

Αοι

Arcangelo Distante
Cosimo Distante

**Visione computazionale
Dai simboli agli oggetti**

Volume III

Prefazione di
Virginio Cantoni





Aracne editrice

www.aracneeditrice.it
info@aracneeditrice.it

Copyright © MMXVIII
Gioacchino Onorati editore S.r.l. – unipersonale

www.gioacchinoonoratieditore.it
info@gioacchinoonoratieditore.it

via Vittorio Veneto, 20
00020 Canterano (RM)
(06) 45551463

ISBN 978-88-255-1647-0

*I diritti di traduzione, di memorizzazione elettronica,
di riproduzione e di adattamento anche parziale,
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

*Non sono assolutamente consentite le fotocopie
senza il permesso scritto dell'Editore.*

I edizione: luglio 2018

Ai miei genitori ed alla mia famiglia, Maria e Maria Grazia – AD

Ai miei genitori, a mia moglie Giovanna e, ai miei figli Francesca e Davide – CD

Indice

Prefazione di Virginio Cantoni	17
Introduzione	19
1 Riconoscimento degli oggetti	27
1.1 Introduzione	27
1.2 Metodi di classificazione	29
1.3 Conoscenza a priori e selezione delle feature	30
1.4 Estrazione caratteristiche significative	32
1.4.1 Spazio delle caratteristiche	34
1.4.2 Interpretazione delle caratteristiche significative	37
1.5 Metodo interattivo	44
1.6 Metodo deterministico	45
1.6.1 Funzioni discriminanti lineari	46
1.6.2 Funzioni discriminanti generalizzate	47
1.6.3 Funzione discriminante lineare di Fisher	48
1.6.3.1 Discriminante lineare di Fisher - 2 classi	49
1.6.3.2 Discriminante lineare di Fisher - C classi	52
1.6.4 Classificatore basato sulla minima distanza	55
1.6.4.1 Prototipi singoli	55
1.6.4.2 Prototipi multipli	56
1.6.5 Classificatore Nearest-Neighbor (NN)	57
1.6.6 Classificatore K-means	58
1.6.6.1 Limitazioni del K-means	60
1.6.7 Classificatore ISODATA	61
1.6.8 Classificatore Fuzzy C-Means-FCM	62
1.7 Metodo statistico	64
1.7.1 Classificatore MAP	64
1.7.2 Classificatore di Massima Verosimiglianza - ML	66
1.7.2.1 Esempio di classificazione Bayesiana non para- metrica	67
1.7.2.2 Calcolo della probabilità di errore Bayesiano	68

	1.7.2.3	Calcolo del minimo rischio per la regola di Bayes	69
	1.7.2.4	Regola di decisione di Bayes con rigetto	73
	1.7.3	Altri criteri di decisione	75
	1.7.4	Classificatore parametrico di Bayes	75
	1.7.5	Stima parametri con Massima Verosimiglianza - ML . .	76
	1.7.5.1	Stima ML per distribuzione Gaussiana con media incognita	77
	1.7.5.2	Stima ML per distribuzione Gaussiana con $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ incognite	77
	1.7.6	Stima parametri della distribuzione con il teorema di Bayes	79
	1.7.6.1	Stima Bayesiana per distribuzione Gaussiana con media incognita	80
	1.7.6.2	Stima Bayesiana della densità condizionale della classe $p(x \mathcal{P})$ per distribuzione normale . . .	82
	1.7.7	Confronto tra apprendimento Bayesiano e stima di massima verosimiglianza	82
1.8		Funzioni discriminanti Bayesiane	83
	1.8.1	Classificatore basato su densità Gaussiana	85
	1.8.2	Funzioni discriminanti per densità Gaussiana	87
	1.8.2.1	Assunzione: $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}$	87
	1.8.2.2	Assunzione: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ (diagonale)	89
	1.8.2.3	Assunzione: $\boldsymbol{\Sigma}_i$ arbitraria (non-diagonale) . . .	90
	1.8.2.4	Conclusioni	91
1.9		Misure di Gaussiane (MoG)	92
	1.9.1	Stima dei parametri della mistura con la Massima Verosimiglianza - ML	94
	1.9.2	Stima dei parametri della mistura con Expectation Maximization - EM	95
	1.9.2.1	Il passo M	96
	1.9.2.2	Il passo E	98
	1.9.3	Teoria EM	99
	1.9.3.1	Diseguaglianza di Jensen	99
	1.9.4	Classificatori non-parametrici	102
	1.9.4.1	Stima della densità di probabilità	103
	1.9.4.2	Finestre di Parzen	106
	1.9.4.3	Classificatore con le finestre di Parzen	112
1.10		Metodo basato sulle reti neurali	113
	1.10.1	Motivazione biologica	113
	1.10.2	Modello matematico della rete neurale	115
	1.10.2.1	Modello del neurone artificiale	116
	1.10.3	Percettrone per la classificazione	119
	1.10.3.1	Apprendimento con il percettrone	120
	1.10.3.2	Confronto tra classificatore statistico e percettrone	122
	1.10.3.3	Algoritmo batch del percettrone	123

