Corso di Visione Artificiale per Ingegneria Informatica a.a. 2004-2005
(pagina web: www.dsi.unifi.it/users/colombo/visart0405.html)

Carlo Colombo (tel. 0554796329, e-mail colombo@dsi.unifi.it)

GIORNALE DI BORDO

Indice generale

I.1/1-3 - *Venerdì 6 maggio 2005 (3 ore) [Introduzione generale]*

Visione artificiale o computazionale (computer vision). Far vedere un computer è difficile (aneddoto di Minsky). Svelare il funzionamento del cervello è una delle scommesse di questo millennio. Il cervello umano dedica gran parte delle sue risorse alla visione. La visione come complesso aggregato di moduli computazionali. Nel vedere un quadro, un grosso sforzo è dedicato a percepire condizioni di pericolo nell'ambiente (reptilian vision), non all'apprezzamento estetico del quadro. I livelli semantici di descrizione del quadro sono molteplici: una versione molto astratta è "San Giorgio e il Drago". Percepire la forma della superficie di supporto del quadro (il piano della tela) è uno dei compiti principali della visione umana e uno dei task caratteristici della CV. Un altro task caratteristico è il riconoscimento di oggetti, di cui quest'anno non ci occuperemo affatto. La percezione della struttura si può ottenere in molti modi: stereo, monocular+motion, analisi delle ombre ("shape from X"), informazioni sul contenuto della scena (presenza di angoli retti, mattonelle quadrate, etc.). Ci sono persone che non hanno la visione stereo, eppure non lo sanno. Immagine=forma+luce. Nel corso ci occuperemo principalmente di geometria della visione. Perdita di una dimensione, e molte altre ambiguità. Il problema è che bisogna estrarre l'informazione dalle immagini, e anche se non c'è texture, il mondo c'è comunque ma non è analizzabile (ad es. blank wall, o stanza al buio)! La prospettiva degli antichi: gli affreschi di Pompei (lo stile tardo è *più* lontano dai canoni prospettici moderni di quello antico). Brunelleschi reinventa la prospettiva. Masaccio: primo affresco in prospettiva (SMN), su direttive di Brunelleschi. Leon Battista Alberti: finestra albertiana, proiezione di un pavimento. I disegni di Dürer. Prospettiva da manuale: Piero della Francesca. Un virtuoso della prospettiva (e anticipatore di Picasso): Paolo Uccello (cfr. Giovanni Acuto). Critiche al modello prospettico: Leonardo. Alberti formalizza il modello della prospettiva (cfr. animazione Java) per consentire di rilevare il punto di vista ottimale da cui osservare il quadro: è un'idea molto moderna, che ci porta al concetto di autocalibrazione (cfr. esempio su affresco Masaccio – che doveva essere alto per la sua epoca, visto che l'altezza del centro ottico è 1,74m). Brunelleschi usava grandi quadrati di cotto nelle sue chiese per consentire la misurazione dello spazio, e fare sentire il visitatore a suo agio nell'ambiente. Meglio ancora: Brunelleschi pare abbia inventato la prospettiva come metodo sintetico per la rappresentazione di architetture e la misura di lunghezze ed angoli del mondo dall'immagine. Per vedere un quadro col massimo effetto prospettico bisogna usare un occhio solo. Centro ottico, punto principale, linea di orizzonte, punti di distanza. In prospettiva non si conservano gli angoli, né le distanze. (Un esempio di proiezione prospettica inversa è quello del videoproiettore: notare la generazione di un trapezoide, nel caso generale – cfr. keystone correction.) Rette parallele convergono verso un unico punto (di fuga). Il punto di fuga è la proiezione prospettica del punto improprio in cui si incontrano tutte le rette dello spazio che hanno una data direzione (e sono quindi parallele tra loro). Retta di fuga = formata da tutti i punti di fuga di una data giacitura di piani nello spazio; essa è la proiezione prospettica della retta impropria del piano. I punti impropri hanno terza coordinata 0, la retta impropria è (0,0,1). La linea di fuga è strettamente legata alla normale al piano. Se conosco due punti di fuga di rette di un piano, posso trovare la linea di fuga del piano (è quella passante per i due punti di fuga). Certi punti di fuga rimangono all'infinito (è il caso in cui le rette parallele nel mondo rimangono parallele anche nell'immagine: cfr. 1-point e 2-point perspective). Coordinate omogenee nel piano. Equazione della retta. Dualità. Retta per due punti. Punto di intersezione di due rette. Il caso particolare di rette parallele: il punto di intersezione è improprio, e rappresenta la direzione comune delle due rette. Anamorfosi: il quadro non è verticale, ma orizzontale. La forte angolazione dell'angolo di vista e del piano di formazione dell'immagine fa sì che il disegno, visto "a volo d'uccello", appaia fortemente distorto (slide madonnaro, "Gli Ambasciatori" di Hans Holbein). Proiezione prospettica definita da un piano e da un centro di proiezione. Calibrare significa riuscire a tracciare il raggio visivo per un punto del piano immagine. Questo consente di triangolare da due viste in un comune sistema di riferimento, qualora si conosca anche la disposizione relativa dei due sistemi di proiezione (geometria epipolare). Alcune applicazioni della CV: shape and camera from a single image (La flagellazione di Piero della Francesca e pavimento di scorcio), shape and camera from stereo (Photobuilder: notare l'approssimazione planare che fa schiacciare l'albero sulla parete dell'edificio), shape and camera from sequences (Pollefeys), video sequenze regularization by camera motion analysis ("2d3": pensare al mondo del cinema, e ai carrelli su binari).

