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

UIBM

DOMANDA DI INVENZIONE NUMERO	**102021000032969**
Data Deposito	**29/12/2021**
Data Pubblicazione	**29/06/2023**

Classifiche IPC

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
G	06	T	7	70

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
G	06	T	7	73

Titolo

SISTEMA DI ELABORAZIONE DI INFORMAZIONE, METODO DI ELABORAZIONE DI INFORMAZIONE, E PROGRAMMA

DESCRIZIONE

del brevetto per invenzione industriale dal titolo:

"SISTEMA DI ELABORAZIONE DI INFORMAZIONE, METODO DI ELABORAZIONE DI INFORMAZIONE, E PROGRAMMA"

di 1) SONY INTERACTIVE ENTERTAINMENT INC.

di nazionalità giapponese

con sede: 1-7-1, KONAN, MINATO-KU

TOKYO 108-0075 (GIAPPONE)

di 2) FONDAZIONE ISTITUTO ITALIANO DI TECNOLOGIA

di nazionalità italiana

con sede: VIA MOREGO 30

16163 GENOVA (GE)

Inventori: SATO Shogo, INADA Tetsugo, SEGAWA Hiroyuki, PASQUALE Giulia, ONYSHCHUK Yuriy, MALAFRONTE Damiano, NATALE Lorenzo, RUZZENENTI Andrea

***

[Campo tecnico]

La presente invenzione riguarda un sistema di elaborazione di informazione, un metodo di elaborazione di informazione, e un programma.

[Stato della tecnica]

È stato sviluppato un metodo per stimare la posa (per essere precisi, la posizione e la posa relative come viste da una fotocamera) di un oggetto fotografato tramite l'uso di un modello di apprendimento automatico (machine learning

model). I dati di addestramento per addestrare questo modello di apprendimento automatico includono un'immagine in CG resa in base ad un modello tridimensionale dell'oggetto. Il suo scopo è, per esempio, assicurare una quantità di dati di addestramento e acquisire facilmente informazioni (per esempio, punti chiave) relative alla posa che fungono da dati di verità di base (ground truth).

Sida Peng et al. hanno pubblicato l'articolo "PVNet: Pixel-Wise Voting Network for 6DoF Pose Estimation" presso la IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) del 2019. In questo articolo, è descritta una tecnologia che implica: immettere un'immagine in un modello di apprendimento automatico; e calcolare le posizioni di punti chiave sull'immagine da utilizzare per la stima di posa in base a un'emissione del modello di apprendimento automatico.

In US2021/0031110A1, è descritta una tecnologia che implica: riconoscere una posa di un oggetto tenuto in una mano da un'immagine in cui l'oggetto è stato fotografato; e utilizzare la posa in un videogioco.

[Riepilogo dell'invenzione]

[Problema da risolvere mediante l'invenzione]

Nella tecnica correlata, per esempio, per via della difficoltà nell'aggiungere un'etichetta appropriata ad un'immagine effettivamente fotografata, un modello di

apprendimento automatico è stato addestrato mediante dati di addestramento includenti immagini in CG, e immagini effettivamente fotografate sono state immesse nel modello di apprendimento automatico che è stato addestrato, per stimare così una posa. Quindi, è probabile che si verifichi un problema ascrivibile a ciò in termini di precisione di stima di posa.

La presente invenzione è stata realizzata alla luce delle circostanze summenzionate e un suo obiettivo è fornire una tecnologia per migliorare la precisione della stima di posa mediante un modello di apprendimento automatico.

[Mezzi per risolvere il problema]

Per risolvere il problema summenzionato, secondo la presente invenzione, viene previsto un sistema di elaborazione di informazione includente: mezzi di acquisizione di regione bersaglio (target) per: acquisire un'immagine immessa; determinare se l'immagine immessa include o meno un'immagine di un oggetto bersaglio immettendo almeno una parte dell'immagine immessa in un modello di classificazione addestrato in base a una pluralità di immagini di apprendimento includente un'immagine in cui l'oggetto bersaglio è stato fotografato e dati di etichetta che indicano se ciascuna della pluralità di immagini di apprendimento include o meno l'oggetto bersaglio; e acquisire una regione bersaglio includente l'immagine

dell'oggetto bersaglio, che è estratta dall'immagine immessa, quando l'immagine immessa include l'oggetto bersaglio; e mezzi di stima di posa per stimare una posa dell'oggetto bersaglio in base a informazioni emesse da un modello di apprendimento automatico, quando la regione bersaglio acquisita è immessa in esso, che è addestrato mediante: una pluralità di immagini di addestramento rese da un modello di forma tridimensionale dell'oggetto bersaglio; e dati di verità di base che sono informazioni relative alla posa dell'oggetto bersaglio nelle immagini di addestramento.

In un aspetto della presente invenzione, i mezzi di acquisizione di regione bersaglio possono essere configurati per estrarre una regione includente un'immagine di un oggetto dall'immagine immessa, il modello di classificazione può includere: un'unità di generazione di caratteristica configurata per generare un valore di caratteristica di un'immagine di almeno una parte della regione estratta; e un classificatore configurato per ricevere un'immissione del valore di caratteristica generato ed emettere informazioni che indicano se la regione estratta ha o meno l'immagine dell'oggetto bersaglio, e il classificatore può essere addestrato mediante dati di addestramento includenti il valore di caratteristica generato dall'immagine cui l'oggetto bersaglio è stato fotografato e i dati di etichetta.

In un aspetto della presente invenzione, l'unità di generazione di caratteristica può essere regolata in modo tale che una distanza tra valori di caratteristica generati da una pluralità di immagini includenti l'oggetto bersaglio diventi minore di una distanza tra il valore di caratteristica generato da un'immagine includente l'oggetto bersaglio e un valore di caratteristica generato da un'immagine includente un oggetto diverso dall'oggetto bersaglio.

In un aspetto della presente invenzione, il modello di apprendimento automatico può essere addestrato mediante dati di addestramento includenti: la pluralità di immagini di addestramento rese dal modello di forma tridimensionale dell'oggetto bersaglio; e i dati di verità di base indicanti le posizioni di punti chiave dell'oggetto bersaglio nelle immagini di addestramento, i mezzi di stima di posa possono essere configurati per acquisire informazioni indicanti le posizioni bidimensionali dei punti chiave dell'oggetto bersaglio nella regione bersaglio immettendo la regione bersaglio acquisita nel modello di apprendimento automatico, e i mezzi di stima di posa possono essere configurati per stimare la posa dell'oggetto bersaglio in base alle informazioni indicanti le posizioni bidimensionali dei punti chiave e le informazioni indicanti le posizioni tridimensionali dei punti chiave nel modello di forma

tridimensionale.

In un aspetto della presente invenzione, il modello di apprendimento automatico può essere addestrato da una pluralità di immagini di addestramento rese da un modello tridimensionale dell'oggetto bersaglio e immagini di verità di base in cui ciascuno dei pixel indica una relazione di posizione rispetto al punto chiave dell'oggetto bersaglio nelle immagini di addestramento, i mezzi di stima di posa possono essere configurati per acquisire un'immagine di posizione in cui ciascuno dei pixel indica la relazione di posizione rispetto al punto chiave dell'oggetto bersaglio immettendo la regione bersaglio acquisita nel modello di apprendimento automatico, i mezzi di stima di posa possono essere configurati per calcolare, in base all'immagine di posizione, la posizione del punto chiave dell'oggetto bersaglio nell'immagine di posizione, e i mezzi di stima di posa possono essere configurati per stimare la posa dell'oggetto bersaglio in base alla posizione calcolata del punto chiave nell'immagine di posizione e nel modello tridimensionale.

In un aspetto della presente invenzione, i mezzi di acquisizione di regione bersaglio possono essere configurati per generare un'immagine di maschera per mascherare una regione diversa dall'immagine dell'oggetto bersaglio nella regione bersaglio, i mezzi di stima di posa possono essere

configurati per mascherare una parte dell'immagine di posizione in base all'immagine di maschera, e i mezzi di stima di posa possono essere configurati per acquisire, in base all'immagine di posizione mascherata, la posizione del punto chiave dell'oggetto bersaglio nell'immagine di posizione.

In un aspetto della presente invenzione, il sistema di elaborazione di informazione può includere inoltre: mezzi di acquisizione di immagini fotografate per acquisire una pluralità di immagini fotografate ottenute fotografando da una pluralità di direzioni rispetto all'oggetto bersaglio; mezzi di generazione di modello di forma per calcolare il modello di forma tridimensionale dell'oggetto bersaglio in base alla pluralità di immagini fotografate; e mezzi di addestramento di classificazione per addestrare, mediante dati di addestramento includenti i dati di verità di base e i dati immessi corrispondenti alla pluralità di immagini fotografate, il modello di classificazione per determinare se l'immagine immessa include o meno l'immagine dell'oggetto bersaglio.

In un aspetto della presente invenzione, i mezzi di generazione di modello di forma possono essere configurati per generare informazioni indicanti la posa dell'oggetto bersaglio nell'immagine fotografata, e il sistema di elaborazione di informazione può includere inoltre mezzi di

addestramento di posa per addestrare il modello di apprendimento automatico mediante: le immagini di addestramento includenti una pluralità di immagini rese dal modello di forma tridimensionale e la pluralità di immagini fotografate; e i dati di verità di base che sono le informazioni relative alla posa dell'oggetto bersaglio nelle immagini di addestramento.

In un aspetto della presente invenzione, il sistema di elaborazione di informazione può includere inoltre mezzi di generazione di dati di addestramento per acquisire, dalla pluralità di immagini fotografate ottenute fotografando, un'immagine fotografata da cui è stata rimossa un'immagine di una mano, e generare un'immagine di addestramento includente l'immagine fotografata da cui è stata rimossa l'immagine della mano e i dati di verità di base.

In un aspetto della presente invenzione, il modello di apprendimento automatico può includere una rete neurale formata da una pluralità di strati inclusi uno strato di immissione, uno strato intermedio e uno strato di emissione, e il modello di apprendimento automatico può essere addestrato in anticipo mediante dati di addestramento preliminare includenti una pluralità di immagini di addestramento relative ad un oggetto diverso dall'oggetto bersaglio e i dati di verità di base, e quindi può essere addestrato mediante la pluralità di immagini di

addestramento relative all'oggetto bersaglio e i dati di verità di base con un parametro fisso per uno strato o una pluralità di strati escluso lo strato di emissione e incluso lo strato di immissione.

Inoltre, secondo la presente invenzione, è fornito un metodo di elaborazione di informazione includente le fasi di: acquisire un'immagine immessa; determinare se l'immagine immessa include o meno un'immagine di un oggetto bersaglio immettendo almeno una parte dell'immagine immessa in un modello di classificazione addestrato in base ad una pluralità di immagini di apprendimento includente un'immagine in cui l'oggetto bersaglio è stato fotografato e dati di etichetta che indicano se ciascuna della pluralità di immagini di apprendimento include o meno l'oggetto bersaglio; acquisire una regione bersaglio includente l'immagine dell'oggetto bersaglio, che è estratta dall'immagine immessa, quando l'immagine immessa include l'oggetto bersaglio; e stimare una posa dell'oggetto bersaglio in base a informazioni emesse da un modello di apprendimento automatico, quando la regione bersaglio acquisita è immessa in esso, che è addestrato mediante: una pluralità di immagini di addestramento rese da un modello di forma tridimensionale dell'oggetto bersaglio; e dati di verità di base che sono informazioni relative alla posa dell'oggetto bersaglio nelle immagini di addestramento.

Inoltre, secondo la presente invenzione, è fornito un programma per fare in modo che un computer esegua i processi di: acquisire un'immagine immessa; determinare se l'immagine immessa include o meno un'immagine di un oggetto bersaglio immettendo almeno una parte dell'immagine immessa in un modello di classificazione addestrato in base ad una pluralità di immagini di addestramento includente un'immagine in cui l'oggetto bersaglio è stato fotografato e dati di etichetta che indicano se ciascuna della pluralità di immagini di apprendimento include o meno l'oggetto bersaglio; acquisire una regione bersaglio includente l'immagine dell'oggetto bersaglio, che è estratta dall'immagine immessa, quando l'immagine immessa include l'oggetto bersaglio; e stimare una posa dell'oggetto bersaglio in base a informazioni emesse da un modello di apprendimento automatico, quando la regione bersaglio acquisita è immessa in esso, che è addestrato mediante: una pluralità di immagini di addestramento rese da un modello di forma tridimensionale dell'oggetto bersaglio; e dati di verità di base che sono informazioni relative alla posa dell'oggetto bersaglio nelle immagini di addestramento.

[Effetti dell'invenzione]

Secondo la presente invenzione, è possibile migliorare la precisione di stima della posa dell'oggetto fotografato.

[Breve descrizione dei disegni]

[Figura 1] Un diagramma per illustrare un esempio di una configurazione di un sistema di elaborazione di informazione secondo una forma di realizzazione della presente invenzione.

[Figura 2] Un diagramma a blocchi funzionali per illustrare un esempio di funzioni implementate nel sistema di elaborazione di informazione secondo l'una forma di realizzazione della presente invenzione.

[Figura 3] Una vista per illustrare un esempio di un'immagine immessa.

[Figura 4] Una vista per illustrare un esempio di punti chiave di un oggetto bersaglio.

[Figura 5] Un diagramma per illustrare schematicamente un esempio di un'immagine di posizione in una regione bersaglio.

[Figura 6] Un diagramma di flusso per illustrare un esempio di elaborazione di un'unità di acquisizione di regione bersaglio e un'unità di stima di posa.

[Figura 7] Un diagramma per illustrare un esempio di una configurazione di un modello di classificazione.

[Figura 8] Un diagramma per illustrare un esempio di un'immagine di maschera per la regione bersaglio.

[Figura 9] Una vista per illustrare una posa dell'oggetto bersaglio che è stato rilevato.

[Figura 10] Un diagramma di flusso per illustrare un

esempio di elaborazione per addestrare un modello di apprendimento automatico e un classificatore.

[Figura 11] Una vista per illustrare lo scatto di fotografie dell'oggetto bersaglio.

[Figura 12] Un diagramma per illustrare un esempio di una configurazione del modello di apprendimento automatico.

[Figura 13] Un diagramma per illustrare una configurazione del modello di apprendimento automatico.

[Forma di realizzazione per realizzare l'invenzione]

Viene ora descritta in dettaglio con riferimento ai disegni una forma di realizzazione della presente invenzione. In questa forma di realizzazione, è fornita la descrizione di un sistema di elaborazione di informazione per immettere un'immagine in cui è stato fotografato un oggetto e per stimarne una posa. Questo sistema di elaborazione di informazione stima la posa tramite l'uso di un modello di apprendimento automatico addestrato mediante dati immessi in base all'immagine in cui è stato fotografato l'oggetto. In questa forma di realizzazione, il sistema di elaborazione di informazione è configurato inoltre per completare l'addestramento in un breve periodo di tempo. Si suppone che un periodo di tempo richiesto sia, per esempio, di diverse decine di secondi per afferrare e ruotare l'oggetto e di diversi minuti per l'apprendimento automatico.

