P_{\text{Range}(Y_{i,j}^r)} = U_{i,j}^r \cdot (U_{i,j}^r)^T$$

Calcolato questo operatore, per ciascuna porzione di immagine $R_{i,j}$ la proiezione $\text{Proj}(\text{IR}_{i,j})$ e l'errore di proiezione $\text{err_Proj}(\text{IR}_{i,j})$ vengono calcolati come

$$\text{Proj}(\text{IR}_{i,j}) = U_{i,j}^r \cdot (U_{i,j}^r)^T \cdot \text{IR}_{i,j}$$

$$\text{err_Proj}(\text{IR}_{i,j}) = \|\text{Proj}(\text{IR}_{i,j}) - \text{IR}_{i,j}\|_2$$

La scena osservata viene segnalata come anomala (ossia viene rilevato un evento) se l'errore di proiezione è superiore alla rispettiva soglia, ossia se vale la relazione

$$\text{err_Proj}(\text{IR}_{i,j}) \geq \text{Thr}_{i,j}$$

Secondo una forma di realizzazione preferita, la soglia $\text{Thr}_{i,j}$ viene determinata automaticamente ed è posta pari all' $r+1$ -esimo valore singolare σ_{r+1} della matrice $Y_{i,j}$, ossia al valore singolare di indice maggiore che soddisfa le relazioni

$$\sum_{i=1}^r \sigma_i \leq \%E \quad \text{e} \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq \sigma_{r+1} \geq 0$$

Vantaggiosamente, la soglia può essere posta pari a $k\sigma_{r+1}$ con $k>1$ in modo da tenere conto di eventuale rumore di fondo presente nelle immagini.

Poiché i quadri di validazione F^{VAL} dovrebbero rappresentare la scena in una situazione di normalità, al limite nessun evento dovrebbe essere rilevato. In caso contrario l'insieme dei quadri di addestramento acquisiti non risulta rappresentativo della scena in una situazione di normalità ed il metodo prevede di selezionare un diverso insieme di apprendimento e ricostruire la matrice $Y_{i,j}^r$.

Per evitare di scartare un buon insieme di quadri di addestramento perché un quadro di validazione (e non quelli di addestramento) non è effettivamente rappresentativo di una situazione di normalità, in una forma preferita e vantaggiosa di implementazione il metodo prevede che si simuli la rilevazione di eventi su una pluralità di quadri di validazione.

Secondo questa forma di implementazione del metodo, l'insieme dei quadri di addestramento viene cambiato se si rilevano un numero di eventi superiore ad una percentuale prefissata, ad esempio il 25%, calcolata sul numero totale di misurazioni. Ad esempio, se la fase di validazione verifica 100 quadri di validazione e rileva 25 eventi, allora il metodo prevede di ricostruire l'insieme di apprendimento.

Secondo un'ulteriore forma di implementazione vantaggiosa, la matrice della normalità $Y_{i,j}$ viene rigenerata a partire da un nuovo insieme di apprendimento se l'errore di proiezione medio $\overline{\text{err_Proj}}$ di una pluralità di immagini di validazione supera o uguaglia la soglia $\text{Thr}_{i,j}$.

L'errore di proiezione medio $\overline{\text{err_Proj}}$ è calcolato come

$$\overline{\text{err_Proj}} = \text{media}[\text{err_Proj1}, \text{err_Proj2}, \text{err_Proj3}, \dots]$$

dove "media" è una funzione che riceve in ingresso gli errori di proiezione err_Proj1 , err_Proj2 , err_Proj3 ... delle immagini di validazione e ne restituisce il valor medio.

In un'altra forma di realizzazione, è invece possibile

calcolare l'errore di proiezione massimo $\overline{\text{err_Proj}}_{MAX}$ tra gli errori di proiezione di una pluralità di immagini di validazione. Se tale errore supera una soglia prefissata, allora viene ripetuta la fase di addestramento costruendo un nuovo insieme di apprendimento.

In una forma di implementazione preferita, il numero di quadri da aggiungere dipende dalla differenza tra l'errore di proiezione medio $\overline{\text{err_Proj}}$ e la soglia.

Alternativamente, si aggiunge un numero di quadri prefissato e dipendente dalla disponibilità di memoria del buffer.

Preferibilmente, i quadri che vengono aggiunti all'insieme di apprendimento non sono tutti consecutivi.