I.2/4-6 - *Martedì 10 maggio 2005 (3 ore) [optical flow & feature extraction]*

Ambiguità della visione: la più evidente deriva dalla perdita della terza dimensione. I problemi di visione sono malposti, e richiedono soluzioni di regolarizzazione (cfr. libro T. Poggio). Illusioni ottiche: il cervello cerca di risolvere problemi visivi anche se mancano dati, col risultato che certe volte sbaglia perché applica ipotesi non corrette (tali ipotesi sono però in genere plausibili in.contesti reali, e non di laboratorio: molte illusioni ottiche sono generate con pattern artificiali, di laboratorio). Il cervello ricava informazioni a partire dalle immagini retiniche, e quando queste non forniscono dati sufficienti è un problema. Esempi: stanza al buio, blank wall, aperture problem. Barber pole (e animazione java): manca un'equazione, e il cervello regolarizza in modo errato. Optical flow e motion field: definizioni. OF constraint equation. Constraint line. Aperture problem rivisitato su un intorno di misura. Tre tipi di punti: intensità costante, gradiente monodirezionale, gradiente multidirezionale. L'OF si stima solo per questi ultimi. Per i secondi: normal flow. Si può regolarizzare ampliando la finestra di stima e facendo ipotesi sull'andamento dell'OF al suo interno (es. OF costante, o variabile linearmente, etc.). OF come risoluzione di un sistema sovradeterminato: pseudo-intersezione di linee di vincolo. Importanza della stima robusta in visione: i LS sono insoddisfacenti in presenza di outlier. Citati weighted LS e RANSAC. Visione umana e selezione dell'informazione: campionamento log-polare+movimentazione dell'occhio. La retina campiona in modo differente da una telecamera, riducendo l'informazione elaborata già a livello di sensore. I tre livelli di elaborazione: image processing, feature extraction (& pattern recognition), computer vision. In realtà i sistemi di visione lavorano sia in modo bottom-up (con un progressivo arricchimento della semantica) che top-down (verifica di ipotesi sui dati grezzi). Esempio di feature extraction: image segmentation. Features: edges (e Canny edgels), corners. Di corners ce n'è di meno che di edges, così come di punti di optical flow ce n'è di meno che di punti di normal flow. E, come si sa, ciò che è più raro è più prezioso, perché porta più informazione. Parentela tra la matematica dei corners (Harris) e quella dell'OF: servono comunque almeno due diverse direzioni di gradiente in un intorno di valutazione. La differenza è che nel caso di un corner non arriviamo a risolvere il sistema lineare, ma ci fermiamo a classificare le proprietà della matrice dei coefficienti attraverso i suoi autovalori (che non è necessario calcolare esplicitamente). Altre features importanti: curve immagine: active contours, snakes. Si lavora su un perimetro anziché su un'area, e dunque si risparmia computazionamente, tanto più che i punti rilevanti ai fini dell'elaborazione sono quelli in cui ci sono variazioni (l'occhio lavora "per differenza" – cfr. anche la background subtraction). Meccanismo di funzionamento degli snakes: minimizzazione di energia: interna (elastica)+esterna (legata ad un "potenziale immagine" opportunamente definito). Applicazioni: rilevazione della curvatura di oggetti nell'immagine (es. occhi, labbra etc. per riconoscimento), inseguimento visivo (ad es. per robotica, guida automatica, HCI – cfr. anche tempo all'impatto). (Video Oxford.)

I.3/7-9 - *Venerdì 13 maggio 2005 (3 ore) [omografie planari]*

Geometria proiettiva del piano: abbiamo già visto punti e rette, vediamo ora le coniche. Matrice 3x3 simmetrica, definita a meno di un fattore di scala: 5 dof. Trasformazioni proiettive di piani: collineazioni, o omografie. Tre esempi: vista di un piano, tra due viste di un piano, rotazioni attorno al centro ottico. Come si trasformano i punti, le rette, le coniche. Omografie: in generale 8 dof. La non linearità dipende tutta dalla terza riga della matrice. Non linearità in prospettiva: la proiezione del baricentro di una figura non equivale al baricentro della figura proiettata. Esempio: il dischetto del centrocampo in un video calcistico non è il centro dell'ellisse ottenuta come proiezione del cerchio di centrocampo. (Nota: è contro-intuitivo che un cerchio si proietti sempre in un'ellisse, mentre un quadrato si proietti in uno strano trapezoide. Ma se la vista è frontoparallel, anche i cerchi periferici restano cerchi.) Anatomia di un'omografia: trasformazione (0) euclidea, (1) metrica, (2) affine, (3) proiettiva. Felix Klein e la definizione di geometria basata sulle proprietà invarianti a classi di trasformazioni. Caratteristiche, gradi di libertà e invarianti delle trasformazioni 0–3. H=PAM. Solo P rende finita la linea all'infinito del piano (0,0,1), e finito un punto improprio (x_3=0); questo significa che qualsiasi trasformazione affine mantiene fissa la linea all'infinito, anche se non in senso "pointwise". Dunque la linea all'infinito è invariante rispetto a trasformazioni affini. Sotto P si conserva solo il cross-ratio. Rettificazione (a meno di una trasformazione) affine: si moltiplica H per P^{-1}, perdendo comunque (via la trasformazione affine) l'informazione di angolo tra rette. La rettificazione equivale alla ricostruzione, per i piani. Rettificazione metrica: basata sul calcolo di H con quattro punti noti (ma questa informazione è in realtà ridondante, visto che basterebbero quattro vincoli lineari su H per la rettificazione metrica – questa è l'informazione che viene dai punti circolari: vedi seguito). Esempio di rettificazione metrica: stima della posizione di giocatori sul campo di calcio: H si può calcolare o tramite punti, o tramite rette (almeno quattro in ogni caso). L'output della rettificazione metrica si chiama anche "ortofoto". Notare che solo i punti del piano si modificano correttamente: gli altri sono ancora più distorti! Punti e rette fisse: legati agli autovalori/autovettori della H. Nel caso generale solo tre punti fissi, e tre rette fisse (ma non pointwise). Nel caso di autovalori ripetuti, dipende dagli autovettori: se distinti, esistono linee di punti fissi (dunque fisse pointwise). Video sugli advertising logos. Esercitazione: VanishingDemo.m, RectDemo.m (per casa l'applicazione del logo al pavimento piastrellato).