La figura 1 è un diagramma per illustrare un esempio di una configurazione di un sistema di elaborazione di informazione secondo l'una forma di realizzazione della presente invenzione. Il sistema di elaborazione di informazione secondo questa forma di realizzazione include un dispositivo di elaborazione di informazione 10. Il dispositivo di elaborazione di informazione 10 è, per esempio, un computer quale una console di gioco o un personal computer. Come illustrato nella figura 1, il dispositivo di elaborazione di informazione 10 include, per esempio, un processore 11, un'unità di memorizzazione 12, un'unità di comunicazione 14, un'unità operativa 16, un'unità di visualizzazione 18 e un'unità fotografica 20. Il sistema di elaborazione di informazione può essere formato da un unico dispositivo di elaborazione di informazione 10 o può essere formato da una pluralità di dispositivi includenti il dispositivo di elaborazione di informazione 10.

Il processore 11 è, per esempio, un dispositivo di controllo di programma, quale una CPU, che funziona secondo un programma installato nel dispositivo di elaborazione di informazione 10.

L'unità di memorizzazione 12 è formata da almeno una parte di un elemento di memorizzazione quale una ROM o una RAM e un dispositivo di memorizzazione esterno, per esempio, un'unità a stato solido. L'unità di memorizzazione 12

memorizza, per esempio, un programma che deve essere eseguito dal processore 11.

L'unità di comunicazione 14 è, per esempio, un'interfaccia di comunicazione per una comunicazione mediante fili o comunicazione senza fili, quale una scheda di interfaccia di rete, e scambia dati con un altro computer o un altro terminale tramite una rete di computer quale Internet.

L'unità operativa 16 è, per esempio, un dispositivo di immissione quale una tastiera, un mouse, un pannello sensibile al tocco o un controller della console di gioco, e riceve l'immissione di un'operazione effettuata da un utente ed emette segnale indicante i suoi dettagli al processore 11.

L'unità di visualizzazione 18 è un dispositivo di visualizzazione, per esempio, uno schermo a cristalli liquidi, e mostra varie immagini secondo le istruzioni fornite dal processore 11. L'unità di visualizzazione 18 può essere un dispositivo per emettere un segnale video ad un dispositivo di visualizzazione esterno.

L'unità fotografica 20 è un dispositivo fotografico, per esempio, una fotocamera digitale. L'unità fotografica 20 in questa forma di realizzazione è una fotocamera in grado di fotografare un'immagine in movimento. L'unità fotografica 20 può essere una fotocamera in grado di acquisire

un'immagine RGB visibile e informazioni di profondità sincronizzate con l'immagine RGB. L'unità fotografica 20 può essere fornita all'esterno del dispositivo di elaborazione di informazione 10 e, in questo caso, il dispositivo di elaborazione di informazione 10 e l'unità fotografica 20 possono essere collegati tra loro tramite l'unità di comunicazione 14 o un'unità di immissione/emissione descritta in seguito.

Il dispositivo di elaborazione di informazione 10 può includere un dispositivo di immissione/emissione audio quale un microfono o un altoparlante. In aggiunta, il dispositivo di elaborazione di informazione 10 può includere, per esempio, un'interfaccia di comunicazione quale una scheda di rete, un'unità di disco ottico per leggere un disco ottico quale un disco DVD-ROM o un Blu-ray (nome commerciale), e l'unità di immissione/emissione (porta di bus seriale universale (USB)) per immettere/emettere dati in/da un dispositivo esterno.

La figura 2 è un diagramma a blocchi funzionali per illustrare un esempio di funzioni implementate nel sistema di elaborazione di informazione secondo l'una forma di realizzazione della presente invenzione. Come illustrato nella figura 2, il sistema di elaborazione di informazione include funzionalmente una unità di acquisizione di regione bersaglio 21, un'unità di stima di posa 25, un'unità di

acquisizione di immagini fotografate 35, un'unità di generazione di modello di forma 36, un'unità di generazione di dati di addestramento 37, un'unità di addestramento di posa 38 e un'unità di addestramento di classificazione 39. L'unità di acquisizione di regione bersaglio 21 include funzionalmente un'unità di estrazione di regione 22 e un modello di classificazione 23. L'unità di stima di posa 25 include funzionalmente un modello di apprendimento automatico 26, un'unità di determinazione di punto chiave 27 e un'unità di calcolo di posa 28.

Queste funzioni sono implementate principalmente dal processore 11 e dall'unità di memorizzazione 12. Più nello specifico, tali funzioni possono essere implementate dal processore 11 che esegue un programma che è installato nel dispositivo di elaborazione di informazione 10 che è un computer, e include istruzioni di esecuzione corrispondenti alle funzioni summenzionate. In aggiunta, questo programma può essere fornito al dispositivo di elaborazione di informazione 10, per esempio, tramite l'intermediazione di un supporto di memorizzazione di informazione leggibile da computer quale un disco ottico, un disco magnetico o una memoria flash o tramite Internet o simili.

Non è necessario che tutte le funzioni illustrate nella figura 2 siano implementate nel sistema di elaborazione di informazione secondo la presente forma di realizzazione, e

nella presente possono essere implementate funzioni diverse dalle funzioni illustrate nella figura 2.

L'unità di acquisizione di regione bersaglio 21 acquisisce un'immagine immessa ottenuta fotografando mediante l'unità fotografica 20 e immette almeno una parte dell'immagine immessa nel modello di classificazione 23, per determinare così se l'immagine immessa include o meno un'immagine di un oggetto bersaglio 51. In aggiunta, quando l'immagine immessa include l'oggetto bersaglio 51, viene acquisita una regione bersaglio 55 includente l'immagine dell'oggetto bersaglio 51, che viene estratta dall'immagine immessa. L'oggetto bersaglio 51 è un oggetto che funge da bersaglio da sottoporre a stima di posa nel dispositivo di elaborazione di informazione 10. L'oggetto bersaglio 51 è un bersaglio di apprendimento preliminare.

La figura 3 è una vista per illustrare un esempio dell'immagine immessa. Nell'esempio della figura 3, l'oggetto bersaglio 51 è un utensile elettrico e, anche nei successivi disegni, si suppone che l'esempio dell'oggetto bersaglio 51 sia l'utensile elettrico, salvo diversamente specificato. L'immagine immessa è stata ottenuta fotografando mediante l'unità fotografica 20, e la regione bersaglio 55 è una regione rettangolare includente l'oggetto bersaglio 51 e le sue prossimità. In un processo di acquisizione della regione bersaglio 55, una o una pluralità

di regioni candidate 56 includenti la regione che non include l'oggetto bersaglio 51 vengono anch'esse estratte come candidate per una regione includente l'oggetto bersaglio 51. I dettagli della regione candidata 56 sono descritti nel seguito.

Quando il modello di classificazione 23 riceve l'immissione di un'immagine, il modello di classificazione 23 emette informazioni che indicano se l'immagine include o meno l'immagine dell'oggetto bersaglio 51. Il modello di classificazione 23 è addestrato mediante dati di addestramento includenti una pluralità di immagini di apprendimento includente un'immagine in cui l'oggetto bersaglio 51 è stato fotografato e dati di etichetta che indicano se ciascuna delle immagini di apprendimento include o meno l'oggetto bersaglio 51. I dettagli del modello di classificazione 23 sono descritti nel seguito.

L'unità di estrazione di regione 22 estrae dall'immagine immessa un'immagine da immettere nel modello di classificazione 23. Più nello specifico, l'unità di estrazione di regione 22 identifica dall'immagine immessa una o una pluralità di regioni candidate 56 in cui è stato fotografato un qualche oggetto mediante una tecnologia di proposta di regione nota, e ciascuna dell'una o della pluralità di regioni candidate 56 viene estratta come immagine da immettere nel modello di classificazione 23.

L'immagine di apprendimento per il modello di classificazione 23 può essere un'immagine di una regione parziale in cui l'oggetto bersaglio 51 è presente nell'immagine ottenuta fotografando, secondo l'unità di estrazione di regione 22. L'unità di estrazione di regione 22 può essere omessa e l'immagine immessa può essere immessa direttamente nel modello di classificazione 23.

L'unità di stima di posa 25 stima la posa dell'oggetto bersaglio 51 in base a informazioni emesse quando la regione bersaglio 55 è immessa nel modello di apprendimento automatico 26. Il modello di apprendimento automatico 26 è addestrato mediante una pluralità di immagini di apprendimento rese da un modello di forma tridimensionale dell'oggetto bersaglio 51 e dati di verità di base che sono informazioni relative alla posa dell'oggetto bersaglio 51 nelle immagini di apprendimento.

Più nello specifico, il modello di apprendimento automatico 26 è addestrato mediante i dati di addestramento includenti una pluralità di immagini di apprendimento rese dal modello di forma tridimensionale dell'oggetto bersaglio 51 i dati di verità di base indicanti le posizioni di punti chiave dell'oggetto bersaglio 51 nelle immagini di apprendimento. Il punto chiave è un punto virtuale nell'oggetto bersaglio 51 ed è un punto da utilizzare per calcolare la posa.

La figura 4 è una vista per illustrare un esempio di punti chiave dell'oggetto bersaglio 51. Le posizioni tridimensionali dei punti chiave dell'oggetto bersaglio 51 sono determinate dal modello di forma tridimensionale dell'oggetto bersaglio 51 (più nello specifico, informazioni sui vertici inclusi nel modello di forma tridimensionale) mediante, per esempio, un algoritmo di punto più lontano noto. Nella figura 4 sono mostrati per semplicità descrittiva tre punti chiave da K1 a K3, ma il numero effettivo dei punti chiave può essere maggiore. Per esempio, in questa forma di realizzazione, il numero effettivo di punti chiave dell'oggetto bersaglio 51 è otto.

Il modello di apprendimento automatico 26 che è stato addestrato emette informazioni indicanti le posizioni bidimensionali dei punti chiave dell'oggetto bersaglio 51 nella regione bersaglio 55 quando il modello di apprendimento automatico 26 riceve un'immissione della regione bersaglio 55. Dalle posizioni bidimensionali dei punti chiave nella regione bersaglio 55 e dalla posizione della regione bersaglio 55 nell'immagine immessa, si ottengono le posizioni bidimensionali dei punti chiave nell'immagine immessa.

I dati indicanti la posizione del punto chiave possono essere un'immagine di posizione in cui ciascuno dei pixel indica una relazione di posizione (per esempio, direzione)

tra ciascuno dei pixel e il punto chiave, possono essere un'immagine, per esempio, una mappa di calore, indicante la posizione del punto chiave o possono essere coordinate di ciascun punto chiave stesso. È fornita principalmente la seguente descrizione di un caso in cui i dati indicanti la posizione del punto chiave sono l'immagine in cui ciascuno dei pixel indica la relazione di posizione tra ciascuno dei pixel e il punto chiave.

La figura 5 è un diagramma per illustrare schematicamente un esempio dell'immagine di posizione nella regione bersaglio 55. L'immagine di posizione può essere generata per ogni tipo di punto chiave. L'immagine di posizione indica una direzione relativa in corrispondenza di ciascuno dei pixel, tra ciascuno dei pixel e il punto chiave. Nell'immagine di posizione illustrata nella figura 5, è illustrato un pattern corrispondente ad un valore di ciascuno dei pixel e il valore di ciascuno dei pixel indica la direzione tra le coordinate di ciascuno dei pixel e le coordinate del punto chiave. La figura 5 è soltanto un diagramma schematico e un valore effettivo di ogni pixel cambia in modo continuo. Sebbene non esplicitamente illustrato nella figura, l'immagine di posizione è un'immagine di campo di vettori indicante una direzione relativa del punto chiave in corrispondenza di ogni pixel, ogni pixel essendo utilizzato come riferimento.

L'unità di determinazione di punto chiave 27 determina la posizione bidimensionale del punto chiave nella regione bersaglio 55 e l'immagine immessa in base all'emissione del modello di apprendimento automatico 26. Più nello specifico, per esempio, l'unità di determinazione di punto chiave 27 calcola candidati per la posizione bidimensionale del punto chiave nella regione bersaglio 55 in base all'immagine di posizione emessa dal modello di apprendimento automatico 26 e determina la posizione bidimensionale del punto chiave nell'immagine immessa dai candidati calcolati per la posizione bidimensionale. Per esempio, l'unità di determinazione di punto chiave 27 calcola un punto candidato per il punto chiave da ogni combinazione di due punti selezionati liberamente nell'immagine di posizione e genera, per una pluralità di punti candidato, un punteggio che indica se vi è o meno una corrispondenza con una direzione indicata da ciascuno dei pixel nell'immagine di posizione. L'unità di determinazione di punto chiave 27 può stimare un punto candidato avente il punteggio massimo come la posizione del punto chiave. L'unità di determinazione di punto chiave 27 ripete inoltre l'elaborazione summenzionata per ogni punto chiave.

L'unità di calcolo di posa 28 stima la posa dell'oggetto bersaglio 51 in base alle informazioni indicanti le posizioni bidimensionali dei punti chiave nell'immagine immessa e a

informazioni indicanti le posizioni tridimensionali dei punti chiave nel modello di forma tridimensionale dell'oggetto bersaglio 51 ed emette dati di posa indicanti la posa stimata. La posa dell'oggetto bersaglio 51 è stimata mediante un algoritmo noto. Per esempio, la posa dell'oggetto bersaglio 51 può essere stimata mediante una soluzione a un problema di n punti di prospettiva (PNP, Perspective-n-Point) relativo alla stima di posa (per esempio, EPnP). In aggiunta, l'unità di calcolo di posa 28 può stimare non solo la posa dell'oggetto bersaglio 51 ma anche la posizione dell'oggetto bersaglio 51 nell'immagine immessa, e i dati di posa possono includere informazioni indicanti la posizione.

Il modello di apprendimento automatico 26, l'unità di determinazione di punto chiave 27 e l'unità di calcolo di posa 28 possono essere quelli descritti nell'articolo "PVNet: Pixe-Wise Voting Network for 6DoF Pose Estimation".

Quando il modello di apprendimento automatico 26 riceve l'immissione dell'immagine bersaglio, il modello di apprendimento automatico 26 può emettere direttamente la posa dell'oggetto bersaglio 51. In questo caso, l'unità di determinazione di punto chiave 27 non è richiesta e l'unità di calcolo di posa 28 ottiene la posa e la posizione dell'oggetto bersaglio 51 nell'immagine immessa in base alla posa e alla posizione dell'oggetto bersaglio 51 che sono state calcolate per la regione bersaglio 55.

