



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

<b>DOMANDA NUMERO</b>	<b>102007901562014</b>
<b>Data Deposito</b>	<b>05/10/2007</b>
<b>Data Pubblicazione</b>	<b>05/04/2009</b>

Titolo

APPARATO DI ACQUISIZIONE ED ELABORAZIONE DELLE INFORMAZIONI RELATIVE AD  
ATTIVITA OCULARI UMANE.

## DESCRIZIONE

dell'invenzione industriale dal titolo: "Apparato di acquisizione ed elaborazione delle informazioni relative ad attività oculari umane"

A nome: UNIVERSITA' DEGLI STUDI DI ROMA "LA SAPIENZA"

di nazionalità italiana

Con sede in P.le Aldo Moro, 5 - 00185 ROMA

Inventori: Fiora PIRRI ARDIZZONE, Anna BELARDINELLI, Andrea CARBONE

\* \* \* \* \*

Campo dell'invenzione

La presente invenzione si riferisce al campo degli apparati per il tracciamento delle attività attentive dell'uomo ai fini dell'addestramento di robot.

Stato della tecnica

Si conoscono sistemi per lo studio dei comportamenti visivi in ambienti reali e per il tracciamento dello sguardo.

Alcuni si avvalgono di dati di profondità, altri degli angoli di rotazione della testa, altri si limitano a tracciare i movimenti oculari, solitamente per movimenti limitati e con soggetti seduti e/o posizionati davanti a schermi, per cui non è necessaria un'informazione di distanza.

La scena viene normalmente ripresa tramite una o più telecamere, solidali con il capo dell'osservatore.

E' necessario, sempre e comunque, includere i movimenti della testa per tenere conto dei meccanismi attentivi originati da fattori non inclusi nel campo di vista dell'osservatore e/o della eventuale telecamera, come eventuali stimoli acustici.

Alcuni sistemi noti posizionano la telecamera il più vicino possibile all'occhio,

cioè al suo campo di vista, per rendere il punto di ripresa coincidente con quello di vista dell'osservatore.

E' noto apparato dotato di una telecamera che punta ad un solo occhio, illuminato da uno o più LED a infrarossi (IR), di un computer portatile per l'elaborazione dei dati immagine, e di un display esterno, fissato o meno al casco che sorregge i sensori. Manca, dunque, il tracciamento dei movimenti della testa e la presenza di una telecamera atta ad acquisire una visione della scena osservata e il tracciamento anche del secondo occhio. Tale apparato prevede una procedura di calibrazione basata su tecniche di "image processing" con le quali, estratta la regione della pupilla, ne viene calcolato il centro per poi ri-proiettarlo su un display posto davanti ad uno degli occhi dell'utente. Tale procedura di calibrazione, alla base del mapping tra le coordinate immagine dei frame dell'occhio e le coordinate sullo schermo, è basata sul calcolo dei coefficienti di due funzioni polinomiali del terzo ordine fissando N punti a locazioni fissate su uno schermo.

Un altro apparato noto è formato da due telecamere rivolte alla scena e quindi può determinare effettivamente il punto di fissazione. Questo è però ottenuto con un complesso sistema motorizzato formato da giunti cardanici (gimbal) che fanno girare le telecamere ai lati degli occhi in modo da renderle parallele all'asse ottico di ciascun occhio. La direzione dell'asse degli occhi, illuminati da LED IR, è riflessa su specchi diecrici ripresi da altre due telecamere, per l'asservimento delle telecamere sulla scena. La complessità di tale apparato si riflette negativamente sul peso dello stesso che lo rende ingombrante e scomodo per l'osservatore.

Un altro apparato, ancora, si basa anch'esso su un sistema indossabile per il

tracciamento dello sguardo. Esso comprende una microcamera CCD dedicata alla ripresa delle immagini dell'occhio mentre una è rivolta alla scena. Le coordinate del centro della pupilla vengono estratte tramite un programma di riconoscimento di immagini. Il metodo su cui si basa prevede una calibrazione su fissazione di punti predeterminati in numero compreso tra 25 e 100 e determina automaticamente la funzione di mapping tra le due camere come una funzione di correlazione adattata tramite regressione quadratica. Una forma di compressione e classificazione delle sequenze viene svolta categorizzando i punti di fissazione.

Dunque, risulta alquanto laboriosa la calibrazione e modesta la precisione nella determinazione del punto guardato anche se l'occhio non sta fissando, denominato anche "point of regard".

Altri apparati ancora, rilevano i movimenti della testa a mezzo di trasponder magnetici fissi, oppure riprendendo con una telecamera l'osservatore, rendendo vincolante per lo stesso l'ambito di operatività.

Quindi, gli apparati forniti dalla tecnica nota non sono in grado di tracciare contemporaneamente i movimenti oculari e della testa dell'osservatore e utilizzano architetture, spesso, ingombranti e complesse.

#### Sommario dell'invenzione

Scopo della presente invenzione è di risolvere i problemi suesposti fornendo un apparato, denominato di seguito per brevità "gaze machine", per lo studio della dinamica di puntamento dello sguardo, che automaticamente acquisisce, processa e memorizza dati relativi ai movimenti di occhi e testa di un osservatore, pur restando leggero, preciso e affidabile.

La presente invenzione, pertanto, si propone di raggiungere gli scopi sopra di-

scussi realizzando un apparato di acquisizione ed elaborazione delle informazioni relative ad attività oculari umane, conformemente alla rivendicazione 1, e comprendente:

- una o più telecamere atte a riprendere in stereoscopia la medesima scena osservata da un osservatore;
- una o più microcamere atte a riprendere gli occhi dell'osservatore;
- almeno un led IR illuminante ciascun occhio;
- una piattaforma inerziale;
- una imbracatura atta ad aggrapparsi stabilmente al capo dell'osservatore, su cui sono montate stabilmente detta una o più telecamere, detta una o più microcamere e detta piattaforma inerziale;
- mezzi di elaborazione di dati atti a realizzare strutture di dati contenenti le informazioni generate da detta una o più microcamere, da detta una o più telecamere e da detta piattaforma inerziale;
- mezzi di rice-trasmissione di detti dati.

Secondo un altro aspetto dell'invenzione detto apparato trova applicazione quando i dati da esso ottenuti vengono utilizzati da una piattaforma robotica in modo da replicare l'attività di osservazione di un osservatore per imitazione.

Vantaggiosamente, l'apparato oggetto del presente trovato, consente di rilevare sia la velocità dello sguardo nella scena, sia la profondità del punto di fissazione osservato, e tracciare i movimenti operati dagli occhi mentre l'osservatore svolge altre attività senso-motorie, ad esempio camminando, manipolando oggetti, e comunque in ambienti dinamici.

In particolare, il tracciamento dei movimenti della testa è svolta tramite sensoristica inerziale che, data la leggerezza, consente libertà di movimento e preci-

sione dei dati, così da evitare di riprendere l'osservatore con una microcamera esterna o di disporre di un transponder magnetico fisso per stimare l'orientazione della testa.

Le rivendicazioni dipendenti descrivono realizzazioni preferite dell'invenzione.

#### Breve descrizione delle Figure

Ulteriori caratteristiche e vantaggi dell'invenzione risulteranno maggiormente evidenti alla luce della descrizione dettagliata di una forma di realizzazione preferita, ma non esclusiva, di un apparato di acquisizione ed elaborazione delle attività oculari umane, illustrato a titolo esemplificativo e non limitativo, con l'ausilio delle unite tavole di disegno in cui:

la Fig. 1 raffigura una vista laterale di una gaze machine secondo l'invenzione;

la Fig. 2 raffigura una vista frontale relativa alla precedente;

la Fig. 3 raffigura una pupilla illuminata con la proiezione di un'immagine di Purkinje;

la Fig. 4a raffigura la rappresentazione semplificata dell'occhio di Le Grand e la

Fig. 4b raffigura i possibili casi di riflessioni delle immagini di Purkinje;

la Fig. 5 rappresenta la successione di elaborazioni compiute sull'immagine dell'occhio al fine di individuarne la pupilla;

la Fig. 6 illustra il calcolo del centro della pupilla CPt, in una sequenza di due frame. Nella sequenza, la figura inferiore illustra l'immagine dell'occhio filtrata, e quella superiore indica l'immagine saturata. Nell'immagine superiore si apprezzano, oltre al centro della pupilla i vettori che corrispondono alla linea di vista e alla congiungente l'immagine Purkinje e la linea di vista;

la Fig. 7 rappresenta un modello di calibrazione utile all'individuazione di una funzione di corrispondenza tra le coordinate dei centri delle pupille, così come

viste in più frame da una microcamera che inquadra gli occhi, e quelle delle corrispondenti immagini-scena riprese da una coppia di telecamere;

la Fig. 8 rappresenta le trasformazioni che devono essere effettuate per ottenere la proiezione di un punto della scena sull'occhio e sulla camera che lo inquadra;

la Fig. 9 illustra schematicamente, la visione di un punto da parte di un osservatore, condiviso da una piattaforma robotica;

la Fig. 10 rappresenta uno schema di una composizione vettoriale propedeutica alla predisposizione delle trasformazioni tra il sistema di riferimento dell'apparato e il sistema di riferimento di una piattaforma robotica istruenda;

la Fig. 11 riporta un diagramma di flusso riferito ad un metodo di elaborazione delle informazioni acquisite dalla gaze machine;

la Fig. 12 riporta una realizzazione preferita di dispositivo munito di una microcamera per ciascun occhio.