I.4/10-12 - *Martedì 17 maggio 2005 (3 ore) [stima di omografie]*

Decomposizione SVD (esempio trasf. affine: valori singolari λ_1 e λ_2). Soluzione di un sistema lineare omogeneo con SVD. Stima di omografie: metodo DLT (usa la rappresentazione omogenea, e porta alla soluzione di un sistema lineare omogeneo). Il metodo DLT non è invariante a trasformazioni di similarità delle coordinate che esprimono i punti: bisogna dunque prima normalizzare i dati (ossia i due set di punti in corrispondenza tra loro). Altri metodi di stima più raffinati si basano su una minimizzazione non lineare: es. Levenberg-Marquardt applicato alla minimizzazione dell'errore di riproiezione, o del cross-transfer error [si darà elaborato]. (Stesso problema ci sarà con la stima della matrice F dello stereo, di cui H è la versione planare). Ancora sulla geometria del piano: coniche duali (line conics), pole-polar relationship e tangenza, proprietà "la linea all'infinito di una conica è la polare del suo centro". Coniche degeneri (rank=2 e rank=1). Punti circolari: forma canonica, proprietà (appartenenti alla linea all'infinito, invarianti a trasformazioni di similarità, intersezione tra linea all'infinito e conica assoluta, intersezione tra linea all'infinito e un qualsiasi cerchio, intersezione tra almeno tre cerchi, 4 dof). I 4 dof dei punti circolari sono quelli che consentono di fare misure nel piano a partire da un'immagine: esempi: stima dell'angolo tra due rette, e rettificazione metrica (si usa la conica duale dei punti circolari). Concetto di stratificazione: a partire dalla rettificazione affine si fa l'upgrade metrico; lo stesso si può fare con la ricostruzione 3D. (Esercizio non fatto: calcolare l'omografia rettificante metrica in termini di vanishing line e circular points – cfr. Liebowitz.) Nota: così come la linea all'infinito è invariante rispetto a trasformazioni affini, e proprio per questo consente di fare misure affini da immagini, così i punti circolari, che sono invarianti rispetto a trasformazioni di similarità, consentono di fare misure metriche. Mosaici: cosa sono, a cosa servono (compressione dell'informazione, aumento della risoluzione). Si va a stimare una trasformazione tra coppie di immagini, e si effettua la registrazione delle immagini rispetto ad un'immagine di reference. La trasformazione non è detto che sia un'omografia: potrebbe bastare un'affinità (video G. Baldi), se gli oggetti sono lontani o la focale è lunga. Viceversa, si viola il modello nel caso in cui si trasli rispetto a una scena 3D: occlusioni, parallasse. Rotazioni pure non danno luogo a parallasse, e per questo motivo non danno informazioni 3D: la scena 3D potrebbe essere dipinta sulla superficie di un cilindro o di una sfera, e il risultato sarebbe il medesimo. L'esempio dei piccioni e delle galline, che per stimare le distanze devono muovere avanti e indietro il capo. La stima automatica di trasformazioni usa tipicamente i corner, e metodi di ricerca di corrispondenze robusti (cfr. RANSAC). La ricerca di corrispondenze si basa tipicamente su correlazione: questa va bene per narrow baseline (ad es. sequenze di immagini), mentre va meno bene per wide baseline (singole foto); in quest'ultimo caso si userà una ricerca invariante rispetto a trasformazioni affini (SIFT: [si darà elaborato]). Demo Stitcher 4.0 + slides Oxford.

I.5/13-15 - *Venerdì 20 maggio 2005 (3 ore) [camera models & viste di piani*]

Punti e piani nello spazio. Trasformazioni proiettive (15 gdl). Quanti gdl ha una retta nello spazio (4) e un piano nello spazio (3). Rappresentazione di una retta nello spazio come linea tra due punti. Geometria della proiezione prospettica: matrice di calibrazione, con significato e importanza relativa dei suoi cinque elementi (si parte dai triangoli simili). Focal length, principal point, aspect ratio, skew. Natural camera (3 dof). Altri modelli di telecamera: radial distortion (più complesso), affine camera (più semplice). Coordinate di telecamera e world. Matrice di proiezione P=K[R **t**], con significato delle sue colonne. Se muovo la telecamera, variano solo i parameti esterni, mentre K è invariata. Il centro è il kernel (nucleo algebrico) della P. Backprojection con la P (tramite pseudoinversa) e con l'inversa della K. Formazione di punti di fuga come limite, e come proiezione di punti sul piano all'infinito. Legame tra K e punti di fuga: **v**=K**d**, con **d** vettore di direzione non omogeneo in coordinate di telecamera; la generalizzazione è **v**=(KR)**d**, con **d** vettore di direzione non omogeneo in coordinate world. La proiezione di punti del piano all'infinito (=direzioni nello spazio) non dipende dalla traslazione, ma solo da calibrazione e rotazione. Esempio: se si trasla, la luna (che è pensabile come appartenente al piano all'infinito) sembra seguirci, perché la sua immagine non cambia: cambia solo se prendiamo un binocolo, o giriamo la testa. La calibrazione consente di fare misure, ma in generale non consente di conoscere tutto di una scena (ad es. la scala relativa). L'importanza della lunghezza focale: se lunga, la proiezione diventa affine, e gli oggetti si ingrandiscono (zoom in) perché il campo di vista si restringe a parità di dimensione del sensore CCD. Viste di piani e omografie: il caso del piano Z=0 e quello del piano non passante per l'origine world (notare che **n** e **d** dipendono dalle coordinate world, e non dalla telecamera utilizzata). Omografia tra due viste dello stesso piano (se ne fa l'anatomia usando la formula di Sherman-Morrison): da questa omografia, se si conoscono le K, si trova (con ambiguità) la posa e la distanza del piano. Dimostrazione che se il centro di proiezione è fisso, la trasformazione è un'omografia comunque si cambi K e/o R, e qualunque sia la scena osservata (no parallax: dunque solo la traslazione porta informazione esterocettiva: cfr. piccioni e galline). Conica assoluta: è equivalente ai punti circolari del piano, e permette di fare misure metriche nello spazio, perché è invariante (ma non pointwise) a trasformazioni di similarità. (La sola conoscenza del piano all'infinito, che è invariante non pointwise a trasformazioni proiettive affini dello spazio, consente di fare misure affini nello spazio.) Relazione tra linea all'infinito, conica assoluta e circular points. L'immagine della conica assoluta (IAC) è una rappresentazione one-one (Cholesky) della K. Si dimostra che $\omega=(KK^T)^{-1}$ usando H=KR e applicando l'eq. di trasformazione delle coniche con $\Omega_\infty=I_{3x3}$ (forma canonica). Esempio di misura usando la K (omega): calcolo dell'angolo tra due raggi visivi corrispondenti a due punti immagine.