La dimensione massima dell'insieme di apprendimento può essere fissata (in fase di installazione o in fase di produzione) in funzione della disponibilità di memoria e della capacità di calcolo a disposizione.

Durante la fase di apprendimento il metodo prevede la memorizzazione di una pluralità di informazioni che servono per la rilevazione di un evento. In particolare vengono memorizzate:

- le matrici $U_{i,j}^r = [u_1 \dots u_r]$
 - le soglie $\text{Thr}_{i,j}$
 - il numero r di componenti principali di ogni regione $R_{i,j}$
- che consente di ottimizzare la matrice $Y_{i,j}$

Terminata la fase di apprendimento, il sistema di video

sorveglianza può iniziare la fase operativa vera e propria, acquisendo immagini correnti dell'area sorvegliata per rilevare eventi.

Attraverso la telecamera 5 viene acquisita un'immagine corrente dell'area sorvegliata, in particolare un quadro F^* del segnale video generato dalla telecamera.

Il quadro F^* viene suddiviso in una pluralità di regioni $R_{i,j}(F^*)$ attraverso la medesima griglia utilizzata in fase di apprendimento.

Per ogni regione $R_{i,j}(F^*)$, si crea il corrispondente vettore $IR_{i,j}(F^*)$ e lo si proietta sulle componenti principali di $Y_{i,j}$, ossia lo si proietta sullo spazio delle immagini di apprendimento, in modo da valutare l'errore di proiezione

$$err_Proj(IR_{i,j}) = \|Proj(IR_{i,j}) - IR_{i,j}\|_2.$$

come illustrato con riferimento alla fase di validazione.

All'interno di ogni regione $R_{i,j}$ viene rilevato un evento se l'errore di proiezione supera il valore della soglia selezionata dal sistema per la medesima regione $R_{i,j}$.

Per eseguire tali operazioni viene fatto uso delle informazioni memorizzate durante la fase di apprendimento.

In figura 4 viene mostrato un esempio di rilevazione di un evento. Nel quadro F^* analizzato viene ripreso un corridoio, lo stesso di fig. 1 e fig.3, in cui sono presenti due persone le

cui sagome occupano una porzione del quadro; nell'esempio di fig.4 dette sagome occupano sette regioni $R_{i,j}$.

Solo queste regioni $R_{i,j}$ (racchiuse in un contorno bianco per meglio risaltare) occupate dalle persone presentano un errore superiore alla soglia accettata e pertanto generano un allarme. Nonostante l'algoritmo sopra descritto preveda il trattamento delle immagini come vettori, è chiaro che non si deve dimenticare il significato fisico di tali operazioni, che possono quindi essere ottenute con modi ed operazioni differenti.

Rilevare un evento attraverso la valutazione dell'errore di proiezione del vettore immagine sullo spazio vettoriale delle immagini $\text{Range}(Y_{i,j}^r)$, corrisponde a confrontare una immagine corrente $R_{i,j}(F^*)$ con una immagine corrispondente ad una combinazione lineare di una pluralità di immagini di riferimento approssimanti, o coincidenti con, rispettive immagini di apprendimento.

Questo concetto risulta più chiaro con l'ausilio della figura 5.

Lo spazio $\text{Range}(Y_{i,j}^r)$ è rappresentato dal piano Γ sul quale giacciono le immagini di riferimento $y_1^r \dots y_s^r$ e tutte le combinazioni lineari $C_1, C_2, C_3 \dots$ di tali immagini di riferimento.

Proiettare una immagine corrente $R_{i,j}(F^*)$ sul piano Γ vuol dire confrontare tale immagine con la combinazione lineare C_i delle

immagini di riferimento più vicina all'immagine $R_{i,j}(F^*)$ nel senso della norma L_2 .

Il confronto non essendo più un confronto pixel per pixel, ma immagine per immagine, con questo intendendo che oltre ai valori dei pixel dell'immagine si considera, attraverso l'analisi delle componenti principali, anche la loro correlazione.

Chiaramente tale confronto, che secondo la forma di realizzazione preferita e vantaggiosa della presente invenzione viene sintetizzato dall'operazione di proiezione del vettore corrispondente all'immagine corrente sullo spazio delle immagini $\text{Range}(Y_{i,j}^r)$, potrebbe alternativamente essere ottenuto con un confronto prima dei valori dei singoli pixel e poi delle relazioni tra i pixel, tra una immagine corrente ed una immagine costruita in fase di apprendimento e corrispondente ad una combinazione lineare di immagini approssimanti o coincidenti con immagini acquisite in una situazione di normalità.

