

Riconoscimento di potenziali ostacoli dall'analisi di sequenze video riprese da un mezzo in movimento

Luca Menci**, Davide Onofrio*, Stefano Tubaro*.

(*) Dipartimento di Elettronica e Informazione,
Politecnico di Milano, P.za L. da Vinci n.32 20133 Milano
tubaro@elet.polimi.it, Davide.Onofrio@polimi.it

(**) menci software S.r.l., via Martiri di Civitella, 11 – Arezzo
luca@menci.com

Riassunto

L'attività illustrata in questo contributo è legata all'analisi di immagini video riprese da un gruppo di camere montate su di un carrello ferroviario allo scopo di riconoscere potenziali ostacoli alla circolazione dei treni. Il sistema è costituito da una tripletta di videocamere poste sulla parte terminale di un vagone ferroviario ai vertici di una struttura triangolare in modo che osservino la sede ferroviaria. Le immagini vengono acquisite in modo continuo e sincrono e registrate su supporto magnetico.

Sebbene vi sia la possibilità di applicare per ogni tripletta di immagini tecniche di stereometria classica per valutare la configurazione 3D della scena occupata, ciò risulta molto difficile per la significativa distorsione prospettica e la grande disparità presente fra le immagini. Si è deciso di operare, almeno in una prima fase di screening, utilizzando una sola delle camere a disposizione, sfruttando la disparità creata dal moto del treno. In questa modalità le linee epipolari, per ciascuna immagine sono una raggiera che ha origine vicino al centro immagine. Anche in questo caso l'accoppiamento fra elementi stereo-corrispondenti è problematico poiché le distorsioni sono significative. Le linee epipolari possono essere rese parallele applicando una omografia che sposti il piano immagine su di un piano parallelo al moto del sistema o su piani perpendicolari ad esso. In questo modo si possono trovare facilmente i punti immagine relativi a strutture 3D che non giacciono sulle superfici laterali di un parallelepipedo nel quale il vagone scorre. Tecniche raffinate di elaborazione hanno portato a risultati significativi ottenuti con tempi di calcolo estremamente brevi. I dati 3D ottenuti possono essere utilizzati per inizializzare la ricostruzione basata sull'uso completo delle triplette di immagini a disposizione.

Abstract

The work here presented concerns the analysis of video sequences acquired, continuously and synchronously, by a camera triplet mounted on a railway truck. The cameras are mounted on the rear side of a wagon and are oriented towards the railway track. After the acquisition, images are processed in order to identify potential obstacles on the train way. Classical stereometric techniques can be applied to the image triplets in order to estimate the shape of the considered 3D scene. However a significant computational power is required to process the data due to the large disparity that characterize each pair of images coming from different cameras.

Therefore we decided to use, at least as a first processing step, only the data coming from a single camera, the central one of the triplet, using the disparity created by train movement. In this case the epipolar lines, relative to a couple of consecutive images, have a configuration similar to a star of rays with the focus near the image center. It is important to notice that in this case disparities are greatly influenced by lens distortion. A homographic transformation can be applied to each image in order to simplify the matching between elements belonging to different frames. In particular the homographic transformation can be chosen in order to guarantee a null disparity for points belonging to a plane parallel to the camera displacement. In this way 3D scene element that are

inside a predefined volume can be easily detected. The 3D data obtained with this approach can be an effective initialization for further 3D analysis based on the use of the full camera triplets.

Introduzione

L'attività svolta in questo studio è incentrata sulla valutazione della fattibilità di un sistema automatico per il rilevamento degli ostacoli e delle sagome limite di passaggio del convoglio in ambito ferroviario. Il sistema prevede l'uso di una terna di telecamere ad alta velocità di otturazione, perfettamente sincronizzate tra loro ed in grado di acquisire triplette di fotogrammi senza "effetto mosso" ad una velocità di crociera dell'ordine dei 120 km orari (fig. 1).

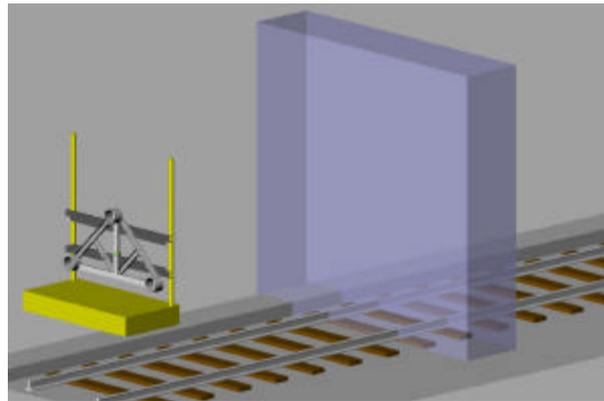


Figura 1 – Struttura di sostegno delle telecamere e volume di studio relativo a ciascuna tripletta di fotogrammi.

