

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI
Corso di Laurea Magistrale in Informatica



Sistema di visione per la guida autonoma di veicoli

Riassunto

Relatore: Prof. Domenico G. Sorrenti

Correlatore: Dr. Axel Furlan

Correlatore: Dr. Daniele Marzorati

Candidato: Roberto FERMI

Matricola: 048091

Anno Accademico 2009 – 2010

Riassunto

L'obiettivo di questa tesi è di realizzare un sistema di visione da utilizzare nei veicoli a guida autonoma per determinare la posizione di un veicolo/robot nel mondo e riconoscere la presenza di ostacoli nel suo campo visivo.

L'utilizzo delle camere per la visione artificiale rappresenta attualmente l'approccio più promettente. Le informazioni generate dalle immagini delle camere consentono di avere una visione globale degli oggetti e di distinguere, a seguito di specifiche elaborazioni, le diverse tipologie di ostacoli presenti, come persone, autovetture, alberi, etc. Di contro, poiché la descrizione del mondo reale tridimensionale è proiettata su di un supporto bidimensionale, si ha la perdita di una dimensione e, in particolare, della profondità. Quando catturiamo un'immagine con la camera, infatti, non siamo più in grado di capire a che distanza siano i diversi punti in essa rappresentati. Se conoscessimo le dimensioni degli oggetti quest'aspetto non costituirebbe un problema, ma ciò non è ipotizzabile in questa applicazione, per le grandi differenze che possono esserci tra diverse istanze reali di oggetti.

Per ovviare a questo problema si tenta di simulare quello che normalmente avviene nella vista umana ovvero si utilizzano due immagini della stessa scena, prese da due punti di vista, cosa che in visione artificiale richiede l'utilizzo di due camere. Le due immagini del mondo catturate dai sensori devono riferirsi allo stesso istante di tempo e quindi il comando di acquisizione delle immagini delle camere deve essere opportunamente sincronizzato, per non rischiare di osservare due scene che hanno poco in comune. I due dispositivi, inoltre, devono essere calibrati: è necessario, in altre parole, stabilire come sono disposti l'uno rispetto all'altro in modo da riuscire a determinare, per i vari punti dell'immagine della prima camera, i punti corrispondenti nella seconda.

Prima di potere svolgere qualsiasi elaborazione di alto livello con le immagini catturate dalle camere (come, ad esempio, il

riconoscimento e la classificazione di oggetti), è necessario assicurarsi che la qualità delle immagini sia sufficiente. La diversa risposta all'intensità luminosa in differenti condizioni di acquisizione, il rumore elettrico, l'ampiezza del range dinamico, il contrasto, sono solo alcuni degli aspetti fondamentali che caratterizzano i sensori e le immagini che verranno esaminati in questa tesi. Per fare analisi matematiche è infatti necessario che, in ogni condizione di utilizzo del sistema, gli oggetti siano ben definiti e ben contrastati, cioè ben distinguibili gli uni dagli altri e discriminabili dallo sfondo. Se in condizioni "controllate", come ad esempio in un laboratorio o in una stanza ben illuminata, questo può essere facilmente raggiunto, in condizioni normali è molto difficile da ottenere. Basti pensare, ad esempio, alla condizione di guida di una vettura in strada durante una giornata soleggiata mentre ci si avvicina all'ingresso di una galleria. In queste situazioni alcune volte è arduo anche per l'occhio umano adattarsi per vedere correttamente i dettagli esposti al sole fuori dal tunnel e contemporaneamente quelli in ombra dentro la galleria. Questa è un contesto di alto range dinamico, dove la dinamica è quella del segnale luminanza immagine. Per una camera stato dell'arte attuale questa è una circostanza in cui è molto difficile fornire immagini di buona qualità; con camere stato-dell'arte non si riuscirà ad avere un'immagine utilizzabile. Se non si dispone di un'immagine sfruttabile, non sarà possibile determinare gli ostacoli che ci precedono e capire cosa accade intorno al veicolo rendendo impraticabile la guida autonoma. Per questo motivo è di essenziale importanza avere sempre un output valido dalle camere.