1.10.4	Funzioni discriminanti lineari ed apprendimento	125
1.10.4.1	Algoritmo del Minimo Errore Quadratico - MSE	127
1.10.4.2	Gradiente discendente di Widrow-Hoff	129
1.10.4.3	Algoritmo di Ho-Kashyap	130
1.10.4.4	Estensione MSE per classi multiple	133
1.10.4.5	Conclusione	134
1.11	Reti neurali	135
1.11.1	Percettroni multistrato - MLP	136
1.11.2	Rete neurale multistrato MLP per la classificazione . . .	137
1.11.3	Algoritmo di backpropagation	139
1.11.3.1	Apprendimento supervisionato	139
1.11.4	Modalità di apprendimento con la backpropagation . . .	144
1.11.5	Generalizzazione della rete MLP	145
1.11.6	Euristiche per migliorare la backpropagation	147
1.11.6.1	Miglioramento dinamica di apprendimento: mo- mento	147
1.11.6.2	Proprietà della funzione di attivazione	148
1.11.6.3	Pre-elaborazione dei dati di input	149
1.11.6.4	Configurazione della rete: neuroni e pesi	149
1.11.6.5	Velocità di apprendimento, criterio di stop e decadimento pesi	150
1.11.6.6	Conclusioni: vantaggi e svantaggi di una rete neurale	151
1.12	Metodi di riconoscimento non-metrici	152
1.13	Albero di decisione	152
1.13.1	Algoritmi per la costruzione di alberi di decisione	154
1.14	Algoritmo-ID3	155
1.14.1	Entropia come misura di omogeneità dei campioni	156
1.14.2	Guadagno di informazione	157
1.14.3	Altri Criteri di partizionamento	162
1.15	Algoritmo-C4.5	164
1.15.1	Potatura di un albero di decisione	165
1.15.2	Algoritmi post-pruning	167
1.15.2.1	Reduced error pruning	168
1.15.2.2	Rule post-pruning	168
1.16	Algoritmo CART	170
1.16.1	Indice di Gini	170
1.16.2	Vantaggi e svantaggi degli alberi di decisione	173
1.16.2.1	Complessità degli alberi di decisione	174
1.17	Metodi gerarchici	175
1.17.1	Clustering gerarchico agglomerativo	177
1.17.1.1	Single linkage o Nearest Neighbor - NN	177
1.17.1.2	Complete linkage o farthest neighbor	178
1.17.1.3	Average distance	178
1.17.1.4	Distanza tra centroidi	178

1.17.2	Clustering gerarchico divisivo	179
1.17.3	Esempio di clustering gerarchico agglomerativo	180
1.18	Metodi sintattici di riconoscimento	182
1.18.1	Tipologia di grammatica	184
1.18.1.1	Esempio 1: Generazione delle stringhe $a^n b^n$ con $n \geq 1$	185
1.18.2	Generazione di linguaggi	186
1.18.3	Tipologie di grammatiche	188
1.18.4	Grammatiche per il riconoscimento	191
1.18.4.1	Analisi sintattica discendente	193
1.18.4.2	Analisi sintattica ascendente	194
1.18.5	Cenni su altri metodi di analisi sintattica	199
1.19	Metodi di riconoscimento delle stringhe	200
1.19.1	String matching	201
1.19.2	Algoritmo string-matching di Boyer-Moore	202
1.19.2.1	Euristica del carattere discordante	204
1.19.2.2	Euristica del buon suffisso	206
1.19.3	Editing distance	211
1.19.4	String matching con errore	216
1.19.5	String matching con simbolo speciale	219
2	Reti neurali RBF, SOM, Hopfield e profonde	221
2.1	Introduzione	221
2.2	Teorema di Cover sulla separabilità dei <i>pattern</i>	222
2.3	Il problema dell'interpolazione	224
2.4	Teorema di Micchelli	227
2.5	Learning e problemi mal-posti	228
2.6	Teoria della regolarizzazione	229
2.7	Architettura della rete RBF	233
2.8	Soluzione della rete RBF	234
2.9	Strategie di learning	236
2.9.1	Centri fissati selezionati casualmente	236
2.9.2	Selezione dei centri mediante tecniche di clustering	238
2.10	Rete neurale di Kohonen	238
2.10.1	Architettura della rete SOM	239
2.10.2	Applicazioni della SOM	247
2.11	Rete Learning Vector Quantization-LVQ	248
2.11.1	Reti LVQ2 ed LVQ3	251
2.12	Reti neurali ricorrenti	252
2.12.1	Rete di Hopfield	253
2.12.2	Applicazione della rete di Hopfield a stati discreti	257
2.12.2.1	Memoria associativa-training	257
2.12.2.2	Memoria associativa-recupero	258
2.12.2.3	Memoria associativa-prestazione	260
2.12.3	Rete di Hopfield a stati continui	262

2.12.3.1	Conclusioni	264
2.12.4	Boltzmann Machine	264
2.13	Deep learning	267
2.13.1	Rete neurale tradizionale profonda	268
2.13.2	Rete neurale convoluzionale - CNN	269
2.13.2.1	Strato di input	271
2.13.2.2	Strato Convoluzionale	271
2.13.2.3	Strato ReLU	273
2.13.2.4	Strato Pooling	274
2.13.2.5	Strato Full Connected-FC	275
2.13.3	Operatività di una rete CNN	276
2.13.3.1	Gradiente discendente stocastico nel deep learning	278
2.13.3.2	Dropout	280
2.13.3.3	Batch Normalization	283
2.13.3.4	Monitoraggio della fase di training	284
2.13.4	Principali architetture di reti CNN	284
3	Analisi della Tessitura	287
3.1	Introduzione	287
3.2	La percezione visiva della tessitura	287
3.2.1	Congettura di Julesz	289
3.2.2	Statistica dei texton	290
3.2.3	Modelli spettrali della tessitura	291
3.3	Analisi della tessitura e sue applicazioni	291
3.4	Metodi statistici della tessitura	293
3.4.1	Statistica del primo ordine	293
3.4.2	Statistica del secondo ordine	294
3.4.3	Statistica di ordine superiore	294
3.4.4	Statistica del secondo ordine con matrice di co-occorrenza	296
3.4.5	Parametri di tessitura basati sulla matrice di co-occorrenza	298
3.4.5.1	Energia o misura del grado di omogeneità della tessitura	299
3.4.5.2	Entropia	300
3.4.5.3	Massima probabilità	300
3.4.5.4	Contrasto	300
3.4.5.5	Momento della differenza inversa	300
3.4.5.6	Valore assoluto	300
3.4.5.7	Correlazione	301
3.5	Caratteristiche di tessitura basata sull'autocorrelazione	303
3.6	Approccio spettrale della tessitura	304
3.7	Tessitura basata sulla metrica dei bordi	307
3.8	Tessitura basata sulle primitive Run Length	309
3.9	Tessitura basata sui modelli MRF SAR e frattali	311
3.10	Tessitura mediante filtraggio spaziale	315
3.10.1	Filtraggio spaziale con filtri di Gabor	320

