

MISURE ATTRAVERSO SISTEMI DI VISIONE

Introduzione

Per introdurre i metodi di misura attraverso sistemi di visione occorre innanzitutto comprendere il processo di formazione dell'immagine. Tale processo può essere inteso come il processo opposto alla visione:

- Formazione dell'immagine = con l'espressione formazione dell'immagine si intende quel processo che permette di ottenere dall'oggetto un'immagine. In questa fase si perdono delle informazioni poiché si passa da una situazione tridimensionale a un'immagine bidimensionale. Generalmente l'immagine viene acquisita da una telecamera
- Visione = con l'espressione visione si intende invece quel processo che permette di dedurre un oggetto dall'analisi di un'immagine.

Visione umana e artificiale

Esistono due principali tipi di percezioni:

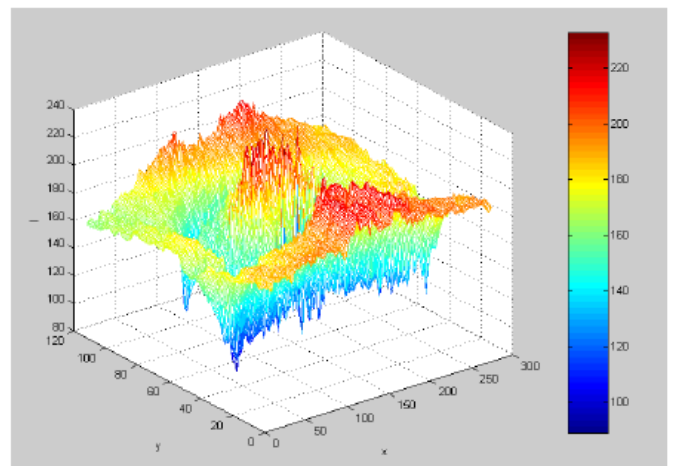
1. Percezione visiva o visione umana:
 - compresa tra i 400 e i 700 nm
 - ha una risoluzione, cioè la capacità di riconoscere diversi livelli di luminosità, pari a 50 livelli
 - gli algoritmi di visione umana sono molto efficaci infatti permettono di riconoscere velocemente i singoli oggetti
 - Gli occhi umani hanno tridimensionalità nativa
2. Visione automatica o artificiale
 - è in funzione del sensore
 - ha una risoluzione in funzione del convertitore
 - gli algoritmi artificiali sono poco efficaci
 - la visione artificiale è generalmente bidimensionale.

Immagini digitali

Analizziamo ora come viene tradotta digitalmente un'immagine. L'immagine può in realtà essere rappresentata in modi diversi:

1. In forma grafica come una superficie che è pari alla dimensione dell'immagine e il cui colore è in funzione dell'intensità luminosa
2. Come un array di numeri
3. Come immagine vera e propria

Nel nostro caso considereremo sempre un'immagine come un array bidimensionale di luminosità. L'intensità luminosa assume valori specifici in funzione delle coordinate spaziali. Le due variabili di posizione x e y potranno assumere solo valori interi positivi inferiori a due valori prefissati in funzione dello standard. Questo significa che l'immagine verrà divisa in una matrice di celle aventi la stessa dimensione ognuna delle quali verrà detta pixel. Altra cosa importante da sottolineare è che il fluire continuo della realtà verrà sostituito con una sequenza di immagini statiche acquisite ad una frequenza, funzione dello standard utilizzato. I valori prefissati dagli standard sono i valori minimi che, sfruttando il tempo di persistenza delle immagini sulla retina, danno come risultato la sensazione di vedere una scena in movimento. L'immagine statica è il risultato dell'integrazione della luce proveniente dalla scena per un tempo, detto shutter, che normalmente ha una durata di $1/50$ di secondo ma può assumere anche valori decisamente inferiori (valori normali sono $1/10000$). Agendo su questo parametro è possibile acquisire immagini di oggetti in movimento. Ogni immagine acquisita è caratterizzata quindi da una matrice aventi valori di luminosità. La maggior parte delle volte in campo industriale si utilizzano telecamere in bianco e nero, per questo motivo non si parla di



- 0 nero
- 255 bianco
- Valori intermedi grigi

- Tutti e tre i colori al massimo bianco
- Tutti colori al minimo nero
- Valori intermedi altri colori