In questa forma di realizzazione, l'oggetto bersaglio 51 viene fotografato e un classificatore 32 e il modello di apprendimento automatico 26 vengono addestrati in un breve periodo di tempo, per esempio, diversi secondi e diversi minuti, rispettivamente, in base all'immagine in cui è stato fotografato l'oggetto bersaglio 51. L'unità di acquisizione di immagini fotografate 35, l'unità di generazione di modello di forma 36, l'unità di generazione di dati di addestramento 37, l'unità di addestramento di posa 38 e l'unità di addestramento di classificazione 39 sono configurate per essere utilizzate per il loro addestramento.

L'unità di acquisizione di immagini fotografate 35 acquisisce immagini fotografate in cui l'oggetto bersaglio 51 è stato fotografato dall'unità fotografica 20 per addestrare il modello di apprendimento automatico 26 incluso nell'unità di stima di posa 25 e/o il modello di classificazione 23 incluso nell'unità di acquisizione di regione bersaglio 21. Si suppone che l'unità fotografica 20 abbia acquisito un parametro intrinseco di fotocamera mediante una calibrazione in anticipo. Questo parametro viene utilizzato per risolvere il problema di PnP.

L'unità di generazione di modello di forma 36, l'unità di generazione di dati di addestramento 37 e l'unità di addestramento di posa 38 eseguono un'elaborazione per addestrare il modello di apprendimento automatico 26 incluso

nell'unità di stima di posa 25.

Più nello specifico, l'unità di generazione di modello di forma 36 estrae una pluralità di vettori di caratteristica indicanti le caratteristiche locali per ciascuna della pluralità di immagini fotografate dell'oggetto bersaglio 51 acquisite dall'unità di acquisizione di immagini fotografate 35. Quindi, da una pluralità di vettori di caratteristica corrispondenti tra loro, che sono stati estratti dalla pluralità di immagini fotografate, e dalle posizioni in cui i vettori di caratteristica sono stati estratti nelle immagini fotografate, l'unità di generazione di modello di forma 36 ottiene le posizioni tridimensionali di punti in corrispondenza dei quali sono stati estratti i vettori di caratteristica e acquisisce il modello di forma tridimensionale dell'oggetto bersaglio 51 in base alle posizioni tridimensionali. Questo metodo è un metodo noto che viene utilizzato anche in un software che implementa un cosiddetto SfM o Visual SLAM e quindi una sua descrizione dettagliata è omessa.

L'unità di generazione di dati di addestramento 37 genera dati di addestramento per addestrare il modello di apprendimento automatico 26. Più nello specifico, l'unità di generazione di dati di addestramento 37 genera dati di addestramento includenti un'immagine di addestramento resa e dati di verità di base indicanti le posizioni dei punti

chiave dal modello di forma tridimensionale dell'oggetto bersaglio 51. L'unità di generazione di dati di addestramento 37 genera anche: dati di verità di base dalla posa dell'oggetto bersaglio 51 ottenuti mediante un metodo DLT o simili quando il modello di forma tridimensionale viene calcolato dalle immagini fotografate; e dati di addestramento formati di una regione includente l'oggetto bersaglio 51 dalle immagini fotografate, e aggiunge i dati di verità di base e i dati di addestramento ai dati di addestramento.

L'unità di addestramento di posa 38 addestra il modello di apprendimento automatico 26 incluso nell'unità di stima di posa 25 mediante i dati di addestramento generati dall'unità di generazione di dati di addestramento 37.

L'unità di addestramento di classificazione 39 addestra il modello di classificazione 23 incluso nell'unità di acquisizione di regione bersaglio 21 in base alle immagini fotografate acquisite dall'unità di acquisizione di immagini fotografate 35. I dettagli dell'unità di addestramento di classificazione 39 sono descritti nel seguito.

Di seguito, è descritta un'elaborazione relativa alla stima di posa. La figura 6 è un diagramma di flusso per illustrare un esempio di elaborazione dell'unità di acquisizione di regione bersaglio 21 e dell'unità di stima di posa 25. L'elaborazione illustrata nella figura 6 può

essere eseguita ripetutamente in modo regolare.

In primo luogo, l'unità di estrazione di regione 22 inclusa nell'unità di acquisizione di regione bersaglio 21 acquisisce un'immagine immessa ottenuta fotografando mediante l'unità fotografica 20 (S101). L'unità di estrazione di regione 22 può acquisire un'immagine immessa ricevendo direttamente l'immagine immessa dall'unità fotografica 20 o può acquisire un'immagine immessa ricevuta dall'unità fotografica 20 e memorizzata nell'unità di memorizzazione 12.

L'unità di estrazione di regione 22 estrae dall'immagine immessa una o una pluralità di regioni candidate 56 in cui un qualche oggetto è stato fotografato (S102). L'unità di estrazione di regione 22 può includere una rete di proposte di regione (RPN, Regional Proposal Network) addestrata in anticipo. La RPN può essere addestrata mediante dati di addestramento irrilevanti per l'immagine in cui è stato fotografato l'oggetto bersaglio 51. Questa elaborazione riduce lo spreco computazionale e garantisce un certo grado di robustezza ad un ambiente.

In questo caso, l'unità di estrazione di regione 22 può inoltre eseguire un'elaborazione di immagine incluse un'elaborazione di rimozione di sfondo (elaborazione di maschera) e una regolazione di dimensione sull'immagine della regione candidata 56 estratta. In aggiunta, l'immagine

elaborata della regione candidata 56 può essere utilizzata per la successiva elaborazione. Con questa elaborazione, è possibile ridurre un gap di dominio dovuto alle condizioni di sfondo e illuminazione e addestrare il classificatore 32 mediante una quantità ridotta di dati di addestramento.

L'unità di acquisizione di regione bersaglio 21 determina se ciascuna delle regioni candidate 56 include o meno l'immagine dell'oggetto bersaglio 51 (S103). Questa elaborazione include un'elaborazione per acquisire l'emissione ottenuta quando l'unità di acquisizione di regione bersaglio 21 immette l'immagine della regione candidata 56 nel modello di classificazione 23.

La figura 7 è un diagramma per illustrare un esempio del modello di classificazione 23. Il modello di classificazione 23 include un'unità di generazione di caratteristica 31 e il classificatore 32.

L'unità di generazione di caratteristica 31 emette, dall'immagine della regione candidata 56, un valore di caratteristica corrispondente all'immagine. L'unità di generazione di caratteristica 31 include una rete neurale convoluzionale (CNN, Convolutional Neural Network) che è stata addestrata. Questa CNN emette, in risposta all'immissione di un'immagine, dati di valore di caratteristica (dati di valore di caratteristica immessi) indicanti un valore di caratteristica corrispondente

all'immagine. L'unità di generazione di caratteristica 31 può estrarre il valore di caratteristica dall'immagine della regione candidata 56 estratta mediante l'RPN, o può acquisire il valore di caratteristica estratto nell'elaborazione dell'RPN, per esempio, una R-CNN più veloce.

Il classificatore 32 è, per esempio, una macchina a vettori di supporto (SVM, Support Vector Machine) ed è un tipo di modello di apprendimento automatico. In risposta all'immissione dei dati di valore di caratteristica immessi indicanti il valore di caratteristica corrispondente all'immagine della regione candidata 56, il classificatore 32 emette un punteggio di discriminazione indicante una probabilità che un oggetto fotografato nella regione candidata 56 appartenga ad una classe positiva nel classificatore 32. Il classificatore 32 è addestrato mediante una pluralità di dati di addestramento di esempio positivo relativi ad esempi positivi e una pluralità di dati di addestramento di esempio negativo relativi ad esempi negativi. I dati di addestramento di esempio positivo sono generati da un'immagine di apprendimento includente l'immagine in cui è stato fotografato l'oggetto bersaglio 51, e i dati di addestramento di esempio negativo sono generati da un'immagine di un oggetto diverso dall'oggetto bersaglio 51, l'immagine essendo fornita in anticipo. I dati di addestramento di esempio negativo possono essere generati

fotografando un ambiente dell'unità fotografica 20, che è stato fotografato dall'unità fotografica 20.

In questo caso, l'apprendimento di metrica può essere eseguito in anticipo sulla CNN dell'unità di generazione di caratteristica 31. È possibile eseguire una regolazione mediante un apprendimento di metrica preliminare in modo tale che i dati di valore di caratteristica indicanti valori di caratteristica vicini tra loro siano emessi dalle immagini in cui vengono fotografati oggetti appartenenti alla classe positiva nel classificatore 32. Il valore di caratteristica indicato dai dati di valore di caratteristica in questa forma di realizzazione è, per esempio, una quantità di vettore normalizzata in modo da avere una norma di 1. Questo apprendimento di metrica può essere eseguito prima che l'immagine dell'oggetto bersaglio 51 sia ottenuta fotografando.

In questa forma di realizzazione, questa CNN viene utilizzata per generare dati di valore di caratteristica indicanti i valori di caratteristica corrispondenti alle immagini sottoposte all'esecuzione di un'elaborazione di normalizzazione. La CNN sottoposta all'esecuzione di un apprendimento di metrica in anticipo viene utilizzata, per aggregare in questo modo i valori di caratteristica di campioni appartenenti ad un'unica classe in una regione compatta indipendentemente dalle condizioni. Di conseguenza,

il dispositivo di elaborazione di informazione 10 in questa forma di realizzazione può determinare un confine di discriminazione appropriato nel classificatore 32 anche da un numero ridotto di campioni. L'unità di generazione di caratteristica 31 può emettere, mediante un altro algoritmo noto per calcolare il valore di caratteristica indicante la caratteristica dell'immagine, i dati di valore di caratteristica indicanti il valore di caratteristica corrispondente ad un'immagine in risposta all'immissione dell'immagine.

Per esempio, quando il punteggio di discriminazione è maggiore di un valore di soglia, l'unità di acquisizione di regione bersaglio 21 determina che la regione candidata 56 rilevante include l'immagine dell'oggetto bersaglio 51.

Dopo che è stato determinato se ciascuna regione candidata 56 include o meno l'immagine dell'oggetto bersaglio 51, l'unità di acquisizione di regione bersaglio 21 determina la regione bersaglio 55 in base ai risultati della determinazione (S104). Più nello specifico, l'unità di acquisizione di regione bersaglio 21 acquisisce, in base alla regione candidata 56 che, secondo quanto determinato, include l'oggetto bersaglio 51, una regione rettangolare includente una regione in prossimità dell'oggetto bersaglio 51 come regione bersaglio 55. L'unità di acquisizione di regione bersaglio 21 può acquisire una regione quadrata

includente una regione in prossimità dell'oggetto bersaglio 51 come regione bersaglio 55 o può semplicemente acquisire la regione candidata 56 come regione bersaglio 55. Non è sempre richiesto che l'unità di acquisizione di regione bersaglio 21 acquisisca la regione bersaglio 55 mediante le fasi di elaborazione di S102 e S103. Per esempio, l'unità di acquisizione di regione bersaglio 21 può acquisire la regione bersaglio 55 eseguendo un'elaborazione di tracciamento di serie temporale sull'immagine immessa acquisita dopo che la regione bersaglio 55 è stata acquisita una volta.

In questo caso, l'unità di acquisizione di regione bersaglio 21 genera un'immagine di maschera per mascherare un'immagine diversa dall'immagine dell'oggetto bersaglio 51 nella regione bersaglio 55. L'unità di acquisizione di regione bersaglio 21 può utilizzare un metodo noto per generare un'immagine di maschera per mascherare lo sfondo dall'immagine della regione bersaglio 55. L'immagine da immettere nel modello di classificazione 23 può essere elaborata in modo da escludere l'immagine dello sfondo, e l'unità di acquisizione di regione bersaglio 21 può anche generare, durante l'elaborazione, un'immagine di maschera regolando una dimensione o simili dell'immagine di maschera generata mediante un metodo noto.

La figura 8 è un diagramma per illustrare un esempio di un'immagine di maschera per la regione bersaglio 55.

Nell'esempio della figura 8, una regione corrispondente all'immagine dell'oggetto bersaglio 51 è mostrata in bianco. L'immagine di maschera viene utilizzata nell'elaborazione relativa alla generazione dell'immagine di posizione eseguita dal modello di apprendimento automatico 26.

L'unità di stima di posa 25 immette un'immagine della regione bersaglio 55 nel modello di apprendimento automatico 26 che è stato addestrato (S105). L'immagine della regione bersaglio 55 immessa in questa fase può essere un'immagine avente la dimensione regolata (ingrandita o ridotta) secondo la dimensione dell'immagine immessa del modello di apprendimento automatico 26. Tramite la regolazione di dimensione (normalizzazione), viene migliorata l'efficienza di apprendimento del modello di apprendimento automatico 26.

L'unità di stima di posa 25 può mascherare lo sfondo dell'immagine della regione bersaglio 55 tramite l'uso dell'immagine di maschera e immettere l'immagine della regione bersaglio 55 avente lo sfondo mascherato nel modello di apprendimento automatico 26. L'unità di stima di posa 25 può anche mascherare l'immagine di posizione emessa dal modello di apprendimento automatico 26 tramite l'uso dell'immagine di maschera. In quest'ultimo metodo di mascheratura, è possibile impedire un'influenza della stima del punto chiave sullo sfondo generando al contempo l'immagine di posizione tramite l'uso di un'immagine vicino

ad un confine tra lo sfondo e l'oggetto bersaglio 51. Questo migliora la precisione della stima di punti chiave. È possibile inoltre mascherare sia l'immagine da immettere nel modello di apprendimento automatico 26 sia l'immagine da emettere dallo stesso.

L'unità di determinazione di punto chiave 27 inclusa nell'unità di stima di posa 25 determina le posizioni bidimensionali dei punti chiave nella regione bersaglio 55 e l'immagine immessa in base all'emissione del modello di apprendimento automatico 26 (S106). Nel caso in cui l'emissione del modello di apprendimento automatico 26 sia l'immagine di posizione, l'unità di determinazione di punto chiave 27 calcola un candidato per la posizione del punto chiave da ogni pixel nell'immagine di posizione e determina la posizione del punto chiave in base al candidato. Nel caso in cui l'emissione del modello di apprendimento automatico 26 sia la posizione del punto chiave nella regione bersaglio 55, la posizione del punto chiave nell'immagine immessa può essere calcolata da tale posizione. Le fasi di elaborazione di S105 e S106 vengono eseguite per ogni tipo di punto chiave.

