

UNIVERSITÀ DEGLI STUDI DI GENOVA

Scuola Politecnica

Corso di Laurea in Ingegneria Elettronica e Tecnologie dell'Informazione

Tesi di Laurea

**Estrazione di parametri per
classificazione di immagini telerilevate ad
alta risoluzione mediante istogrammi di
gradienti orientati**

Feature extraction for high resolution remote sensing image classification using
histograms of oriented gradients



Relatore

Prof. Gabriele Moser

Laureandi

Margherita PICCINI

Correlatore:

Dott. Vladimir Krylov

Simone ROSSI

Eugenio ZUCCARELLI

ANNO ACCADEMICO 2014-2015

"Anyone who has never made a mistake has never tried anything new."

Albert Einstein

Sommario

Gli ultimi decenni hanno visto un crescente sviluppo di tecniche di classificazione automatica di immagini telerilevate digitali, rendendo la tecnologia del telerilevamento sempre più interessante dal punto di vista del monitoraggio e della gestione ambientale (*Earth Observation - EO*). In particolare, questa tesi si è focalizzata su tecniche di classificazione di immagini ad alta risoluzione (VHR), acquisite da sensori ottici multispettrali, finalizzate all'osservazione di aree urbane.

La tesi esplora l'applicazione di una moderna tecnica di estrazione di *feature*, chiamata "istogrammi di gradiente orientati" (HOG), applicata a immagini multispettrali VHR. L'algoritmo, già utilizzato nell'ambito della *human detection*, ma innovativo nel contesto del telerilevamento, è analizzato dettagliatamente in ogni sua fase, evidenziando come a differenti configurazioni di parametri corrispondano variazioni nelle accuratezze.

Si analizzano tre casi di studio rilevanti, riguardanti immagini di aree urbane di Amiens (Francia) acquisite dal sensore SPOT5. Si tratta di un problema di classificazione interessante e alquanto complesso, dal momento che le regioni coinvolte sono ben diversificate, includendo aree spazialmente omogenee (e.g. "acqua"), strutture geometriche ben definite (e.g. zone urbane) e varie zone di suolo con tessiture differenti (e.g. aree vegetate).

I risultati sperimentali hanno confermato l'opportunità dell'estrazione di *feature* aggiuntive, associate all'informazione spaziale dell'intensità dei pixel, soprattutto in presenza di classi spettralmente molto sovrapposte tra loro. Si è verificato che, nei casi di studio in ambito urbano, la classificazione, tramite l'implementazione degli HOG, presenta limitazioni per quanto riguarda le classi caratterizzate da zone omogenee. Tuttavia, si sono registrati incrementi significativi per quelle classi dove è marcata la presenza di strutture geometriche regolari. Ciò suggerisce le potenzialità di HOG come strumenti di estrazione di *feature* per mappature di aree urbane e ne suggerisce l'uso insieme ad altre tecniche di analisi di tessitura.

Summary

The last decades have seen a significant increase in the development of automatic classification techniques for digital remote sensing images, allowing this field to become appealing for the environmental monitoring and management (Earth Observation). In particular, this thesis aims at the classification of very high resolution (VHR) images, obtained by multispectral optical sensors, for the observation of urban areas.

The thesis explores the application of a advanced feature extraction technique, called "histogram of oriented gradients" (HOG), applied to multispectral VHR images. The algorithm, widely used in the human detection area, but new in this context of remote sensing, has been thoroughly analyzed in each phase, highlighting the correspondance between different parameter sets and different accuracy variations.

Three relevant cases, regarding the Amiens (France) urban area and observed through the SPOT5 sensor, have been studied. This is an interesting and challenging classification problem, since the imaged area is well diversified, including spatially homogeneous regions (e.g. "water"), highly defined geometrical structures (e.g. "urban areas") and several ground portions with different textures (e.g. "vegetated areas").

The experimental results have confirmed the opportunity of additional feature extraction, associated with spatial information of the pixel magnitude, in particular in the presence of classes that strongly overlap in terms of spectral response. In the case studies, regarding the urban areas, the classification, through the HOG implementation, has presented some limitations concerning classes identified by homogeneous areas. However, there were remarkable increases regarding the highly geometrical classes. These results suggest the potential of HOG as a feature extraction tool for urban area mapping and the opportunity to use it together with other feature analysis techniques.

Ringraziamenti

Desideriamo ringraziare tutti coloro che ci hanno aiutato nella stesura della tesi con suggerimenti, critiche ed osservazioni.

Un ringraziamento particolare va al Prof. Gabriele Moser, nostro relatore, e al Dottor Vladimir Krylov, che ci hanno supportato e guidato in questi mesi.

Ringraziamo, inoltre, i colleghi e gli amici che ci hanno incoraggiato e accompagnato in questi anni di studio e vorremmo, infine, ringraziare le persone a noi più care, famiglia e amici, a cui questo lavoro è dedicato.

Indice

Sommario	II
Summary	III
Ringraziamenti	IV
Introduzione	1
1 Classificazione di immagini telerilevate ad alta risoluzione	4
1.1 Cenni sul telerilevamento	5
1.1.1 Sensori e piattaforme	5
1.1.2 Tipologie di sensori	6
1.1.3 Il ruolo della risoluzione	7
1.1.4 Telerilevamento tramite sensori ottici multispettrali	8
1.2 Classificazione di immagini telerilevate	11
1.2.1 Premessa	11
1.2.2 Spazi di rappresentazione	11
1.2.3 La classificazione nello spazio delle <i>feature</i>	12
1.3 Il ruolo dell'informazione spaziale	16
2 Iстограмми di gradienti orientati per estrazione di <i>feature</i>	19
2.1 Schema generale dell'algoritmo	20
2.2 Riduzione del rumore	20
2.3 Calcolo dei gradienti	22
2.4 Costruzione degli istogrammi	26
2.5 Normalizzazione dei blocchi	28
2.5.1 Schemi di normalizzazione	28

2.6	Costruzione del vettore delle feature	29
3	SVM - <i>Support Vector Machine</i>	31
3.1	SVM lineare per classificazione binaria	32
3.1.1	Hard Margin SVM	32
3.1.2	Soft Margin SVM	36
3.2	SVM non lineare per classificazione binaria	39
3.3	Estensioni multiclass	42
4	Valutazione delle prestazioni di un classificatore	44
4.1	Stima della probabilità di errore	45
4.2	Matrice di confusione e parametri di accuratezza	46
5	Risultati sperimentali	48
5.1	Descrizione dei <i>dataset</i>	49
5.1.1	Amiens 2006 - 5m - 10 classi	50
5.1.2	Amiens 2006 - 2.5m - 7 classi	53
5.1.3	Amiens 2012 - 2.5m - 7 classi	56
5.2	Applicazione del classificatore SVM	59
5.3	Applicazione del metodo HOG	59
5.3.1	Riduzione del rumore	60
5.3.2	Calcolo dei gradienti	60
5.3.3	Numero di componenti dei vettori delle <i>feature</i>	61
5.3.4	Dimensione delle celle e dei blocchi	61
5.4	Discussione dei risultati sperimentali	63
5.4.1	Amiens 2006 - 5m - 10 classi	63
5.4.2	Amiens 2012 - 2.5m - 7 classi	69
5.4.3	Amiens 2006 - 2.5m - 7 classi	75
6	Conclusioni	78

Elenco delle tabelle

1.1 Principali intervalli di lunghezza d'onda significativi per il telerilevamento passivo	10
2.1 Esempi di operatori derivativi per il calcolo del gradiente	25

Elenco delle figure

1.1	Schema a blocchi (concettuale) di un sistema di telerilevamento.	5
1.2	Metodi di scansione di sensori <i>line scanner</i> (a), <i>whiskbroom scanner</i> (b) e <i>pushbroom scanner</i> (c).	9
1.3	Spettro elettromagnetico	10
1.4	Spazi di rappresentazione per un'immagine multispettrale	12
1.5	Schema funzionale di un classificatore	15
2.1	Schema a blocchi del metodo HOG	20
2.2	Grafico della funzione gaussiana 2D aente deviazioni standard $\sigma_x \sigma_y$, medie nulle e coefficiente di correlazione nullo	22
2.3	Riduzione del rumore nell'immagine di test tramite filtraggio gaussiano con deviazione standard pari a $\sigma = 2$ pixel	22
2.4	Componenti del gradiente dell'immagine di test	24
2.5	Combinazione delle derivate per estrazione di contorni	24
2.6	Schema esemplificativo della distribuzione spaziale di istogrammi, celle e blocchi	30
3.1	<i>Training set</i> binario linearmente separabile	33
3.2	Margine tra i campioni di <i>training</i> di due classi linearmente separabili	34
3.3	<i>Training set</i> binario non linearmente separabile	37
3.4	Esempio di soft margin nel caso di un <i>training set</i> binario non linearmente separabile	38
3.5	Confronto tra Hard Margin SVM e Soft Margin SVM nel caso di un training set binario linearmente separabile	39
3.6	Trasformazione dello spazio delle <i>feature</i> in uno spazio munito di prodotto scalare ed aente dimensione maggiore. Cerchi pieni e vuoti indicano campioni di training di due classi	40

5.1	<i>Dataset</i> con composizione RGB in falso colore (2000×2200 pixel) acquisito su Amiens (Francia) dal sensore SPOT5 HRG	52
5.2	Analisi della distribuzione dei campioni di training di tre classi nelle proiezioni su due diversi sottospazi delle feature	54
5.3	<i>Dataset</i> con composizione RGB in falso colore (4001×4400 pixel) acquisito su Amiens (Francia) dal sensore SPOT5 HRG nel 2006	55
5.4	<i>Dataset</i> con composizione RGB in falso colore (4001×4400 pixel) acquisito su Amiens (Francia) dal sensore SPOT5 HRG nel 2012	58
5.5	Mappa di classificazione ottenuta applicando SVM a <i>feature</i> spettrali e HOG per il <i>dataset Amiens 2006 - 5m - 10 classi</i>	63
5.6	Matrice di confusione associata alla classificazione del dataset <i>Amiens 2006 - 5m - 10 classi</i> mediante l'applicazione di SVM a feature spettrali e HOG	65
5.7	Differenze fra il numero dei campioni di test classificati correttamente classe per classe con e senza estrazione di <i>feature</i> HOG per il data set <i>Amiens 2006 - 5m - 10 classi</i>	67
5.8	Confronto su una porzione dell'area urbana di 250×250 pixel della mappa di classificazione con e senza estrazione di <i>feature</i> HOG sul <i>dataset Amiens 2006</i> a 5 m (i livelli di grigio distinti rappresentano etichette di classi differenti)	68
5.9	Mappa di classificazione ottenuta applicando SVM a feature spettrali e HOG per il dataset <i>Amiens 2012 - 2.5m - 7 classi</i>	69
5.10	Matrice di confusione associata alla classificazione del dataset <i>Amiens 2012 - 2.5m - 7 classi</i> mediante l'applicazione di SVM a feature spettrali e HOG	71
5.11	Differenze tra il numero di campioni di test classificati correttamente classe per classe con e senza estrazione di <i>feature</i> per il <i>dataset Amiens 2012</i> a 2.5 m	72
5.12	Immagine binaria della struttura cartografica di strade ed autostrade della regione di Amiens: strade, autostrade e terreni ad esse collegati sono rappresentati con il bianco e lo sfondo in nero.	73
5.13	Confronto su una porzione della periferia di Amiens di 250×250 pixel della mappa di classificazione con e senza estrazione di <i>feature</i> HOG e con l'aggiunta della cartografia stradale sul <i>dataset Amiens 2012</i> a 2.5 m (i livelli di grigio distinti rappresentano etichette di classi differenti)	74
5.14	Mappa di classificazione ottenuta applicando SVM a feature spettrali e HOG per il dataset <i>Amiens 2006 - 2.5m - 7 classi</i>	75

- 5.15 Matrice di confusione associata alla classificazione del dataset *Amiens 2006*
- *2.5m - 7 classi* mediante l'applicazione di SVM a feature spettrali e HOG 77
- 5.16 Differenze tra il numero di campioni di test classificati correttamente classe
per classe con e senza estrazione di *feature* per il *dataset Amiens 2006* a 2.5 m 77

Introduzione

Gli ultimi decenni hanno visto un crescente sviluppo di tecniche di telerilevamento orientate all'osservazione della Terra (*Earth Observation - EO*), che permettono, oggigiorno, di avere un'ampia disponibilità di immagini digitali, acquisite tramite sensori montati a bordo di satelliti o aviotrasportati. Un ruolo chiave in tale ambito è stato giocato dal rapido aumento di missioni spaziali, volte alla messa in orbita di satelliti dotati di sensori con diverse caratteristiche, quali sensori ottici ad alta o altissima risoluzione spaziale e sensori ottici iperspettrali, capaci di fornire informazioni cartografiche di grande dettaglio. Il dato satellitare, grazie alla multi-temporalità e multispettralità delle sue osservazioni, rende possibile un monitoraggio ripetitivo di vaste aree geografiche, permettendo l'analisi di porzioni di suolo a scala globale, regionale e locale.

Data l'alta disponibilità di tali immagini ad alta risoluzione (*Very High Resolution - VHR*), si è visto un crescente interesse per le procedure automatiche di classificazione, che permettono di classificare vaste porzioni di superficie terrestre in tempi sempre più brevi, risultando di estrema utilità a molteplici applicazioni di carattere ambientale, quali sviluppo urbano, riqualificazione di aree urbane inutilizzate, map-patura agricola, salvaguardia ambientale e gestione delle risorse naturali.

L’incremento della risoluzione spaziale nelle immagini telerilevate VHR ha introdotto, però, nuove problematiche nella classificazione. Se per risoluzioni basse o grossolane (ad esempio, intorno a 30/50 metri) si può sovente evitare di tenere in considerazione l’informazione contestuale, ed accuratezze accettabili sono spesso ottenibili operando con i soli canali spettrali, per alte risoluzioni (2 – 5 metri) dettagli spaziali molto più precisi risultano apprezzabili nell’immagine ed una modellazione opportuna dell’informazione spaziale diventa necessaria per discriminare accuratamente le classi tematiche di interesse. L’alta correlazione fra pixel adiacenti che sussiste a risoluzione così elevata ha portato allo studio di diversi metodi per tenere in considerazione, in fase di classificazione, la distribuzione spaziale delle intensità dei pixel.