Esistono molti tipi di sistemi di acquisizione delle immagini. Noi tratteremo le telecamere del visibile che sono in grado di acquisire immagini bidimensionali o matriciali (creano immagini rettangolari) o lineare in sequenza (scanner e fotocopiatrici si acquisisce una serie di punti). I sensori presenti nelle videocamere per l'acquisizione di immagini attualmente sul mercato che usano tecnologie di traduzione del segnale in intensità luminosa diverse, sono:

-
- Figure 1 is a line graph showing the relative sensitivity of the human eye and a CCD camera across a range of wavelengths from 400 nm to 1000 nm. The y-axis represents 'Relative sensitivity' from 0 to 1.0, and the x-axis represents 'Wavelength nm' from 400 to 1000. The 'Human Eye' curve (blue) starts at 0 at 400 nm, rises to a peak of 1.0 at 555 nm, and then gradually declines to 0 at 700 nm. The 'CCD Camera' curve (red) starts at 0 at 400 nm, rises to a peak of 1.0 at 650 nm, and then declines to 0 at 800 nm. A secondary peak for the CCD camera is visible at 450 nm with a sensitivity of approximately 0.4.
- | Wavelength (nm) | Human Eye Sensitivity | CCD Camera Sensitivity |
|-----------------|-----------------------|------------------------|
| 400 | 0.00 | 0.00 |
| 450 | 0.05 | 0.40 |
| 500 | 0.30 | 0.05 |
| 555 | 1.00 | 0.10 |
| 600 | 0.80 | 0.80 |
| 650 | 0.40 | 1.00 |
| 700 | 0.00 | 0.40 |
| 800 | 0.00 | 0.00 |

- Elevata qualità dell'immagine ripresa
 - Stabilità della geometria dell'immagine nel tempo
 - Ottimo rapporto segnale/disturbo
 - Basso costo
 - Possibilità di ottenere risoluzioni elevate
- CMOS = complementary metal oxide semiconductor . Sono stati introdotti recentemente. Si basano su un wafer di silicio che è sensibile alla luce. Cambia dunque la tecnologia: analizzando le variazioni delle caratteristiche del silicio si può misurare l'intensità di luce. Il costo di questi sensori è modesto ed ha reso possibile la diffusione delle fotocamere nei cellulari e nei giocattoli. Inoltre la possibilità di avere acquisizioni ad alta velocità ha permesso lo sviluppo di telecamere per lo studio degli eventi velocità quali i crash test delle auto. I CMOS presentano una serie di vantaggi rispetto ai CCD cioè:
 - Accesso casuale ai pixel = permette di indirizzare il singolo pixel in quanto tutte le funzionalità di conversione sono integrate attorno ai singoli fotodiodi. Questo permette di definire delle ROI cioè delle regioni di interesse che, al diminuire delle loro dimensioni, permettono di incrementare significativamente il numero di immagini al secondo a valori non compatibili con la tecnologia CCD. Si può dunque chiedere al sensore di rilevare una sola striscia di intensità luminosa e non tutta l'immagine.

- Maggiore dinamica = capacità di vedere contemporaneamente zone molto scure e zone molto chiare. La tecnologia CMOS permette di implementare una funzione di trasferimento logaritmica tra l'intensità della luce incidente e la tensione in uscita così da catturare correttamente delle scene che si caratterizzano per importanti differenze di luminosità nella stessa immagine. Possono quindi avere risposta lineare o logaritmica o lineare e logaritmica (linlog).
- Costo decisamente inferiore
- Livello di integrazione decisamente superiore
- Consumi decisamente inferiori
- Possibilità di integrare funzionalità accessorie nel chip e di utilizzare fonderie di silicio standard
- Caratteristiche anti-blooming e anti-smearing dovuto alla conversione cariche/tensione direttamente nel pixel