L'unità di calcolo di posa 28 inclusa nell'unità di stima di posa 25 calcola la posa dell'oggetto bersaglio 51 da stimare in base alle posizioni bidimensionali determinate dei punti chiave (S107). L'unità di calcolo di posa 28 può

calcolare la posizione dell'oggetto bersaglio 51 insieme alla posa. La posa e la posizione possono essere calcolate mediante la soluzione summenzionata del problema di PNP.

La figura 9 è una vista per illustrare una posa dell'oggetto bersaglio 51 che è stato rilevato. Nella figura 9, per semplicità di descrizione, la posa dell'oggetto bersaglio 51 è rappresentata da assi di coordinate locali 59 indicanti un sistema di coordinate locali dell'oggetto bersaglio 51. La posizione dell'origine degli assi di coordinate locali 59 indica la posizione dell'oggetto bersaglio 51 e le direzioni delle linee degli assi di coordinate locali 59 indicano la posa.

La posa e la posizione stimate dell'oggetto bersaglio 51 possono essere utilizzate per vari scopi. Per esempio, la posa e la posizione possono essere immesse in software applicativo, per esempio, un gioco, al posto di informazioni operative immesse mediante il controller. Quindi, il processore 11 che esegue un codice di esecuzione del software applicativo può generare dati sull'immagine in base alla posa (e alla posizione) e fare in modo che l'unità di visualizzazione 18 emetta l'immagine. Il processore 11 può anche fare in modo che il dispositivo di elaborazione di informazione 10 o un dispositivo di emissione audio collegato al dispositivo di elaborazione di informazione 10 emetta un suono in base alla sua posa (e alla sua posizione). In un

altro caso, il processore 11 può, per esempio, notificare ad un agente AI (quale un robot) la posizione e la posa dell'oggetto, per controllare così un'operazione dell'agente AI per fare in modo che l'oggetto venga afferrato.

Di seguito, viene descritta l'elaborazione di addestramento del modello di apprendimento automatico 26 e del classificatore 32. La figura 10 è un diagramma di flusso illustrare un esempio di elaborazione per addestrare il modello di apprendimento automatico 26 e il classificatore 32.

In primo luogo, l'unità di acquisizione di immagini fotografate 35 acquisisce una pluralità di immagini fotografate in cui è stato fotografato l'oggetto bersaglio 51 (S301).

La figura 11 è una vista per illustrare lo scatto di fotografie dell'oggetto bersaglio 51. L'oggetto bersaglio 51 è tenuto, per esempio, da una mano 53 e viene fotografato dall'unità fotografica 20. In questa forma di realizzazione, si desidera fotografare l'oggetto bersaglio 51 da varie direzioni. Di conseguenza, l'unità fotografica 20 cambia una direzione di scatto di fotografie dell'oggetto bersaglio 51 fotografando periodicamente un'immagine come nel caso di una videografia. Per esempio, la direzione di scatto di fotografie dell'oggetto bersaglio 51 può essere cambiata cambiando la posa dell'oggetto bersaglio 51 mediante la mano

53. In un altro caso, la direzione di scatto di fotografie può essere cambiata posizionando l'oggetto bersaglio 51 su un marcatore AR e muovendo l'unità fotografica 20. Un intervallo di acquisizione tra le immagini fotografate da utilizzare nell'elaborazione descritta nel seguito può essere più lungo di un intervallo fotografico della videografia.

Dopo che sono state acquisite le immagini fotografate, l'unità di acquisizione di immagini fotografate 35 maschera l'immagine della mano 53 da tali immagini fotografate (S302). L'immagine della mano 53 può essere mascherata mediante un metodo noto. Per esempio, l'unità di acquisizione di immagini fotografate 35 può mascherare l'immagine della mano 53 rilevando una regione di colore della pelle inclusa nell'immagine fotografata.

Quindi, l'unità di generazione di modello di forma 36 calcola il modello di forma tridimensionale dell'oggetto bersaglio 51 e la posa nelle rispettive immagini fotografate dalla pluralità di immagini fotografate (S303). Questa elaborazione può essere eseguita mediante il metodo noto summenzionato utilizzato anche in software che implementa un cosiddetto SfM o Visual SLAM. L'unità di generazione di modello di forma 36 può calcolare la posa dell'oggetto bersaglio 51 in base ad una logica per calcolare la direzione di scatto di fotografie della fotocamera mediante questo

metodo.

Dopo che è stato calcolato il modello di forma tridimensionale dell'oggetto bersaglio 51, l'unità di generazione di modello di forma 36 determina le posizioni tridimensionali di una pluralità di punti chiave da utilizzare per stimare la posa del modello di forma tridimensionale (S304). L'unità di generazione di modello di forma 36 può determinare le posizioni tridimensionali di una pluralità di punti chiave mediante, per esempio, un algoritmo di punto più lontano noto.

Dopo che sono state calcolate le posizioni tridimensionali dei punti chiave, l'unità di generazione di dati di addestramento 37 genera dati di addestramento includenti una pluralità di immagini di addestramento e una pluralità di immagini di posizione (S305). Più nello specifico, l'unità di generazione di dati di addestramento 37 genera una pluralità di immagini di addestramento rese dal modello di forma tridimensionale e genera immagini di posizione indicanti le posizioni dei punti chiave nella pluralità di immagini di addestramento. La pluralità di immagini di addestramento consiste in immagini rese dell'oggetto bersaglio 51 osservato da una pluralità di direzioni diverse e l'immagine di posizione viene generata per ogni combinazione dell'immagine di addestramento e del punto chiave. L'immagine di addestramento può essere

un'immagine resa sottoposta ad una cosiddetta casualizzazione di dominio o un'immagine simile ad una fotografia resa insieme all'immagine dello sfondo. L'immagine di addestramento può essere un'immagine avente lo sfondo mascherato.

L'unità di generazione di dati di addestramento 37 proietta virtualmente la posizione del punto chiave sull'immagine di addestramento resa e genera un'immagine di posizione in base a una posizione relativa tra la posizione proiettata del punto chiave e ogni pixel nell'immagine. I dati di addestramento da utilizzare per addestrare il modello di apprendimento automatico 26 includono le immagini di addestramento e le immagini di posizione.

L'unità di generazione di dati di addestramento 37 genera inoltre insiemi di un'immagine di addestramento includente almeno una parte dell'immagine fotografata e un'immagine di posizione e aggiunge tali insiemi ai dati di addestramento (S306). L'unità di generazione di dati di addestramento 37 genera un'immagine di posizione in base alla posa e alla posizione dell'oggetto bersaglio 51 in ogni immagine fotografata. La posa e la posizione dell'oggetto bersaglio 51 sono state calcolate dall'unità di generazione di modello di forma 36. Per esempio, l'unità di generazione di dati di addestramento 37 può generare un'immagine resa dell'oggetto bersaglio 51 nell'immagine fotografata

relativamente alla sua posa e alla sua posizione, e generare un'immagine di posizione per l'immagine fotografata in base all'immagine resa mediante lo stesso metodo di S305. L'immagine di addestramento può essere un'immagine ottenuta mascherando lo sfondo dell'immagine fotografata.

Nelle immagini di addestramento incluse nei dati di addestramento, il numero di immagini rese è maggiore del numero di immagini formate da almeno una regione parziale dell'immagine fotografata. Questo è dovuto al fatto che è possibile generare facilmente immagini osservate da varie direzioni di scatto di fotografie tramite l'uso di un modello di forma tridimensionale, mentre è difficile acquisire immagini fotografate ottenute fotografando da varie direzioni di scatto di fotografie in un breve periodo di tempo. Al contempo, le immagini formate da almeno una regione parziale dell'immagine fotografata vengono utilizzate per addestrare il modello di apprendimento automatico 26, per adattare così il modello di apprendimento automatico 26 non solo ad un'immagine resa ma anche ad un'immagine effettivamente fotografata. Questo consente di migliorare la precisione del modello di apprendimento automatico 26 che è stato addestrato. La pluralità di immagini di addestramento incluse nei dati di addestramento può consistere soltanto in immagini rese nonostante la precisione leggermente ridotta.

Quindi, l'unità di addestramento di posa 38 addestra il

modello di apprendimento automatico 26 mediante i dati di addestramento generati in S305 e S306 (S307).

Vengono descritti dettagli del modello di apprendimento automatico 26 e di un metodo per l'apprendimento. La figura 12 è un diagramma per illustrare un esempio di una configurazione del modello di apprendimento automatico 26. Il modello di apprendimento automatico 26 illustrato nella figura 12 è utilizzato al momento dell'apprendimento preliminare. Il modello di apprendimento automatico 26 utilizza un modello basato su ResNet in un'unità di codificatore e include una pluralità di blocchi da 70a a 70l. L'immagine della regione bersaglio 55 è immessa nel blocco 70a e il blocco 70l corrisponde ad uno strato di emissione. I blocchi da 70a a 70g corrispondono ad un cosiddetto codificatore e mostrano una tendenza per cui una convoluzione comporta una diminuzione dell'area (larghezza e altezza) di una mappa di caratteristiche di ogni strato e un aumento del numero di canali. Al contempo, i blocchi da 70h a 70k corrispondono ad un cosiddetto decodificatore e mostrano una tendenza per cui una deconvoluzione comporta un aumento dell'area della mappa di caratteristiche di ogni strato e una diminuzione del numero di canali. Al posto della deconvoluzione, è possibile utilizzare una combinazione di campionamento in aumento (riscalatura bilineare) e convoluzione. Ciascuno dei blocchi da 70a a 70k può avere

strati Conv2D, BatchNorm e di attivazione. I blocchi illustrati nella figura 12 sono di fatto collegati a blocchi diversi dai blocchi adiacenti, ma la descrizione dei loro dettagli è omessa.

La figura 13 è un diagramma per illustrare una configurazione del modello di apprendimento automatico 26. Il modello di apprendimento automatico 26 illustrato nella figura 13 può eseguire l'emissione per una pluralità di tipi di oggetti e viene utilizzato, per esempio, al momento dell'apprendimento di S307 della figura 10 dopo l'apprendimento preliminare. I blocchi da 70d a 70l e i blocchi da 71d a 71l sono presenti nel successivo stadio del blocco 70c. Per esempio, i blocchi da 70d a 70l sono strati per emettere l'immagine di posizione per un tipo di oggetto bersaglio 51 e i blocchi da 71d a 71l sono strati per emettere l'immagine di posizione per un altro tipo di oggetto bersaglio 51.

Prima di eseguire l'elaborazione della figura 10, l'unità di addestramento di posa 38 addestra i blocchi da 70a a 70l del modello di apprendimento automatico 26 illustrato nella figura 12 mediante dati di addestramento preliminare includenti immagini immesse generate per una pluralità di oggetti campione e le immagini di posizione indicanti le posizioni dei punti chiave degli oggetti campione. La pluralità di oggetti campione include un oggetto

diverso dall'oggetto bersaglio 51.

Quindi, in S307, l'unità di addestramento di posa 38 imposta i parametri appresi per i blocchi da 70a a 70l della figura 12 nei blocchi da 70a a 70l della figura 13. I parametri appresi per i blocchi da 70d a 70l della figura 12 possono essere impostati nei blocchi da 71d a 71l della figura 13. L'unità di addestramento di posa 38 fissa anche i parametri per i blocchi da 70a a 70c e addestra il modello di apprendimento automatico 26 mediante i dati di addestramento generati in S305 e S306. In questo addestramento, l'unità di addestramento di posa 38 addestra il modello di apprendimento automatico 26 regolando i parametri per una rete neurale dei blocchi da 70d a 70l (nonché i blocchi da 71d a 71l nell'esempio della figura 13). I parametri appresi per i blocchi da 70d a 70l della figura 12 vengono utilizzati come valori iniziali dei parametri nell'apprendimento, per velocizzare così l'apprendimento.

Caratteristiche comuni di un gran numero di tipi di oggetti vengono apprese in anticipo e quindi le prestazioni vengono regolate in ogni singola rete per ogni oggetto, per essere in grado così di ridurre un periodo di tempo e una quantità di dati da richiedere per addestrare il modello di apprendimento automatico 26 e migliorare le prestazioni del modello di apprendimento automatico 26. Quando vi è soltanto

un tipo di oggetto bersaglio, i blocchi da 71d a 71l possono essere omessi. I blocchi corrispondenti ai blocchi da 71d a 71l sono forniti corrispondentemente al numero di tipi di oggetti bersaglio. L'unità di addestramento di posa 38 può addestrare i blocchi da 70a a 70l del modello di apprendimento automatico 26 mediante i dati di addestramento preliminare e quindi l'unità di addestramento di posa 38 può fissare i parametri per i blocchi da 70a a 70c e addestrare il modello di apprendimento automatico 26 mediante i dati di addestramento generati in S305 e S306. Quando si utilizza l'immagine di addestramento mascherata, l'unità di addestramento di posa 38 esegue l'addestramento utilizzando soltanto la regione non mascherata come regione effettiva.

L'unità di addestramento di posa 38 può eseguire l'apprendimento preliminare sul modello di apprendimento automatico 26 illustrato nella figura 13. Per esempio, un gruppo di blocchi corrispondente ai blocchi da 70d a 70l dopo una diramazione può essere fornito corrispondentemente al numero di oggetti di una pluralità di campioni e il modello di apprendimento automatico 26 può essere sottoposto all'apprendimento preliminare mediante dati di addestramento corrispondenti alla configurazione del modello di apprendimento automatico 26. I dati di addestramento possono essere, per esempio, dati di addestramento includenti l'immagine di un oggetto campione e dati di verità di base

per il gruppo di blocchi corrispondente all'oggetto. In S307, l'unità di addestramento di posa 38 può impostare parametri casuali o i parametri appresi per oggetti appropriati come valori iniziali per i blocchi da 70d a 70l (o blocchi corrispondenti) dopo una diramazione. In aggiunta, è possibile eseguire un cosiddetto meta-apprendimento come apprendimento preliminare.

Al contempo, dopo che sono state acquisite le immagini fotografate, l'unità di addestramento di classificazione 39 addestra il modello di classificazione 23 mediante i dati di addestramento in base alle immagini fotografate (S308).