Un particolare punto d'interesse nelle applicazioni di visione artificiale è la scelta di adoperare parametri delle camere fissi (come, ad esempio, tempo di esposizione, apertura del diaframma, guadagno digitale, etc.), oppure di farli variare nel tempo, in relazione alle specifiche condizioni dell'ambiente in cui ci stiamo muovendo. La prima tipologia di approccio ha il vantaggio di non dovere fare elaborazioni in background per analizzare la scena, ma generalmente ha dei forti limiti in termini di utilizzo. In questa tesi verrà mostrato come l'uso di particolari tecnologie dei sensori digitali in unione a tecniche di elaborazione delle immagini ha reso possibile

questo scenario, consentendo di usare dei parametri fissi delle camere e, allo stesso tempo, di avere un risultato eccellente in ogni condizione di esposizione, dal buio della notte, al controluce intenso in una giornata soleggiata.

Mostreremo, inoltre, com'è possibile variare un solo parametro delle camere (il tempo di esposizione) con delle analisi in background che tengano conto, oltre che delle particolari condizioni d'illuminazione della scena, anche di specifici indicatori che consentono di definire quanto "*motion blur*", ovvero sfocatura, è presente nelle immagini. Ciò permetterà di rilevare quando il tempo di esposizione è troppo alto per quello che sta avvenendo nel mondo osservato dalle camere. In questi casi, infatti, è importante non ottenere mai un'immagine mossa della scena, sopperendo all'eventuale tempo di esposizione troppo basso (con tempi di esposizioni troppo bassi, si ottengono immagini mediamente più scure) con tecniche di post elaborazione al fine di aumentare la luminosità dell'immagine.

Verrà presentata la metrica CPBD (Cumulative Probability of Blur Detection) che si basa sul concetto di Just Noticeable Blur (JBN), ovvero la minima differenza rilevabile dall'uomo tra due livelli di luminanza dell'immagine. Questa metrica consente di effettuare delle valutazioni psicometriche sulla quantità di sfocatura presente nell'immagine stessa permettendo di supervisionare la variazione di tempo di esposizione delle camere.

Vi sono diverse tecniche di post-elaborazione e la più appropriata, per il nostro progetto, è il pixel-binning, che consiste nel raggruppare (o più precisamente sommare) la quantità di energia immagazzinata in elementi fotosensibili contigui del sensore digitale delle camere, in modo da ottenere una maggiore intensità in scene prive di luce o con scarsa illuminazione, a discapito della risoluzione. Un'altra tecnica molto utile e molto impiegata anche in altri ambiti dell'elaborazione delle immagini, è l'equalizzazione dell'istogramma che permette di ridistribuire equamente i vari livelli di colore nell'intervallo di possibili valori, in modo da aumentare il contrasto dell'immagine stessa. In questa tesi, verrà mostrata una particolare variante di quest'ultima

tecnica che consente di migliorare notevolmente i risultati ottenuti in modo da saturare una percentuale di dati prefissata e stabilita sperimentalmente.

Il lavoro si è articolato come segue:

- **analisi dello stato dell'arte.** Questa prima fase iniziale del progetto ha richiesto la ricerca e l'analisi delle tecnologie presenti sul mercato al fine di identificare quelle più adatte al nostro progetto.
- **realizzazione della testa stereo.** Durante questa fase ci siamo occupati della creazione fisica della testa stereo. Si è reso necessario identificare quale fosse la migliore posizione di montaggio trovando un modo per rendere le camere più solidali possibili ad USAD. Abbiamo quindi proceduto alla calibrazione della testa stereo stessa.
- **analisi della tecnologia LinLog.** E' stato necessario realizzare un programma specifico per leggere i parametri interni non pubblici delle camere riguardanti le impostazioni del LinLog. Successivamente sono stati svolti test sperimentali approfonditi per identificare quali fossero i parametri ottimali da utilizzare.
- **identificazione delle tecniche da usare per l'elaborazione di basso livello delle immagini.** In questa fase si è scelto quali tecniche utilizzare per l'elaborazione delle immagini catturate dalle camere optando per il pixel-binning ed il contrasto dell'istogramma con stretchlim.
- **implementazione nel linguaggio C.** La fase successiva ha riguardato la scrittura del codice C per le tecniche viste precedentemente. E' stato poi necessario integrare il tutto nella libreria DAFNE sviluppata dal D.R. Marzorati che include, tra l'altro, il calcolo della mappa della disparità.
- **identificazione dei parametri fissi per le camere.** Terminata la fase implementativa, ci siamo dedicati ad identificare

sperimentalmente quali fossero i migliori parametri a cui impostare le camere per ottenere il risultato di utilizzare parametri costanti in ogni situazione