4.6.7.2	Parametri intrinseci ed estrinseci di un sistema binoculare	414
4.6.8	Sistema binoculare generale	416
4.7	Algoritmi per la stereovisione artificiale	420
4.7.1	Strutture elementari puntuali	421
4.7.2	Strutture elementari locali e metodi calcolo corrispondenza	423
4.7.2.1	Funzione di <i>matching</i> basata sulla correlazione	424
4.7.2.2	Funzione di <i>matching</i> basata sulla distanza . .	426
4.7.2.3	Funzione di <i>matching</i> basata sulla trasformata rango	427
4.7.2.4	Funzione di <i>matching</i> basata sulla distanza di Hamming	428
4.7.2.5	Funzione di <i>matching</i> basata sul gradiente . .	429
4.7.2.6	Metodi globali per la corrispondenza	430
4.7.3	Strutture elementari sparse	430
4.7.4	Algoritmo PMF di stereovisione	431
5	Shape from Shading	437
5.1	introduzione	437
5.2	La mappa di riflettanza	438
5.2.1	Spazio gradiente	441
5.3	La relazione fondamentale dello Shape from Shading	443
5.4	Algoritmi di Shape from Shading - SfS	446
5.4.1	Shape dalla fotometria stereo con calibrazione	450
5.4.2	Fotometria stereo non calibrata	456
5.4.3	Fotometria stereo con sfera di calibrazione	460
5.4.3.1	Calcolo della direzione delle sorgenti con la fo- tometria stereo inversa	463
5.4.4	Limiti della fotometria stereo	463
5.4.5	Ricostruzione della superficie dalla mappa delle orientazioni	464
5.4.5.1	Metodo locale mediando i gradienti	465
5.4.5.2	Metodo locale basato sui minimi quadrati . . .	466
5.4.5.3	Metodo globale basato sulla trasformata di Fou- rier	467
5.5	Shape from texture	471
5.6	Shape from luce strutturata	474
5.6.1	Shape from luce strutturata con codifica binaria	479
5.6.2	Pattern con codifica binaria Gray	480
5.6.3	Pattern con livelli di grigio	482
5.6.3.1	Pattern con modulazione di fase	482
5.6.4	Pattern con modulazione di fase e codice binario	486
5.6.5	Metodi basati su pattern colorati	487
5.6.6	Calibrazione del sistema di scansione telecamera-proiettore	487
5.7	Shape from (de)Focus	489
5.7.1	Shape from Focus (SfF)	490

5.7.2	Shape from Defocus (SfD)	497
6	Analisi del Movimento	503
6.1	Introduzione	503
6.2	Analogia tra percezione del moto e della profondità valutata con la stereo-visione	506
6.3	Verso la stima del movimento	508
6.3.1	Discretizzazione del movimento	511
6.3.2	Stima del moto - approccio continuo	518
6.3.3	Stima del moto - approccio discreto	520
6.3.4	Analisi del moto dalla differenza di immagini	520
6.3.5	Analisi del moto dalla differenza cumulativa di immagini	521
6.3.6	Ambiguità nell'analisi del movimento	522
6.4	Stima del flusso ottico	523
6.4.1	Metodo di Horn-Schunck	529
6.4.2	Metodo di Horn-Schunck discreto ai minimi quadrati	534
6.4.3	Algoritmo di Horn-Schunck	536
6.4.4	Metodo di Lucas-Kanade	537
6.4.4.1	Variante del metodo di Lucas-Kanade con l'introduzione dei pesi	541
6.4.5	Metodo di BBPW	541
6.4.6	Stima del flusso ottico per moto affine	545
6.4.6.1	Segmentazione per regioni omogenee di moto	547
6.4.7	Stima del flusso ottico per grandi spostamenti	547
6.4.7.1	Raffinamento iterativo del flusso ottico	548
6.4.7.2	Stima del flusso ottico con approccio multirisoluzione	550
6.4.8	Stima del moto mediante allineamento	551
6.4.8.1	Algoritmo di allineamento <i>Compositional</i>	558
6.4.8.2	Algoritmo di allineamento <i>Inverse Compositional</i>	559
6.4.9	Stima del moto con tecniche basate su punti di interesse	561
6.4.9.1	Tracking utilizzando la similarità dei punti di interesse	563
6.4.9.2	Tracking utilizzando la similarità tra grafi di punti di interesse	564
6.4.9.3	Tracking basato sulla corrispondenza probabilistica di punti di interesse	565
6.4.10	Tracking basato sulla dinamica dell'oggetto - Filtro di Kalman	568
6.5	Moto in scene-ambiente complessi	582
6.5.1	Metodo semplice di Background Subtraction-BS	583
6.5.2	Metodo BS con la media o mediano	584
6.5.2.1	Metodo BS media mobile	585
6.5.3	Metodo della media mobile Gaussiana	585
6.5.4	Metodo BS selettivo	587