Dalla descrizione sono chiari i vantaggi della presente invenzione.

L'approccio proposto, basandosi sul confronto di porzioni dell'immagine corrente (ossia gruppi di pixel) con rispettivi modelli costruiti in fase di apprendimento, risulta più affidabile delle soluzioni note basate sul confronto di ogni pixel in quanto permette di tenere conto della correlazione

spaziale tra i pixel.

Inoltre, a differenza delle soluzioni note che utilizzano un modello statistico per rappresentare la situazione di normalità di un pixel, il modello secondo l'invenzione è costituito da un insieme di spazi vettoriali, ognuno relativo ad una regione dell'immagine, ricavati da una fase di apprendimento.

Ne viene quindi che il modello secondo l'invenzione consente di rappresentare la normalità in modo molto più accurato rispetto alle soluzioni note in quanto utilizza immagini invece che parametri statistici quali media e varianza.

Inoltre la PCA è un metodo di compressione dell'informazione che permette la ricostruzione dei dati originali (questo non è possibile con i metodi statistici usati nelle soluzioni esistenti) quindi non compromette la qualità con cui è descritta la normalità e la conseguente affidabilità in fase di rilevazione.

L'approccio proposto presenta poi un elevato grado di automazione dove gli approcci tradizionali richiedono invece scelte arbitrarie in fase di programmazione e set up del sistema.

In particolare la soglia per la rilevazione è calcolata automaticamente sulla base dei dati di apprendimento.

Inoltre l'approccio proposto sceglie in modo autonomo la lunghezza del buffer di apprendimento in funzione delle dinamiche presenti nella scena osservata durante la fase di

apprendimento. In particolare l'algoritmo sceglie una soluzione di compromesso tra spazio di memoria disponibile e dinamica della scena.

Ulteriormente, il numero minimo di componenti principali con cui rappresentare lo spazio vettoriale varia in funzione della dinamica della scena e viene anche questo calcolato dall'algoritmo.

E' chiaro poi che la descrizione qui sopra presenta un sistema di video sorveglianza ed un metodo di rilevazione eventi secondo forme di realizzazione preferite che costituiscono forme esemplificative e non limitative di realizzazione del sistema e di implementazione del metodo.

In particolare, molti modi possono essere utilizzati per selezionare il numero ottimale di componenti $r_{i,j}$ che permette di ridurre il contenuto informativo della matrice $Y_{i,j}$.

Ad esempio, le $r_{i,j}$ componenti principali possono essere scelte come quelle cui è associato un valore singolare maggiore di una percentuale (preferibilmente compresa nell'intervallo 1-4% e ancor più preferibilmente pari al 2%) del valore singolare maggiore (σ_1) della matrice $Y_{i,j}$.

Ad esempio se la matrice $Y_{i,j}$ avesse i seguenti valori singolari: $\sigma_1=100$, $\sigma_2=20$, $\sigma_3=10$, $\sigma_4=1$, $\sigma_5=0,5$, $\sigma_6=0,3$..., accettando valori singolari maggiori al 2% di σ_1 verrebbero considerate solo le prime tre componenti principali u_1 , u_2 , u_3 .

Ulteriormente, l'errore di proiezione, che nella forma

$$\text{err_Proj}(\text{IR}_{i,j}) = \|\text{Proj}(\text{IR}_{i,j}) - \text{IR}_{i,j}\|_2$$

preferita è calcolato come

,

può essere indifferentemente calcolato come una qualsiasi norma

L_p :

$$\text{err_Proj}(\mathbf{IR}_{i,j}) = \|\text{Proj}(\mathbf{IR}_{i,j}) - \mathbf{IR}_{i,j}\|_{L_p}$$

* * * * *

RIVENDICAZIONI

1. Metodo di rilevazione eventi per sistemi di video sorveglianza comprendente una fase di apprendimento in cui in diversi istanti di tempo si acquisiscono immagini di apprendimento di un'area sorvegliata in una situazione di assenza di eventi da rilevare, ed una fase operativa di rilevazione in cui vengono acquisite immagini correnti di detta area, caratterizzato dal fatto di rilevare un evento mediante il confronto di una immagine corrente con una immagine corrispondente ad una combinazione lineare di una pluralità di immagini di riferimento approssimanti, o coincidenti con, rispettive immagini di apprendimento.