- **verifica sperimentale dei risultati.** Successivamente all'integrazione, si è proceduto ad un'eshaustiva fase di verifica sperimentale dei risultati ottenuti. I test sono stati condotti in diverse condizioni di acquisizione, in ambienti chiusi ed all'aperto. Si è scelto, inoltre, di verificare l'affidabilità del sistema anche sotto diverse condizioni metereologiche, come neve, notte e durante il tramonto con sole radente.
- **analisi delle tecniche per il blur-detection.** Nonostante i più che soddisfacenti risultati ottenuti con i parametri costanti delle camere, abbiamo svolto un'approfondita ricerca per identificare quali fossero le tecniche esistenti per la rilevazione della sfocatura (o effetto mosso) nelle immagini.
- **scelta e test della metrica per il blur-detection.** In questa fase abbiamo valutato l'efficacia delle varie tecniche di blur detection individuate in precedenza ed abbiamo scelto la promettente metrica CPBD, implementandone una prima versione in Matlab
- **verifica sperimentale dei risultati per il blur-detection.** Sono stati condotti dei test sperimentali basati sui dataset immagine, catturati nelle fasi precedenti, volti a verificare l'efficacia della metrica CPBM nel rilevare ed identificare le immagini mosse.

Al termine di questo progetto sono stati analizzati i risultati sperimentali raggiunti.

La tecnologia LinLog presentata da PhotonFocus è effettivamente molto promettente e, sotto alcuni punti di vista, i risultati ottenuti sono stati sorprendenti.

Tuttavia, usarla correttamente e, soprattutto, riuscire ad identificare un set di parametri che permetta di ottenere buoni risultati, ha

richiesto sia un notevole sforzo di ricerca, al fine di capirne il funzionamento, sia un considerevole impegno pratico, per testare questa tecnologia in svariate condizioni.

Per ottenere il massimo risultato dalle immagini catturate, si è scelto di sacrificare la risoluzione delle stesse utilizzando la tecnica del pixel-binning, per ottenere delle riprese chiare anche in ambienti con scarsissima illuminazione.

L'equalizzazione dell'istogramma utilizzando la funzione di stretchlim ha permesso di ottenere ottimi risultati nelle immagini, anche quando la tecnica dell'equalizzazione dell'istogramma standard portava ad avere esiti insoddisfacenti.

Per quanto riguarda l'ambizioso progetto di riuscire ad utilizzare dei parametri fissi delle camere per ogni condizione della scena si può affermare che i risultati ottenuti sono ottimi, pur avendo accettato dei compromessi sulla risoluzione delle immagini. Tuttavia, il tempo di esposizione scelto potrebbe non essere adatto a velocità di guida superiori a quelle raggiunte da USAD o in condizioni di scene molto movimentate.

Si è proseguito, quindi, testando la metrica per l'identificazione del mosso delle immagini. Sono stati condotti dei test specifici per questa metrica su immagini direttamente rilevate dalle camere di USAD. Si è evidenziato che la metrica CPBD in oggetto permette di identificare in modo preciso il grado di sfocatura presente in un'immagine.

Vi sono numerosi sviluppi futuri applicabili al progetto presentato in questa tesi ed i principali sono i seguenti:

- la tecnica del pixel-binning può essere resa dinamica ricorrendo ad esempio, ad un binning meno incisivo (come ad esempio un 3x3 oppure un 2x2) al fine di ottenere delle immagini a più alta risoluzione quando l'illuminazione della scena è sufficiente ed incrementare il valore di binning con il diminuire della luminosità;

- l'ottimizzazione algoritmica del calcolo della disparity map può essere migliorata cercando di individuare tecniche che riducano il tempo di computazione;
- è necessario implementare in linguaggio C l'algoritmo per il blur-detection presentato in questa tesi;
- è opportuno implementare un processo in background che effettui la retroazione sul tempo di esposizione delle camere in relazione alla luminosità media dell'immagine ed alla metrica di blur-detection presentata in questa tesi
- le informazioni fornite dalla mappa dello spazio libero possono essere utilizzate per riuscire a guidare in modo autonomo, interfacciandosi con ROS e gli altri sensori di navigazione.