6.5.5	Metodo statistico con misture di Gaussiane	587
6.5.6	Metodo statistico Kernel Density Estimation - KDE	589
6.5.7	Metodo Eigenbackground	590
6.5.8	Ulteriori modelli del background	592
6.6	Struttura analitica del flusso ottico di un corpo rigido	592
6.6.1	Analisi del moto dal campo di flusso	596
6.6.1.1	Moto con pura traslazione	597
6.6.1.2	Moto con pura rotazione	599
6.6.2	Calcolo del tempo di collisione e della profondità	599
6.6.2.1	Calcolo di TTC	601
6.6.2.2	Calcolo della profondità	602
6.6.3	Calcolo del FOE	603
6.6.3.1	Calcolo di TTC e della profondità conoscendo il FOE	605
6.6.4	Stima dei parametri di moto per un corpo rigido	605
6.7	Struttura dal movimento	611
6.7.1	Matrice di proiezione dell'immagine	612
6.7.1.1	Parametri intrinseci di calibrazione	613
6.7.1.2	Parametri estrinseci di calibrazione	614
6.7.2	Metodi di SfM	616
6.7.2.1	Metodo della fattorizzazione	616
7	Calibrazione	623
7.1	Introduzione	623
7.2	Influenza del sistema ottico	625
7.3	Trasformazioni coinvolte nella formazione dell'immagine	626
7.4	Metodi di calibrazione	627
7.4.1	Metodo di Tsai	629
7.4.2	Stima della matrice di proiezione prospettica	634
7.4.2.1	Dalla matrice \mathbf{P} ai parametri intrinseci ed estrinseci	636
7.4.2.2	Parametri intrinseci ed estrinseci con la decomposizione QR della matrice \mathbf{P}	639
7.4.2.3	Stima non lineare della matrice di proiezione prospettica	639
7.4.3	Metodo di Zhang	640
7.4.3.1	Calcolo della matrice omografica	641
7.4.3.2	Dalle matrici omografiche ai parametri intrinseci	643
7.4.3.3	Stima dei parametri estrinseci	645
7.4.3.4	Stima delle distorsioni radiali	646
7.4.3.5	Ottimizzazione non lineare dei parametri di calibrazione e distorsione radiale	647
7.4.3.6	Riepilogo del metodo di autocalibrazione di Zhang	647
7.4.4	Calibrazione telecamere stereo	649
7.5	Stereovisione e geometria epipolare	651

7.5.1	La matrice Essenziale	653
7.5.2	La matrice Fondamentale	657
	7.5.2.1 Relazione tra matrice fondamentale e matrice omografica	660
7.5.3	Stima della matrice Essenziale e Fondamentale	661
	7.5.3.1 Algoritmo degli 8 punti	661
	7.5.3.2 Calcolo delle matrici E ed F da 7 corrispondenze	664
7.5.4	Normalizzazione dell'algoritmo degli 8 punti	665
7.5.5	Decomposizione della matrice Essenziale	666
7.5.6	Rettifica delle immagini stereo	668
	7.5.6.1 Rettifica non calibrata	670
	7.5.6.2 Rettifica calibrata	675
7.5.7	Ricostruzione stereo 3D mediante triangolazione	679
	7.5.7.1 Ricostruzione noti i parametri intrinseci ed estrinseci	679
	7.5.7.2 Ricostruzione noti i parametri intrinseci ed estrinseci con triangolazione lineare	681
	7.5.7.3 Ricostruzione con noti solo i parametri intrinseci	683
	7.5.7.4 Ricostruzione con noti solo i parametri intrinseci e normalizzando \mathbf{E}	684
	7.5.7.5 Ricostruzione con note solo le corrispondenze di punti omologhi	687
	Bibliografia	691
	Ringraziamenti	713

Prefazione

di Virginio Cantoni

Monumentale. Come altro definire l'opera dei due autori? Chiariamo subito che il termine non riguarda l'ampiezza e la qualità dell'opera – oltre 1250 pagine ricche di immagini curate e ben organizzate – ma soprattutto la completezza della trattazione e la descrizione dell'evolversi esplosivo di una disciplina che oggi, finalmente, va occupando il punto focale e critico di tanti settori vitali per la società moderna toccando praticamente tutti i settori applicativi con bioimmagini, remote sensing, assistenza telematica, accessibilità tattile, guida automatica, data mining and video annotation solo per citarne alcune applicazioni già a livello almeno preindustriale. I mattoni su cui si è venuta costruendo l'attuale visione computazionale sono presentati in dettaglio. Modelli diversi partendo dai principi fisici della formazione del segnale visivo, considerando i diversi metodi di acquisizione, i loro limiti e le strategie per migliorarne la qualità alla generazione sulla base della conoscenza delle caratteristiche di trasduzione e conversione in immagine digitale. Segue la trattazione di quella parte della disciplina che un tempo è stata schematicamente chiamata da Michael Duff la visione intermedia. La fase di elaborazione che viene dopo la percezione e prima dell'elaborazione di più "alto" livello con obiettivi concettualmente più elevati allora considerati prossimi al ragionamento umano con l'obiettivo di interpretare il contenuto semantico del segnale visivo. In modo che definirei esaustivo sono quindi presentati i molteplici approcci di rappresentazione e manipolazione dell'informazione visiva in forma digitale: strutture dati (dalla matrice 2D alle strutture piramidali o multigrig); trasformazioni lineari filtraggi di tutti i tipi (nello spazio, in frequenza, ecc.); proprietà peculiari, invarianze, descrittori e rilevatori di varia natura e precisione. Sono quindi introdotti in modo dettagliato i vari approcci alla segmentazione cioè l'identificazione, la localizzazione e la forma delle componenti salienti dell'immagine che da questo momento perde le caratteristiche strutturali di partenza. Infine con il terzo volume si affronta l'analisi del contenuto informativo, lo studio dei componenti salienti, di fatto la semantica del segnale percepito. Accanto ai metodi di successo che si sono affermati specificatamente nel riconoscimento di forme draconianamente descritti in letteratura come approcci statistici, sintattici e strutturali uno spazio significativo è dato all'approccio basato sulle reti neurali che avviate da McCulloch e Pitts (vorrei però ricordare che espressamente nel loro lavoro citato parlano del modello di neurone descritto nella tesi di laurea da Eduardo Caianiello, poi