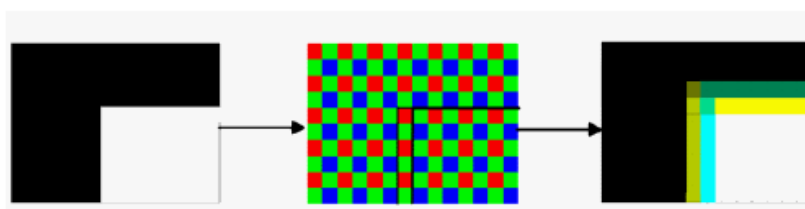
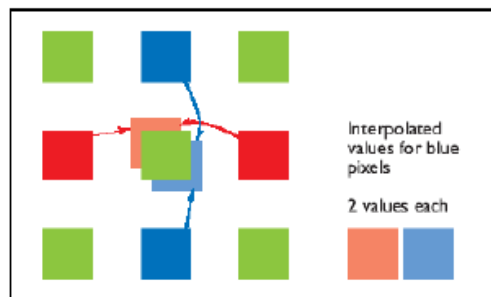
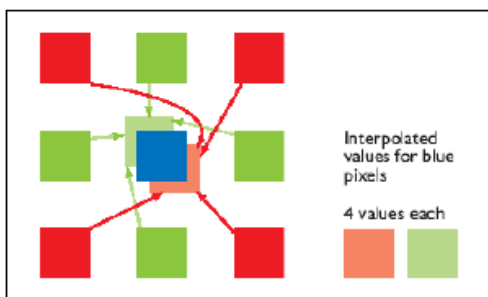
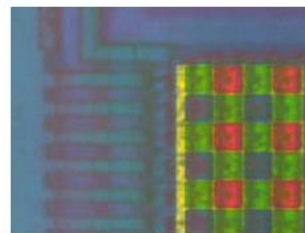
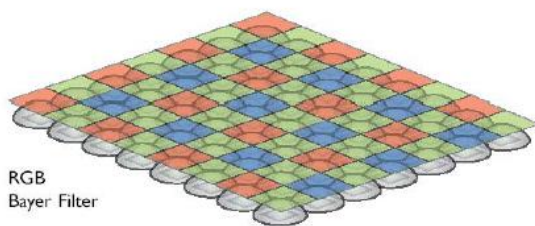
I due principali svantaggi invece sono:

- Rapporto segnale/disturbo
- Peggior sensibilità

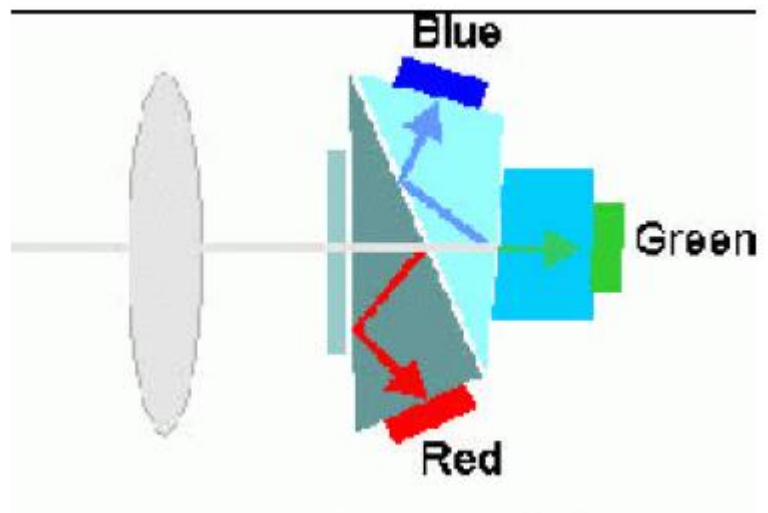
Sensori matriciali a colori

Le immagini a colori possono essere ottenute mediante diversi metodi; in tutti i casi è però fondamentale riconoscere le lunghezze d'onda:

1. Sensore matriciale con filtro a mosaico = ogni pixel ha un filtro di colore diverso. Il pixel potrà acquisire solo l'intensità di luce di una delle tre componenti in funzione del colore del proprio filtro. Generalmente il 50% dei pixel ha un filtro verde mentre il 25% di blu e il restante 25% di rosso. Questo perché il verde è il colore più sensibile all'occhio umano. Dal momento che ciascun pixel acquisisce la luminosità di una sola componente cromatica, l'immagine acquisita non contiene tutte le informazioni (Servono 3 componenti cromatiche per ciascun pixel per avere l'intera immagine a colori). Per ottenere le 3 informazioni cromatiche per ciascun pixel, si usa una interpolazione locale. Questo processo però crea problemi nelle zone dell'immagine dove ci sono evidenti gradienti cromatici locali. In ambito consumer non da problemi.



2. Prismi e tre sensori monocromatici filtrati = divide la luce in tre parti uguali, ogni sensore vede la stessa scena ma filtrata con un colore diverso. Il sistema permette di mantenere la piena risoluzione ma occorre usare tre sensori ed allinearli in modo molto accurato (immagine).
3. Più immagini singolarmente filtrate = prendo una telecamera in bianco nero e metto in sequenza tre filtri. Faccio tre foto ottenendo tre componenti cromatiche, l'oggetto deve essere perfettamente fermo. Ha interesse storico o scientifico.
4. Sensori matriciali RGB = tipo di CMOS che è realizzato a tre strati.