I dati di addestramento utilizzati in S308 includono i dati di addestramento di esempio positivo e i dati di addestramento di esempio negativo. L'unità di addestramento di classificazione 39 immette l'immagine in cui è stato fotografato l'oggetto bersaglio 51 nell'unità di generazione di caratteristica 31 e acquisisce i dati di valore di caratteristica emessi, per generare così una pluralità di dati di addestramento di esempio positivo. Relativamente ad una pluralità di dati di addestramento di esempio negativo, l'unità di addestramento di classificazione 39 immette nell'unità di generazione di caratteristica 31 immagini di campione di esempio negativo memorizzate in anticipo nell'unità di memorizzazione 12 e acquisisce i dati di valore di caratteristica emessi, per generare così una pluralità di

dati di addestramento di esempio negativo. Le immagini di campione di esempio negativo sono immagini ottenute fotografando mediante l'unità fotografica 20 in anticipo o immagini raccolte dal Web. In aggiunta, dati di addestramento di esempio positivo relativi ad un altro oggetto possono essere utilizzati come dati di addestramento di esempio negativo. I dati di addestramento di esempio negativo possono essere generati in anticipo per essere memorizzati nell'unità di memorizzazione 12. In questo caso, l'unità di addestramento di classificazione 39 può semplicemente acquisire i dati di addestramento di esempio negativo memorizzati nell'unità di memorizzazione 12. L'unità di addestramento di classificazione 39 addestra il classificatore 32 incluso nel modello di classificazione 23 mediante questi dati di addestramento.

Il valore di caratteristica da utilizzare per generare i dati di addestramento per il classificatore 32 viene estratto mediante un'elaborazione uguale a quella dell'unità di generazione di caratteristica 31 inclusa nel modello di classificazione 23. Il modello di classificazione 23 è addestrato tramite l'apprendimento del classificatore 32. Il modello di classificazione 23 non è limitato a quello descritto finora e può essere un modello che determina direttamente, dall'immagine, la presenza o l'assenza dell'oggetto bersaglio 51.

In questa forma di realizzazione, l'immagine da immettere nel modello di apprendimento automatico 26 è limitata dall'elaborazione dell'unità di acquisizione di regione bersaglio 21 ad un'immagine della regione in cui l'oggetto bersaglio 51 è presente nell'immagine ottenuta fotografando, l'immagine avendo una probabilità sufficientemente alta che l'oggetto bersaglio 51 sia presente al centro. In aggiunta, il modello di apprendimento automatico 26 dell'unità di stima di posa 25 è addestrato mediante i dati di addestramento generati dal modello di forma tridimensionale, mentre il modello di classificazione 23 dell'unità di acquisizione di regione bersaglio 21 è addestrato in base alle immagini in cui è stato fotografato l'oggetto bersaglio 51.

Le immagini da immettere nel modello di apprendimento automatico 26 sono limitate in modo appropriato, per migliorare così la precisione dell'emissione del modello di apprendimento automatico 26 e migliorare la precisione della stima della posa dell'oggetto bersaglio 51. In aggiunta, il modello di classificazione 23 è addestrato sulla base dell'immagine fotografata, invece dell'immagine basata sul modello di forma tridimensionale, per essere in grado così di selezionare la regione bersaglio 55 in modo più preciso e per migliorare di conseguenza la precisione del modello di apprendimento automatico 26.

In questa forma di realizzazione, l'immagine fotografata per generare il modello di forma tridimensionale per addestrare il modello di apprendimento automatico 26 dell'unità di stima di posa 25 viene utilizzata anche quando viene addestrato il modello di classificazione 23. Questo riduce il tempo e il lavoro richiesti per fotografare l'oggetto bersaglio 51 e riduce anche il tempo richiesto per l'addestramento del modello di apprendimento automatico 26 e del modello di classificazione 23.

Occorre evidenziare che la presente invenzione non è limitata alla forma di realizzazione summenzionata.

Per esempio, il classificatore 32 può essere una SVM di un qualsiasi kernel. Il classificatore 32 può anche essere un classificatore che utilizza un metodo, per esempio, un metodo dei K adiacenti più vicini (K-nearest neighbor), una regressione logistica o un metodo di incremento quale AdaBoost. Inoltre, il classificatore 32 può essere implementato, per esempio, da una rete neurale, un classificatore Naive Bayes, una foresta casuale o un albero di decisione.

Inoltre, le stringhe di caratteri e i valori numerici specifici descritti sopra e i valori numerici e le stringhe di caratteri specifici nei disegni sono soltanto esemplificativi, e la presente invenzione non è limitata a queste stringhe di caratteri e a questi valori numerici.

RIVENDICAZIONI

[Rivendicazione 1]

Sistema di elaborazione di informazione, comprendente:

mezzi di acquisizione di regione bersaglio per: acquisire un'immagine immessa; determinare se l'immagine immessa include o meno un'immagine di un oggetto bersaglio immettendo almeno una parte dell'immagine immessa in un modello di classificazione addestrato in base ad una pluralità di immagini di apprendimento includente un'immagine in cui l'oggetto bersaglio è stato fotografato e dati di etichetta che indicano se ciascuna della pluralità di immagini di apprendimento include o meno l'oggetto bersaglio; e acquisire una regione bersaglio includente l'immagine dell'oggetto bersaglio, che è estratta dall'immagine immessa, quando l'immagine immessa include l'oggetto bersaglio; e

mezzi di stima di posa per stimare una posa dell'oggetto bersaglio in base a informazioni emesse da un modello di apprendimento automatico, quando la regione bersaglio acquisita viene immessa in esso, che è addestrato mediante: una pluralità di immagini di addestramento rese da un modello di forma tridimensionale dell'oggetto bersaglio; e dati di verità di base che sono informazioni relative alla posa dell'oggetto bersaglio nelle immagini di addestramento.

[Rivendicazione 2]

Sistema di elaborazione di informazione secondo la rivendicazione 1,

in cui i mezzi di acquisizione di regione bersaglio sono configurati per estrarre una regione includente un'immagine di un oggetto dall'immagine immessa,

in cui il modello di classificazione include:

un'unità di generazione di caratteristica configurata per generare un valore di caratteristica di un'immagine di almeno una parte della regione estratta; e

un classificatore configurato per ricevere un'immissione del valore di caratteristica generato e per emettere informazioni che indicano se la regione estratta ha o meno l'immagine dell'oggetto bersaglio e

in cui il classificatore è addestrato mediante dati di addestramento includenti valori di caratteristica generati dall'immagine in cui l'oggetto bersaglio è stato fotografato e i dati di etichetta.

[Rivendicazione 3]

Sistema di elaborazione di informazione secondo la rivendicazione 2, in cui l'unità di generazione di caratteristica è regolata in modo tale che una distanza tra valori di caratteristica generati da una pluralità di immagini includenti l'oggetto bersaglio diventi minore di una distanza tra il valore di caratteristica generato da un'immagine includente l'oggetto bersaglio e un valore di

caratteristica generato da un'immagine includente un oggetto diverso dall'oggetto bersaglio.

[Rivendicazione 4]

Sistema di elaborazione di informazione secondo una qualsiasi delle rivendicazioni da 1 a 3,

in cui il modello di apprendimento automatico è addestrato mediante dati di addestramento includenti: la pluralità di immagini di addestramento rese mediante il modello di forma tridimensionale dell'oggetto bersaglio; e i dati di verità di base indicanti le posizioni di punti chiave dell'oggetto bersaglio nelle immagini di addestramento,

in cui i mezzi di stima di posa sono configurati per acquisire informazioni indicanti le posizioni bidimensionali dei punti chiave dell'oggetto bersaglio nella regione bersaglio immettendo la regione bersaglio acquisita nel modello di apprendimento automatico e

in cui i mezzi di stima di posa sono configurati per stimare la posa dell'oggetto bersaglio in base alle informazioni indicanti le posizioni bidimensionali dei punti chiave e informazioni indicanti le posizioni tridimensionali dei punti chiave nel modello di forma tridimensionale.

[Rivendicazione 5]

Sistema di elaborazione di informazione secondo la rivendicazione 4,

in cui il modello di apprendimento automatico è addestrato mediante una pluralità di immagini di addestramento rese da un modello tridimensionale dell'oggetto bersaglio e immagini di verità di base in cui ciascuno dei pixel indica una relazione di posizione rispetto al punto chiave dell'oggetto bersaglio nelle immagini di addestramento,

in cui i mezzi di stima di posa sono configurati per acquisire un'immagine di posizione in cui ciascuno dei pixel indica la relazione di posizione rispetto al punto chiave dell'oggetto bersaglio immettendo la regione bersaglio acquisita nel modello di apprendimento automatico,

in cui i mezzi di stima di posa sono configurati per calcolare, in base all'immagine di posizione, la posizione del punto chiave dell'oggetto bersaglio nell'immagine di posizione e

in cui i mezzi di stima di posa sono configurati per stimare la posa dell'oggetto bersaglio in base alla posizione calcolata del punto chiave nell'immagine di posizione e nel modello tridimensionale.

[Rivendicazione 6]

Sistema di elaborazione di informazione secondo la rivendicazione 5,

in cui i mezzi di acquisizione di regione bersaglio sono configurati per generare un'immagine di maschera per

mascherare una regione diversa dall'immagine dell'oggetto bersaglio nella regione bersaglio,

in cui i mezzi di stima di posa sono configurati per mascherare una parte dell'immagine di posizione in base all'immagine di maschera e

in cui i mezzi di stima di posa sono configurati per acquisire, in base all'immagine di posizione mascherata, la posizione del punto chiave dell'oggetto bersaglio nell'immagine di posizione.

[Rivendicazione 7]

Sistema di elaborazione di informazione secondo una qualsiasi delle rivendicazioni da 1 a 6, comprendente inoltre:

mezzi di acquisizione di immagini fotografate per acquisire una pluralità di immagini fotografate ottenute fotografando da una pluralità di direzioni rispetto all'oggetto bersaglio;

mezzi di generazione di modello di forma per calcolare il modello di forma tridimensionale dell'oggetto bersaglio in base alla pluralità di immagini fotografate; e

mezzi di addestramento di classificazione per addestrare, mediante dati di addestramento includenti i dati di verità di base e i dati immessi corrispondenti alla pluralità di immagini fotografate, il modello di classificazione per determinare se l'immagine immessa

include o meno l'immagine dell'oggetto bersaglio.

[Rivendicazione 8]

Sistema di elaborazione di informazione secondo la rivendicazione 7,

in cui i mezzi di generazione di modello di forma sono configurati per generare informazioni indicanti la posa dell'oggetto bersaglio nell'immagine fotografata e

in cui il sistema di elaborazione di informazione comprende inoltre mezzi di addestramento di posa per addestrare il modello di apprendimento automatico mediante: le immagini di addestramento includenti una pluralità di immagini rese dal modello di forma tridimensionale e la pluralità di immagini fotografate; e i dati di verità di base che sono le informazioni relative alla posa dell'oggetto bersaglio nelle immagini di addestramento.

[Rivendicazione 9]

Sistema di elaborazione di informazione secondo la rivendicazione 8, comprendente inoltre mezzi di generazione di dati di addestramento per acquisire, dalla pluralità di immagini fotografate ottenute fotografando, un'immagine fotografata da cui è stata rimossa un'immagine di una mano, e generare un'immagine di addestramento includente l'immagine fotografata da cui è stata rimossa l'immagine della mano e i dati di verità di base.

[Rivendicazione 10]

Sistema di elaborazione di informazione secondo una qualsiasi delle rivendicazioni da 1 a 9,

in cui il modello di apprendimento automatico include una rete neurale formata da una pluralità di strati includenti uno strato di immissione, uno stato intermedio e uno strato di emissione e

in cui il modello di apprendimento automatico è addestrato in anticipo mediante dati di addestramento preliminare includenti una pluralità di immagini di addestramento relative a oggetti diversi dall'oggetto bersaglio e i dati di verità di base, e viene quindi addestrato mediante la pluralità di immagini di addestramento relative all'oggetto bersaglio e i dati di verità di base con un parametro fisso per uno strato o una pluralità di strati escluso lo strato di emissione e incluso lo strato di immissione.

[Rivendicazione 11]

Metodo di elaborazione di informazione, comprendente le fasi di:

acquisire un'immagine immessa;

determinare se l'immagine immessa include o meno un'immagine di un oggetto bersaglio immettendo almeno una parte dell'immagine immessa in un modello di classificazione addestrato in base ad una pluralità di immagini di apprendimento includente un'immagine in cui l'oggetto

bersaglio è stato fotografato e dati di etichette che indicano se ciascuna della pluralità di immagini di apprendimento include o meno l'oggetto bersaglio;

acquisire una regione bersaglio includente l'immagine dell'oggetto bersaglio, che è estratta dall'immagine immessa, quando l'immagine immessa include l'oggetto bersaglio; e

stimare una posa dell'oggetto bersaglio in base alle informazioni emesse da un modello di apprendimento automatico, quando la regione bersaglio acquisita è immessa in esso, che è addestrato mediante: una pluralità di immagini di addestramento rese da un modello di forma tridimensionale dell'oggetto bersaglio; e dati di verità di base che sono informazioni relative alla posa dell'oggetto bersaglio nelle immagini di addestramento.

[Rivendicazione 12]

Programma per fare in modo che un computer esegua i processi di:

acquisire un'immagine immessa;

determinare se l'immagine immessa include o meno un'immagine di un oggetto bersaglio immettendo almeno una parte dell'immagine immessa in un modello di classificazione addestrato in base ad una pluralità di immagini di apprendimento includente un'immagine in cui è stato fotografato l'oggetto bersaglio e dati di etichette che

indicano se ciascuna della pluralità di immagini di apprendimento include o meno l'oggetto bersaglio;

acquisire una regione bersaglio includente l'immagine dell'oggetto bersaglio, che è estratta dall'immagine immessa, quando l'immagine immessa include l'oggetto bersaglio; e

stimare una posa dell'oggetto bersaglio in base a informazioni emesse da un modello di apprendimento automatico, quando la regione bersaglio acquisita è immessa in esso, che è addestrato mediante: una pluralità di immagini di addestramento rese da un modello di forma tridimensionale dell'oggetto bersaglio; e dati di verità di base che sono informazioni relative alla posa dell'oggetto bersaglio nelle immagini di addestramento.