#### Descrizione in dettaglio di una forma di realizzazione preferita dell'invenzione

Una realizzazione preferita dell'apparato comprende una gaze machine ed una piattaforma robotica atta ad ricevere i dati trasmessi da detta gaze machine allo scopo di allinearsi sul medesimo punto di osservazione dell'osservatore a fini addestrativi della medesima piattaforma robotica.

In particolare una realizzazione preferita della gaze machine comprende un casco munito di piattaforma inerziale 9, due telecamere 8, almeno una microcamera 6, due led 7 ad infrarossi, uno per ciascun occhio, mezzi di elaborazione e di trasmissione dei dati.

Inoltre, detto casco definente l'imbracatura del capo dell'osservatore è munito di mezzi di regolazione per meglio adattarsi ad esso.

In particolare, tali mezzi comprendono una ghiera 4 e cuscinetti 5 posizionati in corrispondenza della nuca.

Al casco 1 è connesso un braccetto 2 orientabile a mezzo di snodi 3, atto a sorreggere la/le microcamere. Si preferisce tale braccetto in carbonio, per essere rigido e leggero al tempo stesso.

Detta almeno una microcamera 6 è in tecnologia CMOS e montata su un sostegno rigido 5.

Poiché la tecnologia CMOS è sensibile agli infrarossi sono presenti ai lati della microcamera, sorretti da altrettanti supporti 5, almeno una coppia di led ad infrarossi 7, atti ad illuminare gli occhi dell'osservatore.

L'uso di detti led ad infrarossi 7 consente di aumentare il contrasto fra la pupilla, zona centrale dell'iride e la sclera. In particolare, l'asse ottico della microcamera e l'asse di emissione dei led ad infrarossi 7 sono volutamente non allineati in modo da favorire la manifestazione dell'effetto, denominato black-pupil, secondo cui la pupilla appare di un colore nero intenso, aumentando il contrasto rispetto alla regione circostante all'iride.

In una realizzazione preferita viene usata una microcamera 6 ed un led ad infrarossi 7 per ciascun occhio.

In un'altra realizzazione preferita, l'unica microcamera è posizionata centralmente in modo da riprendere entrambi gli occhi contemporaneamente. Questa soluzione, ad unica microcamera, consente di determinare con precisione la vergenza delle pupille, in un'unica immagine, e il punto esatto su cui converge lo sguardo. Questo perché il punto di osservazione delle telecamere sulla scena non può mai essere esattamente sovrapposto a quello dell'utente, quindi la proiezione di un solo occhio determina una linea di possibili punti di fissazione,

come illustrato più dettagliatamente nel seguito.

La piattaforma inerziale 9 è un cosiddetto motion tracker, che permette di rilevare gli angoli assoluti di rotazione della testa su almeno 3 gradi di libertà. Detta piattaforma inerziale 9 è solidale con detto casco 1 a mezzo di un supporto 10.

Si preferisce l'utilizzo di una coppia di telecamere 8 atte a realizzare una ripresa stereoscopica della medesima scena fissata dall'osservatore.

Il metodo di acquisizione ed elaborazione delle informazioni viene esplicitato con l'aiuto della figura 11, comprendente i seguenti passi:

- acquisizione dell'immagine della scena e dell'immagine degli occhi;
- identificazione della pupilla ed identificazione del suo centro;
- proiezione del punto di fissazione degli occhi sull'immagine della scena;
- memorizzazione in un'unica struttura di dati delle immagini ricavate al passo precedente insieme alle informazioni riguardanti la posizione della testa;
- trasmissione e ricezione di detti dati e trasformazione del punto osservato nelle coordinate di una piattaforma robot istruenda;
- convergenza della testa definente detta piattaforma robotica sul punto di fissazione dell'osservatore.

In particolare, le telecamere 8 della gaze machine, rivolte alla scena, consentono l'acquisizione di immagini RGB e range, tramite calcolo della disparità tra le due ottiche. Per ogni pixel di ogni frame sono pertanto disponibili i valori di intensità sui tre canali di colore e le coordinate mondo  $[X \ Y \ Z]^T$  rispetto al punto di osservazione corrente.

Al termine di ogni registrazione l'apparato, grazie a detti mezzi di elaborazione, genera vettori di strutture di dati, etichettate in relazione all'istante di acquisi-

zione, comprendenti:

- frame RGB relativi alla scena ripresa dalle telecamere;
- la posizione del punto osservato proiettato sul frame espresso in coordinate di immagine (pixel);
- l'immagine di profondità corrispondente all'immagine a colori;
- l'orientamento della testa espresso in angoli di Eulero.

Quando la gaze machine viene adoperata da un osservatore fermo, per es., seduto, le posizioni dello sguardo lungo i frame della sequenza sono riferite ad un medesimo punto di vista, per cui possono essere ad composte in un unico sistema di riferimento globale, coincidente col sistema di riferimento solidale con la posizione dell'osservatore. In questo caso, è possibile determinare a mezzo di procedure note la direzione e velocità delle saccadi operate dall'osservatore rispetto alle coordinate globali riferite al sistema di riferimento con origine nella posizione dell'osservatore.

I dati restituiti dalla gaze machine vengono trasmessi ad una piattaforma robotica, nota in sé, e da questa ulteriormente elaborati e utilizzati.

Una piattaforma robotica preferita è dotata di una camera stereoscopica ed equipaggiata con un'unità pan-tilt atta a replicare la postura della testa in termini di angoli di beccheggio e imbardata (pitch e yaw), acquisendo i dati forniti dalla gaze machine attraverso mezzi propri di rice-trasmissione di dati ed elaborandoli attraverso mezzi propri di elaborazione, atti ad operare trasformazioni roto-traslazionali, dal sistema di riferimento dell'osservatore, definito globale, a quello proprio della piattaforma robotica, in modo che la piattaforma può replicare, in tempo reale, il processo di acquisizione operato dall'osservatore/tutor, condividendone il punto osservato.

La trasmissione dei dati dall'apparato di acquisizione alla testa robotica è realizzata a mezzo di una connessione wireless, realizzata grazie a detti mezzi di comunicazione connessi con il computer, definente detti mezzi di elaborazione.

I dati sono resi disponibili dalla gaze machine in tempo reale in modo che possono essere adoperati anche da più piattaforme robotiche e/o ulteriori sistemi di elaborazione.

Di seguito viene illustrato un metodo, pure oggetto della presente invenzione, di identificazione della pupilla.

La pupilla, in ambienti dinamici, con condizioni di luce variabile, non presenta le classiche condizioni di distinzione dell'iride e della sclera come in ambienti interni in condizioni di luce stabile, come si vede dalla figura 3, in cui la pupilla è cancellata dal riflesso del cielo.

Per risolvere il problema di identificazione della pupilla si ricorre al riflesso del led sull'occhio e alle immagini Purkinje che si generano per effetto di esso.

Fra le quattro immagini Purkinje, come si vede dalla figura 4, soltanto la prima ha intensità massima quindi valori tra 245 e 255 in qualunque immagine, sia interna che esterna.