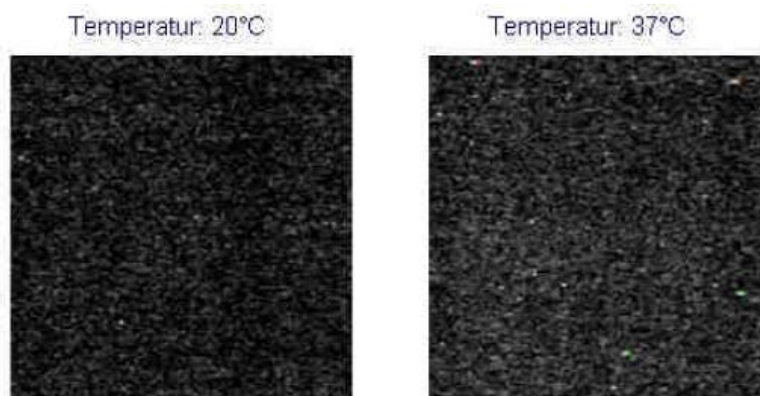
Sensori lineari

I sensori lineari sono caratterizzati dal fatto che ha un migliaio di righe a fronte di un numero di colonne di molto inferiore. Il sensore è formato da tre colonne ciascuna dedicata all'acquisizione di una singola componente cromatica. Questi sensori sono comunemente utilizzati negli scanner e nelle telecamere lineari a colori. Il controllo dell'esposizione è indipendente per ogni canale in modo da permettere già in acquisizione il bilanciamento dei colori. L'area sensibile dei sensori CCD viene misurata in pollici prendendo come riferimento un "tubo equivalente" di una certa altezza e lunghezza. Se il sensore è più grande ha la possibilità di acquisire molta più luce quindi può avere una maggiore sensibilità e di conseguenza avere tempi di esposizione minori. Gli svantaggi del sensore grande è però il costoso e la necessità di un'ottica in grado di focalizzare la luce su una superficie maggiore.

Problemi dell'acquisizione delle immagini

Le immagini acquisite possono presentare una serie di problemi quali:

- Effetto Moiré = noto anche come problema di aliasing spaziale. Si verifica ogni volta che nelle immagini sono presenti delle tessiture tanto piccole che non possono essere risolte a sufficienza dal sensore. Per risolvere questo problema i costruttori di sensori prevedono dei filtri che diffondono la luce incidente in modo che questo effetto non possa verificarsi.
- Rumore termico = questo problema si presenta poiché il sensore è un circuito elettrico. Tutti i dispositivi semiconduttori soffrono infatti della presenza di elettroni indesiderati legati all'effetto termico. Generalmente maggiore è la temperatura maggiore è il rumore termico. È evidente tale problema se si fa una foto con l'obiettivo chiuso: l'immagine non risulterà completamente nera ma presenterà una serie di piccoli punti bianchi che aumentano se si rieflettua una foto a temperature più elevate. Per evitare questo problema si utilizzano dei sensori raffreddanti.



- Blooming = si verifica solo nei CCD. Il blooming è l'effetto per cui il contrasto dell'immagine risulta compromesso a causa della presenza di un oggetto molto luminoso. Fisicamente accade che quando un CCD (non CMOS) cattura troppi elettroni questi tendono a passare anche nei pixel adiacenti ottenendo come risultato un sovradimensionamento dell'area bianca presente nell'immagine. Il pixel si carica completamente e quindi riversa l'eccesso di carica in quei pixel a lui adiacenti creando un effetto onda di disturbo nelle immagini.
- Smear = anche questo problema si verifica esclusivamente nei CCD. È un effetto per cui uno spot molto luminoso in un'immagine genera una linea bianca verticale; è dovuto a motivi costruttivi dei sensori CCD ed al processo di trasferimento verso l'elettronica di condizionamento delle caratteristiche generate nei pixel dall'interazione con la luce. Il metodo più efficace di eliminare lo smear è quello di dotare il sensore di un otturatore meccanico che oscuri il sensore dopo l'esposizione.