Le due filosofie che sono nate come risposta a questo problema hanno visto coinvolti, da una parte, classificatori non contestuali affiancati a metodi di estrazione di *feature* contestuali e, dall’altra, l’uso di metodi di classificazione contestuali, che esplicitamente incorporano l’informazione circa il contesto spaziale di ogni singolo pixel.

Tra i due, l’approccio che si è deciso di seguire in questa tesi è stato il primo: l’uso di una *Support Vector Machine* (SVM), come classificatore non contestuale, affiancato all’algoritmo di *histogram of oriented gradient* (HOG), per estrazione di *feature*. Con questa tesi, infatti, si desidera esplorare l’uso dell’algoritmo HOG, sviluppato nell’ambito della *human detection*, testando l’applicabilità di questo approccio al contesto della classificazione di immagini di aree urbane multispettrali ad alta risoluzione. L’intento delle *feature* HOG è quello di descrivere il comportamento locale del gradiente di un’immagine, cercando di enfatizzare, per quanto possibile, strutture geometriche ben definite; ciò è giustificato dal fatto che, intuitivamente,

un oggetto possa essere identificato grazie al proprio contorno.

Il classificatore SVM è stato scelto in seguito alle sue proprietà di robustezza al numero di *feature* ed alla sua accuratezza in numerose applicazioni.

Questo tesi è organizzata nel seguente modo: nel Capitolo 1 viene introdotto il contesto del telerilevamento, con accenni alle tipologie di sensori utilizzabili, ai tipi di classificatori e allo stato dell'arte delle varie strategie per l'estrazione di informazioni spaziali a fini di classificazione; nel Capitolo 2, invece, si è presentata un'analisi dettagliata dell'algoritmo di estrazione di *feature* HOG introdotto da Dalal e Triggs [6] con particolare riferimento alle modifiche da noi apportate necessarie per l'uso di questo approccio al contesto del telerilevamento di aree urbane. Il Capitolo 3 propone la trattazione matematica della teoria alla base del classificatore SVM. Nel Capitolo 5 si procede con una analisi dei risultati ottenuti con l'implementazione dell'algoritmo HOG e della SVM, valutandone la potenzialità, i punti di forza e debolezza e le situazioni nelle quali è vantaggioso o no utilizzarli, basandoci sugli indici di accuratezza introdotti nel Capitolo 4 e su sperimentazione condotta con tre *dataset* reali particolarmente complessi.

Capitolo 1

Classificazione di immagini telerilevate ad alta risoluzione

Il telerilevamento (*remote sensing*) viene definito come la disciplina che permette di ricavare informazioni su oggetti posti a distanza da uno o più sensori posti non in contatto diretto con gli oggetti stessi. Questa definizione generale comprende un vasto range di sotto-discipline atte ad estrarre informazioni, tra le quali la mappatura di dati di interesse ambientale, l'identificazione di oggetti posti sul fondale marino o fino a tecniche di *human detection*. Nel nostro caso, ci focalizzeremo in particolare nell'ambito dell'Osservazione della Terra (*Earth Observation*, EO) in cui l'oggetto da analizzare sarà una porzione di suolo.

1.1 Cenni sul telerilevamento

1.1.1 Sensori e piattaforme

Nel caso di telerilevamento al fine di *EO*, i sensori sono tipicamente montati su un aereo (sensore aviotrasportato) o su un satellite (sensore satellitare), i quali, a partire dall'onda elettromagnetica incidente su di essi, generano una immagine digitale, che poi può essere elaborata da un calcolatore. I sensori hanno, infatti, lo scopo di acquisire parametri ambientali utili ad una successiva elaborazione delle informazioni, seguendo il procedimento rappresentato dallo schema a blocchi riportato in Figura 1.1.

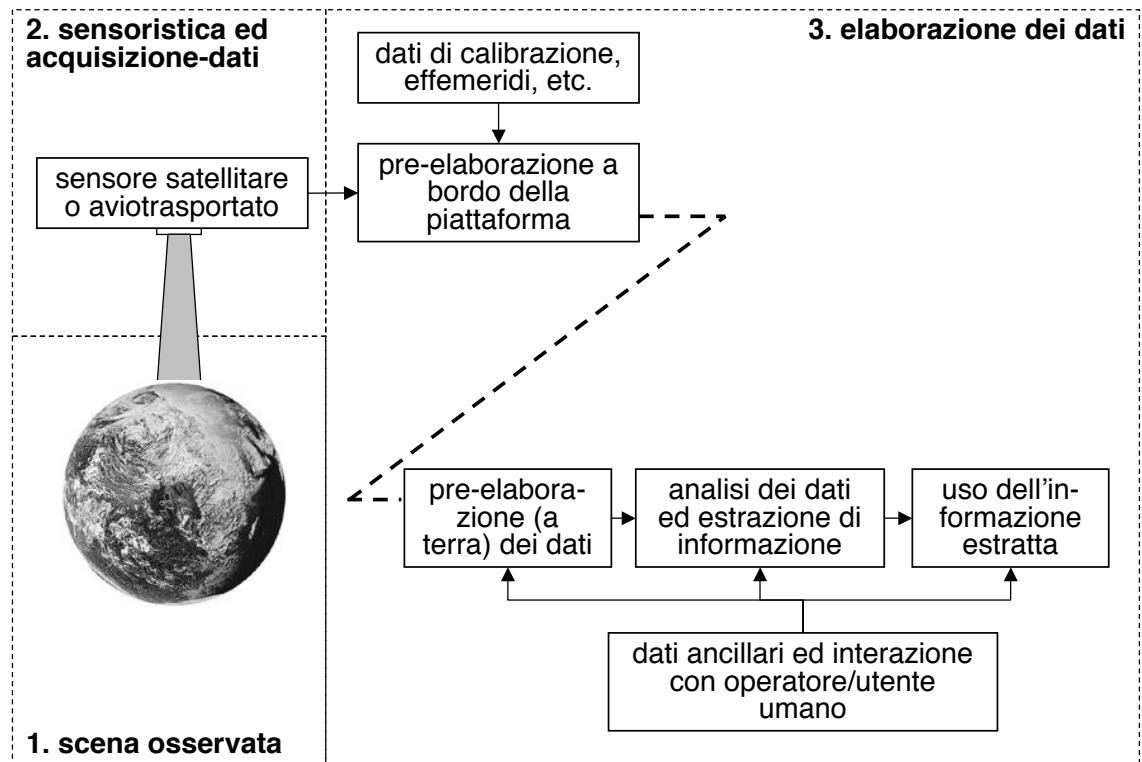


Figura 1.1. Schema a blocchi (concettuale) di un sistema di telerilevamento.

Il procedimento generale è il seguente:

- innanzitutto la radiazione emessa dall'oggetto in esame (ad esempio una porzione di suolo) viene ricevuta in ingresso al sensore, per essere elaborata dal sistema;
- il sistema di elaborazione prevede in taluni casi, una fase iniziale di pre-elaborazione dei dati a bordo della piattaforma, basata su informazioni inerenti ai sensori (ad esempio dati di calibrazione del sensore) o alla piattaforma stessa (ad esempio effemeridi);
- successivamente, vengono effettuate operazioni di pre-elaborazione analoghe ma a terra, aventi fini quali un'ulteriore calibrazione dei dati o la rimozione di distorsioni geometriche o radiometriche dovute al movimento della piattaforma;
- infine il sistema procede all'elaborazione vera e propria dei dati telerilevati al fine di estrarre l'informazione che viene infine fornita all'utilizzatore.

1.1.2 Tipologie di sensori

Generalmente, i sensori per telerilevamento vengono divisi in due macro-famiglie, i sensori passivi e quelli attivi. La prima tipologia non trasmette alcun segnale, bensì, riceve unicamente la radiazione emessa dall'oggetto in esame, ovvero la radiazione elettromagnetica emessa dalla porzione di superficie, la quale può essere spontanea (tipicamente radiazione nell'intervallo dell'infrarosso vicino) oppure la radiazione riflessa o diffusa proveniente dal sole (radiazione infrarossa e/o nello spettro del visibile). I sensori attivi, invece, trasmettono un'onda elettromagnetica nella direzione della superficie in esame e analizzano il segnale (simile ad un eco sonoro) ritrasmesso dalla porzione di suolo stessa. Tali onde elettromagnetiche sono tipicamente segnali laser, usati da sensori *Light Detection And Ranging* (LIDAR), oppure a microonde,

acquisiti tramite sensori *RAdio Detection And Ranging* (RADAR).

La nostra trattazione si baserà unicamente su segnali rilevati tramite sensori passivi, in particolare sensori ottici.

1.1.3 Il ruolo della risoluzione

Un parametro di vitale importanza per un sensore orientato all' *EO* è la risoluzione, con la quale si intende una risoluzione spaziale, temporale e spettrale. La prima rappresenta il più piccolo dettaglio della superficie in esame che risulta distinguibile dopo essere stata estratta dal sensore, in particolare questo parametro è, al meglio, come la grandezza della superficie rappresentata da un singolo pixel. La risoluzione temporale, invece, è definita come la frequenza con cui il sensore osserva una stessa porzione di superficie. Essa è infatti definita come il tempo intercorso tra due passaggi del sensore sulla stessa area geografica, intervallo di tempo che può variare dal mese fino, addirittura, a poche decine di minuti. Infine, la risoluzione spettrale viene definita come il numero di bande (o canali) misurate per ciascun pixel e dalla larghezza di banda di ogni singolo canale. Ad esempio, il sensore iperspettrale AVIRIS campiona la radiazione incidente acquisendo 224 bande distinte, ognuna avente una larghezza pari a 9.3 nm, rendendolo un sensore ad alta risoluzione temporale, a differenza di sensori pancromatici che elaborano solamente l'intervallo di lunghezze d'onda della radiazione visibile.

1.1.4 Telerilevamento tramite sensori ottici multispettrali

La nostra analisi si focalizzerà su sensori passivi multispettrali costituiti da sistemi di elaborazione ottica, i quali indirizzano la radiazione incidente su uno o più foto-rivelatori, che trasducono una grandezza fisica (l'intensità della radiazione elettromagnetica) in una tensione elettrica. Queste tipologie di sensori vengono trasportati da una piattaforma che viaggia ad una quota h e fa percorrere al sensore una direzione di volo (direzione *in-track*), lungo la quale esso effettua uno scan lungo la direzione ortogonale (detta direzione *cross-track*). I sensori possono essere suddivisi, in base alla tecnica di scansione utilizzata, in tre famiglie:

- *Line Scanner*, sensori che montano un solo rivelatore ottico che scandisce l'intera riga per mezzo di uno specchio oscillante;
- *Whiskbroom Scanner*, i quali possiedono un array di rivelatori orientati secondo la direzione *in-track*, che scandiscono, in questo modo, più linee alla volta;
- *Pushbroom Scanner*, contenenti un numero elevato di foto-rivelatori, lungo la direzione *cross-track*, i quali permettono di scandire intere righe dell'immagine senza l'utilizzo di parti mobili.

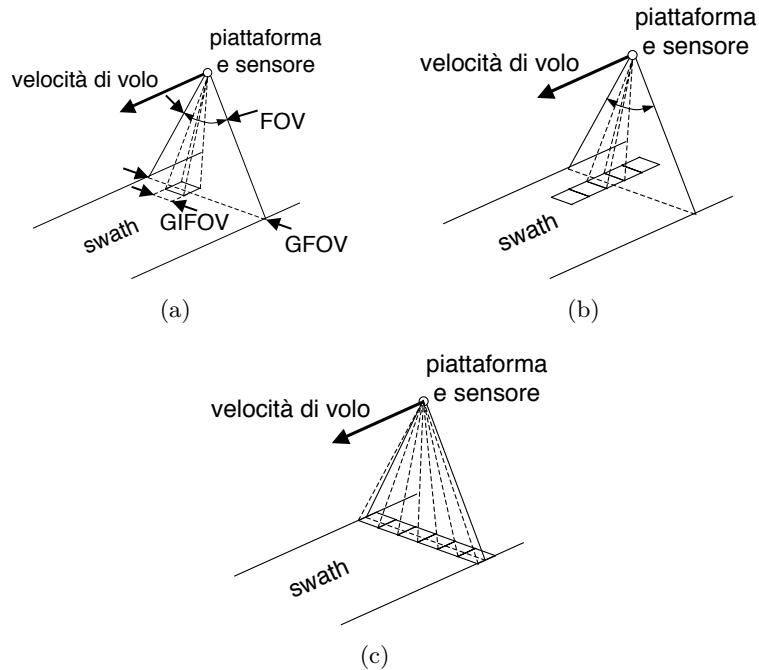


Figura 1.2. Metodi di scansione di sensori *line scanner* (a), *whiskbroom scanner* (b) e *pushbroom scanner* (c).

Parametri chiave dei sensori ottici sono il cosiddetto *Field Of View* (FOV), ovvero l'ampiezza espressa in radianti della zona di suolo osservata in direzione *cross-track* e la corrispondente larghezza a terra della striscia osservata, il *Ground Field Of View* (GFOV). In modo analogo, l'estensione angolare di ogni foto-rivelatore è detta *Istantaneous Field Of View* (IFOV) mentre la sua proiezione viene detta *Ground Instantaneous Field Of View* (GIFOV). Fra queste grandezze esistono delle relazioni espresse dalle seguenti equazioni:

$$GFOV = 2h \tan \frac{FOV}{2} \quad (1.1)$$

$$GIFOV = 2h \tan \frac{IFOV}{2} \quad (1.2)$$

Tali sensori ottici passivi analizzano vari range di frequenze che tendenzialmente vanno dall’infrarosso termico (TIR, 8-9.5 μm , 10-14 μm) allo spettro del visibile (VIS, 0.4-0.7 μm), rilevando la radianza spettrale, grandezza che permette di descrivere la distribuzione spaziale della radiazione elettromagnetica.