I led ad infrarossi sono posizionati a circa 15 cm dall'occhio nel sistema riferimento microcamera, posti lungo gli assi x e y, in modo da garantire che l'immagine Purkinje si posiziona entro il cerchio limbale dell'iride, anche per rotazioni dell'asse ottico fino a 30 gradi.

Si denomina  $L_c$  il punto rappresentativo, sul piano immagine, della posizione di ciascun led ad infrarossi, cioè dove mantiene le caratteristiche di distinzione, come si vede nella figura 3.

Per l'identificazione del centro della pupilla si identifica, quindi, principalmente

l'immagine Purkinje I e quindi si calcolano le opportune trasformazioni come indicato in seguito. Non è necessaria la calibrazione del led, è solo necessario che nell'immagine dell'occhio, centrato rispetto alla microcamera, cioè che guarda nella direzione dell'obiettivo, l'immagine Purkinje I appare internamente alla pupilla, come si vede nel primo tipo di riflessione dell'immagine di Purkinje, vedi figura 4.

Il riflesso degli almeno 2 led tramite le immagini Purkinje dipende dalla posizione della microcamera CMOS rispetto agli occhi, che cambia ogni volta che si indossa la gaze machine, dalla curvatura della cornea e della fascia limbale tipica di ciascun individuo.

Per individuare una funzione di corrispondenza tra le coordinate dei centri delle pupille, così come viste da una microcamera che inquadra gli occhi, e quelle dell'immagine-scena si fa ricorso ad una procedura di calibrazione che necessita di un certo numero di immagini di ciascun occhio in altrettante specifiche posizioni.

Al termine della procedura il metodo preferito di identificazione della pupilla comprende i seguenti passi:

Macropasso 1: L'immagine dell'occhio inquadrata dalla microcamera, di dimensioni  $130 \times 220$ , viene tagliata ad un'immagine  $130 \times 130$  pixel. L'immagine viene quindi saturata, si veda figura 5a, e filtrata, si veda figura 5b, con la derivata di un filtro Gaussiano di dimensioni  $3 \times 3$  e varianza 0.28. Accade che se  $I_x$  rappresenta l'immagine trasformata, allora l'immagine Purkinje I è identificata dalla regione massimale di  $I_x$  che contiene punti minimali. Sia  $P_k$  la regione identificata.

Macropasso 2: date  $M$  immagini dell'occhio che identificano rotazioni estremali

lungo gli assi di rotazione come indicato in figura 5c, con  $M \geq 8$ , si risolve il sistema

$$x = [A^T A]^{-1} A^T b$$

dove la matrice A è una matrice  $M \times 8$  i cui elementi normalizzati sono:

- la posizione di  $Pk_t$  nell'immagine al frame t e la posizione  $Pk_{t+1}$  nell'immagine al frame t + 1,
- la posizione del centro della pupilla  $Cp_t$  nell'immagine al frame t;
- la posizione del punto di intersezione fra la retta L1, che passa per  $Lc$  e  $Pk_t$ ;
- la retta normale alla x dell'immagine e che passa per il centro della pupilla  $Cp_t$ , denominato punto nodale N.

Il vettore b, di dimensione  $M \times 2$ , è costituito dai centri delle pupille  $Cp_{t+1}$  nei frame t+1. Il vettore bidimensionale x, di dimensione  $8 \times 2$ , è il vettore dei coefficienti, l'mse (mean squared error) ottenuto dalla regressione stimata è inferiore a 20.

Macropasso 3: dato il vettore dei coefficienti x, a ciascun passo di computazione si calcolano, per l'immagine corrente, i valori indicati in un'osservazione della matrice A, e si ottiene il vettore T i cui elementi sono appunto quelli di una osservazione della matrice A. Il prodotto interno  $[T] [x]$  restituisce una approssimazione del centro della pupilla, come si vede dalla figura 6 in cui la pupilla è in condizioni estremali e in fase saccadica.

Le posizioni correnti  $Cp_{t+1}$ , calcolate con  $Tx$ , e i valori  $N_{t+1}$  (punto nodale), e  $Pk_{t+1}$  (immagine Purkinje del frame successivo), calcolati come indicato nel passo 2, vengono assegnate alle posizioni corrispondenti per il successivo passo di computazione.

Ad ogni ciclo, definito da detti tre passi di computazione, l'algoritmo fornisce, la

posizione del centro della pupilla, la direzione e l'ampiezza dello spostamento rispetto all'immagine precedente.

Di seguito viene illustrata una procedura di calibrazione funzionale ad una procedura di identificazione della pupilla.

Il processo di calibrazione viene ripetuto ogni volta che un osservatore indossa la gaze machine in quanto la posizione relativa della gaze-machine rispetto agli occhi dell'osservatore influenza l'esito della procedura.

L'osservatore che indossa la gaze machine deve osservare una sequenza di punti distinguibili nello spazio, come quelli in figura 7. E' necessario che ciascun punto sia visibile nella telecamera che inquadra la scena. A tal fine è necessario allestire un pattern di riferimento specifico per effettuare la calibrazione. Il pattern consiste in una griglia contenente un numero minimo di 7 riferimenti planari e non. Gli istanti che corrispondono a ciascuna fissazione sono selezionati direttamente dall'osservatore tramite un qualunque mezzo di interfaccia uomo/elaboratore, per es., a mezzo del tasto di una tastiera.

Per ogni fissazione così eseguita, vengono salvate:

- a) le immagini dalle telecamere che inquadrano la scena;
- b) le immagine dei due occhi che stanno fissando il singolo riferimento.

Una volta collezionate le coppie riferimento/occhi, queste devono subire una serie le operazioni elencati di seguito, per effettuarsi la trasposizione del punto di osservazione sulla scena osservata:

- a) estrazione delle coordinate del centro della pupilla per ciascun occhio e coordinate del riferimento nell'immagine di scena corrispondente;
- b) stima della funzione di mappatura del punto di fissazione tramite calcolo del tensore trifocale.

Dunque, riassumendo, la procedura di calibrazione preferita comprende i seguenti passi:

- si allestisce un modello di riferimento con un numero  $n > 7$  di riferimenti planari e non;
- in un numero  $n$  di istanti l'osservatore fissa un riferimento e lo rileva a mezzo di una interfaccia;
- si effettua l'estrazione delle coordinate del centro della pupilla per ciascun occhio e coordinate del riferimento nell'immagine di scena corrispondente;
- si stima la funzione di mappatura del punto di fissazione tramite calcolo di una matrice denominata tensore trifocale  $T$ .

**Procedura di rototraslazione dal sistema di riferimento globale al sistema di riferimento della microcamera.**

Consideriamo due sistemi di riferimento: il sistema di riferimento dell'occhio, che si assume coincidente con il sistema di riferimento globale e il sistema di riferimento della microcamera che lo inquadra. Essi sono legati da una trasformazione rigida  $(R, t)$  in cui  $R$  è una matrice di rotazione  $3 \times 3$  e  $t$  un vettore di traslazione (si veda la figura 8 per la calibrazione tra i due sistemi di riferimento). Le equazioni di proiezione di un punto  $P = (x, y, z)$  sull'occhio sono date da:

$$u_e = \frac{f_e}{z} x$$

$$v_e = \frac{f_e}{z} y$$

dove viene fatta l'assunzione che il punto  $P_e$ , proiezione di  $P$ , si trova nella regione limbale dell'occhio, assunta planare, e coincide con il centro della pupilla.

Le coordinate del punto  $P_c$ , proiezione di  $P_e$  sulla microcamera, sono determi-

nabili nel modo seguente:

$$k \begin{pmatrix} u_c \\ v_c \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} f_c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & f_c & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & f_c \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} + t_1 / f_c & \mathbf{0} \\ r_{21} & r_{22} & r_{23} + t_2 / f_c & \mathbf{0} \\ r_{31} & r_{32} & r_{33} + t_3 / f_c & \mathbf{0} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ \mathbf{1} \end{pmatrix}$$

Quindi esiste una matrice proiettiva che lega direttamente i punti della scena con i punti dell'immagine della microcamera che inquadra l'occhio.

Effettuando un'analogia procedura per il secondo occhio otteniamo tre piani immagine, quelli dei due occhi e quello della microcamera che inquadra la scena, le cui coordinate delle proiezioni di un punto della scena, sono tra loro legate da una matrice T, denominata tensore trifocale. Di seguito viene illustrato un metodo preferito di calcolo di detto tensore trifocale e, successivamente, come calcolare a partire da T e dalla proiezione di un punto su due immagini, la proiezione del punto sulla terza immagine.