Le camere sono posizionate sulla parte posteriore del convoglio; la frequenza d'acquisizione è tale da consentire una copertura continua del volume di studio; questo è costituito dal volume spazzato dalla sagoma di passaggio del convoglio, opportunamente ampliata rispetto al suo ingombro minimo, durante il moto del treno stesso. Le telecamere sono calibrate con grande accuratezza, sia per quanto riguarda i loro parametri intrinseci (orientamento interno) che per quelli estrinseci (orientamento esterno); il sistema d'acquisizione è rigidamente collegato al telaio della carrozza. I valori misurati sono espressi rispetto ad un sistema di riferimento locale solidale con il telaio della carrozza. È in fase di studio la scelta del sistema più efficiente per il collegamento tra il sistema di riferimento locale (carrozza) ed il sistema di riferimento terreno (binario). I filmati acquisiti vengono archiviati con il loro riferimento metrico. Un sistema rapido per l'analisi dell'immagine si preoccupa di reperire le informazioni in merito all'esistenza di eventuali ostacoli in prossimità dell'area d'interesse; in caso di verificata presenza viene richiamata la sequenza di triplette dalla quale si ricavano accurate misure tridimensionali di controllo. Il sistema consente inoltre la generazione automatica di superfici raster tridimensionali.

Analisi trinoculare

L'approccio trinoculare proposto prevede algoritmi di autocorrelazione ed estrazione di punti e/o entità grafiche tridimensionali dalle immagini provenienti dalle telecamere installate sulla carrozza. Il software è basato su principi di computer vision per quanto riguarda l'analisi dell'immagine, la sua segmentazione, il reperimento di features, la ricostruzione di geometrie dal moto. Per quanto riguarda la correzione delle distorsioni, la triangolazione spaziale e la collimazione sub-pixel ci si affida a tecniche di tipo fotogrammetrico. Obiettivo finale è costituire una semina di punti tridimensionali nell'area di copertura reciproca e contemporanea dei fotogrammi. La semina dei punti ha una densità dipendente dai parametri di calcolo impostati, dalla qualità delle immagini, dalla loro risoluzione, dalla tessitura dei soggetti e dall'illuminazione della scena. Possono essere valutate le informazioni geometriche derivanti dai *frames* precedenti per poter procedere mediante

analisi differenziale sui successivi. Il calcolo è ottimizzato verso la rapidità di calcolo senza compromettere con ciò la precisione di misura.

La procedura è così sintetizzabile:

1. Elaborazione primaria: dedicata alla correzione delle distorsioni geometriche dell'ottica di presa; in conformità al certificato di calibrazione proprio di ogni camera, il software procede alla controdistorsione dell'aberrazione ottica delle immagini che ciascuna di esse produce.
2. Elaborazione secondaria: dedicata alla correzione epipolare dei fotogrammi. In particolare, sulla base dei dati derivanti dall'orientamento esterno (cioè posizionamento ed inclinazione delle fotocamere rispetto alla struttura), per ogni combinazione degli accoppiamenti (camera 1 con la 2, camera 2 con la 3, camera 1 con la 3), vengono calcolate le trasformazioni tali da ricondurre le rette epipolari di ciascuna immagine ad essere parallele alla sua coniugata.
3. Stereo matching trifocale: per ciascuno degli accoppiamenti sopra descritti, nel rispetto del volume d'indagine da analizzare, si procede ad una elaborazione in stereo matching piramidale. Successivamente si integrano i tre risultati parziali al fine di scartare gli outliers (punti palesemente sbagliati) ed infittire la griglia.