Tabella 1.1. Principali intervalli di lunghezza d’onda significativi per il telerilevamento passivo

Nome	Abbreviazione	Lunghezza d’onda [μm]
Visibile	VIS	0.38 – 0.76
Near InfraRed	NIR	0.76 – 1.1
Short Wave InfraRed	SWIR	1.1 – 1.35, 1.4 – 1.8, 2 – 2.5
Mid Wave InfraRed	MWIR	3 – 4, 4.5 – 5
Thermal Infrared	TIR	8 – 9.5, 10 – 14

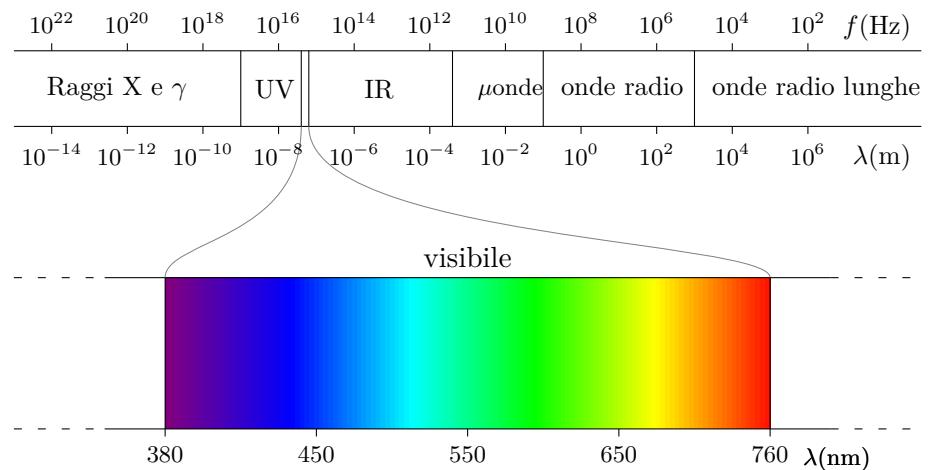


Figura 1.3. Spettro elettromagnetico

1.2 Classificazione di immagini telerilevate

1.2.1 Premessa

Si definisce classificazione di immagini telerilevate, il processo di assegnazione di una "etichetta" a ciascun pixel dell'immagine, in modo tale da renderlo appartenente ad una specifica classe, rappresentativa di una data copertura del suolo. La classificazione è, sovente, il primo step del processo di estrazione di dati di carattere informativo da una immagine telerilevata, che vengono forniti, poi, alle successive fasi di riconoscimento (*matching*) ed interpretazione. Queste tre fasi sono oggetto della disciplina del *Pattern Recognition*, il cui obiettivo è appunto sviluppare tecniche con cui implementare, automaticamente o semi-automaticamente, i tre processi. Oltre alla generazione di mappe tematiche tramite sistemi di telerilevamento, il *pattern recognition* abbraccia vari ambiti quali l'analisi di immagini biomedicali, orientate alla robotica (*Computer Vision*), o per la videosorveglianza.

1.2.2 Spazi di rappresentazione

Esistono tre principali metodologie di rappresentazione dei dati, la rappresentazione nello "spazio-immagine" (*Image Space*), la rappresentazione nello "spazio spettrale" (*Spectral Space*) e quella nello "spazio delle feature" (*Feature Space*). La prima e più immediata consiste nel visualizzare i dati canale per canale, oppure, tramite terne RGB. La seconda tipologia consiste, invece, nel visualizzare per ciascun pixel, i livelli di grigio di ogni canale, per rappresentarli poi in un grafico. La rappresentazione nello "spazio delle feature" si basa sull'assegnazione di assi cartesiani distinti ad ogni banda, e, così facendo, ad ogni pixel viene associato un vettore n-dimensionale, con n il numero di canali. Quest'ultima rappresentazione non solamente evidenzia i livelli

di grigio dei pixel ma anche la distribuzione statistica nelle varie bande, rendendola generalmente più vantaggiosa rispetto alle altre due in quanto a differenti distribuzioni nello spazio delle *feature* corrispondono tipicamente a differenti coperture di suolo.

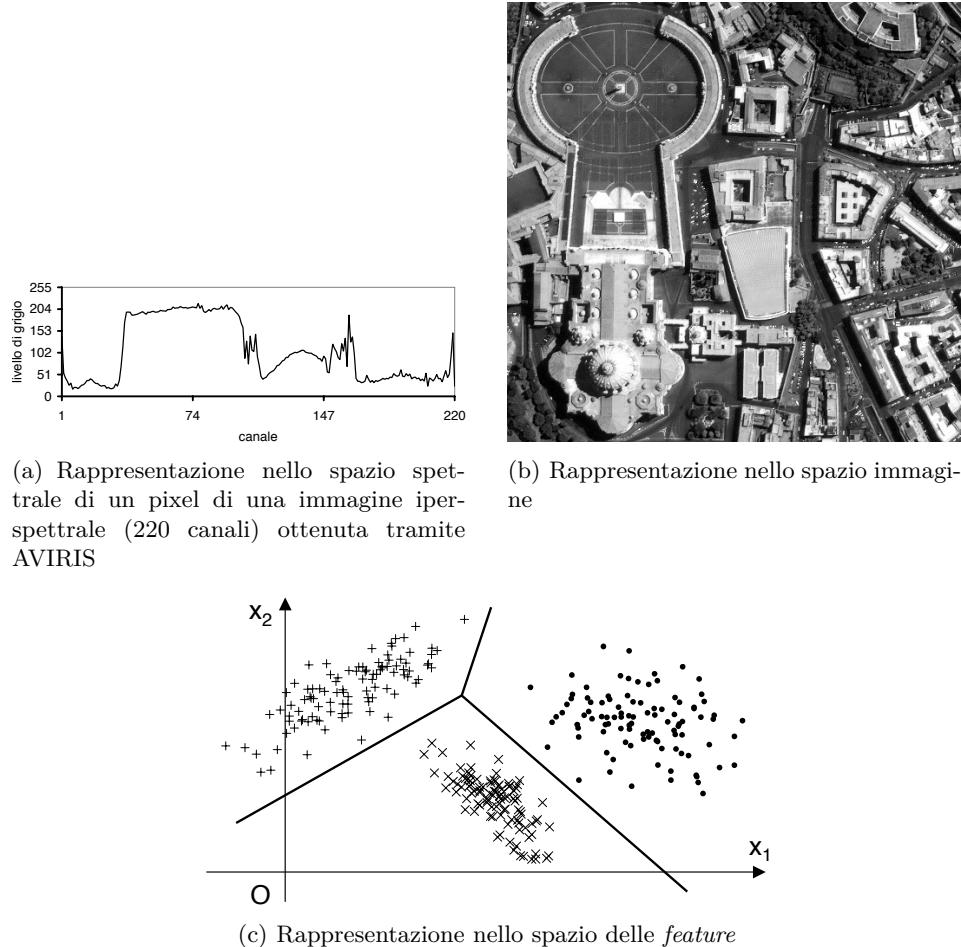


Figura 1.4. Spazi di rappresentazione per un'immagine multispettrale

1.2.3 La classificazione nello spazio delle *feature*

A fini di classificazione, la rappresentazione nello spazio delle *feature* si rivela, solitamente, la più vantaggiosa, in quanto gli andamenti spettrali di pixel appartenenti

a classi distinte possono essere maggiormente separabili. In quest'ottica, classificare significa quindi partizionare lo spazio delle "feature" in opportuni sottoinsiemi, ciascuno associato ad una data classe.

I classificatori possono essere divisi in due macro-famiglie, classificatori supervisionati (*supervised*) e non supervisionati (*unsupervised*). La prima tipologia prevede principalmente due fasi, una fase iniziale di addestramento (*training*) e una di verifica (*test*); nel primo step, il sistema ottimizza parametri del classificatore tramite un insieme di pixel preclassificati (*training set*), fino a raggiungere una adeguata accuratezza. Successivamente, il sistema viene testato in modo analogo alla fase di training, attraverso un insieme di pixel preclassificati ma differenti rispetto a quelli coinvolti nella fase di *training*.

Nei classificatori non supervisionati, invece, non viene utilizzato alcun *training set*, solitamente perché non sono note né facilmente identificabili le classi coinvolte nell'applicazione in esame.

Nei capitoli a seguire la nostra attenzione sarà focalizzata sul concetto di classificazione supervisionata, che si basa, in generale, su tre assunti:

- si ha una conoscenza a priori esaustiva su un sottoinsieme di pixel preclassificati (*training set*);
- le classi esistono in numero finito e sono note *ex ante*;
- ogni pixel è rappresentabile tramite un insieme di valori che vengono raccolti nel cosiddetto vettore delle *feature* (*feature vector*).

Concetti chiave e definizioni

Data un’immagine telerilevata, a ciascun pixel $(m, n) \in \mathbb{Z}^2$ può essere associato un vettore d-dimensionale delle *feature* $\mathbf{x}(m, n) \in \mathbb{R}^d$, le cui componenti (le d *feature*) possono essere non solamente i livelli di grigio del pixel (m, n) nelle varie bande, ma anche parametri aggiuntivi come ad esempio i cosiddetti parametri di tessitura (o *feature* di tessitura).

Le *feature* di tessitura vengono estratte al fine di analizzare le differenze nella distribuzione spaziale dei livelli di grigio dei pixel, permettendo di distinguere coperture di suolo differenti. In generale, se i pixel estratti da classi differenti sono situati in zone disgiunte dello spazio delle feature, la accuratezza di classificazione è tanto maggiore quanto più sono separate le regioni. Se infatti, in una stessa regione si trovano classi distinte, esse risulteranno sovrapposte e quindi distinguerle risulterà largamente più difficoltoso.

Dato un insieme $\Omega = \{\omega_1, \dots, \omega_c\}$ costituito da C classi distinte, note a priori, si assume che ogni oggetto o entità da analizzare sia appartenente ad una ed una sola classe (nel nostro caso ad una data copertura del suolo). A ciascun pixel è associata, quindi, anche un’etichetta di classe $y(m, n) \in \Omega$ e l’immagine $y(m, n)$ (con $m, n \in \mathbb{Z}$) è detta mappa di classificazione.

Il *training set* è l’insieme dei pixel di cui è nota a priori l’etichetta di classe ed è definito dall’insieme delle coppie $\{\mathbf{x}_i, y_i\}$, $i = 1, \dots, N$ con N il numero di vettori di *training* scelti. L’insieme di tali etichette rappresenta un’ulteriore immagine, detta *mappa di training*, che evidenzia alcuni pixel dell’immagine assegnati a ciascuna

classe, distinguendoli dai "pixel di sfondo", cioè quei pixel per i quali l'etichetta di classe è incognita.

Il processo di classificazione, come già accennato precedentemente, è diviso in due passi: *learning* e *prediction* (fase di addestramento e fase di predizione). Lo scopo della fase di *learning* è quello di costruire un modello di classificatore, tramite la definizione di una *regola di decisione* (o anche detta *funzione discriminante*).

In termini generali, la *regola di decisione* è una funzione $f : \mathbb{R}^d \rightarrow \mathbb{R}$ modellata sulla base dell'andamento statistico del *training set* in modo tale che f valutata per un pixel incognito restituisca una stima dell'etichetta da associargli. Le funzioni discriminanti possono essere definite con vari metodi, cui corrispondono varie tecniche di classificazione (nel Capitolo 3 ne viene descritto uno dei possibili).

Conclusa questa fase, si procede con la classificazione vera e propria: si scandisce l'intera immagine valutando ogni pixel tramite la *regola di decisione*, realizzando, in questo modo, una stima della mappa di classificazione.

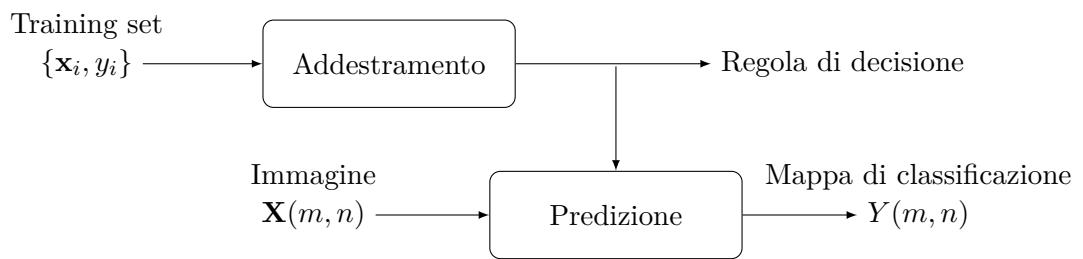


Figura 1.5. Schema funzionale di un classificatore

1.3 Il ruolo dell'informazione spaziale

Una ulteriore distinzione che permette di differenziare tra loro le tipologie di classificatori è quella inherente al ruolo dei pixel adiacenti nella analisi della copertura al suolo. Esistono, infatti, classificatori cosiddetti non contestuali, in cui la copertura viene analizzata senza tener conto dei pixel adiacenti, risparmiando così un alto costo computazionale, ma trascurando la forte correlazione tra pixel limitrofi. Infatti, la probabilità che una regione dell'immagine sia formata da pixel tutti appartenenti alla stessa classe è molto elevata; si pensi solamente a zone boschive o fluviali.

Queste due tipologie di classificatori sono chiaramente differenti anche in base all'ambito in cui vengono applicate; mentre le tecniche di classificazione supervisionata non contestuale risultano molto efficaci e largamente consolidate per le immagini con una risoluzione piuttosto grossolana, esse mostrano forti limiti per le immagini *Very High Resolution*. Una maggiore risoluzione spaziale comporta, infatti, una maggiore eterogeneità e una corrispondente buona definizione delle strutture geometriche quali strade e edifici, caratteristiche che rendono necessario l'utilizzo di classificatori contestuali. A tal fin, un ruolo chiave viene giocato principalmente da tre approcci metodologici, l'estrazione di parametri di tessitura, metodi basati su regioni e oggetti e i *Markov Random Field* (MRF).

Estrazione dei parametri di tessitura

L'estrazione di *texture* ha come obiettivo quello di rilevare, in una determinata regione dell'immagine, strutture ripetitive nella distribuzione spaziale dei pixel quali zone urbane o boschive. Ciò fornisce una sorgente di dati complementare per le applicazioni in cui l'informazione relativa unicamente allo spettro dell'immagine risulta non

sufficiente ai fini della classificazione. Le principali tecniche di estrazione di parametri di tessitura si basano sui semivariogrammi, che consistono in una statistica del secondo ordine delle intensità dei pixel, sulla morfologia matematica, sull'uso della matrice di co-occorrenza dei livelli di grigio, o su statistiche associate a confronti iistogrammi locali. Il calcolo di tali *feature* coinvolge tipicamente l'uso di algoritmi basati su finestre mobili.