### Metodo di individuazione del tensore trifocale.

Dati  $n > 7$  punti in corrispondenza tra le tre immagini l'algoritmo di calcolo del tensore trifocale procede attraverso i seguenti passi:

- si trovano tre matrici di trasformazione H, H' e H'' da applicare alle tre immagini;
- ciascun punto  $x^i$  dell'immagine i-esima viene trasformato nel punto  $\hat{x}^i = H_j^i x^j$  al fine di operare una normalizzazione;
- si calcola T linearmente in termini dei punti trasformati risolvendo l'insieme di equazioni della forma  $At = 0$ , utilizzando la pseudoinversa e il metodo dei minimi quadrati. La matrice A è una matrice di 27 righe tali che per ciascuna equazione del tipo  $x^i x^{ij} x^{nk} \mathcal{E}_{jqs} \mathcal{E}_{krt} T_i^q r = 0_{st}$  si hanno nove equazioni, una per ogni scelta di s e t.

Nella predetta equazione  $\epsilon_{jqs}$  e  $\epsilon_{qrt}$  sono i tensori definiti per  $r, s, t = 1, 2, 3$  nel seguente modo:

$$\epsilon_{rst} = \begin{cases} 0 & \text{a meno che } r, s, t \text{ non sono distinti} \\ 1 & \text{se } rst \text{ è una permutazione pari di } 123 \\ -1 & \text{se } rst \text{ è una permutazione dispari di } 123 \end{cases}$$

e  $0_{st}$  è anch'esso un tensore.

- Si calcola il tensore trifocale applicando la trasformazione inversa di denormalizzazione per ottenere il tensore originale:

$$T_i^{jk} = H_i^r (H^{r-1})_s^j (H^{r-1})_t^k \hat{T}_r^{st}$$

### **Procedura di trasferimento di corrispondenze da una coppia di viste ad una terza.**

La proiezione del punto di fissazione degli occhi sull'immagine della scena, al fine di rendere possibile la convergenza di detta testa robotica con il punto di osservazione dell'osservatore, viene realizzata a mezzo di una rototraslazione che dapprima converte il sistema di riferimento associato a detta microcamera 6 nel sistema di riferimento degli occhi e successivamente nel sistema di riferimento di detta telecamera 8 riprendente la scena osservata dal soggetto.

Per cui, siano  $x \rightarrow x'$  una corrispondenza esatta tra la prima e la seconda vista calcolate nel modo usuale. Sia  $F_{21}$  la matrice fondamentale che lega le prime due viste, determinabile a partire dal tensore trifocale  $T$ . Per ottenere la corrispondenza nella terza vista è sufficiente seguire i seguenti passi:

- si calcola la linea  $l'$  che passa attraverso  $x'$ , perpendicolare ad

$$l'_e = F_{21}x = (l_1 l_2 l_3)T$$

- si determinano le coordinate del punto trasferito  $x''^k = x^i l'_j T$

Dunque, il presente apparato permette di realizzare un'interfaccia uomo-robot utilizzabile per diversi compiti di apprendimento automatico applicato alla visio-

ne artificiale, semplificando l'acquisizione di dati e di pattern di scansione. Questi possono essere utilizzati non solo nell'ambito della robotica cognitiva ma anche in sistemi di sorveglianza, per sistemi teleoperati per disabili o per applicazioni di informatica grafica e realtà virtuale. In altri ambiti è pensabile l'utilizzo per indagini di marketing, analisi di usabilità di sistemi ed interfacce, studi sull'attenzione umana di vario genere, scopi terapeutici per soggetti affetti da disfunzioni cognitive legate alla visione ed in generale qualunque applicazione in cui interessa conoscere cosa osserva un utente.

I modi particolari di realizzazione qui descritti non limitano il contenuto di questa domanda che copre tutte le varianti dell'invenzione definite dalle rivendicazioni.

## RIVENDICAZIONI

1. Apparato di acquisizione ed elaborazione di informazioni relative ad attività oculari umane attuate da un osservatore comprendente:

- una o più telecamere (8) atte a riprendere in stereoscopia una medesima scena osservata dall'osservatore;

- una o più microcamere (6) atte a riprendere occhi dell'osservatore;

- almeno un led IR (7) illuminante ciascuno di detti occhi;

- una piattaforma inerziale (9);

- una imbracatura (1) atta a fissarsi stabilmente ad una testa dell'osservatore, su cui sono montate stabilmente detta una o più telecamere, detta una o più microcamere e detta piattaforma inerziale;

- mezzi di elaborazione di dati atti a realizzare strutture di dati contenenti le informazioni generate da detta una o più microcamere, da detta una o più telecamere e da detta piattaforma inerziale;

- mezzi di rice-trasmissione di detti dati;

essendo detti dati ricevuti da una piattaforma robotica atta a ricavarne da essi i punti di una scena osservati dall'osservatore in modo da addestrare la piattaforma robotica.

2. Apparato di acquisizione secondo la rivendicazione 1, in cui detta imbracatura definisce un casco (1) munito di mezzi di regolazione per meglio adattarsi alla testa dell'osservatore.

3. Apparato di acquisizione secondo la rivendicazione 2, in cui detto casco è munito di un braccetto (2) orientabile a mezzo di snodi (3), atto a sorreggere dette una o più microcamere (6) ed un led (7) per ciascun occhio.

4. Apparato di acquisizione secondo la rivendicazione 1, in cui dette una o più

microcamere (6) sono in tecnologia CMOS.

5. Apparato di acquisizione secondo la rivendicazione 1, in cui detta piattaforma inerziale (9) è un motion tracker, atto a rilevare angoli assoluti di rotazione della testa dell'osservatore su almeno 3 gradi di libertà.

6. Apparato di acquisizione secondo la rivendicazione 1, in cui dette telecamere (8) sono atte ad acquisire immagini RGB e range tramite un calcolo della disparità tra due ottiche di dette telecamere.

7. Apparato di acquisizione secondo la rivendicazione 1, in cui detti mezzi di elaborazione, al termine di ogni registrazione sono atti a generare vettori di strutture di dati etichettate in relazione all'istante di acquisizione, comprendenti:

- fotogrammi RGB relativi alla scena ripresa dalle telecamere;
- la posizione del punto osservato proiettato sui fotogrammi RGB espresso in coordinate di immagine (pixel);
- l'immagine di profondità corrispondente all'immagine a colori;
- l'orientamento della testa espresso in angoli di Eulero.

8. Metodo di acquisizione ed elaborazione delle informazioni, mediante l'apparato della rivendicazione 1, comprendente i seguenti passi:

- acquisizione di una immagine di una scena e di una immagine di occhi;
- identificazione di una pupilla di detti occhi ed identificazione del suo centro;
- proiezione di un punto di fissazione degli occhi su detta immagine della scena;
- memorizzazione in un'unica struttura di dati dell'immagine ricavata al passo precedente insieme alle informazioni riguardanti una posizione di una testa di osservatore;
- trasmissione e ricezione di detti dati e trasformazione del punto di fissazio-

ne nelle coordinate di una piattaforma robot istruenda;

- convergenza della testa definente detta piattaforma robotica sul punto di fissazione dell'osservatore.