2. Metodo secondo la rivendicazione 1, caratterizzato dal fatto che dette immagini di riferimento sono ottenute utilizzando un metodo di analisi delle componenti principali di un insieme di apprendimento costituito da una pluralità di immagini di apprendimento.

3. Metodo secondo la rivendicazione 2 in cui detto insieme di apprendimento è organizzato in una matrice della normalità le cui colonne contengono i pixel di dette immagini di apprendimento e dette immagini di riferimento sono le colonne di una matrice $(Y_{i,j}^r)$ approssimante detta matrice della normalità $(Y_{i,j})$.

4. Metodo secondo la rivendicazione 3 caratterizzato dal fatto che detto metodo di analisi delle componenti principali

comprende un'operazione di scomposizione in valori singolari di detta matrice della normalità, detta scomposizione determinando tre matrici

$$\begin{aligned}U_{i,j} &= [u_1 \dots u_s], \\ \Sigma_{i,j} &= \text{diag}(\sigma_1, \dots, \sigma_s), \text{ e} \\ V_{i,j} &= [v_1 \dots v_s],\end{aligned}$$

tali che la matrice della normalità $Y_{i,j}$, può essere scritta come:

$$Y_{i,j} = U_{i,j} \cdot \Sigma_{i,j} \cdot V_{i,j}^T$$

5. Metodo secondo la rivendicazione 4 in cui detta matrice approssimante detta matrice della normalità è

$$Y_{i,j}^r = U_{i,j}^r \cdot \Sigma_{i,j}^r \cdot (V_{i,j}^r)^T$$

dove

$$\begin{aligned}U_{i,j}^r &= [u_1 \dots u_r] \\ \Sigma_{i,j}^r &= \text{diag}(\sigma_1 \dots \sigma_r) \\ V_{i,j}^r &= [v_1 \dots v_r]\end{aligned}$$

il numero r di colonne delle matrici $U_{i,j}^r, \Sigma_{i,j}^r, V_{i,j}^r$ essendo inferiore al numero s delle colonne della matrici $U_{i,j}, \Sigma_{i,j}, V_{i,j}$.

6. Metodo secondo una delle rivendicazioni da 3 a 5 in cui detta matrice approssimante la matrice della normalità è tale che la somma dei suoi valori singolari è inferiore ad una percentuale prefissata, preferibilmente pari al 75%, della somma dei valori singolari della matrice della normalità.

7. Metodo secondo una delle rivendicazioni da 3 a 5 in cui i

valori singolari di detta matrice approssimante la matrice della normalità sono superiori ad una percentuale prefissata, preferibilmente compresa tra il 2% ed il 4%, del valore singolare massimo di detta matrice della normalità.

8. Metodo secondo una delle rivendicazioni precedenti caratterizzato dal fatto che detto confronto comprende l'operazione di proiettare un vettore immagine, corrispondente a detta immagine corrente, sullo spazio vettoriale definito dall'insieme di tutte le combinazioni lineari delle immagini di riferimento.

9. Metodo secondo la rivendicazione 8 caratterizzato dal fatto di rilevare un evento se è soddisfatta la relazione

$$\text{err_Proj}(\text{IR}) \geq \text{Thr}$$

dove Thr è una soglia ed $\text{err_Proj}(\text{IR})$ è un errore di proiezione calcolato come la norma della differenza tra i vettori IR e Proj(IR), IR essendo detto vettore immagine e Proj(IR) essendo il vettore corrispondente alla proiezione di IR su detto spazio vettoriale.

10. Metodo secondo la rivendicazione 9 quando la rivendicazione 8 dipende da una delle rivendicazioni da 2 a 7, in cui detta matrice della normalità e detta matrice approssimante hanno in comune una pluralità di valori singolari e caratterizzato dal fatto che detta soglia è presa uguale al maggiore dei valori singolari non comuni.

11. Metodo secondo la rivendicazione 9 quando la rivendicazione

8 dipende da una delle rivendicazioni da 2 a 7, in cui detta matrice della normalità e detta matrice approssimante hanno in comune una pluralità di valori singolari e caratterizzato dal fatto che detta soglia è presa uguale al maggiore dei valori singolari non comuni, moltiplicato per un numero reale prefissato maggiore di uno.

12. Metodo secondo la rivendicazione 10, caratterizzato dal fatto che detta soglia è presa uguale ad una percentuale prefissata, compresa tra il 2% ed il 4%, del valore singolare maggiore di detta matrice della normalità.