Tecniche region-based

Gli approcci basati invece sulle regioni (*region-based methods*) si basano su tecniche che puntano a suddividere le immagini in segmenti o regioni omogenee. In generale, una buona tecnica di segmentazione possiede:

- pixel nella stessa categoria aventi livelli di grigio simili che formano una regione connessa;
- pixel adiacenti che sono in categorie differenti hanno valori differenti.

In generale, una tecnica di segmentazione punta a suddividere l'immagine in un insieme di regioni (anche dette segmenti) che spesso corrispondono a oggetti o porzioni di oggetti a se stanti. Una famiglia di tecniche tra le più utilizzate in tale ambito è quella degli algoritmi di crescita delle regioni (*region-growing*), i quali, progressivamente, espandono un insieme di segmenti a partire da un determinato insieme di punti detti semi (*seed points*). In particolare, questo approccio alla segmentazione esamina i pixel adiacenti di un seme e determina se i pixel adiacenti devono essere aggiunti alla regione. Il processo è così iterato fino alla generazione della intera regione.

Markov random field

I *Markov Random Fields*, che generalizzano il concetto di catena markoviana monodimensionale ad un sistema 2D, massimizzano l'accuratezza tramite la dipendenza che sussiste tra pixel adiacenti. Essi offrono, infatti, una soluzione computazionalmente efficiente per restringere la zona di interesse dall'intera immagine (elaborazione globale) ad un intorno del pixel (elaborazione locale). In particolare, siano i e j due pixel dell'immagine, si ha allora che se la funzione di probabilità $P(Y) > 0$ per ogni configurazione Y e se la seguente condizione è garantita per tutti i pixel i dell'immagine, si ha allora che:

$$P(y_i|y_j, j \neq i) = P(y_i|y_j, j \sim i) \quad (1.3)$$

con $i \sim j$ che indica che i pixel i e j sono vicini rispetto ad un sistema di intorni definito sul reticolo dei pixel.

Ciò indica che la distribuzione di probabilità delle etichette di ciascun pixel i , condizionata ai valori di tutti gli altri pixel dell'immagine, può essere ristretta alla distribuzione delle etichette di i condizionato solamente alle etichette dei pixel adiacenti. Si può chiaramente osservare come tale definizione sia una generalizzazione dei processi markoviani monodimensionali, in cui la probabilità di transizione da uno stato ad un altro dipende unicamente dallo stato precedente.

Capitolo 2

Iistogrammi di gradienti orientati per estrazione di *feature*

In questo capitolo verrà presentato il metodo di estrazione di *feature* tramite istogrammi di gradienti orientati (*Histogram of Oriented Gradients*, HOG). Questo metodo, introdotto per la prima volta da Dalal e Triggs [6], si basa sulla valutazione di istogrammi calcolati in base alla direzione ed alle intensità dei gradienti dell’immagine in ingresso. L’idea di base è che la forma di un oggetto possa essere rappresentata piuttosto bene mediante la distribuzione di modulo e direzione dei gradienti locali.

2.1 Schema generale dell’algoritmo

Dopo aver calcolato il gradiente dell’immagine pixel per pixel, l’immagine stessa viene divisa in piccole regioni spaziali denominate "celle". Per ogni cella viene costruito un istogramma sulla base della direzione del gradiente dei pixel in essa contenuti. Per una migliore invarianza agli effetti dovuti all’illuminazione viene effettuata una normalizzazione degli istogrammi delle celle, ottenuta sulla base di una regione spaziale di dimensione maggiore denominata "blocco". I singoli istogrammi, combinati insieme, danno origine a vettori delle *feature* che vengono poi utilizzati da un classificatore, al fine di classificare l’immagine di partenza. Lo schema a blocchi del descrittore HOG è rappresentato in Figura 2.1.

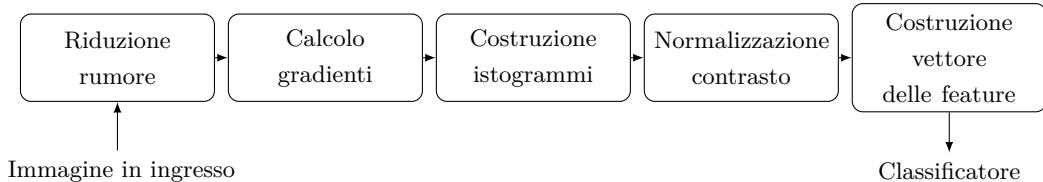


Figura 2.1. Schema a blocchi del metodo HOG

Di seguito verranno descritte in maniera dettagliata le singole fasi che costituiscono l’approccio HOG, evidenziando le scelte progettuali effettuate e analizzando come i diversi parametri impattino sulla prestazione.

2.2 Riduzione del rumore

La strumentazione di un sensore passivo multispettrale introduce, in generale, disturbi nella misura della radianza, dovuti ad esempio ad una variazione di emittanza o illuminazione solare o, in generale, all’aleatorietà intrinseca nel processo di misura

remota. Tali contributi sono indicati complessivamente con il termine di rumore. Questo può essere modellato come rumore additivo gaussiano¹ Per limitarne gli effetti sull’immagine di ingresso, è opportuno applicare un filtraggio per riduzione del rumore.

Una tipologia di filtraggio può essere ottenuta tramite la convoluzione con una gaussiana bidimensionale (Figura 2.2) il che, essendo un filtro passabasso, riduce le componenti di rumore in alta frequenza. Lo stesso tipo di filtraggio può essere applicato anche sulle immagini in uscita dall’algoritmo HOG.

La funzione gaussiana 2D di deviazioni standard σ_x σ_y e coefficiente di correlazione nullo è data da:

$$H_\sigma(x, y) = \frac{1}{\sigma_x \cdot \sqrt{2\pi}} \cdot e^{-\frac{x^2}{2(\sigma_x)^2}} \cdot \frac{1}{\sigma_y \cdot \sqrt{2\pi}} \cdot e^{-\frac{y^2}{2(\sigma_y)^2}} \quad (2.1)$$

La sua versione discreta si ottiene campionandola su una finestra quadrata di dimensioni $K \times K$ con $K > 3 \cdot 2\sqrt{\sigma}$ per rendere trascurabili gli effetti del troncamento. La maschera così ottenuta viene moltiplicata per $\frac{1}{c}$, dove il fattore di normalizzazione c è scelto in modo tale che

$$\sum_j \sum_k g_{jk} = 1, \quad (2.2)$$

dove g_{jk} è il generico elemento della gaussiana campionata.

¹Come conseguenza del teorema del limite centrale, la distribuzione di probabilità di ogni pixel può essere modellata come gaussiana, visto l’elevato numero di contributi che concorrono alla definizione del pixel stesso e che sono originati nelle cella di risoluzione corrispondenti.

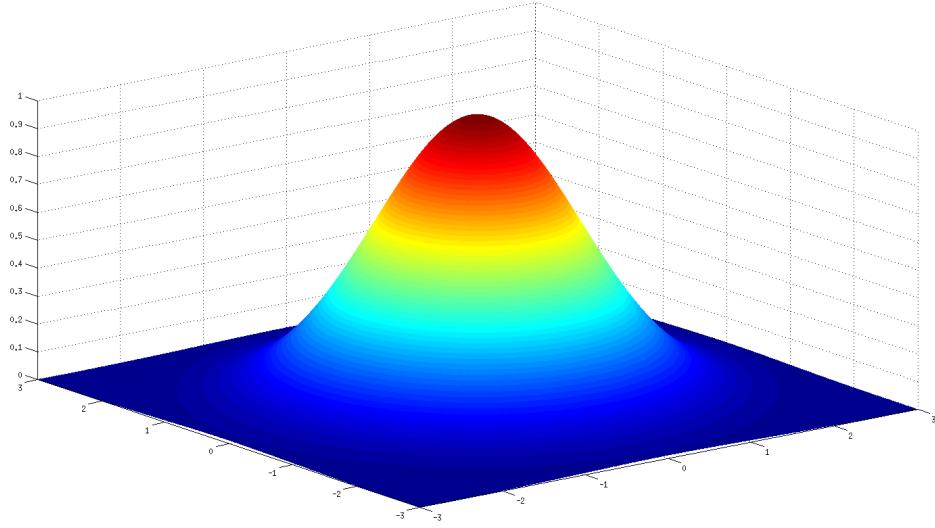
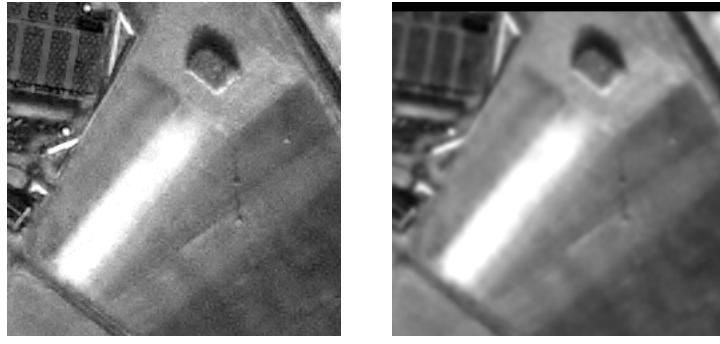


Figura 2.2. Grafico della funzione gaussiana 2D avente deviazioni standard σ_x σ_y , medie nulle e coefficiente di correlazione nullo



(a) Immagine originale con rumore spazialmente uniforme (b) Risultato di filtraggio gaussiano

Figura 2.3. Riduzione del rumore nell’immagine di test tramite filtraggio gaussiano con deviazione standard pari a $\sigma = 2$ pixel

2.3 Calcolo dei gradienti

Il gradiente di un’immagine misura la sua variazione direzionale di intensità. Matematicamente, il gradiente di una funzione differenziabile di due variabili reali associa

ad ogni punto del dominio un vettore 2D con componenti date dalle derivate parziali, calcolate sulle direzioni orizzontali e verticali. Poiché la funzione intensità di un’immagine è conosciuta solo in punti discreti, si assume che le sue derivate siano calcolate su di una funzione continua campionata nei punti dell’immagine.

Dal punto di vista matematico, detta $F(x, y)$ una funzione continua e derivabile, il suo gradiente è dato da:

$$\nabla F = \frac{\partial F}{\partial x} \hat{x} + \frac{\partial F}{\partial y} \hat{y} \quad (2.3)$$

dove:

- $\frac{\partial F}{\partial x} = G_x$ è la componente del gradiente calcolato lungo la direzione x;
- $\frac{\partial F}{\partial y} = G_y$ è la componente del gradiente calcolato lungo la direzione y.

Nel caso di immagini bidimensionali su griglie discrete di pixel, sovente, approssimazioni di queste derivate parziali possono essere definite al variare del grado di robustezza al rumore. Il seguente schema evidenzia uno dei metodi di approssimazione più comune (Figura 2.5).

La stima di G_x e G_y si ottiene dall’espressione della derivata monodimensionale limitata ad un piccolo intorno (spesso solo 3 pixel). L’approccio più semplice utilizza una risposta all’impulso monodimensionale di tipo $\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$.

Dal momento che la derivata, soprattutto quando calcolata su un intervallo così breve, è molto sensibile al rumore, si usa mediare in direzione ortogonale prima di effettuare la differenza per stabilizzarla. Con l’operatore di Prewitt, ad esempio, G_y viene stimata mediante la differenza in orizzontale della media in verticale calcolata su tre punti. Nella Tabella 2.3 sono riassunti le varianti più comuni.

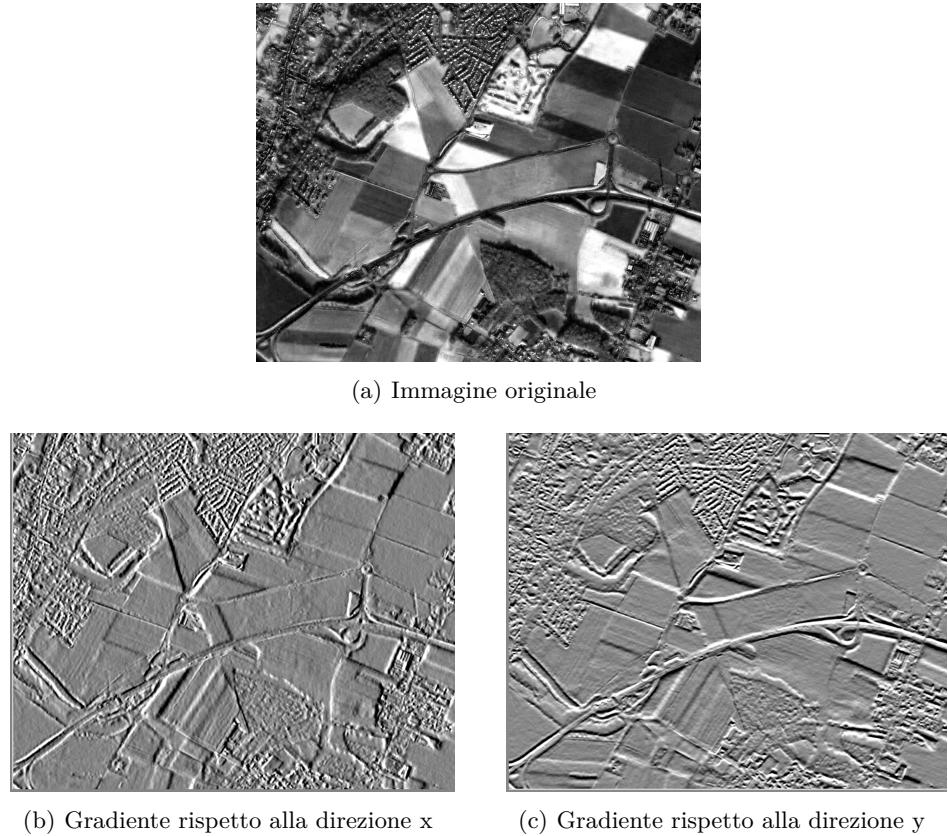


Figura 2.4. Componenti del gradiente dell’immagine di test

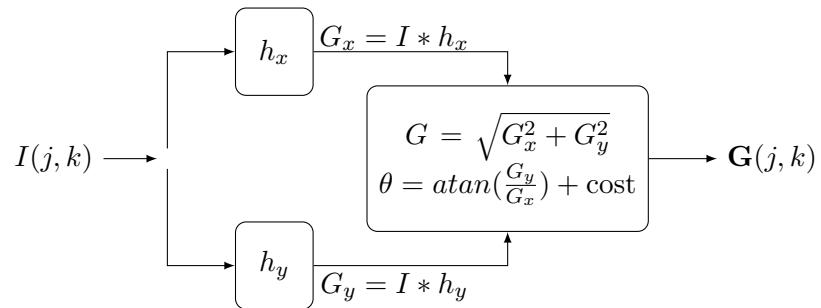


Figura 2.5. Combinazione delle derivate per estrazione di contorni

Questi operatori forniscono risultati accettabili per immagini poco rumorose, infatti è opportuno che l’immagine, come introdotto precedentemente, in ingresso sia filtrata al fine di limitare gli effetti del rumore.