I.6/16-18 - *Martedì 24 maggio 2005 (3 ore) [Camera geometry]*

Ripreso RectDemo.m con la rettificazione di Lab1.jpg rispetto al piano del pavimento: tutto il resto è altamente deformato! Cenno alla rappresentazione plane+parallax. MosaicDemo.m con Lab1.jpg e Lab2.jpg: prima si registrano le immagini, e poi si fondono. La fusione non andrebbe fatta solo a livello geometrico, ma anche a livello cromatico: anche qui bisogna avere un certo numero di corrispondenze di colore, per poi stimare una trasformazione di colore da applicare ad ambo le immagini. Il mosaico automatico: si parte dalle corrispondenze "putative", e si usa un metodo di stima robusta di H, per rimuovere gli outlier (ad es. RANSAC, il cui numero di iterazioni è funzione della percentuale di outlier – tale metodo, che riprenderemo verso la fine del corso, raccoglie il consenso su un campione casuale minimale di 4 punti usato per la stima di H). Per stimare bene un parametro, bisogna raccogliere osservazioni nella maniera più sparpagliata possibile: interpolazione vs estrapolazione. L'errore non si distribuisce uniformemente, bensì in funzione della distribuzione delle osservazioni. Altro criterio generale: bisogna cercare di adoperare il modello più adatto: è metodologicamente sbagliato stimare un'omografia completa se sappiamo a priori che la trasformazione è affine (l'errore affliggerà tutti i parametri, dando luogo ad un errore complessivo maggiore). Ancora sul mosaico: la registrazione globale può dare luogo, nella "richiusura" di una traiettoria sul piano immagine, a disallineamenti dovuti all'accumulo di errori tra coppie di frame successivi: sarà necessario ridistribuire l'errore di stima su tutte le trasformazioni, ricalcolandole tutte, magari facendo in modo da rispettare vincoli di chiusura. Mosaici: planari, cilindrici (per cui serve la focale). Riepilogo sulla matrice di calibrazione: $\mathbf{d}=K^{-1}\mathbf{x}$ è il raggio visivo backprojected dal pixel $\mathbf{x}$, e sappiamo che l'angolo tra i raggi $\mathbf{d}_1$ e $\mathbf{d}_2$ si calcola con ω ed $\mathbf{x}_1$ e $\mathbf{x}_2$. Ma c'è di più: se al posto di $\mathbf{x}_1$ e $\mathbf{x}_2$ ho due punti di fuga $\mathbf{v}_1$ e $\mathbf{v}_2$, allora misuro gli angoli tra due direzioni qualsiasi dello spazio: tale angolo coincide infatti con quello formato dalle backprojection dei punti di fuga. In particolare, il vincolo di perpendicolarità è $\mathbf{v}_1^T\omega\mathbf{v}_2=0$, che fissa un gdl per la K. Se esplicito l'equazione di omega in funzione di K, trovo l'espressione degli elementi di omega in funzione dei parametri interni. In particolare, se i pixel sono quadrati (ossia skew zero e aspect ratio 1 -> natural camera), si ha che $\omega_{12}=0$ e che $\omega_{11}=\omega_{22}$. Altri tre vincoli su omega si possono scrivere se si ha il triangolo di fuga: e si può quindi calibrare la natural camera. Da ω si passa a K fattorizzando con Cholesky. CalibrateDemo2.m applicata a Cubo1.jpg: l'immagine è 800x600, e dunque il p.p. deve venire all'incirca (400,300). Caprile & Torre: "il p.p. è l'ortocentro del triangolo di fuga" (un'interpretazione geometrica leggermente più complicata sussiste per la focale). Calibrazione con tre omografie di tre piani disposti in modo generale nello spazio (cfr. HZ): si trovano sei images of circular points, per cinque dei quali passa una e una sola ω (ricordata la relazione sul piano all'infinito tra la AC e la $\mathbf{L}_\infty$ – il disegno non si potrebbe fare, perché la conica è a punti tutti immaginari…). Matematicamente, i vincoli si scrivono isolando le parti reali e immaginarie. Zhang: calibrazione con pattern planare che va mosso davanti alla telecamera: l'approccio è simile a quello appena visto. Stima automatica dei corner, gestisce anche la distorsione radiale. Si stimano prima i parametri di distorsione radiale, e ci si riconduce al modello pinhole. Come giudicare i risultati sperimentali: ci si basa sul concetto di ground truth, ossia sui risultati di un metodo che si ritiene più preciso di quello che si sta valutando (es.:

calibrazione Zhang vs calibrazione Tsai). Cenno al metodo Monte Carlo per la misura della dipendenza dal rumore. Consegnato l'articolo di Zhang e il software: uno degli studenti racconterà la prossima volta l'articolo mostrando il funzionamento del software.