FIG.1

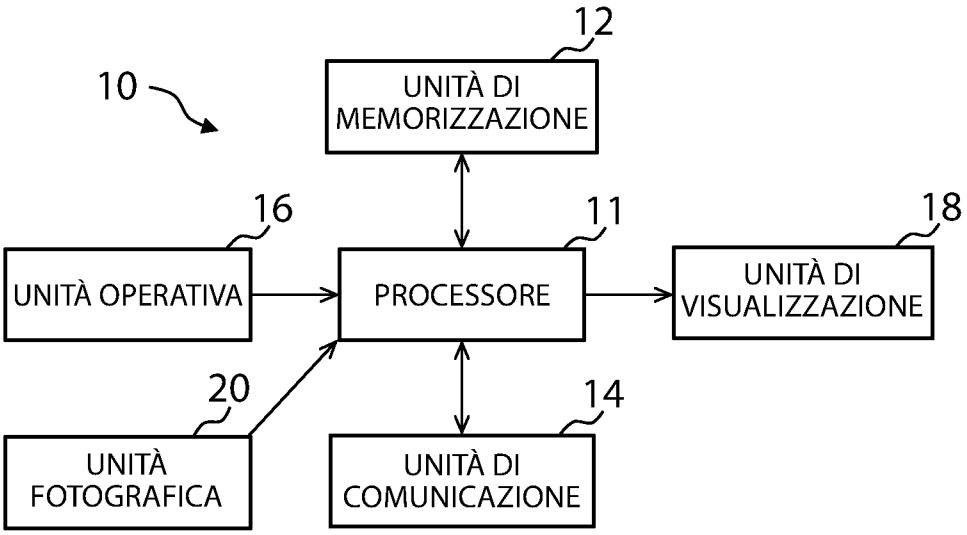

10
12
UNITÀ DI MEMORIZZAZIONE
16
UNITÀ OPERATIVA
11
PROCESSORE
18
UNITÀ DI VISUALIZZAZIONE
20
UNITÀ FOTOGRAFICA
14
UNITÀ DI COMUNICAZIONE

FIG.2

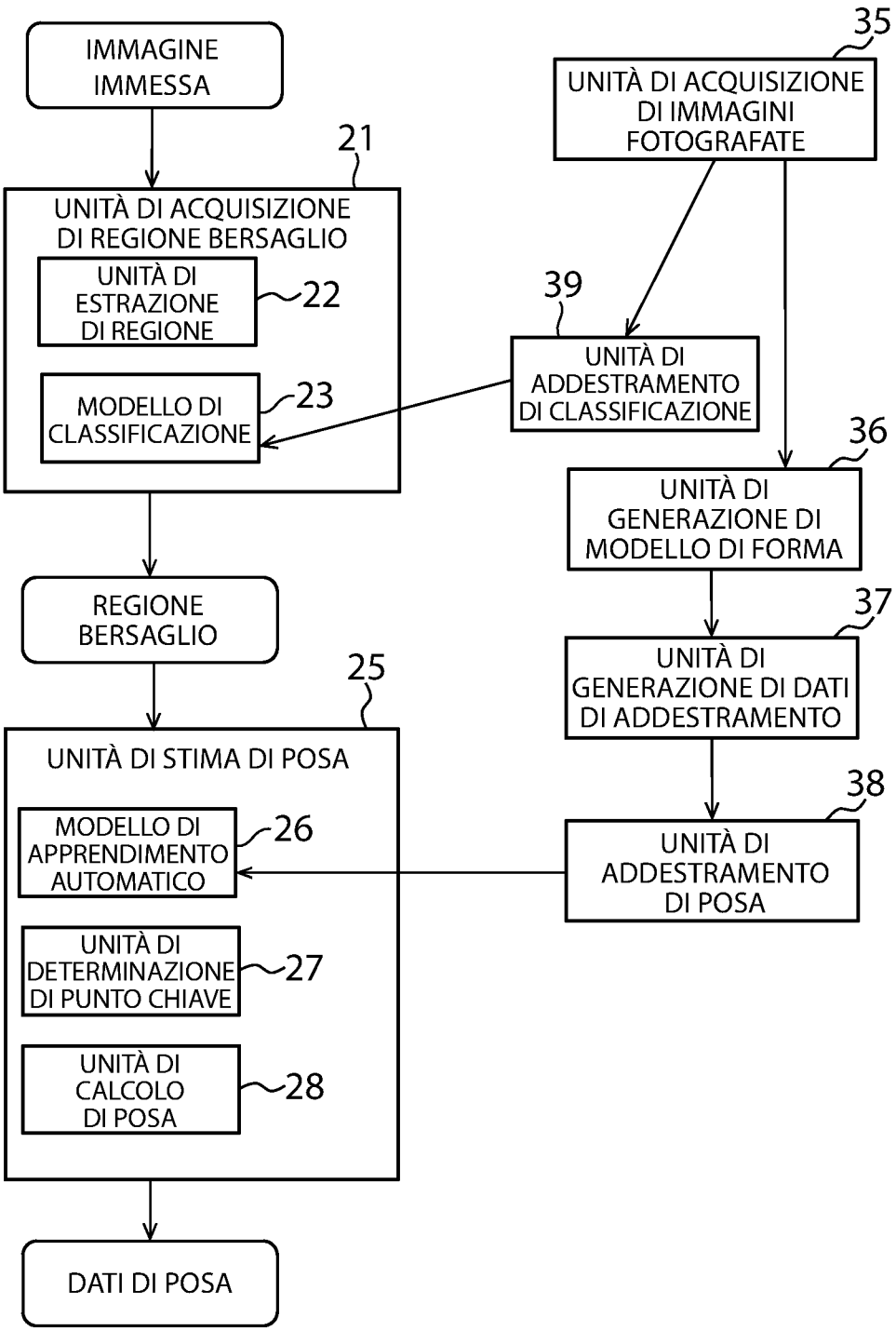


IMMAGINE IMMESSA
21
UNITÀ DI ACQUISIZIONE DI REGIONE BERSAGLIO
UNITÀ DI ESTRAZIONE DI REGIONE
22
MODELLO DI CLASSIFICAZIONE
23
REGIONE BERSAGLIO
25
UNITÀ DI STIMA DI POSA
MODELLO DI APPRENDIMENTO AUTOMATICO
26
UNITÀ DI DETERMINAZIONE DI PUNTO CHIAVE
27
UNITÀ DI CALCOLO DI POSA
28
DATI DI POSA
35
UNITÀ DI ACQUISIZIONE DI IMMAGINI FOTOGRAFATE
39
UNITÀ DI ADDESTRAMENTO DI CLASSIFICAZIONE
36
UNITÀ DI GENERAZIONE DI MODELLO DI FORMA
37
UNITÀ DI GENERAZIONE DI DATI DI ADDESTRAMENTO
38
UNITÀ DI ADDESTRAMENTO DI POSA