fondatore del gruppo italiano di riconoscimento di forme) che oggi sono alla base del deep learning. Questa evoluzione della machine learning sta portando l'apprendimento delle macchine verso traguardi impensati raggiungendo risultati efficienti in settori della visione computerizzata che con i vecchi approcci erano proibitivi. Infine voglio sottolineare il costante riferimento che gli autori portano verso le soluzioni della visione umana, forse non sono più i tempi di David Marr nei quali la neurofisiologia poteva offrire modelli di elaborazione del segnale visivo facilmente applicabili per le soluzioni automatiche ma ai livelli di elaborazione più alti le soluzioni antropomorfe ancora possono ispirare nuovi modelli concettuali!

Virginio Cantoni

IAPR e IEEE Fellow

Dipartimento di Ingegneria Industriale e dell'Informazione

Università di Pavia

Introduzione

Negli ultimi 20 anni diverse ricerche interdisciplinari nei settori della *fisica, dell'informatica e della cibernetica, dell'Elaborazione numerica di Segnali ed Immagini, delle tecnologie elettriche ed elettroniche*, hanno portato allo sviluppo di Sistemi Intelligenti.

I cosiddetti *Sistemi Intelligenti (o Agenti Intelligenti)* rappresentano la frontiera tuttora più avanzata e innovativa della ricerca in campo informatico ed elettronico, in grado di influenzare direttamente la qualità della vita, la competitività e le modalità di produzione delle imprese, di monitorare e valutare l'impatto ambientale, di rendere più efficienti le attività di servizio e di gestione delle amministrazioni pubbliche, e la sicurezza delle persone.

Lo studio di un sistema intelligente, indipendentemente dal campo d'impiego, può essere semplificato in tre componenti essenziali:

1. la *prima* interagisce con l'ambiente per l'acquisizione dei dati del dominio di interesse, utilizzando sensori adeguati (per l'acquisizione di Segnali ed Immagini);
2. la *seconda* analizza ed interpreta i dati rilevati dalla prima componente utilizzando anche tecniche di apprendimento per costruire/aggiornare rappresentazioni adeguate della realtà anche complessa nella quale il sistema opera (Visione Computazionale);
3. la *terza* sceglie le azioni più appropriate per raggiungere gli obiettivi assegnati al sistema intelligente (scelta dei Modelli Decisionali Ottimali) interagendo con le prime due componenti, e con gli operatori umani, nel caso di soluzioni applicative basate su paradigmi cooperativi uomo-macchina (le attuali evoluzioni dell'automazione compresa quella industriale).

In questo scenario di avanzamento della conoscenza per lo sviluppo di Sistemi Intelligenti, s'inquadra il contenuto informativo di questo manoscritto nel quale sono riportate le esperienze di ricerca pluriennali e di docenza degli autori, e degli approfondimenti scientifici esistenti in letteratura. In particolare, il manoscritto articolato in tre parti (volumi), tratta gli aspetti del sottosistema sensoriale per percepire l'ambiente in cui un sistema intelligente è immerso e capace di agire anche autonomamente.

Il *primo volume* descrive l'insieme dei processi fondamentali della visione artificiale che dall'energia portano alla formazione dell'immagine digitale. So-

no analizzati i fenomeni di propagazione della luce (cap.1 e 2), la teoria di percezione del colore (cap.3), l'impatto del sistema ottico (cap.4), gli aspetti di trasduzione da energia luminosa (il flusso ottico) a segnale elettrico (dei fotorecettori), e gli aspetti di trasduzione del segnale elettrico (con valori continui) in valori discreti (pixel), ovvero la conversione del segnale da analogico a digitale (cap.5). Questi primi 5 capitoli sintetizzano il processo di acquisizione della scena 3D, in forma simbolica, rappresentata numericamente dai pixel dell'immagine digitale (proiezione 2D della scena 3D).

Il cap.6 descrive le proprietà geometriche, topologiche, qualità e informazioni percettive dell'immagine digitale. Sono definite le metriche, le modalità di aggregazione e di correlazione tra pixel, utili per definire strutture simboliche della scena di più alto livello rispetto al pixel. L'organizzazione dei dati per i diversi livelli di elaborazione è descritta nel cap.7 mentre nel cap.8 è riportata la rappresentazione e descrizione delle strutture omogenee della scena.

Con il cap.9 inizia la descrizione degli algoritmi di elaborazione immagine, per il miglioramento delle qualità visive dell'immagine, basati su operatori puntuali, locali e globali. Sono riportati algoritmi operanti nel dominio spaziale e nel dominio delle frequenze evidenziando con degli esempi le differenze significative tra i vari algoritmi anche dal punto di vista del carico computazionale.

Il *secondo volume* inizia con il capitolo che descrive gli algoritmi di estrazione dei contorni basati su operatori locali nel dominio spaziale e su tecniche di filtraggio nel dominio delle frequenze.

Nel cap.2 sono presentate le trasformate lineari fondamentali che hanno immediata applicazione nel campo dell'elaborazione dell'immagine, in particolare, per estrarre le caratteristiche essenziali contenute nelle immagini. Tali caratteristiche, che sintetizzano efficacemente il carattere informativo globale dell'immagine, sono poi utilizzate per gli altri processi di elaborazione dell'immagine: classificazione, compressione, descrizione, ecc. Le trasformate lineari sono anche utilizzate, come operatori globali, per migliorare le qualità visive dell'immagine (*enhancement*), per attenuare il rumore (*restoration*), oppure per ridurre la dimensionalità dei dati (*data reduction*).