13. Metodo secondo una delle rivendicazioni precedenti caratterizzato dal fatto che detta fase di apprendimento comprende una fase di addestramento ed una fase di validazione, detta fase di addestramento essendo atta a selezionare un insieme di apprendimento costituito da dette rispettive immagini di apprendimento, e detta fase di validazione essendo atta a verificare che detto insieme di apprendimento sia un modello rappresentativo di detta area in una situazione di assenza di eventi da rilevare.

14. Metodo secondo la rivendicazione 13, caratterizzato dal fatto che detta fase di validazione comprende almeno un'operazione di simulazione di detta fase operativa, detta simulazione prevedendo il confronto di una immagine di validazione con una combinazione lineare di dette immagini di riferimento, detta immagine di validazione essendo un'immagine

di apprendimento non appartenente a detto insieme di apprendimento.

15. Metodo secondo la rivendicazione 13 o 14, caratterizzato dal fatto di ripetere la fase di addestramento cambiando l'insieme di apprendimento, se detta fase di validazione rileva un evento.

16. Metodo secondo la rivendicazione 13 o 14, in cui detta fase di validazione comprende le operazioni di:

- proiettare una pluralità di vettori immagine, corrispondenti ad una pluralità di immagini di validazione, sullo spazio vettoriale definito dall'insieme di tutte le combinazioni lineari di dette immagini di riferimento,
- per ogni proiezione rilevare un errore di proiezione definito come la norma della differenza tra un vettore immagine ed il vettore corrispondente alla proiezione del vettore immagine su detto spazio vettoriale.

17. Metodo secondo la rivendicazione 16, caratterizzato dal fatto di

- calcolare l'errore di proiezione medio $\overline{\text{err_Proj}}$ come

$$\overline{\text{err_Proj}} = \text{media}[\text{err_Proj1}, \text{err_Proj2}, \text{err_Proj3}, \dots]$$

dove "media" è una funzione che riceve in ingresso gli errori di proiezione err_Proj1 , err_Proj2 , err_Proj3 ... delle immagini di validazione e ne restituisce il valor medio,

- ripetere la fase di addestramento cambiando l'insieme di apprendimento, se l'errore di proiezione medio $\overline{\text{err_Proj}}$ di una

pluralità di immagini di validazione supera o uguaglia una soglia.

18. Metodo secondo la rivendicazione 16, caratterizzato dal fatto di

- determinare l'errore di proiezione massimo $\overline{\text{err_Proj}}_{\text{MAX}}$ di tutte le proiezioni di detta pluralità di vettori immagine,
- ripetere la fase di addestramento cambiando l'insieme di apprendimento, se l'errore di proiezione massimo $\overline{\text{err_Proj}}_{\text{MAX}}$ supera o uguaglia una soglia.

19. Metodo secondo la rivendicazione 13 o 14, caratterizzato dal fatto di ripetere la fase di addestramento cambiando l'insieme di apprendimento, se detta fase di validazione rileva una percentuale di eventi superiore ad una soglia prefissata.

20. Metodo secondo una delle rivendicazioni precedenti caratterizzata dal fatto di effettuare operazioni di elaborazione delle immagini sia in fase di apprendimento sia in fase di rilevazione eventi.

21. Metodo secondo la rivendicazione 20, in cui dette operazioni di elaborazione delle immagini comprendono un'operazione di ridurre in scala di grigi dette immagini.

22. Metodo secondo la rivendicazione 20 o 21, in cui dette operazioni di elaborazione delle immagini comprendono un'operazione di filtraggio passa-basso preferibilmente con un kernel gaussiano.

23. Metodo secondo una delle rivendicazioni precedenti in cui

dette immagini di apprendimento e dette immagini correnti sono gruppi di pixel ottenuti mediante suddivisione, attraverso una griglia predefinita, di quadri o semiquadri di un segnale video.

24. Prodotto informatico caricabile in un'area di memoria di un elaboratore elettronico, comprendente porzioni di codice atte ad implementare il metodo secondo una delle rivendicazioni da 1 a 23 quando eseguite da detto computer.

25. Sistema di video sorveglianza comprendente almeno un'unità di acquisizione immagini operativamente connessa ad un unità di elaborazione di immagini, detta unità di elaborazione immagini essendo atta ad eseguire almeno una porzione del, in particolare tutto il, metodo secondo una delle rivendicazioni da 1 a 23.