I.7/19-21 - *Venerdì 28 maggio 2005 (3 ore) [Camera calibration]*

Fotogrammetria e computer vision: analogie e differenze. Ad es., non sempre si può adoperare un pattern (foto d'archivio, video amatoriali). Calibrazione fotogrammetrica (Tsai 1987): usa pattern 3D. Stima rette con Hough, e stima corner come intersezione tra rette. Calibrazione di P + distorsione radiale (tipicamente 4 parametri: centro e i primi due termini della serie): in tutto 15 parametri. Con la distorsione radiale, le rette non rimangono tali in proiezione. Si applica prima il metodo DLT stimando gli 11 gdl di P, come se non ci fosse d.r.; poi si raffina con un metodo non lineare (es. Levenberg & Marquardt). Metodi di stima non lineare: più accurati, ma devono essere inizializzati ("first guess": tipicamente ottenuta come soluzione di un problema lineare, qui DLT), sono più lenti perché iterativi, e si possono avere minimi locali (Ci sono meno minimi locali se la parametrizzazione è non minima). Usualmente, i metodi lineari minimizzano un errore algebrico, quelli non lineari un errore geometrico. N.B.: come già detto, col metodo DLT bisogna normalizzare i dati! Finite camera, o modello "pinhole": è il modello caratterizzato dalla sola P (non tiene conto delle lenti, che se non di buona qualità danno luogo alla d.r.). In realtà la proiezione inverte l'immagine "upside-down", ma il modello matematico che si usa più spesso è quello non invertente, che tanto porta la stessa informazione. Due sistemi reali (presentazione di D. Comanducci): varianti di un sistema di acquisizione low cost basato su lama laser (desktop laser scanner). Sistemi a luce (bianca o laser) strutturata: semplificano il low-level processing e l'estrazione di features. Ipotesi: velocità del tavolo rotante costante. Sistema 1 (LIMA3D): calibrazione con pattern di Tsai, calcolo dei coefficienti del piano laser in coordinate world, calibrazione turntable (stima centro di rotazione con equazione pole-polar centro/ellisse, stima #frame per giro completo), ricostruzione con backprojection del profilo immagine, e intersezione di ciascun raggio retroproiettato con il piano laser. (N.B. addendum alla lezione precedente: la backproiection di una retta immagine è un piano passante per il centro ottico). Sistema 2 (ICVS): autocalibrazione con metodo pubblicato nella rivista PAMI (gennaio 2005), dopo aver generato il SOR virtuale: si deve trascurare la d.r., e si deve fare l'ipotesi di natural camera. SOR: simmetria nello spazio e simmetria dopo la proiezione: $\mathbf{l}_s$ è l'immagine dell'asse di simmetria 3D, il profilo si distorce, e le sezioni trasversali, che nello spazio sono parallele, si piegano dando luogo a una convergenza delle rette per punti omologhi in $\mathbf{v}_\infty$. Omologia armonica (caso particolare di omografia): ha solo 4 gdl: $\mathbf{l}_s$ e $\mathbf{v}_\infty$. I vincoli di calibrazione sono 4, di cui 3 l.i. (passaggio di **i** e **j** per omega, e relazione pole-polar tra $\mathbf{v}_\infty$ e $\mathbf{l}_s$ – altro addendum per le lezioni precedenti: la linea di fuga di un piano e il punto di fuga della normale al piano sono in pole-polar attraverso la IAC). Come viene stimata l'omologia armonica (che serve per calibrare): analisi multirisoluzione e RANSAC (cfr. POCV). Una volta calibrato, la ricostruzione si effettua per rettificazione, senza cioè conoscere esplicitamente l'equazione del piano laser nel 3D, ma lavorando sull'immagine: si trova infatti la linea di fuga del piano laser a partire dall'immagine della retta intersezione tra piano laser e piano turntable, dalla IAC e dalla linea di fuga del piano turntable. Trovata la linea di fuga del piano laser, si trovano i punti circolari di questo piano per intersezione di tale linea con la IAC. Modelli 3D ottenuti: il problema delle occlusioni: si risolve variando il punto di vista della telecamera, e fondendo ("registrazione" 3D) tra di loro i modelli ottenuti. Allo scopo si usa tipicamente la ICP (Iterative Closest Point), una tecnica della geometria computazionale usata spesso in computer graphics. (Addendum per la lezione sul mosaico: è un capitolo della teoria della registrazione 2D). Resezione della telecamera, ossia fattorizzazione di P per ottenere K,R,**t** (e **C**). La matrice KR si fattorizza col metodo RQ (Givens), dove R è upper triangular (K), e Q è ortogonale (R). L'ambiguità di segno si risolve imponendo det K>0. Il punto C_{inhom}=-R**t** si calcola facendo P$\mathbf{C}_{hom}$=0 (con SVD, o in forma chiusa, con la formula dei determinanti). Relazione di uno studente (Lippi) su articolo Zhang: l'articolo offre una dimostrazione alternativa per arrivare alle equazioni di calibrazione, e mostra la rappresentazione dei punti circolari in forma non canonica. La matrice K non varia nel tempo, mentre il frame world (solidale al piano di riferimento) cambia, e con lui R e **t**. (Le equazioni per la distorsione radiale contengono probabilmente errori di stampa!) Zhang fa coincidere il centro della distorsione radiale col principal point. Notare la differenza tra coordinate immagine ideali (in unità di lunghezza focale) e reali (in pixel): tra di loro c'è la K.

I.8/22-23 - *Martedì 31 maggio 2005 (2 ore) [Ricostruzione da singola vista]*

Un'altra relazione fondamentale che coinvolge la K: $\mathbf{n}=K^T\mathbf{l}_\infty$, dove **n** è la normale a un piano in coordinate non omogenee di telecamera e $\mathbf{l}_\infty$ è la linea di fuga dello stesso piano. Si usa per ricostruire scene planari a tratti. Algoritmo: (1) da tre direzioni ortogonali dello spazio si calibra una natural camera; (2) si estraggono tutte le regioni planari e le loro linee di fuga; (3) si individuano almeno due punti in comune tra coppie di regioni, per fissare il fattore di scala della rappresentazione; (4) si calcolano le orientazioni dei vari piani, e si dispongono le regioni nello spazio. Volendo, si possono ricavare (ed applicare) le singole texture mediante rettificazione metrica. Demo Rec1stDemo.m e Rec2ndDemo2.m.