9. Metodo secondo la rivendicazione 8, in cui detta fase di identificazione della pupilla comprendente i seguenti macropassi:

Macropasso 1:

- si ottiene una immagine dell'occhio inquadrata da una microcamera, di dimensioni 130×220;
- viene ritagliata da detta immagine un'altra immagine di 130×130 pixel;
- quest'ultima immagine viene saturata e filtrata con la derivata di un filtro Gaussiano di dimensioni 3 × 3 e varianza 0,28, così che si rende possibile ottenere l'immagine Purkinje I identificando una regione massimale della immagine ottenuta, contenente punti minimali;
- allora si contrassegna col simbolo Pk la regione identificata;

Macropasso 2:

- rilevate un numero  $M \geq 8$  di immagini dell'occhio che identificano rotazioni estremali lungo gli assi di rotazione, si risolve il sistema  $x = [A^T A]^{-1} A^T b$  dove la matrice A è una matrice  $M \times 8$  i cui elementi normalizzati sono:
  - posizione della regione identificata  $Pk_t$  nell'immagine al fotogramma t e una posizione  $Pk_{t+1}$  identificata nell'immagine al fotogramma t + 1,
  - posizione del centro della pupilla  $Cp_t$  nell'immagine al fotogramma t;
  - la posizione del punto di intersezione fra la retta L1, che passa per Lc e  $Pk_t$ ;
  - la retta normale alla x dell'immagine e che passa per il centro della pupilla  $Cp_t$ , punto nodale N;
- essendo il vettore b, di dimensione  $M \times 2$ , è costituito dai centri delle pupille

$C_{p_{t+1}}$  nei fotogrammi  $t+1$ ; essendo il vettore bidimensionale  $x$ , di dimensione  $8 \times 2$ , è il vettore dei coefficienti, così che l'mse, cioè il cosiddetto "mean squared error", ottenuto dalla regressione stimata risulta inferiore a 20;

Macropasso 3:

- dato il vettore dei coefficienti  $x$ , a ciascun passo di computazione si calcolano, per l'immagine corrente, i valori indicati in un'osservazione della matrice  $A$ , e si ottiene il vettore  $T$  i cui elementi sono appunto quelli di una osservazione della matrice  $A$ ; sicché il prodotto interno  $[T] * [x]$  restituisce una approssimazione del centro della pupilla, in condizioni estremali e in fase saccadica;

- dunque, le posizioni correnti  $C_{p_{t+1}}$ , calcolate con  $Tx$ , e i valori  $N_{t+1}$ , e  $PK_{t+1}$ , calcolati come indicato nel macropasso 2, vengono assegnate alle posizioni corrispondenti per il successivo passo di computazione;

dunque, rendendosi possibile, ad ogni ciclo, definito da detti tre macropassi di computazione, la posizione del centro della pupilla, la direzione e l'ampiezza dello spostamento rispetto all'immagine precedente.

10. Metodo secondo la rivendicazione 8, comprendente una ulteriore e preventiva procedura di calibrazione comprendente le seguenti fasi:

- si allestisce un modello di riferimento con un numero  $n > 7$  di riferimenti planari e non;

- in un numero  $n$  di istanti l'osservatore fissa un riferimento e lo rileva a mezzo di una interfaccia;

- si effettua l'estrazione delle coordinate del centro della pupilla per ciascun occhio e coordinate del riferimento nell'immagine di scena corrispondente;

- si stima la funzione di mappatura del punto di fissazione tramite calcolo di una matrice denominata tensore **trifocale T**.

11. Metodo secondo la rivendicazione 10, in cui il calcolo di detto tensore trifocale nella procedura di calibrazione, comprendente i seguenti passi:

- si legano un sistema di riferimento di un primo occhio ed un sistema di riferimento di una microcamera che lo inquadra con una trasformazione rigida (R, t) in cui R è una matrice di rotazione 3 x 3 e t un vettore di traslazione, sicché le equazioni di proiezione di un punto P = (x, y, z) sull'occhio sono date da:

$$u_e = \frac{f_e}{z} x$$

$$v_e = \frac{f_e}{z} y$$

dove viene fatta l'assunzione che un punto Pe, corrispondente alla proiezione di P, si trova nella regione limbale dell'occhio, assunta planare, e coincidente con il centro di una pupilla dell'occhio;

- si ricavano le coordinate di un punto Pc, corrispondente alla proiezione di Pe sulla microcamera a mezzo della seguente equazione matriciale:

$$k \begin{pmatrix} u_e \\ v_e \\ 1 \end{pmatrix} = \begin{pmatrix} f_e & 0 & 0 \\ 0 & f_e & 0 \\ 0 & 0 & f_e \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} + t_1 / f_e & 0 \\ r_{21} & r_{22} & r_{23} + t_2 / f_e & 0 \\ r_{31} & r_{32} & r_{33} + t_3 / f_e & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

- si effettua un'analogha procedura per un secondo occhio;

- si ottengono tre piani immagine, quelli dei primo e secondo occhi e quello della microcamera che inquadra la scena, le cui coordinate delle proiezioni di un punto della scena, sono tra loro legate da detta matrice T.

12. Metodo secondo la rivendicazione 11, in cui detta trasformazione dal sistema di riferimento di detto primo occhio al sistema di riferimento della microcamera, comprendente i seguenti passi:

- si individua un numero  $n > 7$  punti in corrispondenza tra detti tre piani immagine;
- si definiscono tre matrici di trasformazione  $H$ ,  $H'$  e  $H''$  da applicare ai tre piani immagini;
- ciascun punto  $x^i$  di un'immagine  $i$ -esima viene trasformato nel punto  $\hat{x}^i = H_j^i x^j$  al fine di operare una normalizzazione;
- si calcola detto tensore trifocale  $T$  linearmente in termini dei punti trasformati risolvendo l'insieme di equazioni nella forma  $At = 0$ , utilizzando la pseudoinversa e il metodo dei minimi quadrati; essendo  $A$  una matrice di 27 righe tali che per ciascuna equazione del tipo  $x^i x^{ij} x^{nk} \varepsilon_{jqs} \varepsilon_{krt} T_i^q r = 0_{st}$  si hanno nove equazioni, una per ogni scelta di  $s$  e  $t$ , ove  $\varepsilon_{jqs}$  e  $\varepsilon_{krt}$  sono i tensori definiti per  $r, s, t = 1, 2, 3$  nel seguente modo:

$$\varepsilon_{rst} = \begin{cases} 0 & \text{a meno che } r, s, t \text{ non sono distinti} \\ 1 & \text{se } rst \text{ è una permutazione pari di } 123 \\ -1 & \text{se } rst \text{ è una permutazione dispari di } 123 \end{cases}$$

e  $0_{st}$  è anch'esso un tensore.

- si calcola il tensore trifocale  $T$  applicando la trasformazione inversa di denormalizzazione per ottenere il tensore originale a mezzo della seguente equazione:

$$T_i^{jk} = H_i^r (H'^{-1})_s^j (H''^{-1})_t^k \hat{T}_r^{st}.$$

13. Metodo secondo la rivendicazione 8, in cui detta proiezione del punto di fissazione degli occhi su detta immagine della scena, al fine di rendere possibile la convergenza di detta testa robotica con il punto di osservazione dell'osservatore, viene realizzato a mezzo di una rototraslazione che dapprima converte un sistema di riferimento associato a mezzi atti ad acquisire dette immagini degli occhi e poi a convertirlo in un sistema di riferimento associato a mezzi atti ad acquisire dette immagini della scena.

14. Metodo secondo le rivendicazioni 9 e 13, in cui detta rototraslazione è ottenuta a mezzo di una funzione del tipo  $x \rightarrow x'$ , che è una corrispondenza esatta tra un punto in un primo sistema di riferimento ed in un secondo sistema di riferimento, allora, data  $F_{21}$ , una matrice fondamentale che lega i due sistemi di riferimento, determinabile a partire da detto tensore trifocale  $T$ , per ottenere la trasformazione in un terzo sistema di riferimento, si eseguono i seguenti passi:

- si calcola una linea  $l'$  che passa attraverso  $x'$ , perpendicolare a mezzo della seguente equazione  $l'_e = F_{21}x = (l_1 l_2 l_3)T$ ;

- si determinano le coordinate di un punto trasferito a mezzo della seguente equazione  $x'^k = x^i l'_j T$ .