Tabella 2.1. Esempi di operatori derivativi per il calcolo del gradiente

Gradiente per riga Gradiente per colonna

	Roberts	$\begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
	Prewitt	$\frac{1}{3} \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$	$\frac{1}{3} \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$
	Sobel	$\frac{1}{4} \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$	$\frac{1}{4} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$

Siano G_x e G_y le immagini gradiente generate da:

$$G_x(j, k) = Img(j, k) * h_x(j, k) \quad (2.4)$$

$$G_y(j, k) = Img(j, k) * h_y(j, k) \quad (2.5)$$

dove $*$ rappresenta la convoluzione e dove h_x e h_y rappresentano rispettivamente le risposte all'impulso degli operatori descritti precedentemente. Modulo e direzione del gradiente si ottengono, per ogni pixel dell'immagine, combinando G_x e G_y rispettivamente come:

$$G(j, k) = \sqrt{(G_x(j, k))^2 + (G_y(j, k))^2} \quad (2.6)$$

$$\theta(j, k) = \begin{cases} atan\left(\frac{G_y(j, k)}{G_x(j, k)}\right) & \text{per } G_y(j, k) \geq 0 \\ atan\left(\frac{G_y(j, k)}{G_x(j, k)}\right) + \pi & \text{per } G_y(j, k) < 0 \end{cases} \quad (2.7)$$

Per alcune applicazioni, il segno del gradiente, e quindi il valore di θ compreso tra $[0, 2\pi]$ è rilevante per il problema di classificazione. Nella maggior parte delle applicazioni però, il segno del gradiente fornisce informazioni secondarie e irrilevanti; dunque $\theta(j, k)$ può essere calcolato nell'intervallo $[0, \pi]$.

2.4 Costruzione degli istogrammi

Il passo successivo al calcolo dei gradienti è quello di costruire gli istogrammi. A tale fine l'immagine di ingresso viene divisa in "celle" che possono essere di due diverse forme geometriche: rettangolari (R-HOG), di dimensioni $N \times N$ pixel, o circolari (C-HOG), di raggio N pixel.

Per ogni cella viene creato un istogramma accumulando all'interno dei corrispondenti intervalli di quantizzazione i voti dei gradienti di ciascun pixel della cella, pesati in base al modulo del gradiente. Gli *orientation bin* sono spaziati uniformemente nell'intervallo $[0, 2\pi]$ (gradiente con segno) o $[0, \pi]$ (gradiente senza segno).

Si considerino n_θ *angle bin* in $[0, \pi]$ (o, come discusso in precedenza, potenzialmente, in $[0, 2\pi]$). I descrittori HOG racchiudono le statistiche locali dei gradienti (modulo e direzione) in quanto ogni pixel dà il suo voto ad uno specifico *angle bin* il cui peso è proporzionale all'intensità del gradiente in quel determinato pixel.

Detto n_θ il numero di *angle bin* e posto $\phi_k = \frac{\pi \cdot k}{n_\theta}$, con $k = 0 \dots n_\theta$, e dette n_x e n_y il numero di celle rispettivamente sulla riga e sulla colonna dell'immagine, si costruisce una matrice tridimensionale di dimensione $n_x \times n_y \times n_\theta$ data da:

$$V(i, j, k) = f[G(i, j)] \delta(\phi_{k-1} < \theta(i, j) \leq \phi_k), \quad \text{con } k = 1, \dots, n_\theta \quad (2.8)$$

dove $\delta(x)$ restituisce 1 quando l'argomento è vero, 0 quando è falso e f è una funzione del modulo del gradiente (lineare, radice quadrata, quadrato o una forma saturata tra quelle riportate per rappresentare la presenza o assenza di contorni).

Un istogramma con n_θ canali (*orientations bins*) ad esempio è costruito nel modo seguente.

- i voti di tutti i gradienti della celle che hanno un angolo compreso nell'intervallo $[0, \frac{\pi}{n_\theta})$ sono accumulati nel primo canale;
- i voti di tutti i gradienti della celle che hanno un angolo compreso nell'intervallo $[\frac{\pi}{n_\theta}, \frac{\pi}{n_\theta} \cdot 2)$ sono accumulati nel secondo canale;
- i voti di tutti i gradienti della celle che hanno un angolo compreso nell'intervallo $[\frac{\pi}{n_\theta} \cdot 2, \frac{\pi}{n_\theta} \cdot 3)$ sono accumulati nel terzo canale;
- ⋮
- fino al canale n_θ dove sono accumulati i gradienti delle celle che hanno angolo compreso tra $[\frac{\pi}{n_\theta} \cdot (k - 1), \frac{\pi}{n_\theta} \cdot k)$

Formalmente, data la c -esima cella, l'istogramma è costruito nel seguente modo ($k = 1, \dots, n_\theta$):

$$H(c, k) = \sum_{(i,j) \in c} V(i, j, k) = \sum_{(i,j) \in c} f[G(i, j)] \delta(\phi_{k-1} < \theta(i, j) \leq \phi_k) \quad (2.9)$$

Il voto dato da ciascun pixel è proporzionale all'intensità del gradiente di quel punto, poiché risulta importante associare ad ogni orientazione del gradiente in un dato intervallo un voto che tenga conto dell'importanza del gradiente in un determinato pixel. Infatti, il gradiente calcolato attorno a un bordo risulta molto più significativo di quello calcolato in una zona uniforme dell'immagine ed è essenziale

per estrarre informazioni utili ad avere una descrizione dettagliata delle strutture geometriche presenti.

2.5 Normalizzazione dei blocchi

L'intensità del gradiente utilizzata nei descrittori HOG è sensibile ai cambiamenti locali di luminosità. Per questo motivo è essenziale, per il raggiungimento di una buona discriminazione fra classi nell'immagine, eseguire una normalizzazione dell'istogramma, così da introdurre una migliore invarianza a diverse condizioni di luminosità, contrasto, ombre.

In questa fase ogni istogramma è dunque normalizzato separatamente in base ad un fattore di normalizzazione calcolato sulla base di raggruppamenti di celle circostanti, denominati "blocchi". Ognuno di questi blocchi è composto da $M \times M$ celle.

2.5.1 Schemi di normalizzazione

Si possono valutare diversi schemi per calcolare il valore di normalizzazione.

Detto $\mathbf{v} = \sum_{k=1}^{n_\theta} H(c, k) \hat{v}_k$ il vettore descrittore non ancora normalizzato e sia $\|\mathbf{v}\|_k$ la sua norma k -esima, con $k = 1, 2$, si possono utilizzare i seguenti schemi come proposto da Dalal e Triggs [6]:

- L1-Norm : $\mathbf{v} \rightarrow \frac{\mathbf{v}}{\|\mathbf{v}\|_1 + \varepsilon}$
- L2-Norm : $\mathbf{v} \rightarrow \frac{\mathbf{v}}{\|\mathbf{v}\|_2^2 + \varepsilon^2}$

dove ε è una costante introdotta per evitare la divisione per zero e sufficientemente piccola da non alterare significativamente il risultato.

Un altro metodo per effettuare la normalizzazione è quello proposto da Torrione e Morton [17].

Detta N l’insieme di celle comprese nel blocco di interesse, il valore di normalizzazione può essere calcolato come segue:

$$H(c, k) = \frac{H_1(c, k)}{\left(\sum_{c_i \in N} \sqrt{\|H_1(c_i)\|_2^2 + \varepsilon^2} \right)} \quad (2.10)$$

dove $H_1(c, k) = \sum_{(i,j) \in c} V(i, j, k)$ e $H_1(c)$ rappresenta il vettore colonna $[H_1(c, 1), \dots, H_1(c, n_\theta)]^T$.

2.6 Costruzione del vettore delle feature

A questo punto si procede alla costruzione del descrittore vettoriale (*feature vector*), che verrà classificato mediante un classificatore. Il descrittore avrà le stesse dimensioni dell’immagine originale con un numero di bande pari a $C + C \times B$, dove C è il numero di canali dell’immagine da classificare e B il numero di *bin* scelto nella fase di costruzione degli istogrammi.

Tenendo ben presente che per immagini multispettrali l’HOG viene calcolato separatamente per ogni canale spettrale, la costruzione dei vettori delle *feauture* avviene nel modo seguente: per ogni pixel vengono concatenati l’immagine originale a C canali con i C istogrammi, costituiti da B canali ciascuno, relativi a ciascuna cella e alla banda corrispondente.

Matematicamente, detto \mathbf{X}_i il vettore delle *feature* corrispondente all’ i -esimo pixel dell’immagine:

$$\mathbf{X}_i = [\mathbf{X}_{RGB_i}, \mathbf{f}(\mathbf{X}_{RGB_i})] \quad (2.11)$$

con

$$\mathbf{f}(\cdot) = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{n_\theta}) \quad (2.12)$$

dove \mathbf{f}_j con $j = 1, 2, \dots, n_\theta$ è la componente j -esima dell'istogramma e \mathbf{X}_{RGB} è l'immagine originale.

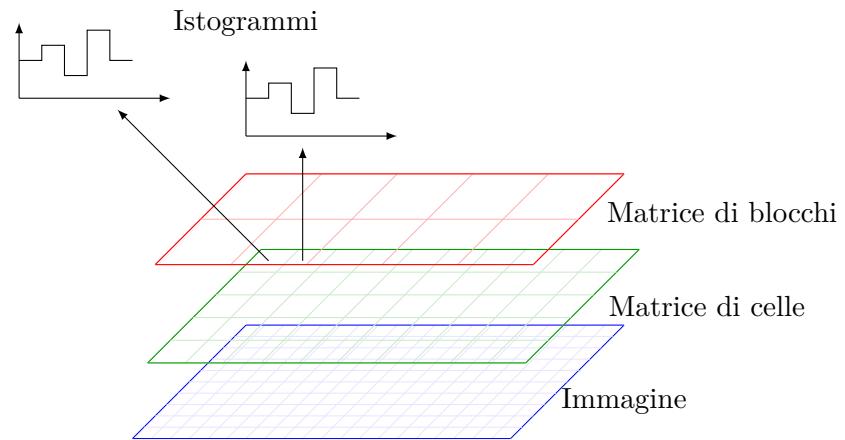


Figura 2.6. Schema esemplificativo della distribuzione spaziale di istogrammi, celle e blocchi

Capitolo 3

SVM - *Support Vector Machine*

Una SVM (*Support Vector Machine*) è un modello di apprendimento supervisionato associato ad algoritmi largamente utilizzati per l'analisi statistica di dati e il *pattern recognition* (tra cui anche il problema della classificazione). L'SVM è un modello abbastanza recente; anche se la sua formulazione risale agli anni '60, solo negli ultimi 15/20 anni si è registrato un incremento significativo nell'uso di algoritmi SVM per classificazione [18]. La sua fortuna risiede nel fatto che possa essere facilmente applicato a molti campi ed alle sue proprietà analitiche in termini di capacità di generalizzazione e robustezza all'aumento del numero di *feature*.

Questo capitolo servirà per introdurre la teoria matematica su cui si basa un classificatore SVM e sarà impostato nel seguente modo: si inizierà descrivendo dettagliatamente il modello più semplice di SVM (SVM lineare per classificazione binaria), poi si proseguirà con la descrizione di come è possibile estendere questo modello per una classificazione non-lineare e, infine, verranno introdotte le possibili modifiche che permettono la classificazione multiclass.

3.1 SVM lineare per classificazione binaria

Dato un *training set* linearmente separabile (\mathbf{x}_i, y_i) , $i = 1, \dots, N$, con $\mathbf{x}_i \in \mathbb{R}^d$ e $y_i \in \{-1, 1\}$, l'obiettivo è addestrare un classificatore affinché

$$f(\mathbf{x}_i) : \begin{cases} > 0 & \text{per } y_i = +1 \\ < 0 & \text{per } y_i = -1 \end{cases} \quad (3.1)$$

ovvero $y_i f(\mathbf{x}_i) > 0$ per una corretta classificazione. La funzione f è detta *regola di decisione* ed è costruita in modo tale che, preso un qualsiasi *campione incognito* \mathbf{u} da classificare, il valore di f valutato in \mathbf{u} restituisca una stima dell'etichetta y_u da associare. In equazioni:

$$f \text{ tale che } \begin{cases} \text{Se } f(\mathbf{u}) > 0 \text{ allora } y_u = +1 \\ \text{Se } f(\mathbf{u}) < 0 \text{ allora } y_u = -1 \end{cases} \quad (3.2)$$

3.1.1 Hard Margin SVM

Un insieme di elementi in \mathbb{R}^d è linearmente separabile se esiste almeno un iperpiano (che in generale avrà dimensione $d - 1$) in grado di separare, nello spazio vettoriale dei campioni in ingresso, quelli che richiedono un'etichetta positiva da quelli che richiedono un'etichetta negativa.

Si prenda, per esempio, la situazione proposta in Figura 3.1: qui si può facilmente evincere che esiste almeno un iperpiano (in questo caso una retta) che divida lo spazio in due semispazi, ciascuno dei quali contiene campioni di una sola classe.

Il problema nasce dal fatto che possono esistere infiniti iperpiani e la scelta di quale usare per l'addestramento possa avere ripercussioni notevoli sulla fase di classificazione. L'SVM risolve questo problema cercando l'iperpiano che massimizza il margine tra i due insiemi di elementi.