I.9/24-26 - *Venerdì 3 giugno 2005 (3 ore) [Metrologia e geometria due viste di un piano]*

Cenni all'uso dell'omografia RANSAC nel progetto EyeMouse (interazione uomo-macchina basata sul controllo oculare). Glitch, componente riflessa vs componente diffusa (l'eliminazione dei riflessi e la ricerca di "intrinsic images"). Def. di cross-ratio e suo uso per fare metrologia. Il teorema di Desargues e le omologie planari. Esempio: ombre. Parametrizzazione dell'omologia planare (5 dof, 3 punti per stimarla): essa generalizza quella armonica (4 dof, 2 punti), vista coi SOR. Nei SOR ci sono anche infinite omologie planari, legate a coppie di cross-sections. Omologie planari e autovalori/autovettori: l'asse è una retta di punti fissi, il vertice è l'autovettore corrispondente all'autovalore singolo, l'ultimo parametro è un cross-ratio. Stima dell'omologia planare: (1) attraverso la stima dei singoli parametri; (2) stimando 8 gdl di un'omografia generale, e forzandoli a 5. Ancora su autovalori e autovettori: (1) il caso delle trasformazioni di similarità 2D (**i** e **j** autovettori con autovalori dipendenti dall'angolo di rotazione, il terzo è il polo di rotazione con autovalore 1); (2) il caso delle rotazioni 3D (un autovalore è 1, con l'autovettore corrispondente pari alla direzione **a** dell'asse istantaneo); (3) rotazioni coniugate nel 3D TRT^{-1}, con stessi autovalori che nel caso precedente, ma con l'autovettore corrispondente all'autovalore unitario pari a T**a**: dunque se T coincide la calibrazione K, nel caso non calibrato si riesce comunque a conoscere dagli autovalori l'entità della rotazione – cfr, caso (1) –, ma non l'asse di rotazione; (4) omografia $H=K'(R+\mathbf{tn}^T/d)K^{-1}$ tra due viste di un piano: qui non c'è più un autovalore 1, ma dei tre valori singolari di $A=(R+\mathbf{tn}^T/d)$, quello di mezzo è 1, e autovettore di A^TA pari a $\mathbf{n}\times\mathbf{q}$, con $\mathbf{q}=R^T\mathbf{t}/d$. Problema della fattorizzazione di A per l'ottenimento della geometria relativa tra telecamere, e la posa del piano relativa alla prima telecamera. Si ricorda che la normale è conoscibile se è nota, oltre alla calibrazione, anche la vanishing line del piano, che qui però non è data. Risultato: (a) **t** non è comunque mai conoscibile al netto di d (speed-scale ambiguity); (b) si ottengono – utilizzando il positive depth constraint per entrambe le telecamere – due soluzioni valide per **n** e **q**, da ognuna delle quali è facilmente ricavabile la R, e dunque **t**/d. (Il 7/6/5 ho concluso il discorso, sviluppando meglio il positive depth constraint e il fatto che ci sono due soluzioni possibili per **n** e **q**.) L'ambiguità nasce dal fatto che non conosciamo l'orientazione "vera" dei versori ottenuti dalla SVD. Anticipazione: con sole viste di piani è impossibile ottenere la geometria epipolare: servono misure nel 3D (cfr. lez. 14/6).

I.10/27-29 - *Martedì 7 giugno 2005 (3 ore) [Geometria stereo]*

Altro caso dove si può usare il positive depth constraint: rimozione dell'ambiguità nel camera recovery a partire dalla matrice essenziale (slide): solo una delle 4 soluzioni ha depth positiva per entrambe le telecamere. Geometria stereo (slides Visione2005). Piano epipolare, epipoli, rette epipolari. La "correlazione" punto-retta epipolare. Epipoli all'infinito, e loro rette epipolari. Accenno alla stereo pair rectification: serve a velocizzare la ricerca delle corrispondenze. La geometria stereo è catturata da P e P' o, più economicamente, da K, K' e F (matrice fondamentale). F è proprio la "correlazione" punto-retta epipolare. Proprietà della F: fattorizzabilità in termini di epipolo e omografia planare (per qualsiasi piano), oppure in termini di epipolo e matrici di proiezione (usando P' e la pseudoinversa P^+). La matrice antisimmetrica a rango 2 $[\mathbf{e}']_x$. F ha rango 2: il suo null vector (kernel) è l'epipolo **e**. I molti modi di scrivere la F con una coppia di telecamere canoniche. Le relazioni simmetriche (coinvolgono F^T). Nel caso di traslazione pura (R=I, K'=K), l'epipolo è l'intersezione del vettore **t** col piano immagine (esempio differenziale: optic flow e FOE, focus of expansion). L'equazione scalare che collega punti omologhi attraverso la F: è lineare nei suoi parametri. F ha sette gdl: bastano 7 corrispondenze per stimarla, ma è meglio usare l'8-point algorithm (è un algoritmo lineare, che richiede la normalizzazione dei dati per evitare malcondizionamenti). Esempio di stima di F con 8 punti: il programma Imagemodeler della RealViz. I punti non devono essere tutti su un piano: altrimenti non si riesce a estrarre la parte legata all'epipolo (cfr. lez. 14/6). La F serve (1) a costruire P e P' per poter triangolare, e trovare la **X** che ha come immagini **x** e **x**' (**X** si chiama in gergo "preimage" di **x** e **x**'); (2) a cercare le corrispondenze su rette anziché sull'intera immagine. Nota: F NON permette di conoscere **x**' dato **x**: questo dipende non solo dalle telecamere ma anche dalla scena! Baseline: wide (buona la triangolazione, ma difficile la ricerca – si useranno le tecniche SIFT – e maggiori le occlusioni) e narrow (peggiore la triangolazione, perché i raggi ottici sono quasi paralleli, con maggiore incertezza sulla loro intersezione). La vista umana ha una narrow baseline, ed infatti non riusciamo a misurare bene le distanze oltre un certo range – tipicamente, quello corrispondente alla manipolazione degli oggetti! Coordinate ideali di telecamera e matrice essenziale E. E ha solo 5 gdl (visto il fattore di scala arbitrario). Relazione tra E ed F. Con E si ricostruisce in modo metrico, con F in modo proiettivo (ossia a meno di una trasformazione proiettiva dello spazio).