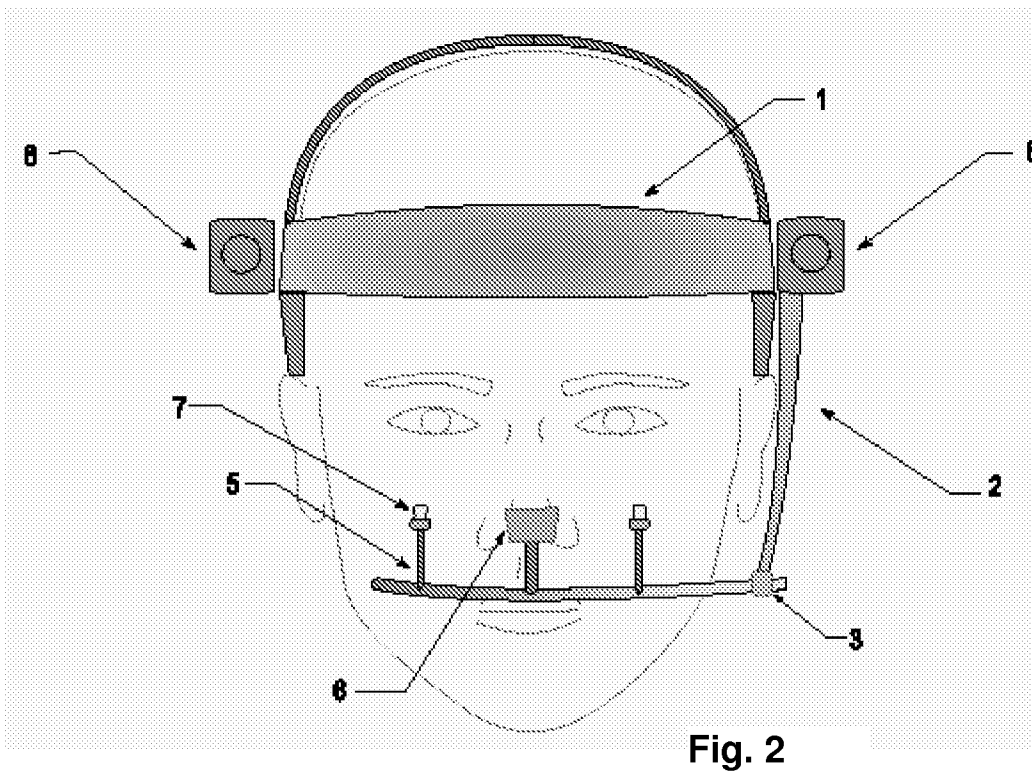
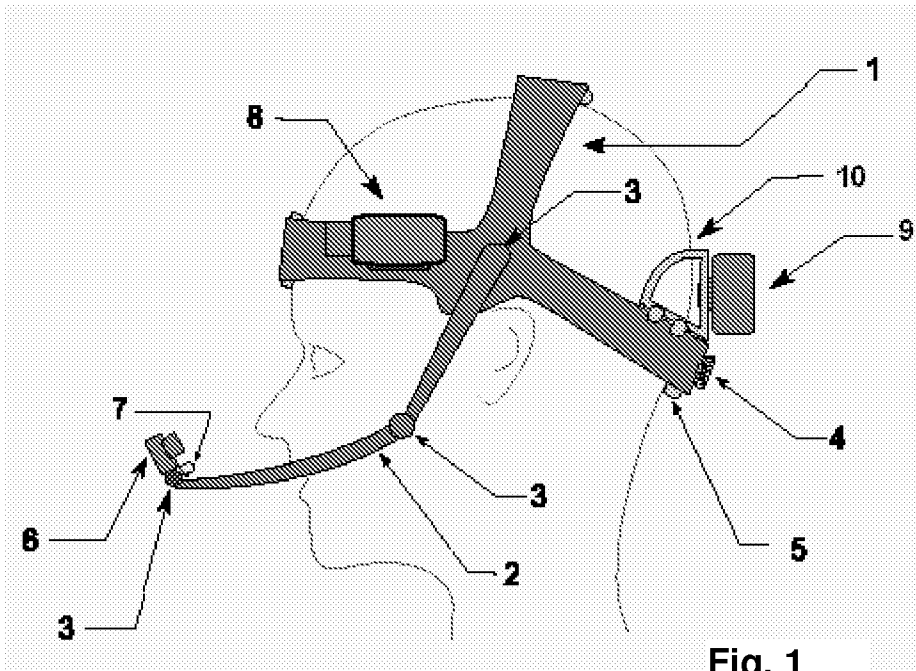
Roma, 5 Ottobre 2007

Per UNIVERSITA' DEGLI STUDI DI ROMA "LA SAPIENZA"

Il mandatario

Ing. Bruno Cinquantini

della NOTARBARTOLO & GERVASI SPA



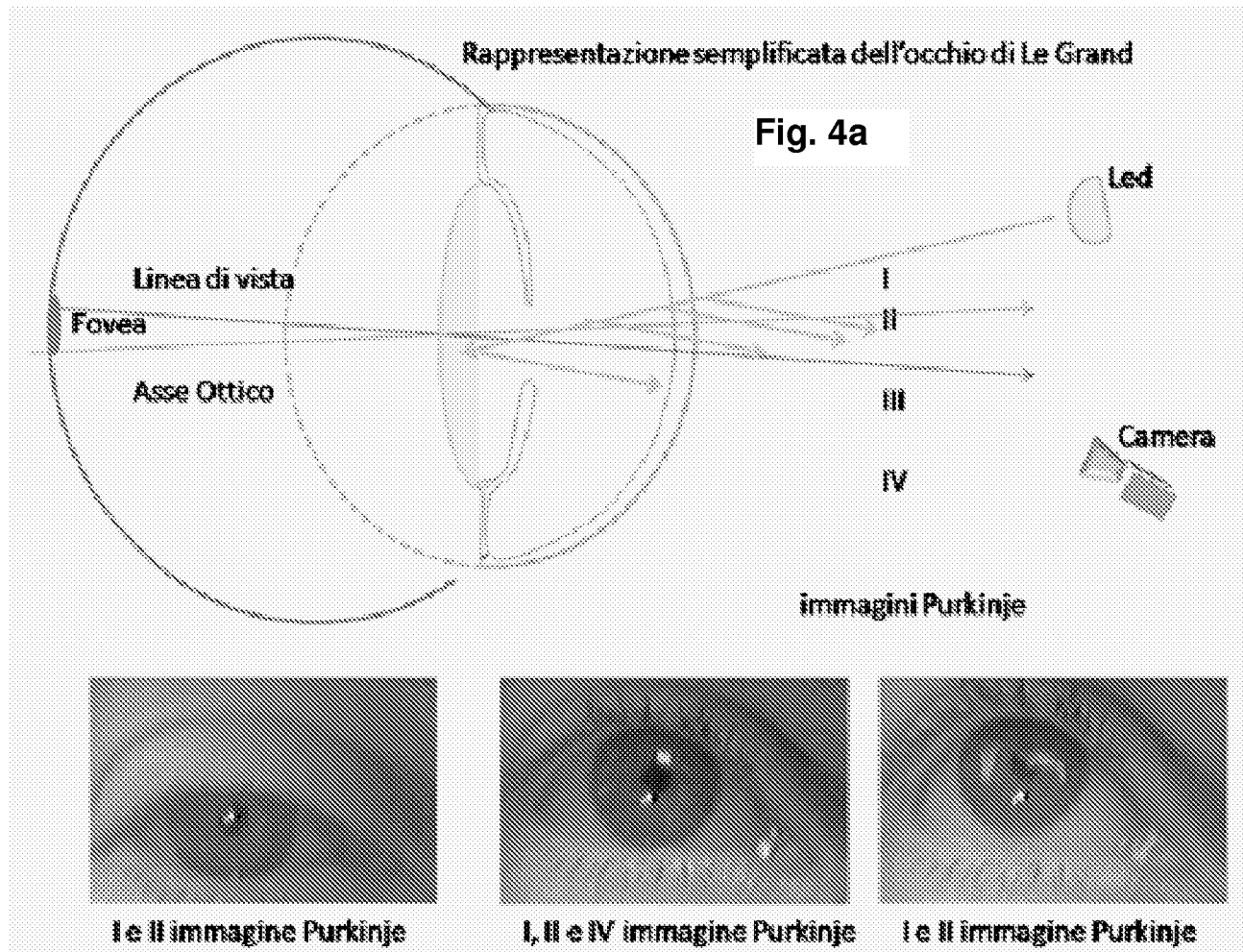
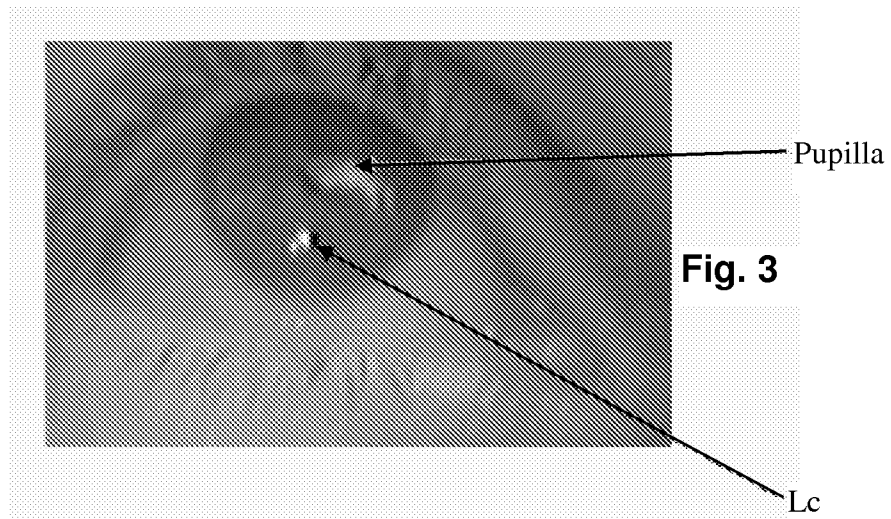