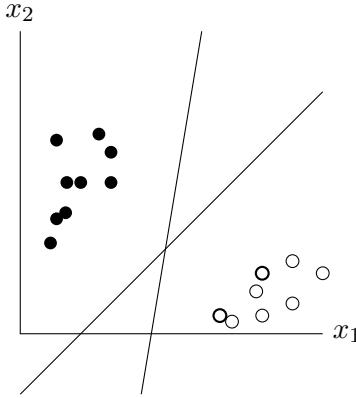


Figura 3.1. *Training set* binario linearmente separabile

In termini matematici, un iperpiano in \mathbb{R}^d ha forma:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (3.3)$$

dove $\mathbf{w} \in \mathbb{R}^d$ è la normale all'iperpiano e $b/\|\mathbf{w}\|$ è la distanza dell'origine dall'iperpiano.

Alla luce di ciò, si può riscrivere la *regola di decisione* presentata nell'equazione (3.2) nel seguente modo:

$$\begin{cases} \text{Se } \mathbf{w} \cdot \mathbf{u} + b > 0 \text{ allora } y_u = +1 \\ \text{Se } \mathbf{w} \cdot \mathbf{u} + b < 0 \text{ allora } y_u = -1 \end{cases} \quad (3.4)$$

Dato che $\mathbf{w} \cdot \mathbf{x} + b = 0$ e $c(\mathbf{w} \cdot \mathbf{x} + b) = 0$ definiscono la stessa regola di decisione per $c > 0$, si ha libertà di scegliere la normalizzazione di \mathbf{w} . In questo caso, si può scegliere il fattore di normalizzazione in modo tale che $\mathbf{w} \cdot \mathbf{x}_i + b \geq 1$ e $\mathbf{w} \cdot \mathbf{x}_i + b \leq -1$, per gli elementi di *training* della prima classe e della seconda rispettivamente. Per convenienza matematica, dato che $y_i \in \{-1, 1\}$, le due identità precedenti possono essere riscritte nel seguente modo:

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, \dots, N \quad (3.5)$$

o, analogamente,

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0, \quad \forall i = 1, \dots, N \quad (3.6)$$

con l'uguaglianza valida per gli elementi sul bordo (i punti rossi in Figura 3.2).

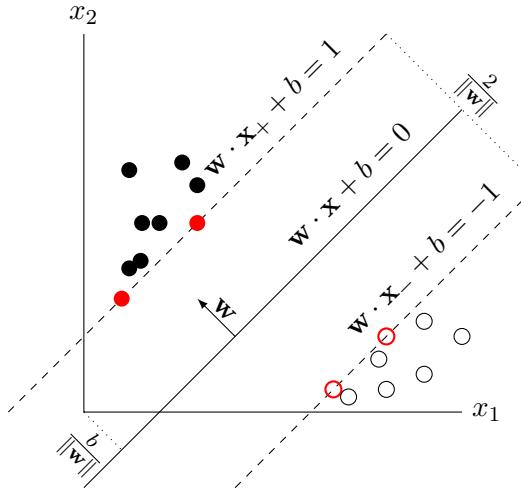


Figura 3.2. Margine tra i campioni di *training* di due classi linearmente separabili

Geometricamente, sotto queste ipotesi, si dimostra che il margine risulta essere:

$$\frac{\mathbf{w}}{\|\mathbf{w}\|} \cdot (\mathbf{x}_+ - \mathbf{x}_-) = \frac{2}{\|\mathbf{w}\|} \quad (3.7)$$

dove \mathbf{x}_+ e \mathbf{x}_- sono i campioni delle due classi (rispettivamente) più vicini all'iper piano separatore (vedi Figura 3.2 ed equazione (3.6)).

L'obiettivo ora è quindi massimizzare questo valore, che equivale a minimizzare $\|\mathbf{w}\|$, il quale, a sua volta, per convenienza matematica, può essere espresso nel seguente termine:

$$\min \frac{1}{2} \|\mathbf{w}\|^2$$

vincolato a $y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad \forall i = 1, \dots, N \quad (3.8)$

Essendo questo un problema di calcolo di estremi vincolati si possono usare i *moltiplicatori di Lagrange*, che permettono di lavorare su un problema duale, ovvero l’ottimizzazione (minimizzazione rispetto a \mathbf{w} e a b) della seguente funzione:

$$\min_{\mathbf{w}, b} L = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_i \alpha_i [y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1] \quad (3.9)$$

dove si sottintende, da questo punto in poi, che la sommatoria sia per ogni $i = 1, \dots, N$.

Procedendo con il calcolo del gradiente di L si ottengono i seguenti risultati:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0 \Rightarrow \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \quad (3.10)$$

$$\frac{\partial L}{\partial b} = - \sum_i \alpha_i y_i = 0 \Rightarrow \sum_i \alpha_i y_i = 0 \quad (3.11)$$

L’equazione (3.10) suggerisce che il vettore \mathbf{w} non sia altro che una combinazione lineare di alcuni vettori di *training* (per alcuni di loro α_i sarà pari a 0), mentre l’equazione (3.11) sarà utile più avanti.

A questo punto si introduce il cosiddetto *problema lagrangiano duale*: invece di *minimizzare* rispetto a \mathbf{w} e b , si può *massimizzare* rispetto ad α con vincoli le relazioni ottenute precedentemente per \mathbf{w} e b (equazioni (3.10) e (3.11))¹.

Dato che l’equazione (3.10) restituisce una espressione per \mathbf{w} , possiamo sostituire questo risultato nella lagrangiana L (eq.ne (3.9)), ottenendo:

$$\begin{aligned} L(\mathbf{w}, b) &= \frac{1}{2} \left(\sum_i \alpha_i y_i \mathbf{x}_i \right) \cdot \left(\sum_j \alpha_j y_j \mathbf{x}_j \right) + \\ &- \sum_i \left[\alpha_i y_i \mathbf{x}_i \cdot \left(\sum_j \alpha_j y_j \mathbf{x}_j \right) \right] - \sum_i \alpha_i y_i b + \sum_i \alpha_i \end{aligned} \quad (3.12)$$

¹Il teorema di Kunh-Tucker assicura che la soluzione di questo problema è la stessa del problema originario; per maggiori informazioni si consulti [2] e [5]

A questo punto, cambiando l'ordine delle sommatorie nel secondo membro e notando che il penultimo addendo è sempre nullo (vedi equazione (3.11)), si ottiene la versione finale della lagrangiana duale:

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (3.13)$$

L'equazione appena riportata è un problema di programmazione quadratica (*quadratic programming*, QP) che assicura l'esistenza e l'unicità di una e una sola soluzione.

Per completare il discorso sulla SVM lineare, si consideri nuovamente la *regola di decisione* introdotta nell'equazione (3.4). Avendo ora una definizione formale per \mathbf{w} , questa può essere riscritta nel seguente modo:

$$\begin{cases} \text{Se } \sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{u} + b > 0 \text{ allora } y_u = +1 \\ \text{Se } \sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{u} + b < 0 \text{ allora } y_u = -1 \end{cases} \quad (3.14)$$

Il motivo dell'uso di questo espediente matematico risiede nel fatto che adesso sia l'ottimizzazione della lagrangiana (3.13) sia la *regola di decisione* (3.14) dipendono esclusivamente da un prodotto scalare tra due vettori, $\mathbf{x}_i \cdot \mathbf{x}_j$ e $\mathbf{x}_i \cdot \mathbf{u}$ rispettivamente, e questo semplificherà notevolmente la trattazione della SVM non lineare.

3.1.2 Soft Margin SVM

Prima di introdurre l'estensione della SVM che permetta la classificazione non-lineare, è interessante discutere di come sia possibile usare una SVM lineare anche per situazioni in cui il *training set* non sia linearmente separabile.

Si prenda, come esempio, la situazione proposta di seguito (Figura 3.3).

Si nota chiaramente come, in questo caso, non esista alcun iperpiano separatore.

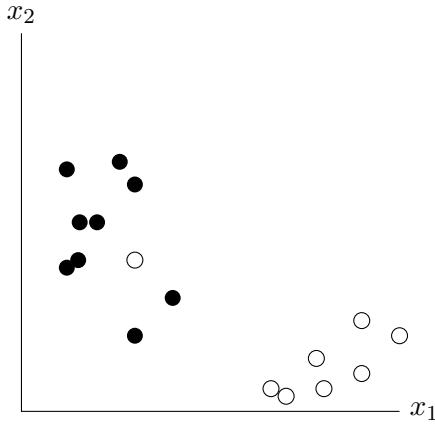


Figura 3.3. *Training set binario non linearmente separabile*

Nonostante ciò, modificando leggermente il modello di SVM introdotto fino a questo punto, si può dimostrare che è ancora possibile utilizzare una SVM lineare.

La chiave di questo nuovo modello sta nell'introduzione di una *variabile di slack*, in modo tale da rilassare quei vincoli rigidi che non permetterebbero l'uso di una SVM lineare.

Precedentemente, i margini erano definiti dai vincoli:

$$y_i (\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 \quad y_i \in \{-1, 1\} \quad (3.15)$$

Nella soluzione proposta, invece, sono definiti da:

$$y_i (\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \xi_i \quad \xi_i \geq 0, y_i \in \{-1, 1\} \quad (3.16)$$

La situazione che si è andata a definire, graficamente, appare in Figura 3.4.

Il nuovo vincolo permette al margine funzionale di essere minore di 1. L'errore che si commette è, però, $C\xi_i$, sia per i punti che ricadrebbero nella parte corretta dell'iperpiano separatore ($0 < \xi_i \leq 1$), sia per quelli che sarebbero nel lato sbagliato ($\xi_i > 1$). Si sono, quindi, "rilassati" i vincoli in modo tale da classificare

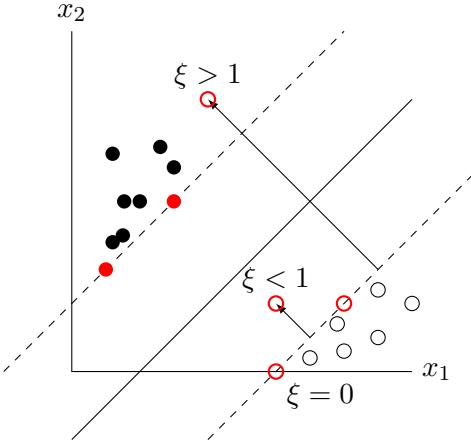


Figura 3.4. Esempio di soft margin nel caso di un *training set* binario non linearmente separabile

dati non-separabili, con una penalità linearmente proporzionale all’entità dell’errore commesso sul *training set*.

Il nuovo problema di ottimizzazione è il seguente:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \quad (3.17)$$

$$\text{vincolato a} \quad y_i (\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \xi_i \quad (3.18)$$

$$\xi_i \geq 0, \quad \forall i = 1, \dots, N \quad (3.19)$$

La costante C (con $C \geq 0$) gestisce il compromesso fra la minimizzazione dei due contributi alla funzione obiettivo: se $C = 0$ gli errori sul *training set* non sono penalizzati; se C è molto grande, il termine legato al margine è minoritario e gli errori sul *training set* sono molto penalizzati.

Allo stesso modo del caso lineare, possono essere usati i moltiplicatori di Lagrange; la funzione da ottimizzare quindi è la seguente:

$$L = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i + \sum_i \alpha_i [1 - \xi_i - y_i (\mathbf{x}_i \cdot \mathbf{w} + b)] - \sum_i \beta_i \xi_i \quad (3.20)$$

dove β_i sono i moltiplicatori di Lagrange necessari per vincolare la positività delle *variabili di slack*.

È interessante notare come l'applicazione di una soft-margin SVM possa avere prestazioni migliori della hard-margin SVM anche per dataset linearmente separabili. Per chiarificare il motivo, un esempio è riportato nella figura successiva: il *training set* permetterebbe l'uso di una hard-margin SVM, ma il risultato avrebbe un margine piuttosto limitato, confrontato con quello che si otterrebbe in caso di soft-margin SVM.

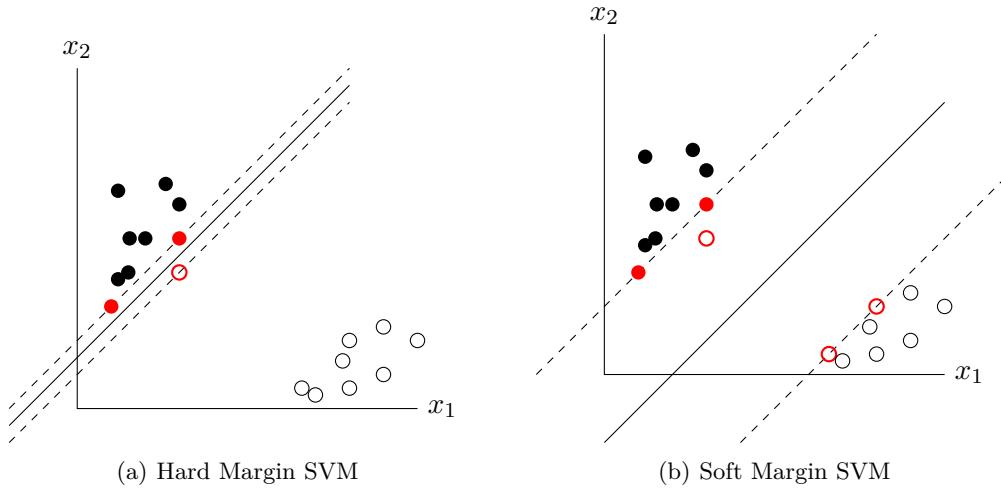


Figura 3.5. Confronto tra Hard Margin SVM e Soft Margin SVM nel caso di un training set binario linearmente separabile

3.2 SVM non lineare per classificazione binaria

Nonostante l'introduzione della variante Soft-Margin della SVM possa rimuovere l'ipotesi restrittiva di separabilità lineare del *training set*, può capitare che anche

questa strategia sia difficilmente applicabile o produca risultati non soddisfacenti (si rimanda al Capitolo 4 dove si discuterà come avere una stima delle prestazioni di un classificatore).