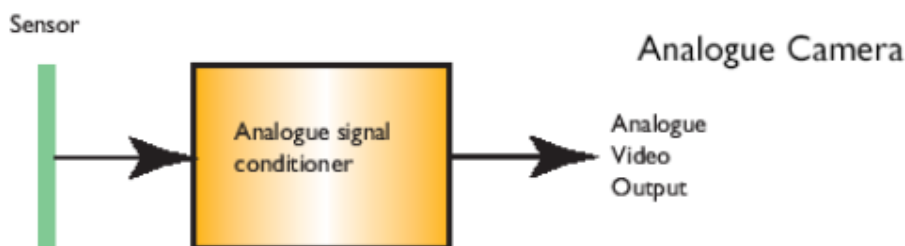
Trigger

La cosa interessante di molte telecamere industriali è la possibilità di usare un ingresso trigger; in questo modo quando queste ricevono un segnale in tensione di un valore compreso in un certo range scattano un'immagine. Lo scopo del trigger è quello di acquistare l'immagine in modo sincronizzato con altri eventi. Usando il trigger si può effettuare un'analisi più mirata e ottimizzare l'acquisizione delle immagini focalizzando ad esempio la luce.

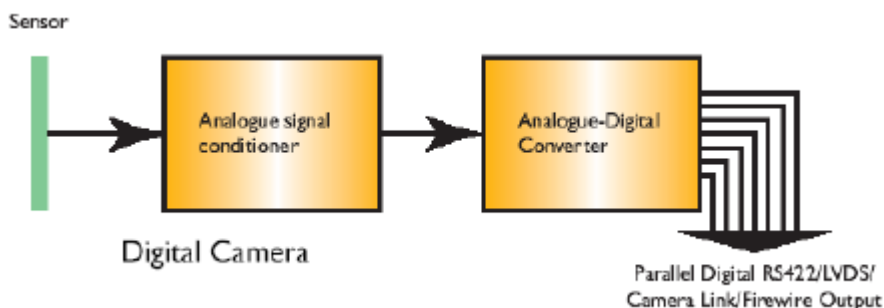
Tipi di telecamere

Esistono due diversi tipi principali di telecamere:

- Analogiche = rappresentano una soluzione molto economica anche se poco diffusa ormai in ambito industriale. Oggi vengono usate solo nei sistemi di sorveglianza.



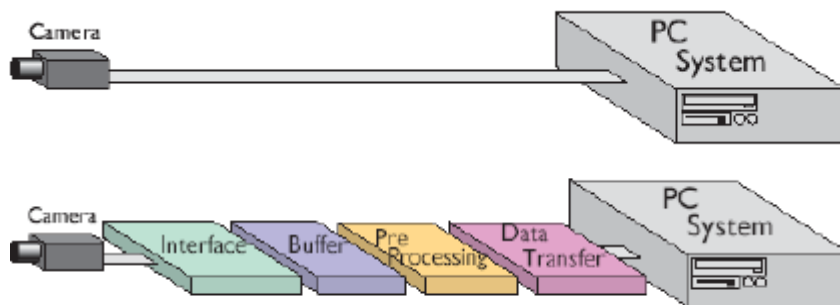
- Digitale = la sezione di acquisizione è simile a quella delle telecamere analogiche ma integrata con un convertitore analogico digitale. Questa soluzione permette di ottenere immagini di qualità superiore e di utilizzare un cablaggio più semplice ed economico. Le telecamere digitali possono avere diversi tipi di bus di uscita, attraverso il bus inviano le immagini acquisite al computer di controllo o al sistema di acquisizione e memorizzazione delle immagini. Vediamo ora i bus digitali più diffusi per collegare una telecamera al computer o al sistema di acquisizione delle immagini.



Le telecamere vengono interfacciate dunque verso un elaboratore in due diversi modi:

- Direttamente tramite:
 - USB = il vantaggio di questo collegamento diretto è che si ha una diretta alimentazione

- Firewire = standard di comunicazione come bus per l'interfaccia di periferiche a personal computer. Vantaggi come plug and play. Lo svantaggio è che il tempo di trasferimento dell'immagine dalla telecamera all'elaboratore non è deterministico.
- Gigabit ethernet = dette anche TCP-IP. Le telecamere si collegano all'elaboratore tramite una rete Ethernet. Può essere fatto con una struttura di rete oppure avere una telecomandata collegata direttamente al pc. La differenza è insita nel tempo. Il primo caso va bene quando si devono acquisire poche immagini al secondo. Se invece occorre acquisire velocemente conviene usare la seconda configurazione. Il vantaggio di questa soluzione è l'utilizzo della stessa tecnologia delle reti; ciò permette di realizzare topologie anche molto complesse e di coprire anche notevoli distanze. Come già detto lo svantaggio principale è rappresentato dal tempo di acquisizione di un'immagine non deterministica in quanto dipende dalla banda disponibile e dal traffico di rete
- Tramite hardware dedicato:
 - Camera link = è la soluzione più completa ma anche la più costosa. Essendo uno standard fatto apposta per il trasferimento delle immagini si interfaccia direttamente con la telecamera. Non si passerà più per la CPU. È in grado di serializzare dei dati paralleli in modo da poterli trasmettere con un numero limitato di cavi.