Nel cap.3 sono descritte le trasformazioni geometriche delle immagini, necessarie in diverse applicazioni della visione artificiale, sia per correggere eventuali distorsioni geometriche introdotte durante l'acquisizione (per esempio, immagini acquisite mentre gli oggetti oppure i sensori sono in movimento, come nel caso di acquisizioni da satellite e/o aereo), oppure per introdurre voluti effetti geometrici visivi. In tutte e due i casi, l'operatore geometrico deve essere in grado di riprodurre nel modo più fedele possibile l'immagine con lo stesso contenuto informativo iniziale attraverso il processo di ricampionamento dell'immagine.

Nel cap.4 *Ricostruzione dell'immagine degradata (image restoration)* sono descritte un insieme di tecniche che eseguono correzioni quantitative sull'immagine per compensare le degradazioni introdotte durante il processo di acquisizione e trasmissione. Tali degradazioni sono rappresentate dall'effetto nebbia o sfocamento causato dal sistema ottico e dal moto dell'oggetto o dell'osservato-

re, dal rumore causato dal sistema opto-elettronico e dalla risposta non lineare dei sensori, dal rumore casuale dovuto alla turbolenza atmosferica o, più in generale, dal processo di digitalizzazione e trasmissione. Mentre le tecniche di *enhancement*, tendono a ridurre in misura qualitativa, le degradazioni presenti nell'immagine, migliorandone la qualità visiva anche quando non si ha nessuna conoscenza del modello di degradazione, le tecniche di *restoration* sono utilizzate invece per eliminare o attenuare in modo quantitativo le degradazioni presenti nell'immagine, partendo anche dalle ipotesi di conoscenza dei modelli di degradazione.

Il cap.5, *Segmentazione delle immagini*, descrive diversi algoritmi di segmentazione, ovvero il processo di dividere l'immagine in regioni omogenee, dove tutti i pixel che corrispondono ad un oggetto della scena sono raggruppati insieme. Il raggruppamento dei pixel in regioni è basato in relazione ad un criterio di omogeneità che li distingue tra loro. Sono riportati algoritmi di segmentazione basati su criteri di similarità degli attributi dei pixel (colore, tessitura, ecc.) oppure basati su criteri geometrici di prossimità spaziale dei pixel (distanza Euclidea, ecc.). Tali criteri non sono sempre validi, e in diverse applicazioni è necessario integrare altre informazioni in relazione alla conoscenza a priori del contesto applicativo (dominio dell'applicazione). Il raggruppamento dei pixel, in quest'ultimo caso, si basa confrontando le regioni ipotizzate con le regioni modellate a priori.

Il cap.6 *Rilevatori e descrittori di punti di interesse*, descrive gli algoritmi più utilizzati per rilevare automaticamente le strutture significative (noti come punti di interesse, corner, feature) presenti nell'immagine corrispondenti a parti fisiche stabili della scena. L'abilità di tali algoritmi è quella di rilevare ed identificare parti fisiche della stessa scena in modo ripetibile, anche quando le immagini sono acquisite in condizioni di variabilità di illuminazione e cambiamento del punto di osservazione con eventuale cambiamento del fattore di scala.

Il terzo volume descrive gli algoritmi di visione artificiale che rilevano gli oggetti della scena, tentano la loro identificazione, la ricostruzione 3D, il loro assetto e localizzazione rispetto all'osservatore, e l'eventuale loro movimento.

Il cap.1 *Riconoscimento degli oggetti*, descrive gli algoritmi fondamentali della visione artificiale per riconoscere automaticamente gli oggetti della scena caratteristica essenziali di tutti i sistemi di visione degli organismi viventi. Mentre un osservatore umano esegue il riconoscimento anche di oggetti complessi, apparentemente in modo agevole e tempestivo, per una macchina di visione il processo di riconoscimento risulta difficile, necessita di un notevole tempo di calcolo e non sempre i risultati sono ottimali. Fondamentali per il processo di riconoscimento degli oggetti, diventano gli algoritmi di selezione ed estrazione delle feature. In varie applicazioni si può avere una conoscenza a priori della popolazione degli oggetti da classificare poichè si conoscono i pattern (feature significative) campioni dai quali si possono estrarre informazioni utili per la decisione di associare (*decision making*) ciascun individuo della popolazione ad una determinata classe. Questi pattern campioni (*training set*)

sono utilizzati dal sistema di riconoscimento per apprendere le informazioni significative sulla popolazione (estrazione dei parametri statistici, caratteristiche rilevanti, ecc.). Il processo di riconoscimento confronta le feature degli oggetti incogniti con le feature dei pattern modelli, allo scopo di identificarne in modo univoco la classe di appartenenza. Diversi sono stati negli anni i settori disciplinari (apprendimento automatico, analisi d'immagini, riconoscimento oggetti, ricerca dell'informazione, bioinformatica, biomedicina, analisi di dati intelligente, data mining, ...) ed i settori applicativi (robotica, telerilevamento, visione artificiale,...) per i quali diversi ricercatori hanno proposto diversi metodi di riconoscimento e sviluppato differenti algoritmi basati su diversi modelli di classificazione. Sebbene gli algoritmi proposti hanno un'univoca finalità, si differenziano per la proprietà attribuita alle classi di oggetti (i cluster) e per il modello con cui sono definiti tali classi (connettività, distribuzione statistica, densità,...). La diversità di discipline, specialmente tra quelle dell'estrazione automatica dei dati (*data mining*) e dell'apprendimento automatico (*machine learning*), ha portato a sottili differenze, soprattutto nell'utilizzo dei risultati e nelle terminologie, talvolta contraddittorie forse causate dai diversi obiettivi. Per esempio, nel data mining l'interesse dominante è l'estrazione automatica dei raggruppamenti, nella classificazione automatica è fondamentale il potere discriminante delle classi di appartenenza dei pattern. Gli argomenti di questo capitolo si sovrappongono tra aspetti legati al machine learning e quelli del riconoscimento basato su metodi statistici. Per semplicità gli algoritmi descritti sono ripartiti in base ai metodi di classificazione degli oggetti in supervisionati (basati su modelli deterministici, statistici, neurali, e non metrici quali i modelli sintattici e alberi decisionali) e non-supervisionati, ovvero metodi che non utilizzano nessuna conoscenza a priori per estrarre le classi di appartenenza dei pattern.