L'intento della classificazione SVM non-lineare è quello di trasformare lo spazio dei vettori di *training* non linearmente separabili \mathbb{R}^d in uno spazio \mathcal{H} (la cui dimensionalità sarà maggiore di d o anche, potenzialmente, infinita) in cui i campioni siano disposti in modo tale da permettere l'utilizzo di una SVM lineare. Questo passo si giustifica tramite il *Teorema di Cover sulla separabilità*, il quale afferma che un problema di classificazione complesso, formulato attraverso una trasformazione non-lineare dei dati in uno spazio ad alta dimensionalità, ha maggiore probabilità di essere linearmente separabile che in uno spazio a bassa dimensionalità.

Si ricerca quindi una funzione di trasformazione:

$$\Phi : \mathbb{R}^d \rightarrow \mathcal{H} \quad (3.21)$$

Graficamente, in Figura 3.6 è riportato un esempio della situazione in esame.

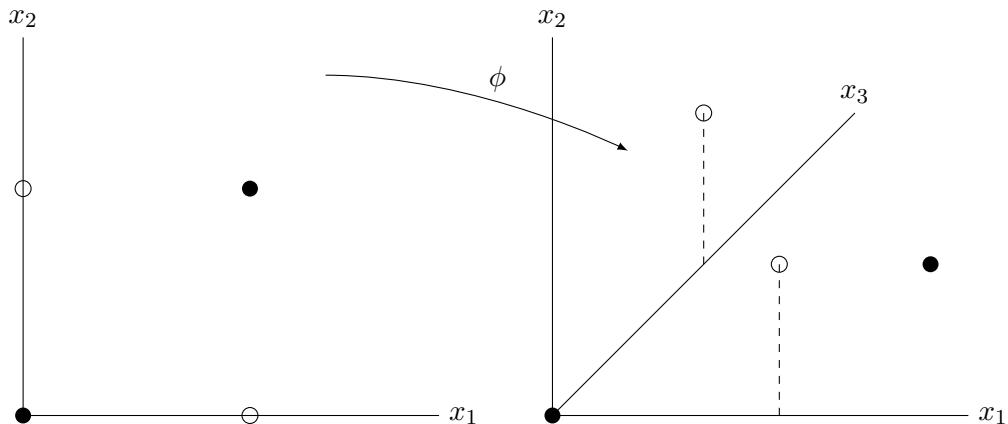


Figura 3.6. Trasformazione dello spazio delle *feature* in uno spazio munito di prodotto scalare ed avente dimensione maggiore. Cerchi pieni e vuoti indicano campioni di training di due classi

Alla luce di ciò, la lagrangiana (3.13), per la fase di addestramento, e la *regola di decisione* (3.14), per la fase di classificazione, possono essere riscritte come segue:

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \quad (3.22)$$

$$f(\mathbf{u}) = \sum_i \alpha_i y_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{u}) + b \quad (3.23)$$

Come già accennato in precedenza, operativamente, entrambe le due fasi sono caratterizzate solo dal prodotto scalare dei vettori trasformati; questo suggerisce che non sia necessaria la conoscenza di Φ , in quanto è sufficiente definire un *kernel*:

$$K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \quad \text{tale che} \quad K(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}) \quad (3.24)$$

Esistono condizioni necessarie e sufficienti affinché una data funzione K sia un *kernel*, ovvero affinché esistano uno spazio \mathcal{H} e una funzione di trasformazione Φ , tali che valga l'equazione (3.24) per $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$.

Le condizioni di Mercer sono condizioni necessarie e sufficienti per definire per quali famiglie di *kernel* K esiste la coppia $\{\mathcal{H}, \Phi\}$ con le proprietà presentate precedentemente.

Teorema 1 *Sia $K(\mathbf{x}, \mathbf{y})$ una funzione continua che può essere espansa nella serie*

$$K(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{\infty} \Phi(\mathbf{x})_i \cdot \Phi(\mathbf{y})_i \quad (3.25)$$

Affinché tale espansione sia valida e per la sua convergenza assoluta, è necessario e sufficiente che la condizione

$$\int \int K(\mathbf{x}, \mathbf{y}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0 \quad (3.26)$$

sia vera per ogni $g(\cdot)$ che soddisfa

$$\int g^2(\mathbf{x}) d\mathbf{x} < \infty \quad (3.27)$$

Questo espediente matematico è detto ***kernel trick*** in quanto, dato un *kernel* che soddisfi tali condizioni, un classificatore SVM risulta identificato dal *kernel* stesso, senza alcuna necessità di definire esplicitamente né Φ né \mathcal{H} .

Per completezza, vengono riportati ora i problemi di ottimizzazione della lagrangiana e la *regola di decisione* usando il *kernel trick*:

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (3.28)$$

$$f(\mathbf{u}) = \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{u}) + b \quad (3.29)$$

Alcuni dei *kernel* più famosi e usati sono i seguenti:

- **Polinomiale:** $K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x} \cdot \mathbf{y})^d$ con $d > 0$
- **Radiale Gaussiano (RBF):** $K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right)$ il quale introduce una dimensione dello spazio delle *feature* trasformate infinita.

3.3 Estensioni multiclassse

Un problema con più classi si affronta mediante SVM esprimendolo come combinazione di problemi binari. Assumendo di operare con C classi, tale operazione si effettua tipicamente in due modi possibili:

- l'approccio *one-against-one* (OAO) calcola mediante SVM binaria una *regola di decisione* f_{ij} per ogni coppia di classi, per un totale di $C(C - 1)/2$ funzioni discriminanti, e classifica un campione incognito $\mathbf{u} \in \mathbb{R}^d$ mediante "votazione"

(se $f_{ij}(\mathbf{u}) > 0$ si dà un voto alla classe i -esima, altrimenti alla j -esima; si assegna infine \mathbf{u} alla classe che ha ricevuto più voti);

- l'approccio *one-against-all* (OAA) calcola mediante SVM binaria una *regola di decisione* f_i per ogni decisione del tipo "classe i contro classe non- i ", per un totale di C funzioni discriminanti e assegna $\mathbf{u} \in \mathbb{R}^d$ alla classe i -esima se $f_i(\mathbf{u}) \geq f_j(\mathbf{u})$ per ogni $j = 1, \dots, C$ con $i \neq j$.

Capitolo 4

Valutazione delle prestazioni di un classificatore

Dato un classificatore supervisionato, addestrato su un *training set*, è fondamentale saper valutare l'accuratezza che può essere ottenuta quando tale classificatore è applicato a campioni incogniti.

A tale scopo, è essenziale valutare la probabilità di errore P_e del classificatore per decidere, ad esempio, se le *feature* utilizzate siano sufficienti a discriminare bene le classi o se sia necessario estrarne altre (come parametri di tessitura nel caso in cui i canali spettrali non siano abbastanza discriminanti).

4.1 Stima della probabilità di errore

In presenza di C classi $\omega_1, \omega_2, \dots, \omega_C$, detta P_i la probabilità *a priori*, la probabilità di errore P_e si può esprimere nel modo seguente:

$$P_e = P\{\hat{Y} \neq Y\} = \sum_{i=1}^C P\{\hat{Y} \neq \omega_i | Y = \omega_i\} P\{Y = \omega_i\} \quad (4.1)$$

dove Y è l'etichetta di classe corrispondente alla realtà a terra e \hat{Y} è l'etichetta stimata dal classificatore. Dal momento che tale espressione è calcolabile solo in pochissimi casi semplici, per valutarla si adotta generalmente un approccio empirico, che stima la P_e come la percentuale dei pixel di test classificati erroneamente.

Solitamente la P_e viene valutata su un insieme di campioni pre-etichettati (*test set*) diverso rispetto a quello utilizzato per addestrare il classificatore (*training set*). Questa tecnica, detta *hold-out*, permette una misura delle prestazioni priva di *bias* in quanto eseguita su istanze non utilizzate in fase di apprendimento.

Per poter individuare e possibilmente evitare il fenomeno dell'*overfitting*,¹ una condizione fondamentale per stimare P_e come frequenza relativa degli errori sul *test set* è che i campioni pre-etichettati siano indipendenti e identicamente distribuiti (i.i.d.). Per questa ragione, è buona norma prelevare i campioni di *training* e di *test* in regioni dell'immagine spazialmente disgiunte tra loro.

E' altrettanto importante effettuare un'analisi qualitativa dell'intera mappa, mediante foto-interpretazione, come complemento alla valutazione quantitativa delle prestazioni di classificazione sui campioni di test.

¹Si parla di *overfitting* quando la funzione discriminante è strettamente dipendente dai campioni di *training* specifici utilizzati per calcolarla ed è quindi particolarmente inefficace quando applicata a campioni incogniti

4.2 Matrice di confusione e parametri di accuratezza

La P_e fornisce una valutazione complessiva delle prestazioni del classificatore, senza però differenziare gli errori commessi in corrispondenza di classi diverse. Per una valutazione più dettagliata la matrice di confusione (*confusion matrix*) è la tipologia di osservazione statistica maggiormente utilizzata: il risultato della classificazione sulle aree campione viene confrontato con la verità al suolo. Questa è una matrice $C \times C$, il cui elemento e_{ij} è il numero di pixel di test della classe ω_i che il classificatore ha assegnato alla classe ω_j . Sulla diagonale $i = j$ della matrice di confusione si leggono dunque i numeri di pixel di test classificati in modo corretto. Questo tipo di analisi statistica consente non solo di quantificare il successo ottenuto dalla procedura, ma anche di focalizzare i punti critici del processo di classificazione, ovvero le classi meno distinguibili tra loro.

L'accuratezza della classificazione può essere valutata con diversi parametri numerici, tra cui i più utilizzati sono i seguenti:

- La *Producer Accuracy* (PA) di una classe ω_i è la parte di pixel ben classificati rispetto al numero totale di pixel di test di ω_i , in particolare essa si può esprimere come:

$$PA_i = \frac{e_{ii}}{\sum_{j=1}^C e_{ij}} \quad (4.2)$$

Analogamente, si può definire *Omission Error* (OE) di ω_i la frazione complementare di pixel classificati erroneamente.

- L' *average accuracy* (AA) delle C classi è la media delle C *producer accuracy* ed è data da:

$$AA = \frac{1}{C} \sum_{i=1}^C PA_i \quad (4.3)$$

- L'*overall accuracy* (OA) è la percentuale di pixel classificati correttamente sull'intero test set ed è data da:

$$OA = \frac{1}{t} \sum_{i=1}^C e_{ij} \quad (4.4)$$

dove t è il numero di pixel di test.

- Il parametro " κ ", rappresenta una modifica dell'OA finalizzata a tenere conto in modo più completo della distribuzione degli errori tra le diverse classi, ed è dato da:

$$\kappa = \frac{OA - \frac{1}{t^2} \sum_{i=1}^C \sum_{j=1}^C \sum_{k=1}^C e_{ij} e_{ki}}{1 - \frac{1}{t^2} \sum_{i=1}^C \sum_{j=1}^C \sum_{k=1}^C e_{ij} e_{ki}} \quad (4.5)$$

dove t rappresenta ancora il numero di pixel di test utilizzati.

Capitolo 5

Risultati sperimentali

In questo capitolo verranno presentati tre casi di studio e verrà fatta un'analisi delle prestazioni al variare dei parametri dell'algoritmo HOG, evidenziando i valori ottimali individuati. Successivamente verranno presentati i risultati ottenuti valutando anche gli aspetti per cui questo algoritmo ha dato risultati parzialmente soddisfacenti.

5.1 Descrizione dei *dataset*

In questa tesi sono stati analizzati tre casi rilevanti, tutti relativi alla classificazione dell'area urbana di Amiens (Francia). Si tratta di un problema di classificazione interessante e particolarmente complesso, in quanto le regioni coinvolte sono caratterizzate sia da aree omogenee (come corsi d'acqua) sia da strutture geometriche ben definite (come edifici e strade) e zone di suolo con *pattern* regolari (come campi coltivabili e non).

L'analisi condotta ha coinvolto l'uso di immagini ad elevata risoluzione spaziale, acquisite nell'ambito di un progetto europeo dedicato allo sviluppo di tecnologie ICT innovative per l'identificazione di aree urbane attualmente non usate e potenzialmente riqualificabili.

Il *dataset* per ogni esperimento è costituito da un'immagine telerilevata, dalla mappa di *training* e dalla mappa di *testing* corrispondenti. Le immagini telerilevate sono state acquisite dal sensore passivo SPOT5 HRG a tre canali, corrispondenti alle lunghezze d'onda del verde (G, 495 – 570 nm), del rosso (R, 620 – 750 nm) e del vicino infrarosso (NIR, *Near InfraRed* 0.75 – 1.4 μm). Sebbene i canali spettrali di SPOT5 HRG avrebbero risoluzione spaziale nativa di 10 m, il sensore stesso acquisisce anche un canale pancromatico, associato all'intero intervallo di lunghezza d'onda (dal visibile al vicino infrarosso) e caratterizzato da risoluzione spaziale di 5 m. In fase di pre-elaborazione, tecniche di *pansharpening*¹ sono state applicate al fine di fondere le informazioni fornite dai dati multispettrali e pancromatici e generare un'immagine G-R-NIR a risoluzione spaziale di 5 m. Inoltre, SPOT5 HRG

¹Pansharpening è un processo di fusione di immagini pancromatiche ad alta risoluzione con immagini multispettrali a risoluzione inferiore per creare una singola immagine multispettrale ad alta risoluzione.

permette anche acquisizioni di coppie stereo di immagini. L'applicazione di tecniche di super-risoluzione ad una di tali coppie permette di generare un'ulteriore immagine volta a stimare la distribuzione spaziale della radianza ricevuta ad una risoluzione di 2.5 m. Sono stati usati a fini di sperimentazione dati ottenuti da entrambe le tipologie di pre-elaborazione e quindi caratterizzati da pixel di dimensione pari a 2.5 e 5 m.

5.1.1 Amiens 2006 - 5m - 10 classi

L'immagine *Amiens6–5m* è stata acquisita nel 2006 e ha pixel di dimensione spaziale pari a 5m, coprendo approssimativamente un'area di $10\ km \times 11\ km$ (2000×2200 pixel).