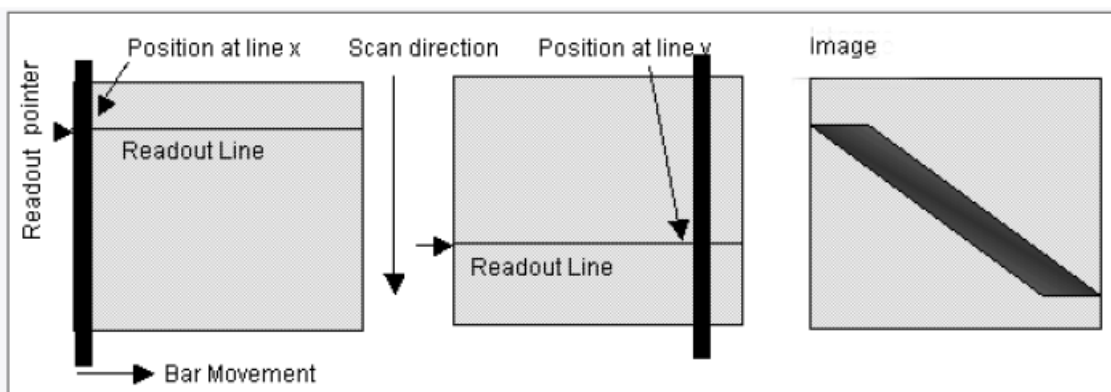


Global shutter and rolling shutter

Riprendendo i sensori CMOS dobbiamo sottolineare che non tutti sono costruiti nella stessa maniera. La differenza tra i tipi di CMOS è insita nella sequenza con cui acquisiscono le immagini. Le due categorie principali:

- Global shutter = tutti i pixel del sensore vengono esposti alla luce
- Rolling shutter = per motivi di costo anziché leggere tutti i pixel insieme si interrogano i pixel del sensore uno alla volta. Quindi non vengono esposti tutti i pixel contemporaneamente ma lo si fa in rapidissima sequenza.

Generalmente se la scena è ferma si può utilizzare un global shutter o un rolling shutter indifferentemente. Se invece la scena da riprendere è in movimento si dovrebbe prediligere un'acquisizione con una global shutter poiché l'acquisizione con la rolling shutter potrebbe comportare una ripresa distorta; nel momento infatti in cui inizia la scansione l'immagine da riprendere può modificarsi e dunque l'immagine scansionata risulta essere distorta da quella iniziale.



Pre-processing delle immagini

Introduciamo ora il concetto di pre-processing cioè il procedimento necessario per preparare le immagini alle successive analisi da cui si otterranno le misure, Il pre processamento serve dunque per:

- Evidenziare i dettagli di interesse nell'immagine
- Ridurre l'effetto del rumore
- Eliminare distorsioni dell'immagine dovute ad esempio ad aberrazioni ottiche oppure a prospettiva

Il pre-processamento può essere di vari tipi:

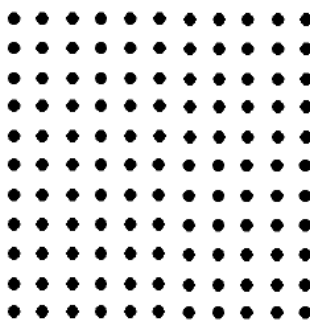
- Sul singolo pixel = si elabora un pixel alla volta indipendentemente dai valori dei pixel circostanti. Esempio di queste operazioni è ottenere il negativo di un'immagine.
- Sull'intorno del pixel = per elaborare un generico pixel dell'immagine si tiene conto anche dei valori di luminosità dei pixel vicini. Generalmente questa elaborazione viene effettuata quando si desiderano eliminare i disturbi ad alta frequenza. Si può anche usare per creare un filtro Sobel cioè uno strumento che rileva i contorni orizzontali e verticali separatamente in un'immagine a scala di grigi. Questo filtro serve quindi per identificare i bordi degli oggetti.
- Sull'intera immagine = in questo caso la trasformata di Fourier in 2D permette di individuare componenti periodiche e rimuoverle o amplificarle. Questa trasformata considera dunque l'immagine come una funzione a due gradi di libertà che esprime la luminosità di ogni punto e la sviluppa in serie. Generalmente viene utilizzata per rimuovere un disturbo