Analisi monocamera

Nel paragrafo precedente sono state analizzate le caratteristiche e le potenzialità del sistema di analisi trinoculare. L'elaborazione, per ogni istante di acquisizione, dei dati provenienti dalle tre telecamere ottiene un insieme di dati ed informazioni molto ricco, ma è dispendioso in termini di potenza di calcolo. Appare quindi ragionevole pensare di effettuare una prima stima dei potenziali ostacoli presenti sulla (o attorno) sede ferroviaria attraverso l'analisi dei dati provenienti da una singola telecamera (fig. 2.a) sfruttando i dati sul moto del sistema di ripresa forniti dalle apparecchiature a bordo del vagone. L'algoritmo messo a punto si basa quindi sulla stima ed analisi del flusso ottico tra due *frame* successivi della sequenza di immagini analizzata.

Il flusso ottico è definito, nel caso di telecamera fissa, come la proiezione nel piano immagine della telecamera dello spostamento tridimensionale subito dai punti della scena in due istanti successivi (Horn&Schunck, 1993). Nel nostro caso si può invece considerare la scena osservata stazionaria mentre il sistema di ripresa è in movimento. In questo caso il vettore di moto associato a ciascun punto dell'immagine è, in buona sostanza, inversamente proporzionale alla distanza fra l'oggetto considerato ed il sistema di ripresa.

L'Algoritmo proposto

Una sequenza di immagini a toni di grigio è rappresentata da una funzione intensità: $I(t; \mathbf{x})$, ove con \mathbf{x} si indica la coordinata bidimensionale che individua un punto sull'immagine. Per misurare il flusso ottico è d'uso supporre che un punto 3d proiettato nel piano immagine della camera abbia una intensità costante nel tempo, cioè:

$$\frac{d}{dt} I(t; \mathbf{x}(t)) = \frac{\partial I}{\partial t} + \frac{d\mathbf{x}}{dt} \cdot \nabla I = 0 \quad [1]$$

Utilizzando un approccio a correlazione si cerca di trovare corrispondenze tra finestre appartenenti a *frame* successivi (Falkenhagen, 1997). Se $W + \{\mathbf{x}_0\}$ è una finestra in un *frame* attorno al punto \mathbf{x}_0 , il metodo impiegato consiste nel cercare un vettore $\mathbf{v}(\mathbf{x}_0, t)$ tale che l'errore:

$$E = \int_{W + \{\mathbf{x}_0\}} |(I(t+1; \mathbf{x} + \mathbf{v}) - I(t; \mathbf{x}))|^2 d^2x \quad [2]$$

sia minimo (Nishihara, 1990). Il flusso ottico è calcolato quindi spostando, sull'immagine relativa all'istante $t+1$, il centro della finestra W di osservazione e cercando la posizione in corrispondenza della quale l'errore definito in [2] sia minimo. Come già indicato, i moduli dei vettori del campo così ottenuto sono inversamente proporzionali alle distanze dei particolari osservati dal piano immagine della telecamera (Ancona&Poggio, 1993). Dato il particolare tipo di moto cui è soggetta la camera, i vettori di moto sono orientati in modo sostanzialmente radiale con fuoco di espansione

localizzato nel punto di fuga delle rette che descrivono i binari. Ciò rende, da una parte, problematica la stima dei vettori di moto e dall'altra parte rende ardua la loro interpretazione in termini di elevazione dei particolari osservati dal piano ferroviario o comunque di ingombro all'interno della sagoma definita dalle dimensioni dei vagoni. Prima di effettuare la stima del flusso ottico è quindi applicata alle immagini una trasformazione omografica (Hartley, Zisserman, 2000) tale da annullare le deformazioni prospettiche per i particolari posti su di un piano preselezionato, ad esempio quello su cui giacciono i binari (ovviamente nel caso di una tratta piana, fig. 2.b). In questo modo il flusso ottico associato alle immagini così modificate ha un'unica direzione e le proiezioni dei punti 3d appartenenti al piano preselezionato sono caratterizzati da spostamento (fra le due immagini considerate) nullo. La ricerca del minimo della relazione [2] risulterà più efficiente. Il modulo poi di tale flusso in ogni punto sarà direttamente legato alla distanza dal piano di riferimento (fig. 2.c). Nel caso in esame, il procedimento illustrato relativamente al piano ferroviario è esteso ad altri 3 piani che delimitano una sorta di corridoio di transito. In questo modo l'analisi efficiente dei flussi ottici dà immediatamente la distanza da tali riferimenti e quindi la potenziale presenza di ostacoli al transito del materiale ferroviario, e può essere una sorta di campanello di allarme per un'analisi più accurata con il sistema trinoculare.