FIG.3

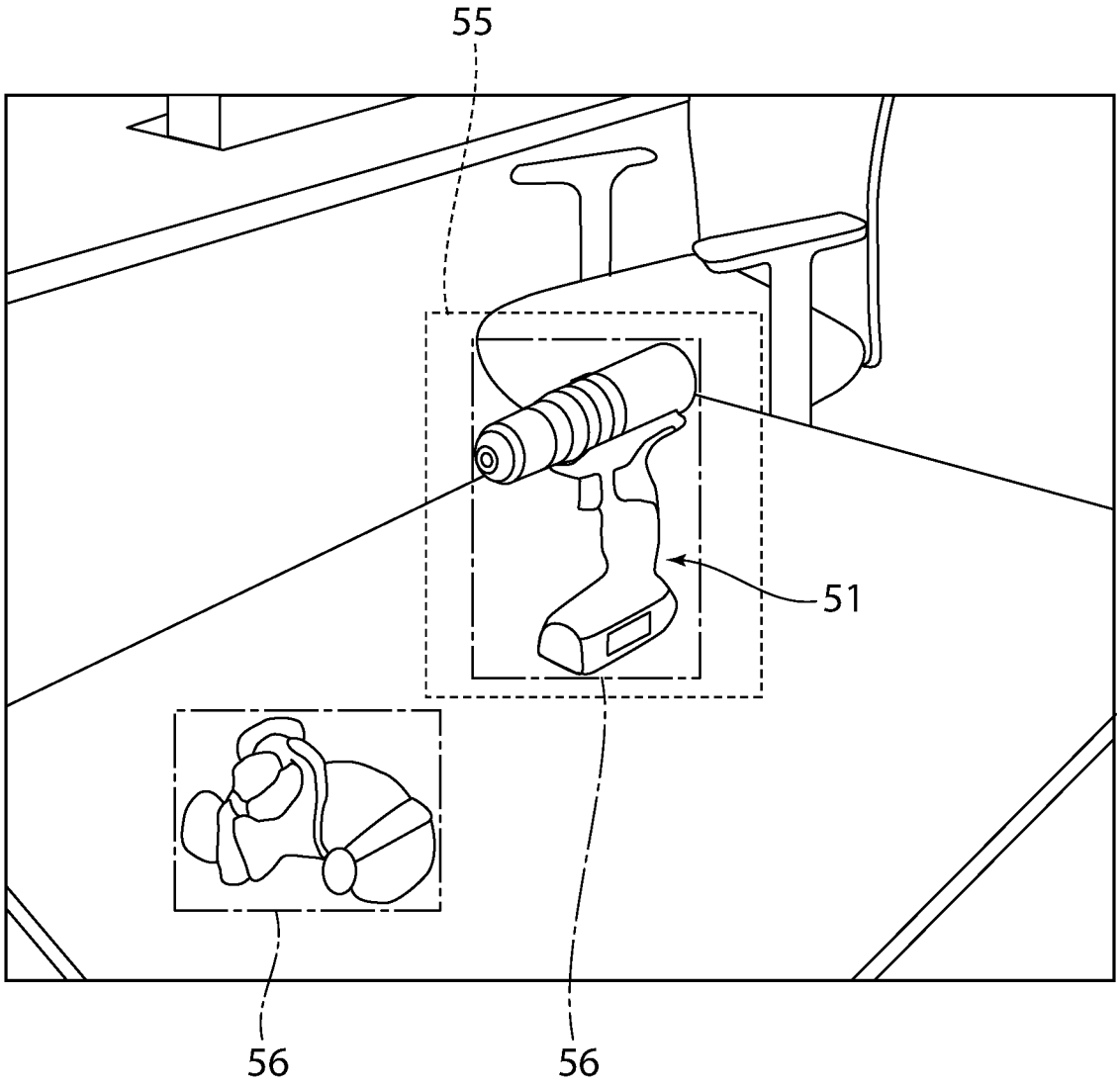


FIG.4

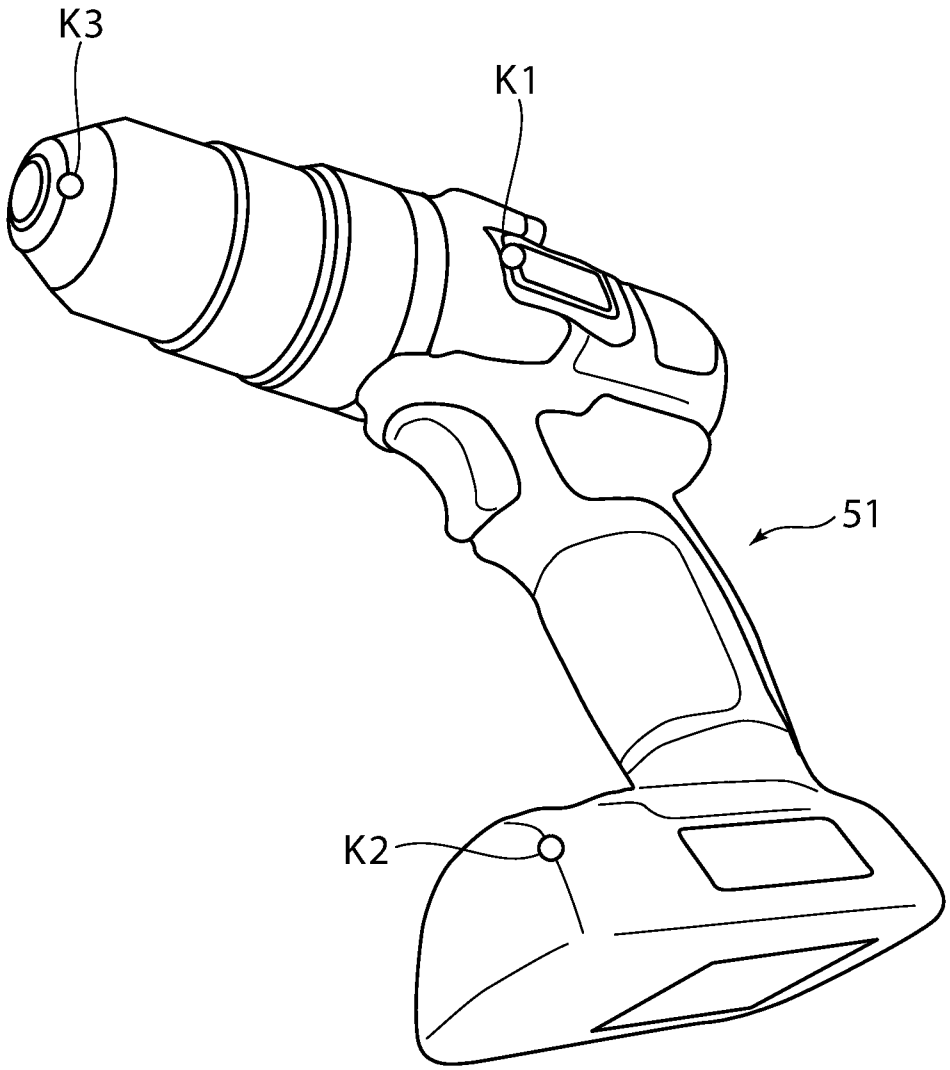


FIG.5

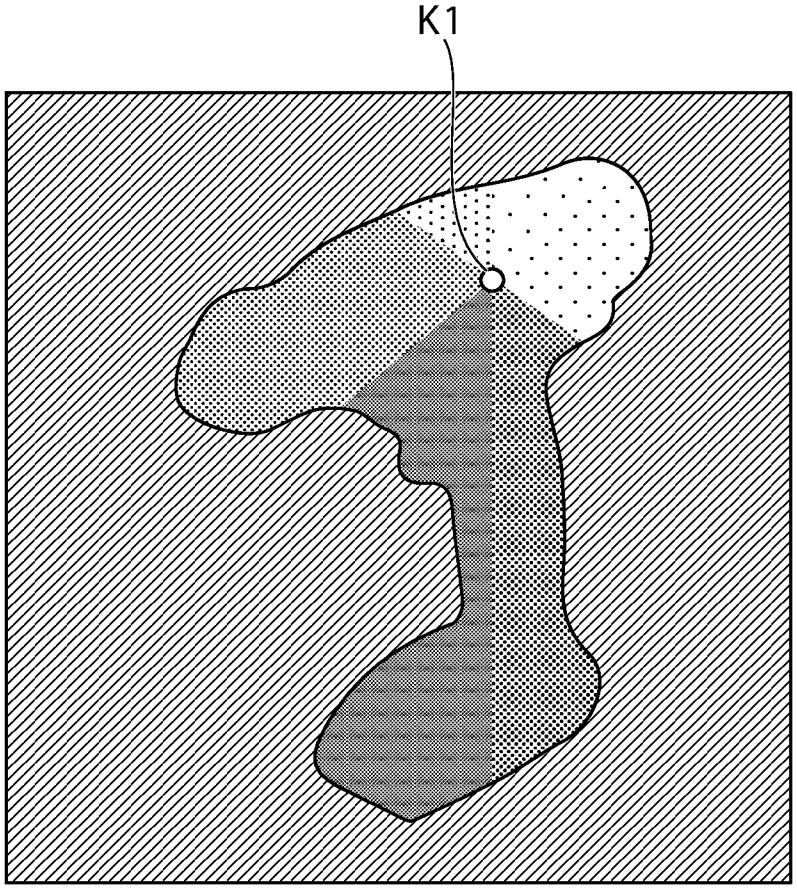


FIG.6

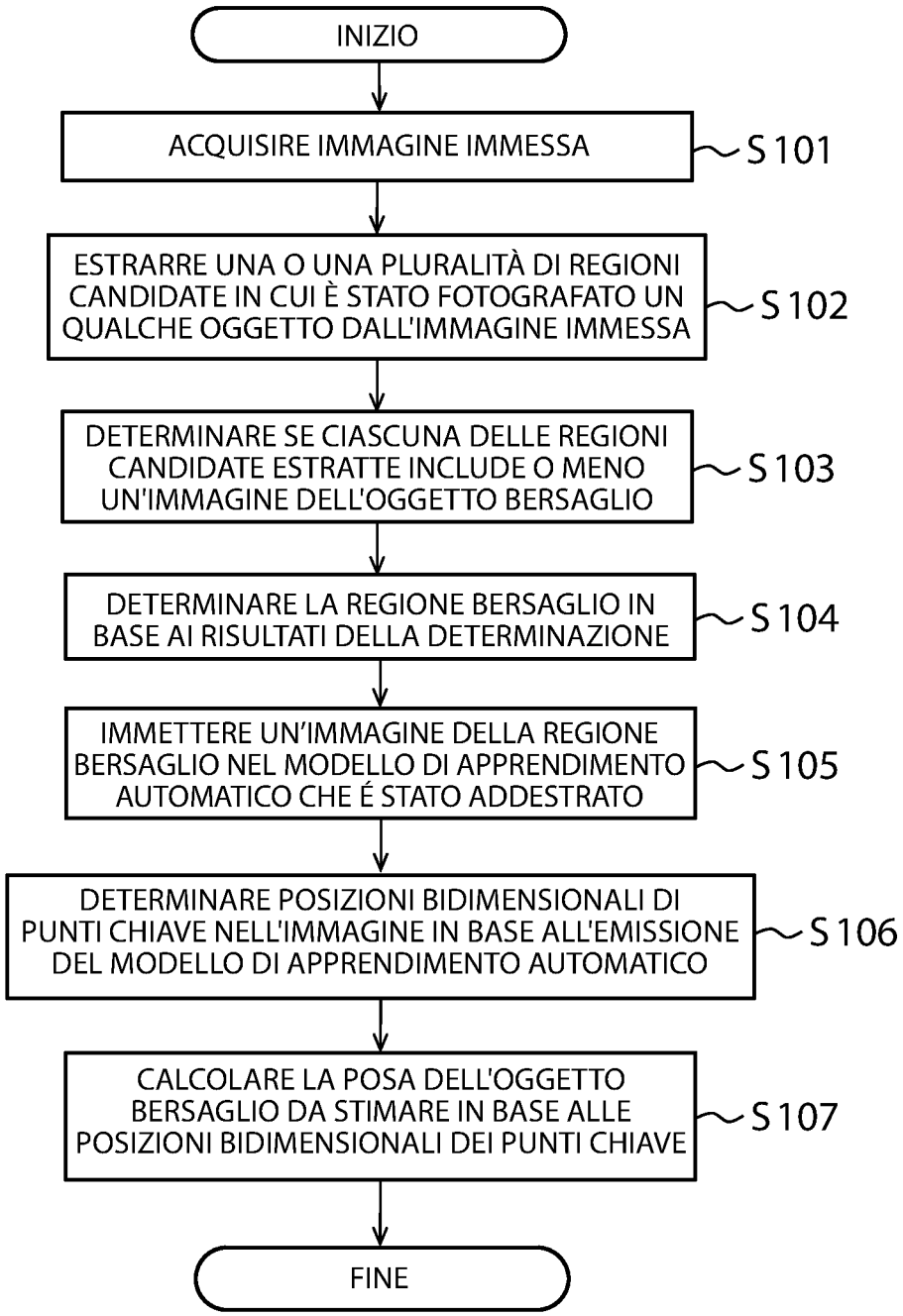

INIZIO
ACQUISIRE IMMAGINE IMMESSA
S 101
ESTRARRE UNA O UNA PLURALITÀ DI REGIONI CANDIDATE IN CUI È STATO FOTOGRAFATO UN QUALCHE OGGETTO DALL'IMMAGINE IMMESSA
S 102
DETERMINARE SE CIASCUNA DELLE REGIONI CANDIDATE ESTRATTE INCLUDE O MENO UN'IMMAGINE DELL'OGGETTO BERSAGLIO
S 103
DETERMINARE LA REGIONE BERSAGLIO IN BASE AI RISULTATI DELLA DETERMINAZIONE
S 104
IMMETTERE UN'IMMAGINE DELLA REGIONE BERSAGLIO NEL MODELLO DI APPRENDIMENTO AUTOMATICO CHE É STATO ADDESTRATO
S 105
DETERMINARE POSIZIONI BIDIMENSIONALI DI PUNTI CHIAVE NELL'IMMAGINE IN BASE ALL'EMISSIONE DEL MODELLO DI APPRENDIMENTO AUTOMATICO
S 106
CALCOLARE LA POSA DELL'OGGETTO BERSAGLIO DA STIMARE IN BASE ALLE POSIZIONI BIDIMENSIONALI DEI PUNTI CHIAVE
S 107
FINE

FIG.7

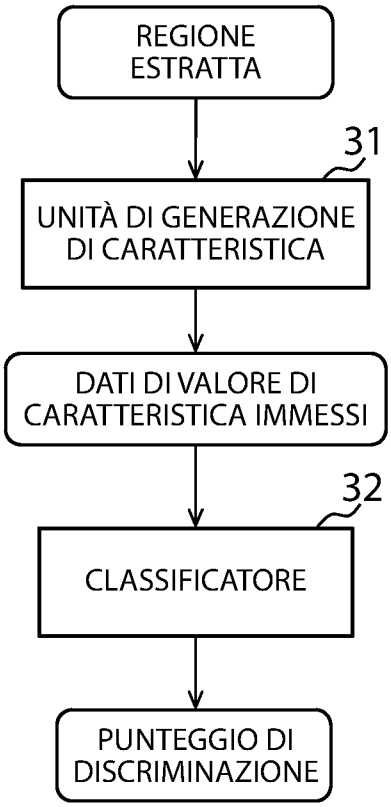

REGIONE ESTRATTA
31
UNITÀ DI GENERAZIONE DI CARATTERISTICA
DATI DI VALORE DI CARATTERISTICA IMMESSI
32
CLASSIFICATORE
PUNTEGGIO DI DISCRIMINAZIONE