I.11/30-32 - *venerdì 10 giugno 2005 (3 ore) [Ricostruzione stereo]*

Camera matrices given F and given E. Conoscere P e P' (ad es. con Tsai) significa fissare 11+11=22 gdl. La conoscenza di F rimuove 7 gdl: ne rimangono 22-7=15, che sono quelli di una trasformazione proiettiva dello spazio H_{4x4}. Dunque ci sono ∞^{15} telecamere compatibili con la F, ed anche ∞^{15} soluzioni di ricostruzione (P -> PH; $\mathbf{X}$->$H^{-1}\mathbf{X}$). Forma della coppia di telecamere: P=[I | O], P'=[$[\mathbf{e}']_x$F+ $\mathbf{e}'\mathbf{v}^T$ | k**e**']. Similmente, conoscendo E (5 gdl), K e K' (5+5=10 gdl), P e P' sono fissabili a meno di una trasformazione metrica, o di similarità (22-15=7 gdl), il che vuol dire che è nota (a meno della distanza t) la posizione relativa, ma non quella assoluta delle telecamere dello spazio. N.B.: la triangolazione con E non si fa con le coordinate in pixel **x** e **x**', ma con le coordinate ideali $\mathbf{x}$^=$K^{-1}\mathbf{x}$ e $\mathbf{x}'$^=$K'^{-1}\mathbf{x}'$. La non conoscenza di t (modulo del vettore **t**) comporta che la ricostruzione sia ottenibile solo a meno di un fattore di scala. Esempio: torre di Pisa e suo modellino. (Ricostruzione 3D da due viste: i raggi non si incontrano realmente: bisogna stabilire un criterio di ottimizzazione, meglio se geometrico e non algebrico.) Ricostruzione stratificata: dalla ricostruzione proiettiva a quella metrica e Euclidea. Se conosco le coordinate di 5 punti nello spazio fisso esattamente la H_{4x4}, e dunque ricostruisco in modo euclideo. Altro uso di F (oltre a quello di trovare le telecamere): rettificazione stereo (con esposizione dei risultati, senza dimostrazione). Si trovano due omografie deformanti la coppia stereo che mantengano la stessa F ma che diano luogo a coppie di rette epipolari identiche. E' questo un caso particolare di un moto di telecamera di pura traslazione, senza R né variazione di K: la F in questo caso ha due soli gdl, e vale $[\mathbf{e}']_x$, e l'epipolo è il FOE. (Nel caso di $\mathbf{t}=(1,0,0)^T$ è tra l'altro superfluo ipotizzare K costante.) Laboratorio: il programma Photobuilder.

I.12/33-35 - *martedì 14 giugno 2005 (3 ore) [Plane+parallax e stima robusta (RANSAC)]*