Risultati del processing monoculare

Come già indicato, riportando per ogni punto il modulo dei vettori del flusso ottico, ottenuto come descritto in precedenza, si ottiene un'immagine (di disparità) in cui zone più chiare corrispondono a oggetti più sporgenti rispetto ai quattro piani di riferimento. I risultati ottenuti sono incoraggianti; alcuni di questi sono mostrati nelle figg. 3-4. L'accuratezza dei dati così ottenuti è fortemente legata, da un parte alla qualità del processing delle immagini, ma d'altra dalla precisione con cui è noto il moto della camera (comprese oscillazioni e beccheggi del vagone). In questa prima fase sperimentale, poiché le informazioni di moto non erano disponibili esse sono state stimate a partire dalla posizione sulle immagini dei bulloni che legano fra loro binari e traversine (la distanza fra le traversine è fissa e nota).

Qualora in futuro le informazioni di moto del vagone non siano disponibili (o non siano abbastanza precise) sarà possibile sviluppare un'applicazione automatica per la stima di tale moto direttamente dall'analisi delle immagini basandosi appunto su informazioni relative alla geometria della sede ferroviaria.

Conclusioni

In questo contributo è stata presentata una tecnica per l'analisi di immagini provenienti da una camera montata su di un carrello ferroviario. La disparità presente fra coppie di immagini dovute al moto della camera è stata efficacemente utilizzata per stimare gli ingombri degli elementi 3D presenti nelle vicinanze della sede ferroviaria. L'uso di una appropriata trasformazione delle immagini permette di contenere i costi computazionali pur a fronte di buone qualità delle informazioni 3D ottenute. Sebbene in questo lavoro siano stati utilizzati i dati provenienti da una sola camera, il sistema hardware considerato è basato su tre camere sincrone montate rigidamente su di un apposito vagone. I dati ottenuti dall'analisi monocamera potrebbero quindi essere utilizzati come inizializzazione per il processamento delle immagini provenienti dalla tripletta di camere. Ciò permetterebbe di migliorare ulteriormente la qualità dei dati 3D contenendo in ogni caso il costo computazionale dell'elaborazione. La possibilità di acquisire dati ad una velocità compatibile con le normali velocità di transito dei convogli ferroviari, la rapidità di elaborazione e la snellezza dei dati, rendono questa ipotesi competitiva anche rispetto ad eventuali tecnologie in grado di acquisire nuvole di punti in tempo reale.

Bibliografia

- Horn B.K.P., Schunck B.G. (1993), "Determining optical flow: a retrospective", *Artificial Intelligence*, 59(1-2), pp.81-87.
- Ancona N., Poggio T. (1993), "Optical Flow from 1D Correlation: Application to a Simple Time-to-Crash Detector". *Proceedings of the Fourth International Conference on Computer Vision*, Berling, Germany, May 1993.
- Camus Ted A. (1995), *Real-Time Optical Flow, Ph.D. Thesis.*
- Falkenhagen L. (1997), "Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints". *Proc. International Workshop on SNHC and 3D Imaging*, Rhodes, Greece.
- Hartley R., Zisserman A.(2000), "Multiple View Geometry in Computer Vision". Cambridge University Press.
- Nishihara H. K. (1990), *Real-Time Implementation of a Sign-Correlation Algorithm for Image-Matching. Technical Report*, Teleos Research, February 1990.