Fig. 4b

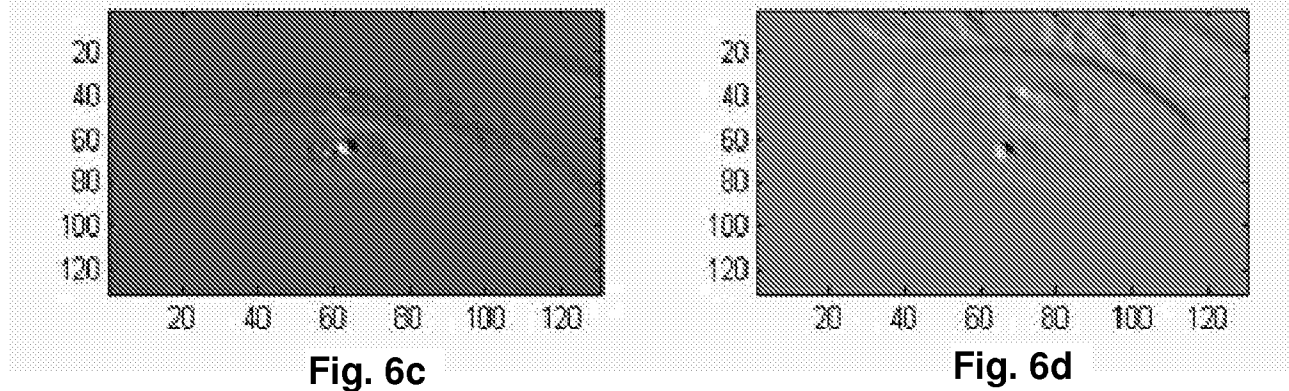
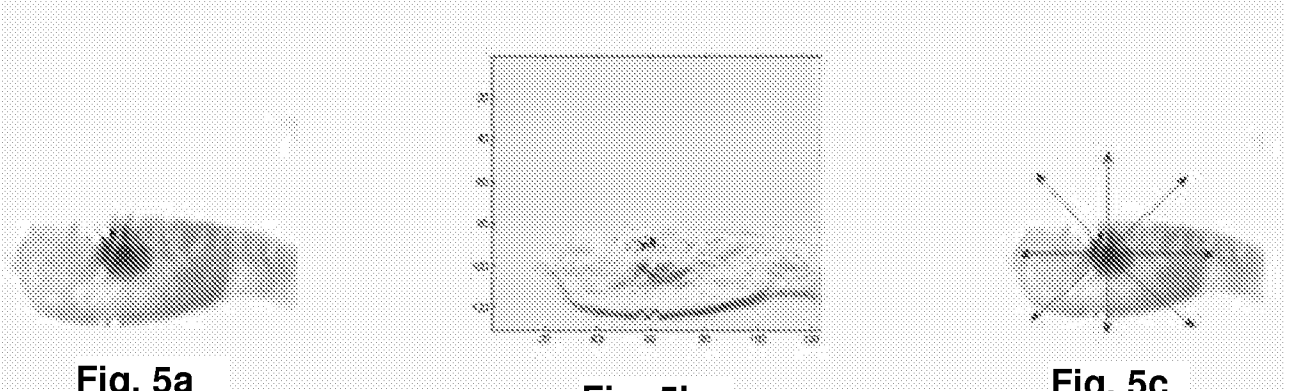