Nel cap.2 *Reti neurali RBF, SOM, Hopfield e profonde* sono descritte quattro tipologie differenti di reti neurali: *Radial Basis Functions* - RBF, *Self-Organizing Maps* - SOM, la rete di Hopfield e le reti neurali profonde. RBF utilizza un approccio differente nel disegno di una rete neurale basato sullo strato *hidden* (unico nella rete) composto da neuroni in cui vengono definite funzioni a base radiale, da qui il nome di *Radial Basis Functions*-RBF, e che effettua una trasformazione non-lineare dei dati di input forniti alla rete. Questi neuroni costituiscono delle basi per i dati di input (vettori). La ragione per cui si adopera una trasformazione non-lineare nello strato hidden, seguita da una lineare in quello di output, permette ad un problema di classificazione di pattern di operare in uno spazio a dimensione molto più grande (nella trasformazione non lineare dallo strato di input in quello hidden) ed è più probabile di essere linearmente separabile rispetto ad uno spazio a dimensione ridotta. Da questa osservazione deriva la ragione per cui lo strato hidden è generalmente più grande di quello di input (ovvero il numero di neuroni hidden è maggiore della cardinalità del segnale di input). La rete SOM invece ha un modello di apprendimento non supervisionato ed ha l'originalità di raggruppare autonomamente dati di input sulla base della loro similarità senza valutare l'errore di

convergenza con informazioni esterne sui dati. Utile quando sui dati non si ha una conoscenza esatta per classificarli. Si ispira alla topologia del modello della corteccia del cervello considerando la connettività dei neuroni ed in particolare il comportamento di un neurone attivato e l'influenza con i neuroni vicini che rinforzano i legami rispetto a quelli più lontani che man man diventano più deboli. Con la rete di Hopfield il modello di apprendimento è supervisionato e con la capacità di memorizzare informazione e di recuperarla attraverso contenuti anche parziali dell'informazione originale. Presenta una sua originalità basata su fondamenti fisici che ha rivitalizzato l'intero settore delle reti neurali. Alla rete è associata una funzione energia da minimizzare durante la sua evoluzione con una successione di stati, fino a raggiungere uno stato finale corrispondente al minimo della funzione energia. Questa sua caratteristica consente di essere utilizzata per risolvere e impostare un problema di ottimizzazione in termini di funzione obiettivo da associare ad una funzione energia. Il capitolo si chiude con la descrizione delle reti neurali convoluzionali (CNN), ormai le più diffuse a partire dal 2012, basate sull'architettura di apprendimento profondo (*deep learning*).

Nel cap.3 *Analisi della Tessitura* sono riportati gli algoritmi che caratterizzano la tessitura presente nelle immagini. La tessitura risulta una componente importante per il riconoscimento degli oggetti. Nel campo dell'elaborazione delle immagini si è consolidato con il termine tessitura una qualunque disposizione geometrica e ripetitiva dei livelli di grigio (o di colore) di una immagine. In tale contesto la tessitura diventa una componente strategica aggiuntiva per risolvere il problema del riconoscimento oggetti, la segmentazione delle immagini ed i problemi di sintesi. Alcuni degli algoritmi descritti si basano sui meccanismi della percezione visiva umana della tessitura. Sono utili per lo sviluppo di sistemi per l'analisi automatica del contenuto informativo di un'immagine ottenendo un partizionamento dell'immagine in regioni con differenti tessiture.

Nel cap.4 *Paradigmi per la Visione 3D* sono riportati gli algoritmi che analizzano immagini 2D per ricostruire una scena tipicamente di oggetti 3D. Un sistema di visione 3D ha il problema fondamentale tipico dei problemi inversi, ossia da singole immagini 2D, che sono solo una proiezione bidimensionale del mondo 3D (acquisizione parziale), deve poter ricostruire la struttura 3D della scena osservata ed eventualmente definire una relazione tra gli oggetti. La ricostruzione 3D avviene partendo dalle immagini 2D che contengono solo informazioni parziali del mondo 3D (perdita di informazione dalla proiezione 3D→2D) e possibilmente utilizzando i parametri geometrici e radiometrici di calibrazione del sistema di acquisizione. Sono illustrati i meccanismi della visione umana basati anche sulla predizione e conoscenza a priori del mondo. Nel settore della visione artificiale, la tendenza attuale è quella di sviluppare sistemi 3D orientati per specifici domini ma con caratteristiche che vanno nella direzione di imitare alcune funzioni del sistema visivo umano. Sono descritti metodi di ricostruzione 3D che utilizzano più telecamere osservando la scena da più punti di vista, oppure sequenze d'immagini tempo-varianti acquisite da una singola telecamera. Sono descritte le teorie della visione, dalle leggi Gestalt al

paradigma della visione di Marr ed i modelli computazionali della stereovisione.