Superfici critiche per F: se tutti gli **X** giacciono su una di queste superfici, stimare F in modo univoco è impossibile. La superficie critica più nota per F è il piano (dimostrazione: data l'omografia del piano H_π, ci sono ∞^2 soluzioni $F(\mathbf{u})=[\mathbf{u}]_x H_\pi$, con u qualsiasi, mentre un teorema dice che H è l'omografia di un piano compatibile con una coppia di viste – e dunque compatibile con F – se e solo se $F=[\mathbf{e}']_x H$). N.B. La condizione di compatibilità di una H con la F si può esprimere in questo modo: $H^TF+F^TH=0$, ossia H^TF deve essere skew-symmetric. Peraltro, data F, non è possibile risalire direttamente alla H di un dato piano (ad es. H_∞ per il piano all'infinito), perché $[\mathbf{e}']_x$ non ha rango pieno, ed esiste un'intera famiglia di omografie compatibile con una data F: $H(\mathbf{v})=H+\mathbf{e}'\mathbf{v}^T$. Relazione tra due viste: rappresentazione "plane plus parallax": considerando un piano π e la H_π ad esso associata, le immagini **x** e **x**' di un punto 3D qualsiasi **X** sono così legate (tralasciando i fattori di scala): $\mathbf{x}' = H_\pi\mathbf{x} + \rho_\pi\mathbf{e}'$, con ρ_π "profondità proiettiva", o "parallasse indotta dal piano π". Notiamo che **x**', $\mathbf{x}'_\pi=H_\pi\mathbf{x}$ e **e**' sono tre punti allineati, ed appartenenti alla linea epipolare F**x**. Il segno di ρ_π indica in quale dei due semispazi generati dal piano si trova **X**. La rappresentazione plane+parallax consente di stimare F con soli 6 punti, che però non possono più essere in posizione qualsiasi. In particolare, 4 punti dovranno giacere su π, per consentire la stima di H_π, ed altri due punti dovranno essere esterni a π, in modo da consentire il calcolo di **e**' come intersezione delle rette epipolari ($[\mathbf{x}'_5]_x H_\pi\mathbf{x}_5$) e ($[\mathbf{x}'_6]_x H_\pi\mathbf{x}_6$). Esempio: bastano i 6 vertici di un diedro retto per calibrare le telecamere (3 dof) e stimare la matrice fondamentale come detto sopra. Importanza dei fattori di scala nelle relazioni proiettive quando ci siano in gioco le somme. Esempio: ricaviamo l'equazione plane+parallax tenendo conto dei fattori di scala. Si considerino **x**,**x**' ed **e**' normalizzati (terza componente unitaria). L'elemento k_{33} della K sia anch'esso 1. In tale caso, il fattore di scala per **x** è la profondità $\lambda={}^cZ$ (e similmente $\lambda'={}^cZ'$). Cons. il riferimento world coincidente con quello della prima telecamera, cosicché $X= {}^cX$, $Y= {}^cY$ e $Z={}^cZ$. Otteniamo $(\lambda'/\lambda)\mathbf{x}'=H_\infty\mathbf{x}+(\lambda'_e/\lambda)\mathbf{e}'$ [*], con $H_\infty=K'RK^{-1}$ omografia del piano all'infinito. Considerando che (cfr. lez. 3/6) $H_\pi=H_\infty+K'\mathbf{t}\mathbf{n}_\pi{}^T/d_\pi K^{-1}$, si ottiene $(\lambda'/\lambda)\mathbf{x}'=H_\pi\mathbf{x}+(\delta/d_\pi)(\lambda'_e/\lambda)\mathbf{e}'$, con $\delta=d_\pi-\lambda\mathbf{n}_\pi{}^TK^{-1}\mathbf{x}$ distanza con segno del punto **X** dal piano π. Osservazioni: (1) se **X** è un punto all'infinito, si ha $\mathbf{x}'=H_\infty\mathbf{x}$, dove **x** e **x**' sono punti di fuga. (2) se $\pi=\pi_\infty$, allora $H_\pi=H_\infty$ e si ritorna alla relazione [*], in cui l'immagine sulla seconda telecamera dipende da due termini: il primo dipendente dai parametri di calibrazione e rotazione, ma indipendente dalla struttura 3D della scena (non dipende infatti da **X** ma solo da **x**), il secondo dipendente invece dalla traslazione e dalla struttura della scena (attraverso λ, variabile da punto a punto). Da ciò si ricavano le seguenti considerazioni: (2a) H_∞ non è di nessun aiuto per determinare la struttura, ma può servire per determinare la calibrazione e/o la rotazione (cfr. seguito); (2b) per ricavare informazioni di profondità, è necessario che la telecamera compia una traslazione **t**, tale da generare un termine di parallasse non nullo. Se supponiamo che R=I e che K'=K, allora $H_\infty=I$, e $(\lambda'/\lambda)\mathbf{x}'=\mathbf{x}+(\lambda'_e/\lambda)\mathbf{e}'$: nella lez. del 10/6 questa espressione non è stata ricavata, ma se ne è discusso il significato: le rette epipolari coincidono, la F vale $[\mathbf{e}']_x$, lo spostamento sull'immagine del punto **x** è di tipo radiale (con centro nell'epipolo) e di intensità inversamente proporzionale alla profondità (depth) del punto – esempio: moto apparente degli oggetti osservati dal finestrino di un treno in corsa: i punti lontani si muovono lentamente, quelli vicini velocemente. (3) La quantità $\rho_\pi(\mathbf{X})=(\delta/d_\pi)(\lambda'_e/\lambda)$ prende il nome di "struttura affine relativa" (relative affine structure: si tratta in realtà di una quantità proiettiva, non affine!): essa – fatta eccezione per la costante λ'_e – non dipende dalla telecamera P'. Infatti: δ dipende dalla posizione relativa tra **X** e π, d_π dipende dalla posizione relativa di P e π, e infine λ dipende dalla posizione relativa di **X** e P. Se non ci fosse dipendenza da **X** sia nel numeratore (δ) che nel denominatore (λ), sarebbe possibile ricavare la struttura in modo metrico, e non solo proiettivo. [Addendum: si può fissare il fattore di scala per l'omografia H_π (cioè h_o tale che $h_o H^\wedge{}_\pi= H_\pi$, con $H^\wedge{}_\pi$ l'omografia stimata) imponendo che per un punto arbitrario $\mathbf{X}_o$ non su π si abbia $\rho_\pi(\mathbf{X}_o)=1$: questo comporta che $(\lambda'_e/ d_\pi)=(\lambda_o/\delta_o)$. Dopo questa scalatura è $\rho_\pi(\mathbf{X})=(\delta/\delta_o)/(\lambda/\lambda_o)$, quantità effettivamente invariante rispetto alla seconda vista. La relative affine structure ρ_π e la relative depth (λ'/λ) si possono calcolare in maniera simile a quanto fatto per h_o: si fa un prodotto vettoriale e a seguire un prodotto scalare. Fissata dunque una nuova telecamera P'', e la coppia omografia-epipolo corrispondente, si può generare l'immagine corrispondente a questa nuova vista utilizzando l'immagine di P e la struttura affine relativa ("view synthesis").] Importanza della H_∞: permette di calibrare da due immagini (K'=K, skew=0), perché si può scrivere l'equazione $H_\infty \omega^{-1} H_\infty^T = \omega^{-1}$, che fissa quattro su cinque gdl di calibrazione. Per calcolare anche lo skew, serve una terza vista. Nota K, si ottiene poi subito la rotazione R. Come si calcola H_∞: (1) da 3 punti di fuga + F [cfr. HZ^2 p. 339]; (2) considerando 4 corrispondenze di punti sufficientemente lontani da essere considerabili sul piano all'infinito; (3) da 4 corrispondenze qualsiasi tra due viste ruotate (no traslazione). Stima robusta di omografie (ma anche di F, ellissi etc.) con RANSAC – RANdom SAmple Consensus. I parametri: *p* (probabilità di avere almeno un campione casuale su N privo di outlier), s (numero minimo di misure per effettuare la stima), ε (percentuale prevista di outlier). Il numero di iterazioni N è funzione di questi parametri. Per ogni iterazione, selezionato casualmente un campione di *s* misure, si calcola il consenso sulla base di una "distanza" (o "fit") di ciascuna delle *n* misure con quanto stimato: se c'è fit, allora tale misura è considerato un inlier per la soluzione, e "vota" per essa. Al termine del ciclo, si sceglie la soluzione più votata, e la si ricalcola utilizzando tutti gli inlier. Dopo di ciò si può ancora migliorare la soluzione con un "guided matching" (ricerca di nuove misure guidata dalla soluzione ottenuta), cui segue un nuovo ricalcalo con il set di misure esteso.

I.13/36-38 - *venerdì 17 giugno 2005 (3 ore) [Presentazione elaborati + autovalutazione]*

Quella che segue è una lista di possibili elaborati, a cui se ne potranno aggiungere altri sia su iniziativa del docente che su proposta degli studenti.

1. Depth from defocus [Nayar, Favaro]
2. Shape from shadows [Perona]
3. Dense vs sparse methods for motion and stereo [Anandan & Irani]
4. SIFT & wide baseline stereo [Tuytelaars]
5. Mosaics with global alignment [Fusiello]
6. Intrinsic images [Ikeuchi]
7. View synthesis [Irani, Shashua]
8. Visual modeling from sequences [Pollefeys]
9. Stereo with graphics commodity hardware [Pollefeys]