Fig. 7

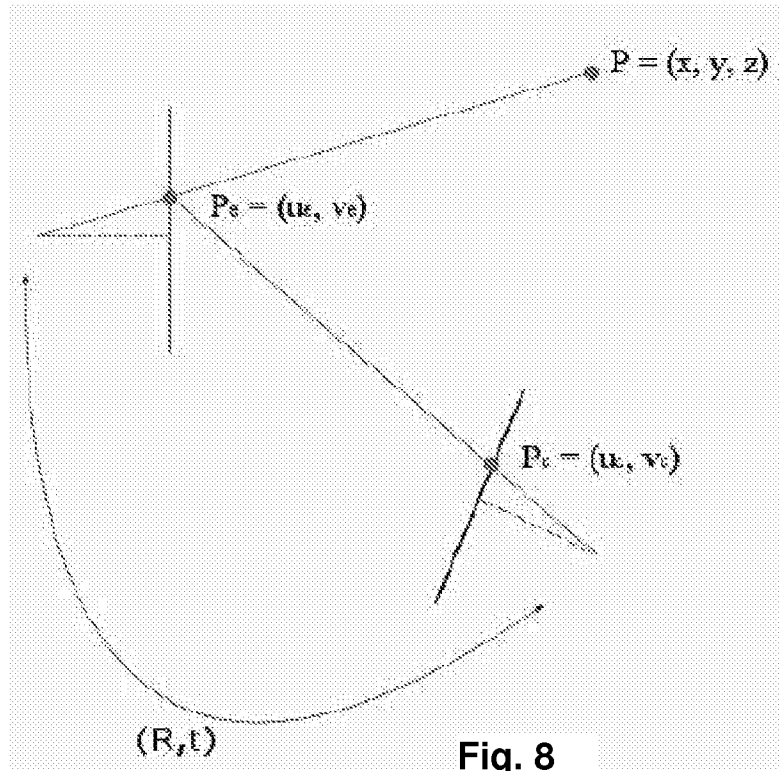


Fig. 8

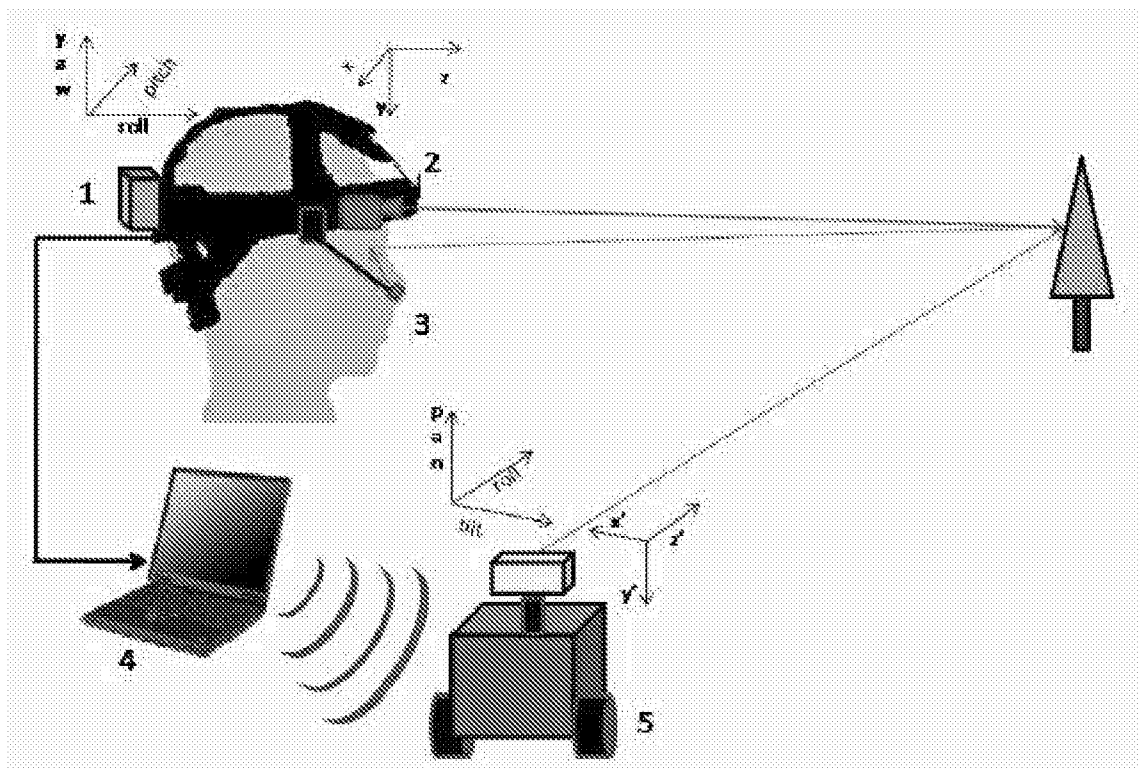


Fig. 9

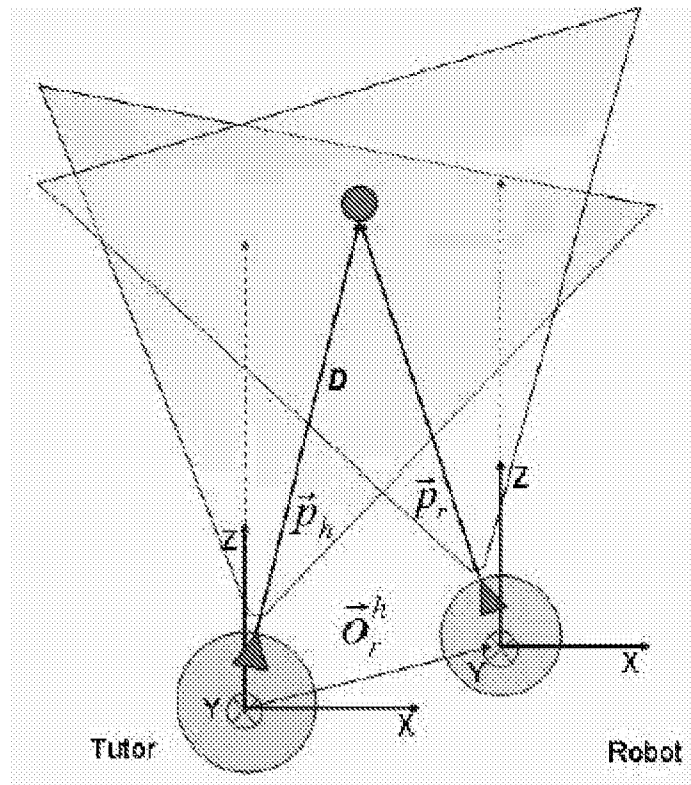


Fig. 10

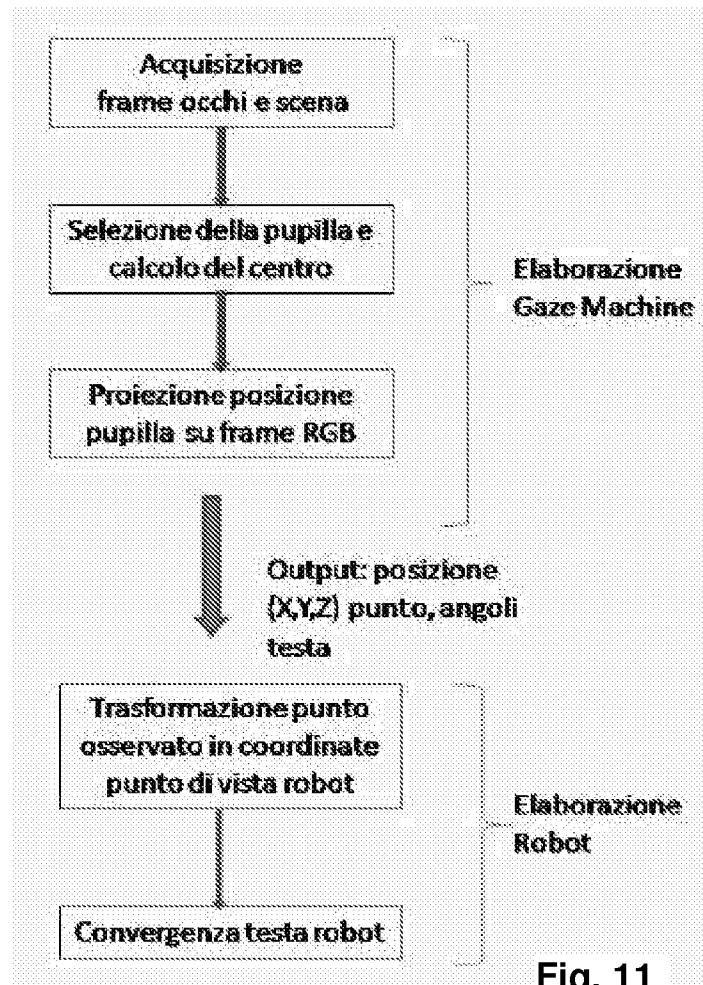


Fig. 11

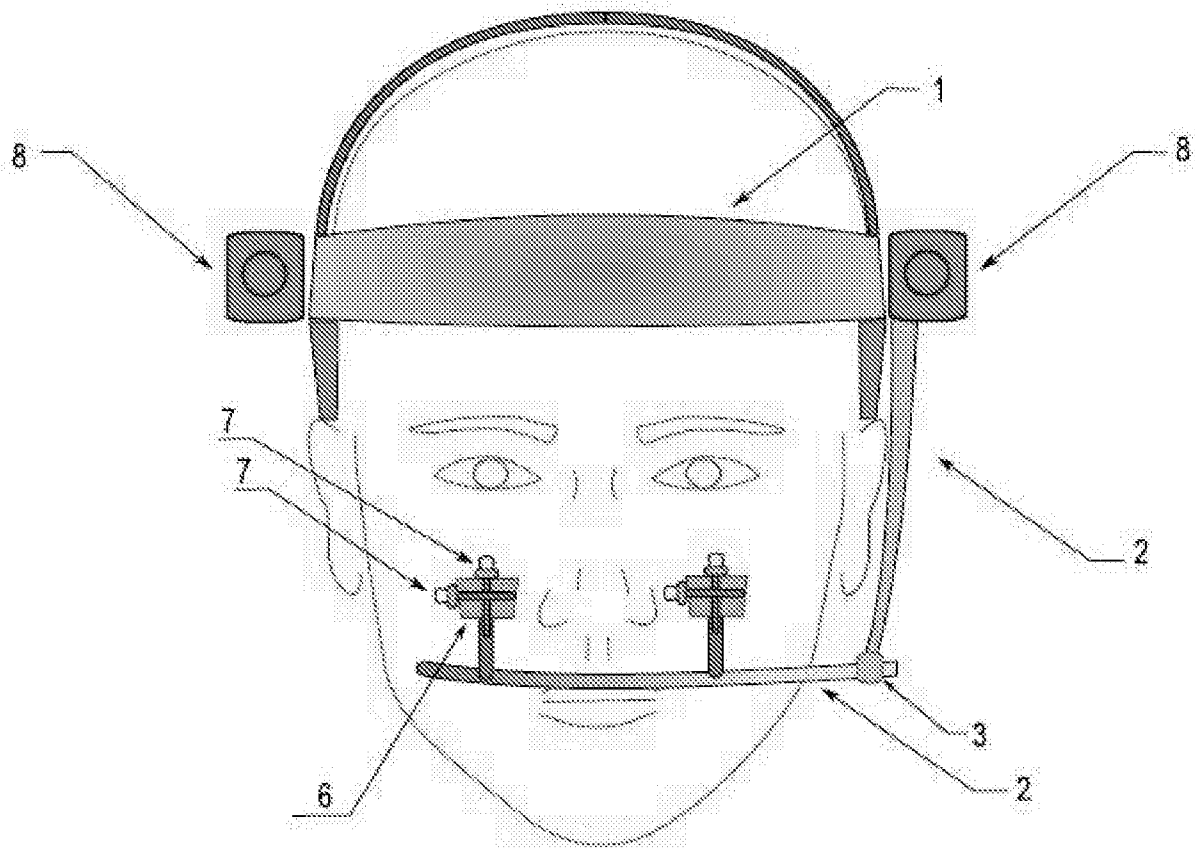


Fig. 12