26. Metodo secondo gli insegnamenti innovativi della presente descrizione e dei disegni annessi, i quali mostrano esempi di implementazione preferiti e vantaggiosi di detto metodo.

27. Sistema secondo gli insegnamenti innovativi della presente descrizione e dei disegni annessi, i quali mostrano esempi di realizzazione preferiti e vantaggiosi di detto sistema.

* * * * *

Videotec S.p.A.

p.i. Ing. Roberto DINI

(No. Iscr. Albo 270 BM)

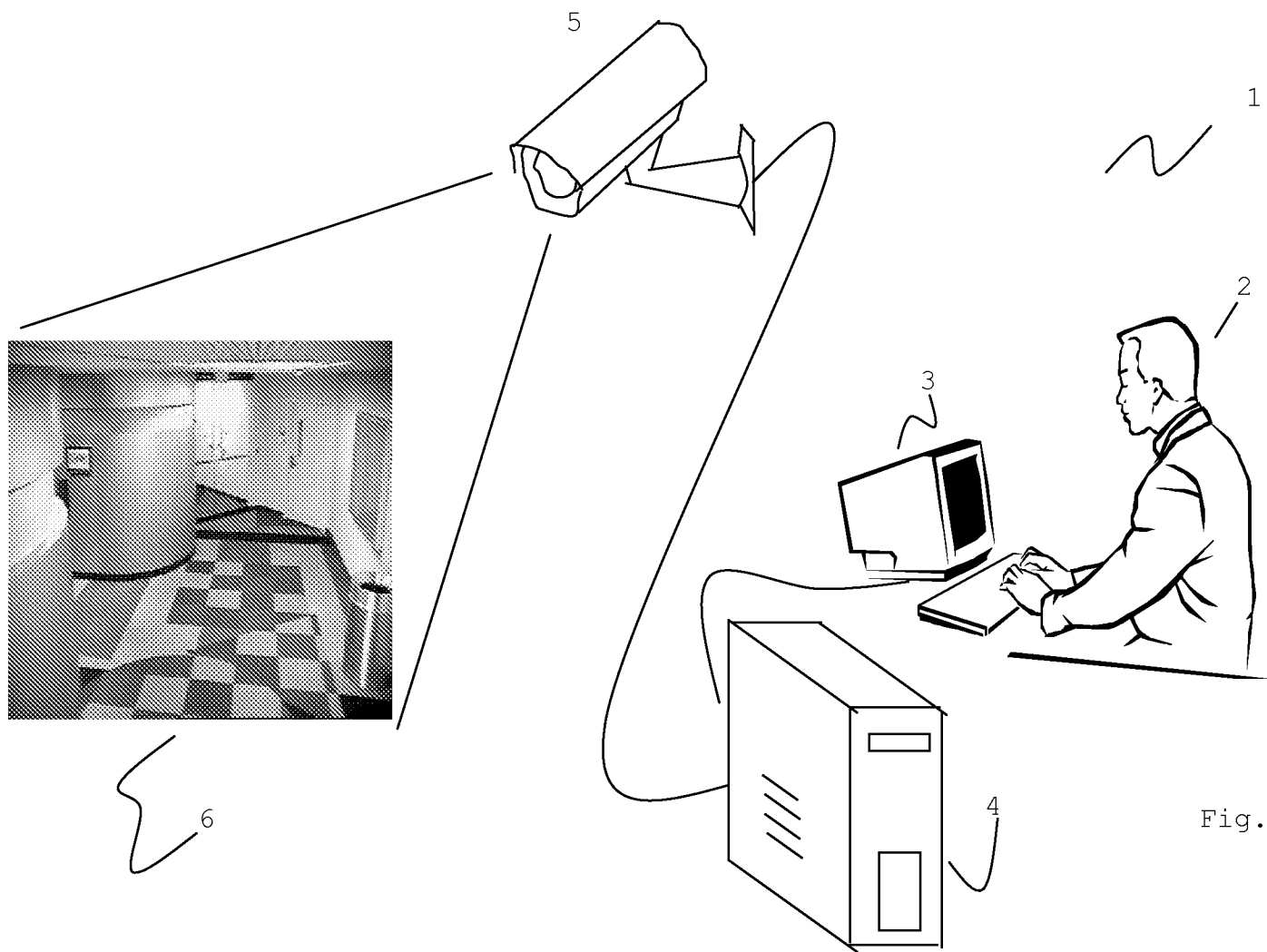


Fig.1

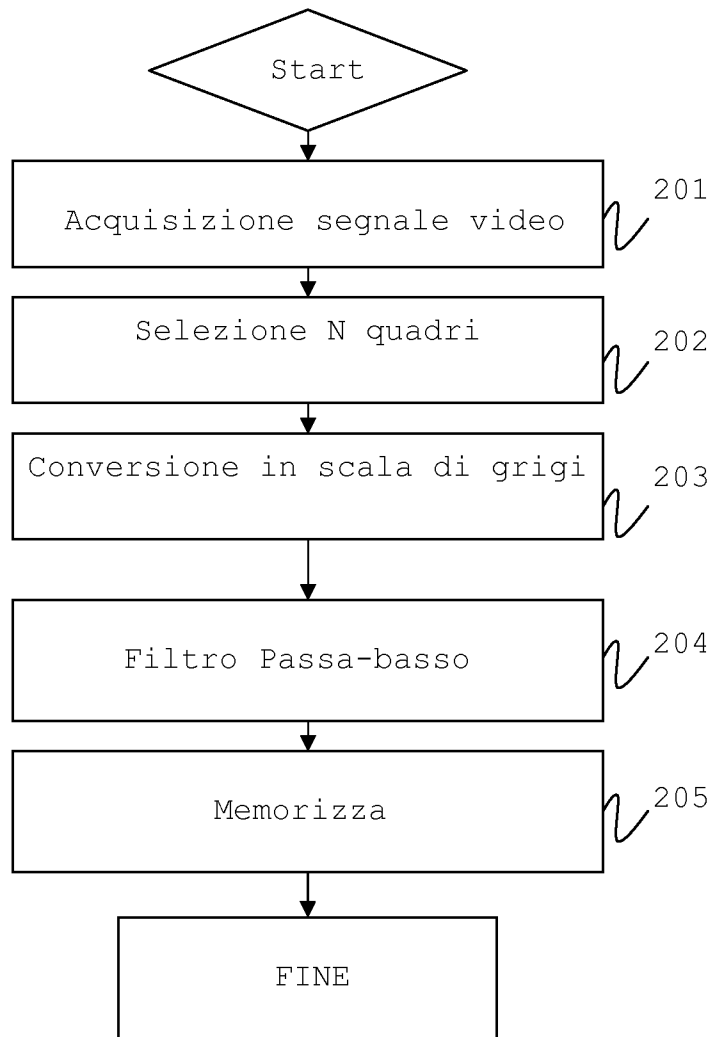


Fig.2a

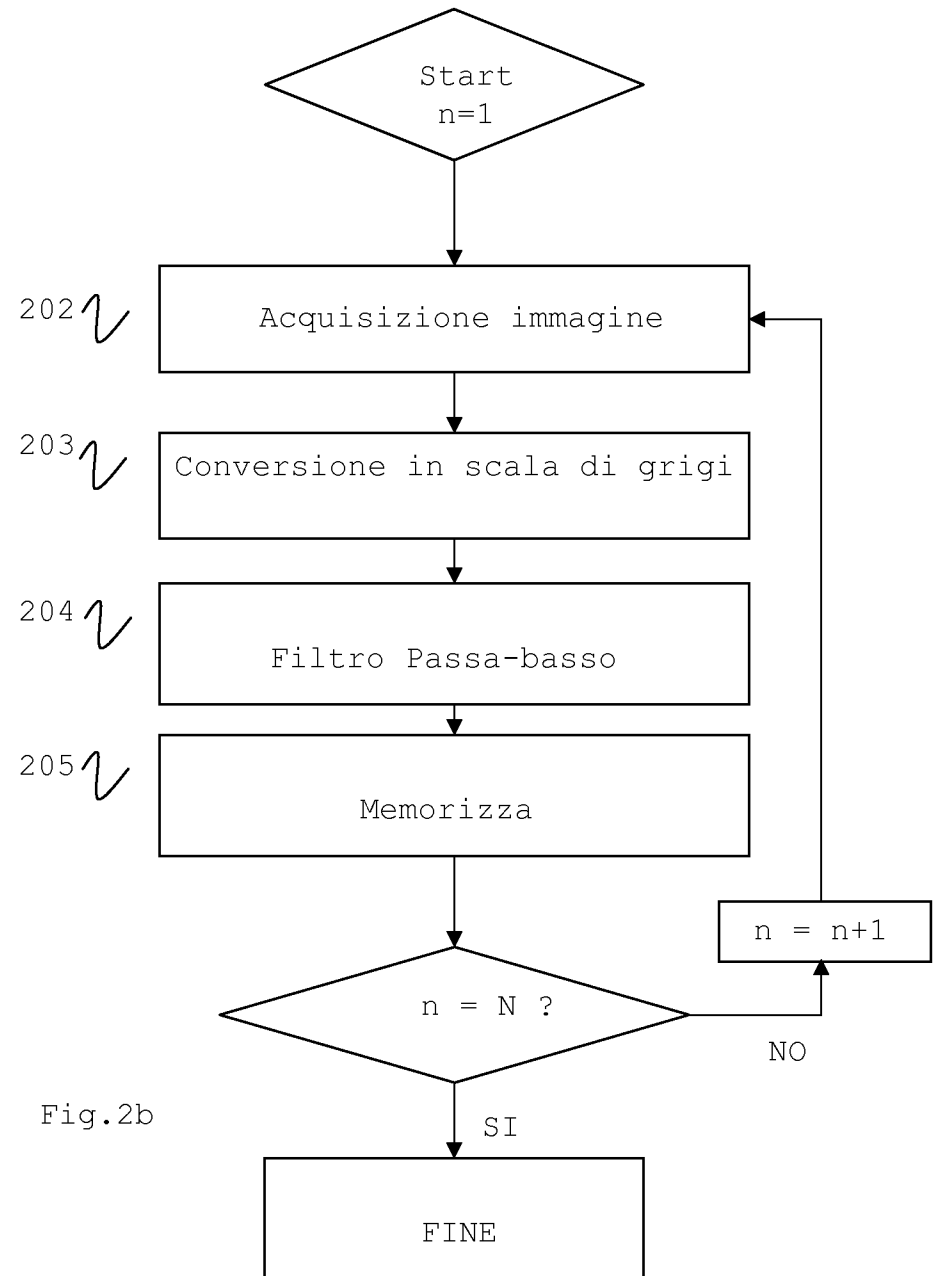


Fig.2b

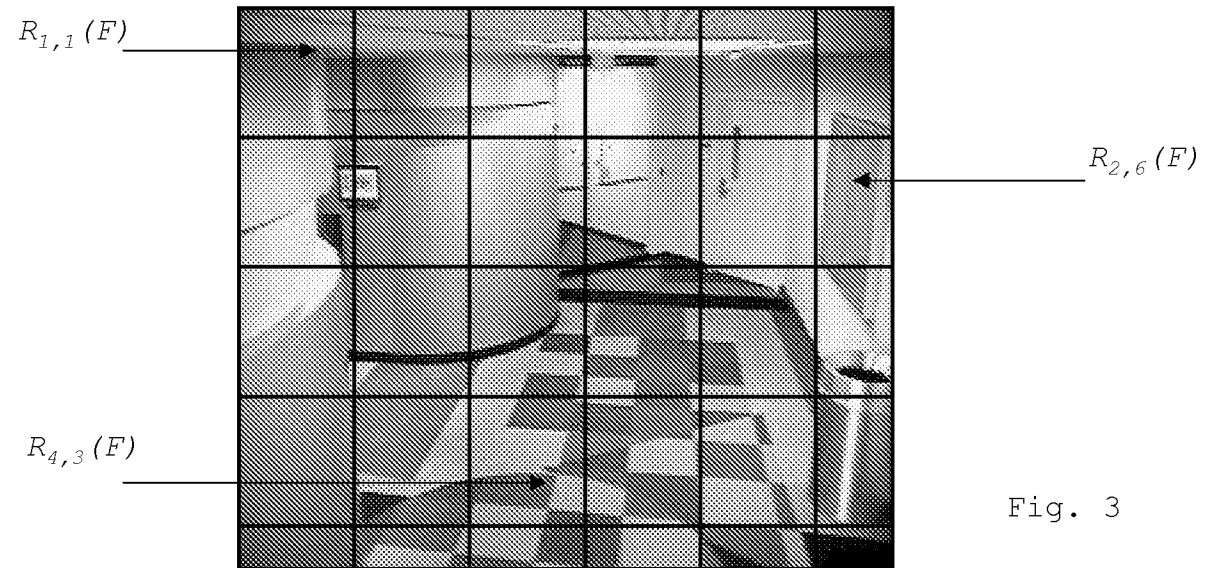


Fig. 3



Fig. 4