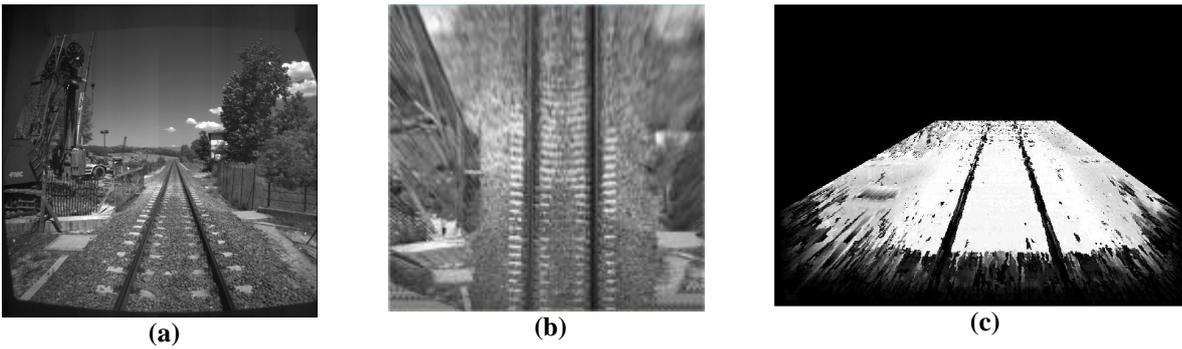


Figura 2: (a) Una generica immagine della sequenza acquisita da una delle telecamere poste sul vagone per il rilievo. (b) Risultato dell'applicazione dell'omografia di trasformazione all'immagine (a), si noti come il punto di vista sia ora "dall'alto". (c) Immagine di disparità rispetto al piano orizzontale ottenuta da due *frame* successivi della sequenza e riportata nel punto vista originale della telecamera.

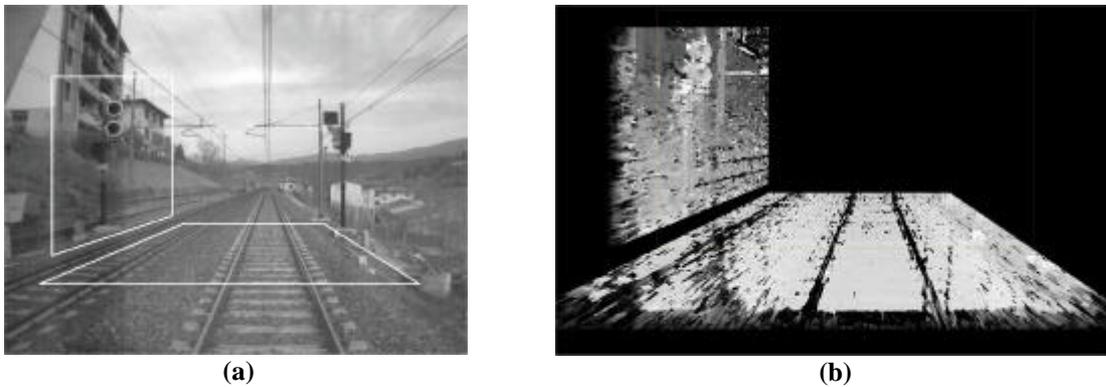


Figura 3: (a) Una immagine della sequenza "Rignano" in cui sono evidenziati due dei piani di riferimento. (b) I piani evidenziati in (a) elaborati, si noti come il semaforo sul piano di sinistra appaia chiaro in conseguenza della sua sporgenza rispetto al piano di riferimento.

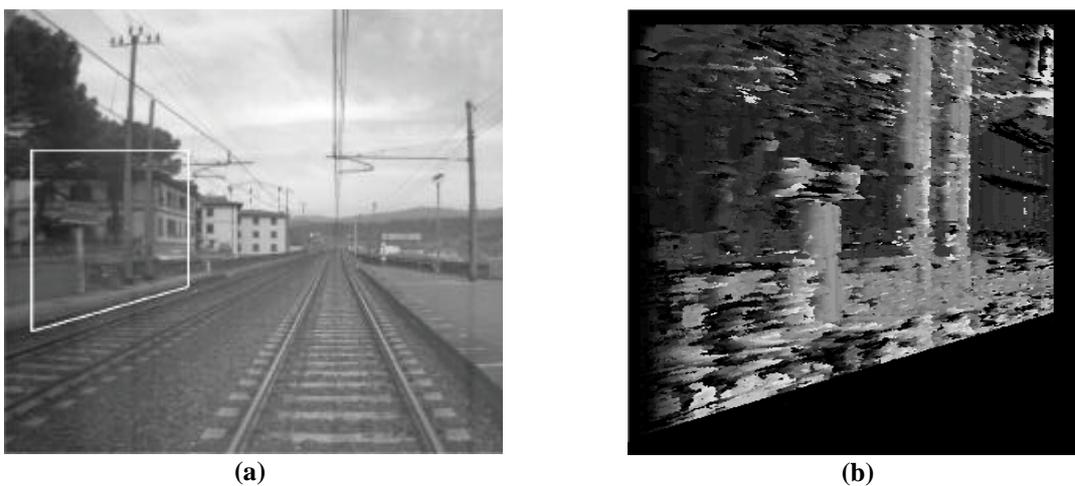


Figura 4: (a) Un'altra immagine sempre dalla sequenza "Rignano" in cui è evidenziato uno dei piani di riferimento. (b) Si noti come il cartello con il nome della località venga correttamente evidenziato.