Nel cap.5 *Shape from Shading*-(SfS) sono riportati gli algoritmi per ricostruire la forma (*shape*) della superficie 3D visibile utilizzando soltanto le informazioni di variazione di luminosità (*shading*, ossia le sfumature di livello di grigio o di colore) presenti nell'immagine. Il problema inverso, di ricostruire la forma della superficie visibile dalle variazioni di luminosità presenti nell'immagine è noto come il problema dello *Shape from Shading*. La ricostruzione della superficie visibile non deve intendersi rigorosamente come ricostruzione 3D della superficie. Infatti, da un singolo punto di osservazione della scena, un sistema di visione monoculare non può stimare una misura di distanza tra osservatore e oggetto visibile, pertanto con gli algoritmi di SfS si ha una ricostruzione non metrica ma qualitativa della superficie 3D. E' descritta la teoria dello SfS basata sulla conoscenza della sorgente luminosa (direzione e distribuzione), il modello di riflettanza della scena, il punto di osservazione e la geometria della superficie visibile, che insieme, concorrono nel processo di formazione dell'immagine. Sono derivate le relazioni tra i valori d'intensità luminosa dell'immagine e la geometria della superficie visibile (in termini di orientazione della superficie punto per punto) sotto alcune condizioni di illuminazione e del modello di riflettanza. Sono inoltre descritti altri algoritmi di ricostruzione 3D della superficie basato sul paradigma Shape from xxx, dove con xxx può essere la tessitura, luce strutturata proiettata sulla superficie da ricostruire, oppure immagini 2D della superficie focalizzata o defocalizzata.

Nel cap.6 *Analisi del Movimento* sono riportati gli algoritmi di percezione della dinamica della scena, in analogia a quanto avviene, nei sistemi di visione di diversi esseri viventi. Con gli algoritmi di analisi del movimento è possibile derivare il moto 3D, quasi in tempo reale, dall'analisi di sequenze d'immagini 2D tempo-varianti. Paradigmi sull'analisi del movimento hanno dimostrato che la percezione del movimento deriva dalle informazioni degli oggetti valutando la presenza di occlusioni, tessitura, contorni, ecc. Sono descritti gli algoritmi per la percezione del movimento che avviene nella realtà fisica e non al movimento apparente. Sono analizzate diverse metodologie di analisi del movimento da quelli con carico computazionale limitato come quelli basati sulla differenza delle immagini tempo-varianti a quelli più complessi basati sul flusso ottico considerando contesti applicativi con diversi livelli di entità di movimento e scene-ambiente con diversa complessità. Nel contesto di corpi rigidi, dall'analisi del moto, derivato da una sequenza di immagini tempo-variante, sono descritti gli algoritmi che, oltre al movimento (di traslazione e rotazione), stimano la ricostruzione della struttura 3D della scena e la distanza di tale struttura dall'osservatore. Informazioni utili, nel caso di osservatore mobile (robot o veicolo), per stimare anche il tempo di collisione. Sono infatti descritti i metodi per risolvere il problema della ricostruzione 3D della scena acquisendo una sequenza di immagini con una singola telecamera i cui parametri intrinseci rimangono costanti anche se non noti (telecamera non calibrata) insieme alla non conoscenza del moto. I metodi proposti rientrano nella problematica di risolvere un problema inverso. Sono descritti algoritmi per ricostruire la strut-

tura 3D della scena (ed il moto), ovvero per calcolare le coordinate di punti 3D della scena di cui è nota la loro proiezione 2D in ciascuna immagine della sequenza tempo-variante.

Infine, nel cap.7 *Calibrazione* sono riportati gli algoritmi per calibrare il sistema di acquisizione immagini (normalmente una telecamera e stereovisione) fondamentali per rilevare informazioni metriche (rilevare dimensioni di un oggetto oppure determinare misure accurate di distanza oggetto-osservatore) della scena dall'immagine. Sono descritti i vari metodi di calibrazione della telecamera che determinano i relativi parametri intrinseci (lunghezza focale, la dimensione orizzontale e verticale del singolo fotorecettore del sensore oppure il rapporto d'aspetto, la dimensione della matrice del sensore, i coefficienti del modello di distorsione radiale, le coordinate del punto principale ovvero il centro ottico) ed i parametri estrinseci che definiscono la trasformazione geometrica per passare dal sistema di riferimento del mondo a quello della telecamera. La geometria epipolare introdotta nel cap.19 è descritta in questo capitolo per risolvere il problema della corrispondenza di punti omologhi in un sistema di visione stereo con le due telecamere calibrate e non. Con la geometria epipolare è semplificata la ricerca dei punti omologhi tra le immagini stereo introducendo la matrice Essenziale e la matrice Fondamentale. Sono inoltre descritti gli algoritmi per la stima di tali matrici, noti a priori i punti corrispondenti di una piattaforma di calibrazione. Con la geometria epipolare il problema della ricerca dei punti omologhi si riduce a mappare un punto di una immagine sulla corrispondente linea epipolare nell'altra immagine. E' possibile semplificare il problema della corrispondenza attraverso una ricerca monodimensionale punto a punto tra le immagine stereo. Ciò è realizzato con la procedura di allineamento delle immagini, noto come rettifica delle immagini stereo. I diversi algoritmi sono stati descritti alcuni basati sui vincoli della geometria epipolare (telecamere non calibrate dove la matrice fondamentale include i parametri intrinseci) e sulla conoscenza o meno dei parametri intrinseci ed estrinseci di telecamere calibrate. Il capitolo 7 termina con il paragrafo della ricostruzione 3D della scena in relazione della conoscenza a disposizione del sistema stereo di acquisizione. Sono descritte le procedure di triangolazione per la ricostruzione 3D della geometria della scena senza ambiguità, date le proiezioni 2D dei punti omologhi delle immagini stereo, noti i parametri di calibrazione del sistema stereo. Se sono noti solo i parametri intrinseci la geometria 3D della scena è ricostruita stimando i parametri estrinseci del sistema a meno di un fattore di scala non determinabile. Se i parametri di calibrazione del sistema stereo non sono disponibili ma sono note solo le corrispondenze tra le immagini stereo la struttura della scena è recuperata attraverso una trasformazione omografica incognita.