Una volta eseguito il pre processamento avviene la fase di elaborazione delle immagini che ha lo scopo di estrarre, dall'enorme mole di informazioni a disposizione, quella significativa. Ad esempio:

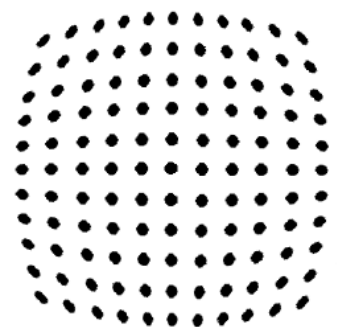
- Presenza di un oggetto
- Orientazione di un marchio o di un qualsiasi oggetto
- Misura di una certa caratteristica = le immagini sono di fatto delle proiezioni di un mondo tridimensionale. Utilizzando una singola immagine non si possono effettuare misure tridimensionali ma solo monodimensionali o bidimensionali. Per misure tridimensionali occorrono almeno due immagini acquisite da punti di vista non coincidenti

L'esecuzione di misure su immagini necessita della massima attenzione verso le possibili distorsioni delle stesse che possono essere introdotte a più livelli della catena di misura. Rimane comunque la necessità di inserire una operazione di taratura in modo da:

1. Potere esprimere le misure effettuate in pixel in coordinate reali mondo. Occorre dunque tarare le telecamere e quindi determinare un fattore di conversione tra i pixel e i millimetri. La taratura della telecamera è legata alla situazione specifica di applicazione.
2. Compensare eventuali aberrazioni ottiche = il fenomeno dell'aberrazione può essere dovuto ad un'ottica angolare o alla prospettiva. Dunque anche la prospettiva genera una distorsione. Per compensare le aberrazioni si possono applicare delle deformazioni all'immagine con la tecnica detta del warping: si pone davanti alla telecamera una griglia di punti e si acquisisce l'immagine di una griglia indeformato costituita da punti equispaziati. Il foglio viene poi appoggiato sulla superficie che si vuole fotografare e poi si acquisisce l'immagine che sarà caratterizzata da una griglia distorta e deformata. Si prende dunque questa seconda immagine e si applica una deformazione di warping così da distorcere la griglia acquisita per riottenere la griglia regolare iniziale. I parametri del warping che servono per ottenere la griglia regolare sono quelli che si dovranno applicare anche alle altre immagini acquisite dalla stessa telecamera nella stessa posizione per compensare le aberrazioni e la prospettiva.



Griglia da acquisire



Griglia acquisita (Distorta)

Misurare con le immagini

Le coordinate dei singoli punti sono ricavate con metodi ottici e quindi senza contatto nel senso meccanico del termine. È ad esempio possibile fare le seguenti misure:

- Misura di un punto singolo = misura lo spostamento lungo una direzione, uno strumento che effettua tale misura è il laser a triangolazione.
- Misura dei punti contenuti in un piano
- Misura dei punti su una superficie nello spazio (misure di forma 3D) = mettendo insieme più laser è possibile realizzare un sistema di misura che è in grado di misurare molti più punti lungo una linea, con questo metodo è possibile ricavare un intero profilo dell'oggetto.
- Misure di posizione/spostamento di oggetti nello spazio = estendono l'idea precedentemente esposta è possibile proiettare molte linee parallele e ottenere la misura di molti punti nello spazio. Dunque analizzando la deformazione delle ombre proiettate sull'oggetto si può risalire alla sua geometria.
- Misure di campi di spostamento e campi di deformazione, ovvero misure del moto e della deformazione di un elevato numero di punti appartenenti a uno o più oggetti.