FIG.8

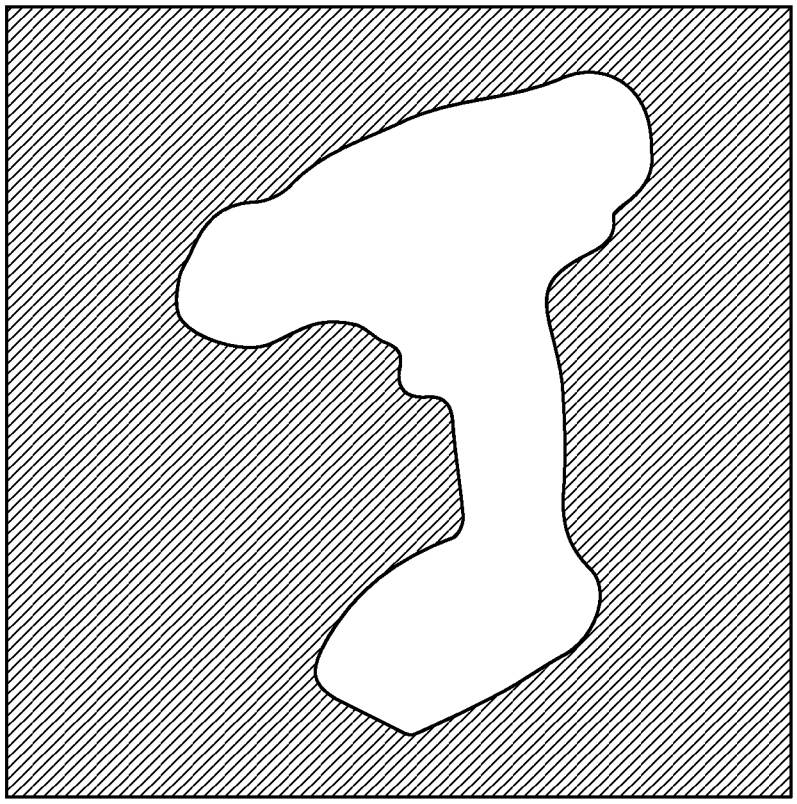


FIG.9

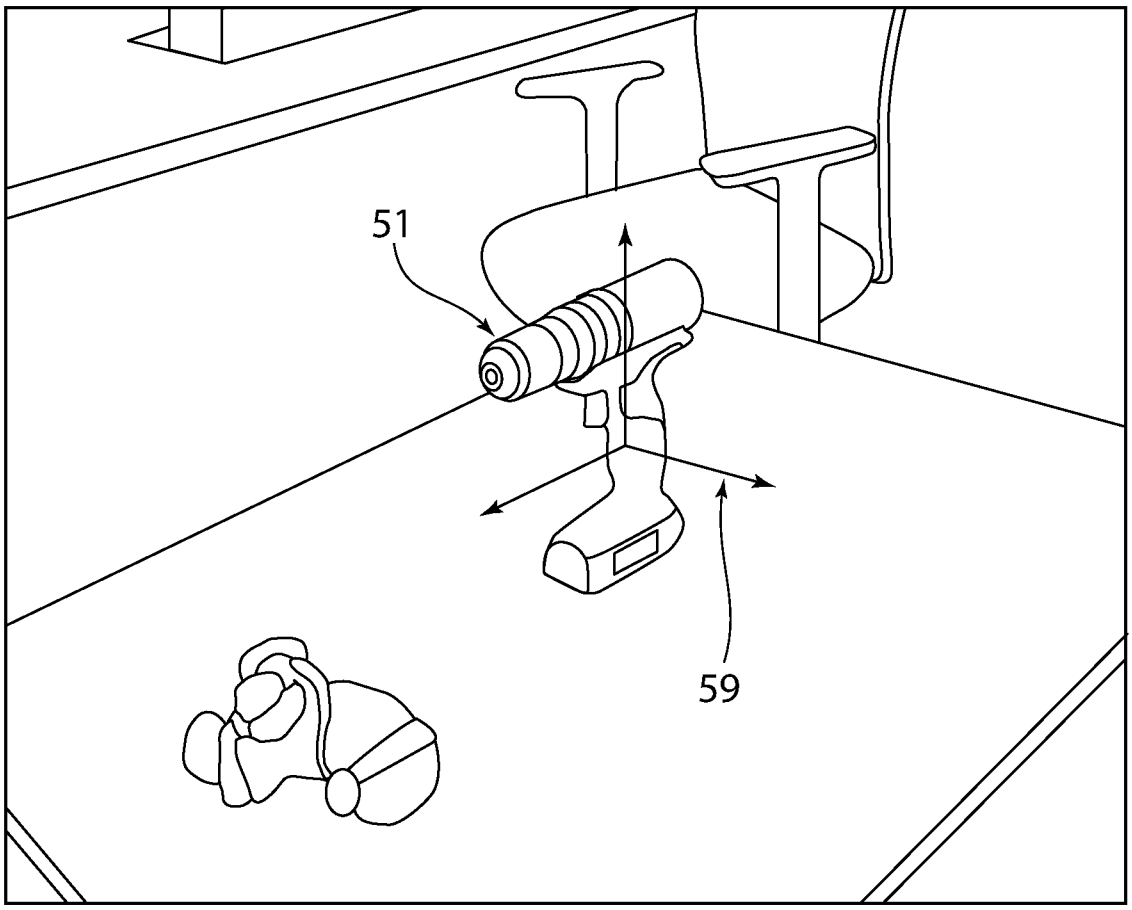


FIG.10

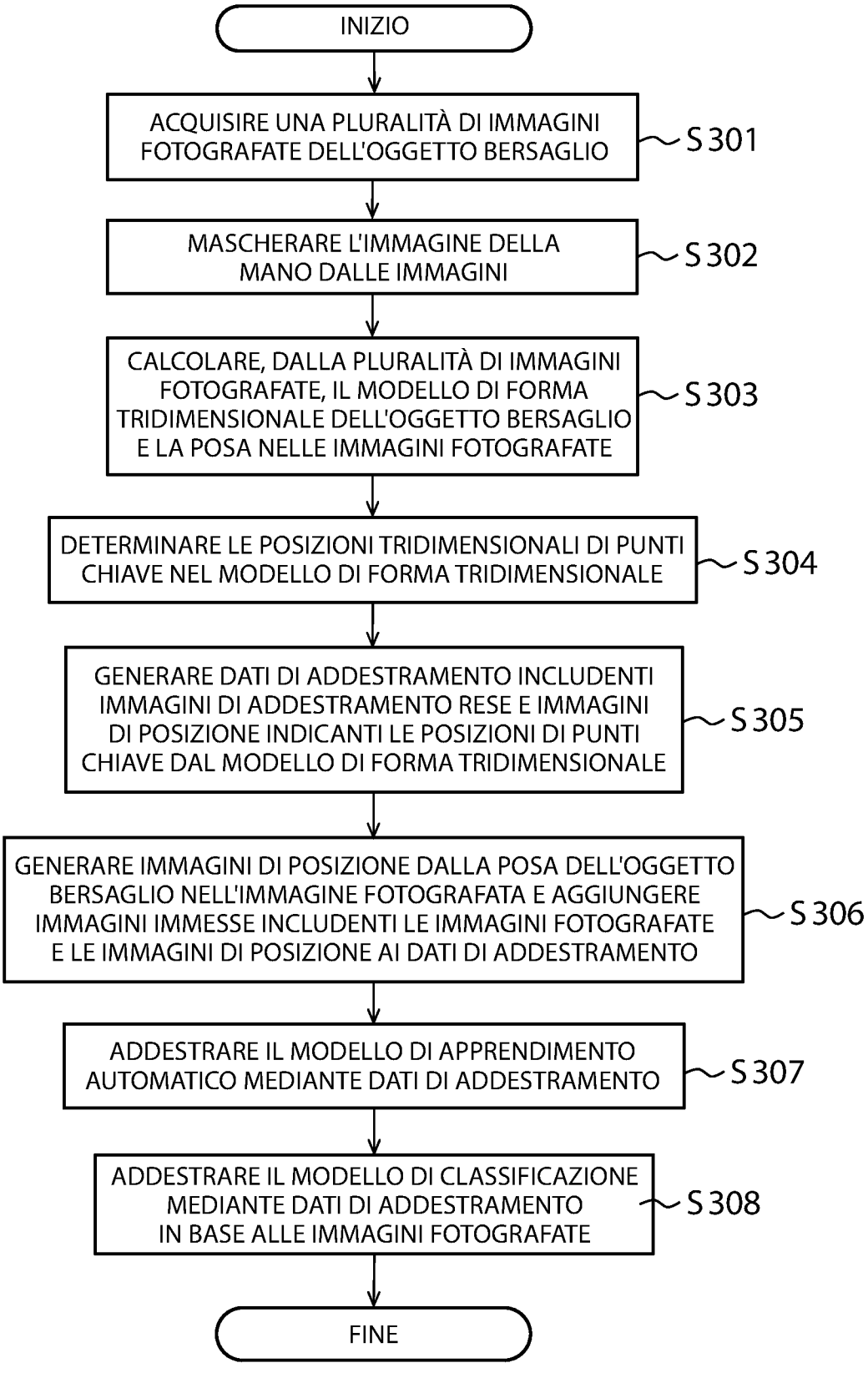

INIZIO
ACQUISIRE UNA PLURALITÀ DI IMMAGINI FOTOGRAFATE DELL'OGGETTO BERSAGLIO
S 301
MASCHERARE L'IMMAGINE DELLA MANO DALLE IMMAGINI
S 302
CALCOLARE, DALLA PLURALITÀ DI IMMAGINI FOTOGRAFATE, IL MODELLO DI FORMA TRIDIMENSIONALE DELL'OGGETTO BERSAGLIO E LA POSA NELLE IMMAGINI FOTOGRAFATE
S 303
DETERMINARE LE POSIZIONI TRIDIMENSIONALI DI PUNTI CHIAVE NEL MODELLO DI FORMA TRIDIMENSIONALE
S 304
GENERARE DATI DI ADDESTRAMENTO INCLUDENTI IMMAGINI DI ADDESTRAMENTO RESE E IMMAGINI DI POSIZIONE INDICANTI LE POSIZIONI DI PUNTI CHIAVE DAL MODELLO DI FORMA TRIDIMENSIONALE
S 305
GENERARE IMMAGINI DI POSIZIONE DALLA POSA DELL'OGGETTO BERSAGLIO NELL'IMMAGINE FOTOGRAFATA E AGGIUNGERE IMMAGINI IMMESSE INCLUDENTI LE IMMAGINI FOTOGRAFATE E LE IMMAGINI DI POSIZIONE AI DATI DI ADDESTRAMENTO
S 306
ADDESTRARE IL MODELLO DI APPRENDIMENTO AUTOMATICO MEDIANTE DATI DI ADDESTRAMENTO
S 307
ADDESTRARE IL MODELLO DI CLASSIFICAZIONE MEDIANTE DATI DI ADDESTRAMENTO IN BASE ALLE IMMAGINI FOTOGRAFATE
S 308
FINE

FIG.11

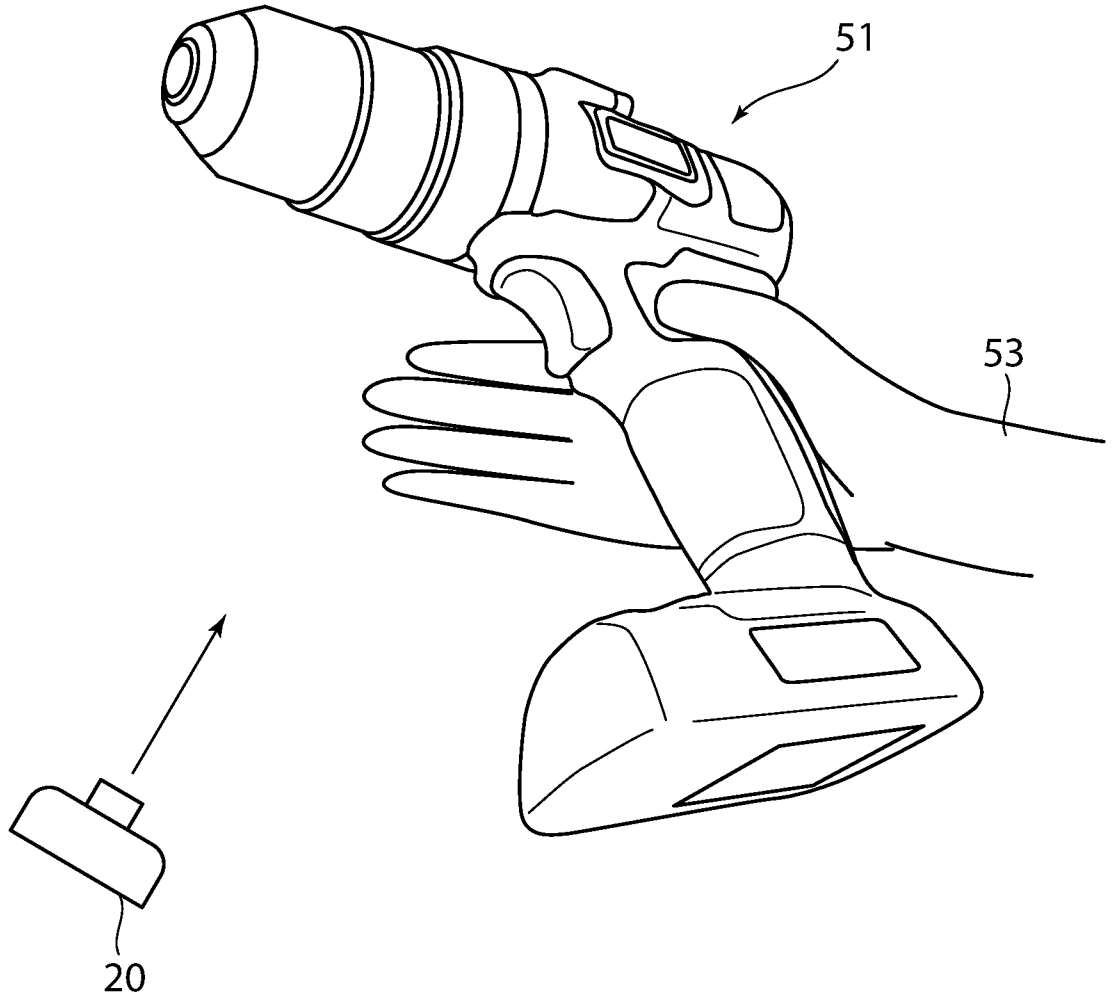


FIG.12

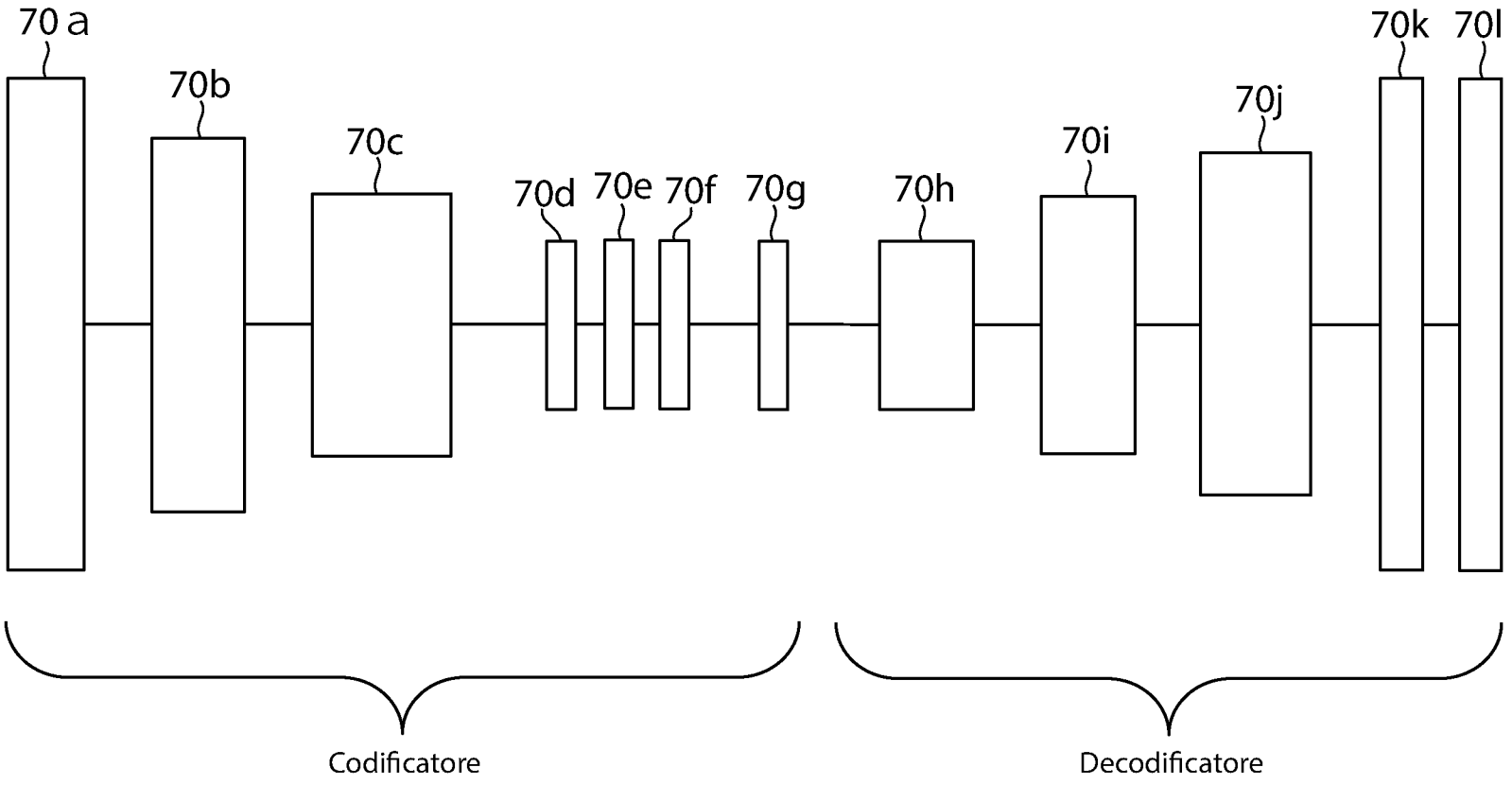
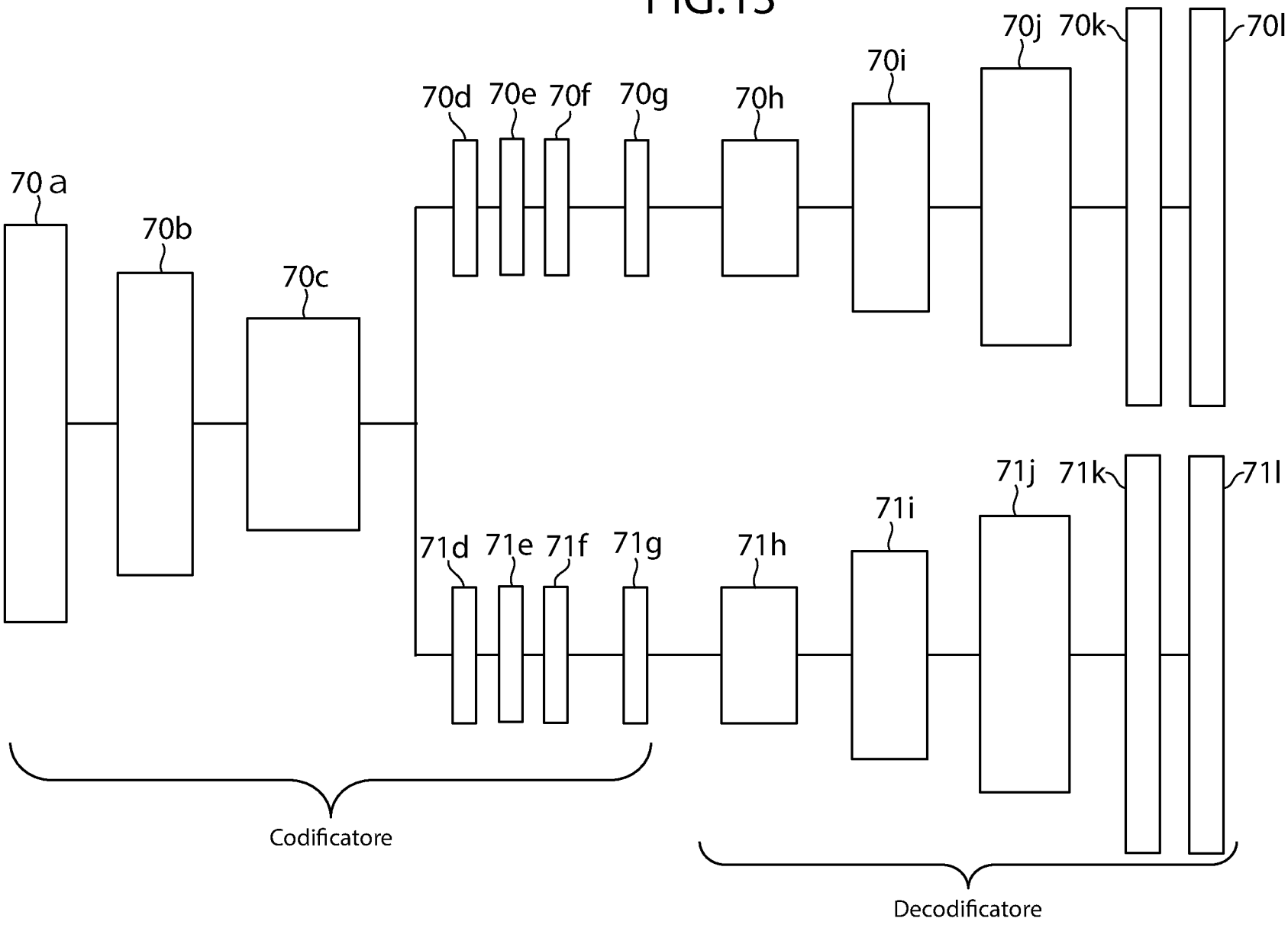

70 a
70b
70c
70d
70e
70f
70g
70h
70i
70j
70k
70l
Codificatore
Decodificatore

FIG.13


70 a
70b
70c
70d
70e
70f
70g
70h
70i
70j
70k
70l
71d
71e
71f
71g
71h
71i
71j
71k
71l
Codificatore
Decodificatore