Misure bidimensionali

Analizziamo ora più nello specifico gli algoritmi di ricerca/misure di spostamento bidimensionali:

- Edge detection = algoritmi che permettono di estrarre delle informazioni dalle immagini. Prevedono la ricerca all'intero dell'immagine di uno o più dettagli le cui coordinate esprimono la posizione degli oggetti di interesse. Un edge può rappresentare ad esempio il bordo di un oggetto.
- Blob analysis = Identificando i blob si ottengono informazioni sulla posizione ad esempio del baricentro, sul suo spostamento (misure di posizione o spostamento) ma anche area (misura di estensione ad esempio di un difetto, o di una cellula in immagini biologiche, o semplicemente oggetti su un nastro trasportatore...), e molte altre informazioni, fra cui l'allungamento, il momento di inerzia.
- Pattern matching a livelli di grigio = permette di ricercare all'interno dell'immagine tutti gli oggetti che hanno la stessa distribuzione di grigi.
- Pattern matching geometrico = ricerca oggetti con lo stesso contorno dei modelli e dunque cerca nelle immagini delle forme predefinite. Cerca i contorni.
- DIC cioè digital image correlation = è un algoritmo di misura di campi di spostamento/deformazioni/velocità nei fluidi. È una tecnica di analisi delle immagini che consente di misurare l'intero campo di spostamenti degli oggetti inquadrati. Gli spostamenti analizzati possono essere di qualsiasi tipologia: traslazione o rotazione. L'idea della DIC è quella di dividere l'immagine iniziale in molte sotto aree (di solito quadrate) e di cercare in una successiva immagine, dove è andata a spostarsi ciascuna sotto area dell'immagine iniziale. Una caratteristica molto interessante è che si ottiene un intero campo di spostamenti: non solo lo spostamento di un punto ma lo spostamento di migliaia (o anche decine di migliaia) di punti nell'immagine. Affinché la ricerca di ciascuna sotto area fra immagine iniziale e immagine successiva funzioni bene, è necessario che l'oggetto della misura abbia una superficie non uniformemente grigia ma con molte variazioni di luminosità. In genere vengono realizzate delle piccole macchioline così da ottenere diverse variazioni di luminosità. Misurando gli spostamenti, attraverso una derivazione numerica, si può stimare la deformazione, che ricordiamo essere la derivata spaziale degli spostamenti stessi. Per aumentare il numero di punti di misura per unità di area (senza ridurre troppo la dimensione delle sotto aree, per non ridurre la qualità della misura) si usa spesso la tecnica di sovrapporre parzialmente i subset vicini. È importante sottolineare che i subset, nel passaggio dalla prima immagine alle successive, possono traslare ma anche ruotare e subire deformazioni. Per questo motivo, per rendere più efficace la ricerca, anche il generico subset (sotto area da cercare nell'immagine successiva) si può deformare (usando delle funzioni di forma). Con il DIC si possono ricercare anche delle forme nell'immagine grazie alla funzione shape function. Esiste anche nella versione tridimensionale.

Misure tridimensionali

Le misure tridimensionali prevedono la ricerca all'interno delle immagini riprese da almeno due telecamere disposte stereoscopicamente di uno o più target le cui coordinate esprimono la posizione degli oggetti di interesse. Il target è spesso rappresentato da una sfera in quanto rappresenta l'unica figura geometrica invariante rispetto all'orientazione intesa come rotazione attorno ad un qualsiasi asse passante per il suo baricentro. La determinazione del baricentro avviene separatamente nelle due immagini stereoscopiche e quindi, per triangolazione, viene ricavata la sua coordinata tridimensionale. Analizziamo ora più nello specifico gli algoritmi di ricerca/misure di spostamento tridimensionali:

- Stereoscopia = permette di effettuare misure in 3D mediante una triangolazione utilizzando immagini fatte da punti di vista differenti. Conoscendo la posizione relativa tra le due telecamere si può ricavare la posizione dell'oggetto misurato. Le telecamere

Pose estimation

Una tecnica per stimare la posizione e l'orientazione nello spazio di un corpo rigido è la "pose estimation". Il concetto base è che analizzando come la telecamera vede un oggetto di geometria nota (analizzando dunque l'immagine) si può ricavare la posizione e l'orientazione dell'oggetto stesso. Matematicamente parlando all'oggetto che ha una sua matrice di coordinate si può applicare sia una matrice di traslazione che una di rotazione che combinate contribuiscono a fornire il prodotto tra la proiezione e la matrice dell'immagine stessa. Esistono diversi tipi di algoritmi di pose estimation:

- LuHuger = si riducono i tempi di calcolo così da cercare soluzioni rapide che potrebbero avere però delle ambiguità
- Pisinger-Mayer = ha incertezza maggiore ma non ha ambiguità
- Lager-Pismay = effettua più calcoli e quindi ha incertezza bassa
- Metodo poli = fornisce risultati molto precisi ma è più pesante computazionalmente