L'insieme delle classi $\Omega = \{\omega_1, \omega_2, \dots, \omega_{10}\}$ che costituisce questo primo esperimento è il seguente:

1. Area urbana ad alta densità
2. Area urbana a bassa densità
3. Strade
4. Area verde urbana
5. Suolo nudo
6. Terreno coltivabile
7. Aree vegetate
8. Alberi
9. Corsi d'acqua

10. Specchi d'acqua

Nella Figura 5.1 vengono presentate le immagini caratterizzanti il primo *dataset*.

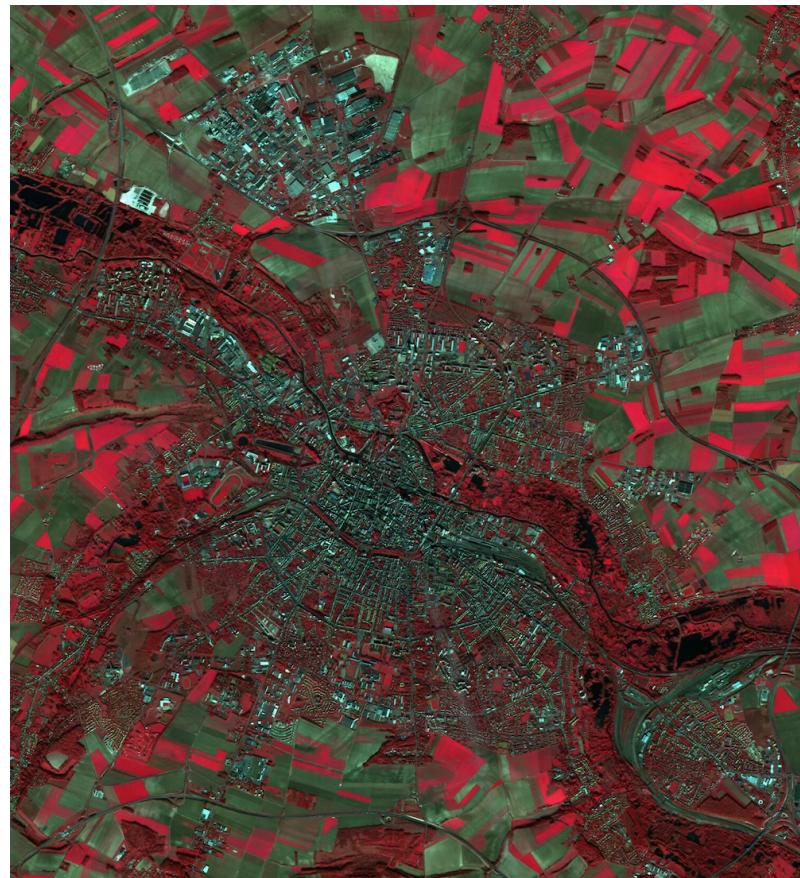
Distribuzione dei pixel di training (TR) e test (TE) classe per classe.

TR:

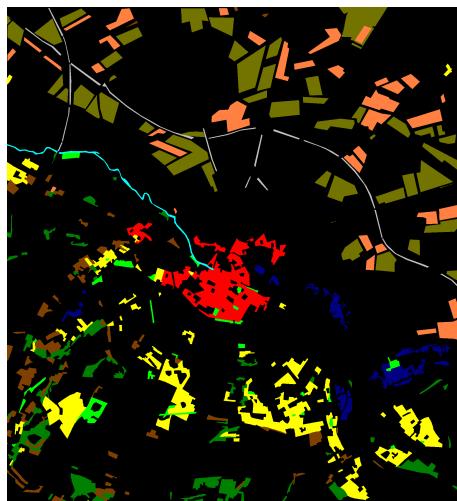
Class 1 : 72483
Class 2 : 149917
Class 3 : 17945
Class 4 : 25121
Class 5 : 233481
Class 6 : 113030
Class 7 : 62854
Class 8 : 90718
Class 9 : 7903
Class 10: 34833
Total : 808285

TE:

Class 1 : 70299
Class 2 : 68435
Class 3 : 10019
Class 4 : 16012
Class 5 : 115570
Class 6 : 48615
Class 7 : 54582
Class 8 : 50701
Class 9 : 3012
Class 10: 27924
Total : 465169



(a) Immagine telerilevata



(b) Mappa di *training*

Urbano ad alta densità	Terreno coltivabile
Urbano ad bassa densità	Aree vegetate
Strade	Alberi
Area verde urbana	Corsi d'acqua
Suolo nudo	Specchi d'acqua

(c) Legenda classi della mappa di *training*

Figura 5.1. *Dataset* con composizione RGB in falso colore (2000×2200 pixel) acquisito su Amiens (Francia) dal sensore SPOT5 HRG

5.1.2 Amiens 2006 - 2.5m - 7 classi

L’immagine *Amiens6–2.5m* è stata acquisita sempre nel 2006, ma ha pixel di dimensione spaziale pari a 2.5 m e copre sempre approssimativamente un’area di $10\text{ km} \times 11\text{ km}$ (4001×4400 pixel).

L’insieme delle classi $\Omega = \{\omega_1, \omega_2, \dots, \omega_7\}$ che costituisce il secondo *dataset* è il seguente:

1. Edifici
2. Strade e marciapiedi
3. Aree vegetate
4. Suolo nudo
5. Terreno coltivabile
6. Alberi
7. Acqua

Nella Figura 5.3 vengono presentate le immagini caratterizzanti il secondo *dataset*. Nella Figura 5.2 sono riportate, per una migliore comprensione, la distribuzione dei pixel di *training* di tre classi diverse (nero = strade, blu = acqua, verde = aree vegetate), rispetto a due distinte coppie di *feature* spettrali. Tali grafici evidenziano come alcune classi siano spettralmente molto sovrapposte e confermano l’opportunità dell’estrazione di *feature* aggiuntive associate alla distribuzione spaziale delle intensità dei pixel invece che all’informazione spettrale da essi apportata. Inoltre è importante notare come l’immagine a 2.5 m sia visibilmente più sfocata di quella a 5 m . Ciò è legato all’efficacia solo parziale del metodo di super-risoluzione che, in fase

di pre-elaborazione, fu applicato. Pertanto, la risoluzione spaziale di tale immagine, ossia la dimensione del più piccolo dettaglio distinguibile, si ritiene peggiore di 2.5 m. Ciò rende il processo di classificazione ulteriormente complesso.

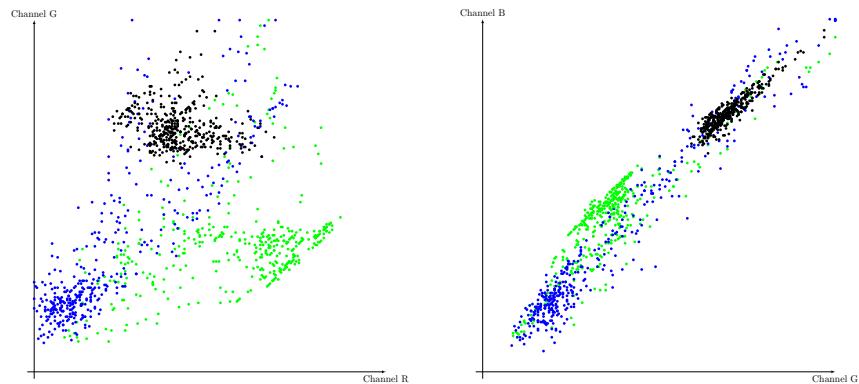
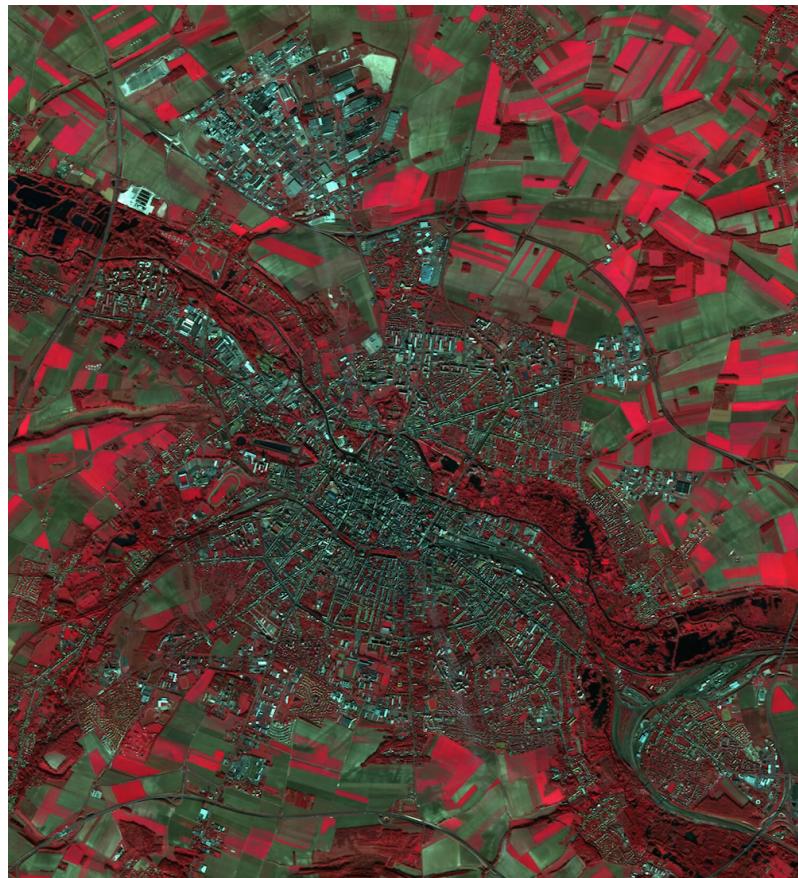
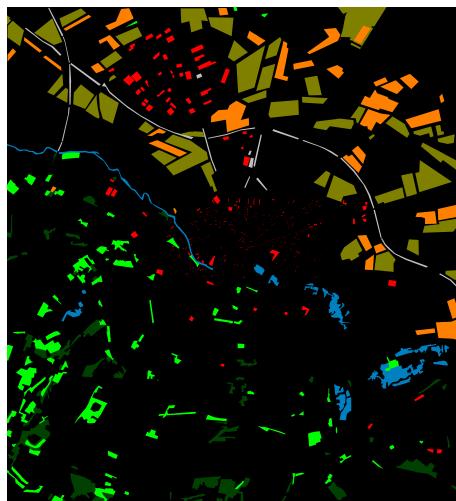


Figura 5.2. Analisi della distribuzione dei campioni di training di tre classi nelle proiezioni su due diversi sottospazi delle feature



(a) Immagine telerilevata



(b) Mappa di *training*

Edifici	Terreno coltivabile
Strade e marciapiedi	Alberi
Aree Vegetate	Acqua
Suolo nudo	

(c) Legenda classi della mappa di *training*

Figura 5.3. *Dataset* con composizione RGB in falso colore (4001×4400 pixel) acquisito su Amiens (Francia) dal sensore SPOT5 HRG nel 2006

5.1.3 Amiens 2012 - 2.5m - 7 classi

L’immagine *Amiens12–2.5m* è stata acquisita nel 2012 e ha pixel di dimensione spaziale pari a 2.5 m , coprendo sempre un’area di circa $10\text{ km} \times 11\text{ km}$ (4001×4400 pixel).

L’insieme delle classi $\Omega = \{\omega_1, \omega_2, \dots, \omega_7\}$, che costituisce il *dataset*, è lo stesso del precedente. Per uniformità vengono ugualmente riportate:

1. Edifici
2. Strade e marciapiedi
3. Aree vegetate
4. Suolo nudo
5. Terreno coltivabile
6. Alberi
7. Acqua

Nella Figura 5.4 vengono presentate le immagini caratterizzanti il primo *dataset*; si osservi con attenzione la composizione dell’immagine di *training*.

Anche in questo caso si nota come l’immagine a 2.5 m sia sfocata, rendendo il processo di classificazione largamente complesso. Infatti, la risoluzione spaziale di tale immagine è da ritenersi peggiore di 2.5 m , sempre a causa della parziale efficacia del metodo di super-risoluzione applicato in fase di pre-elaborazione.

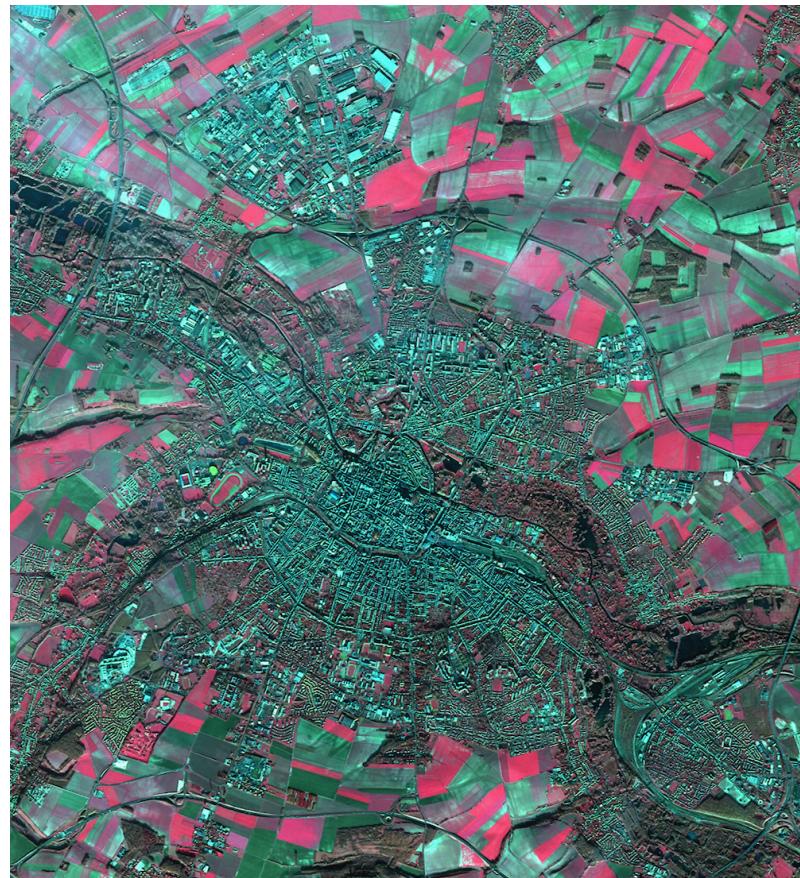
Distribuzione dei pixel di training (TR) e test (TE) classe per classe.

TR:

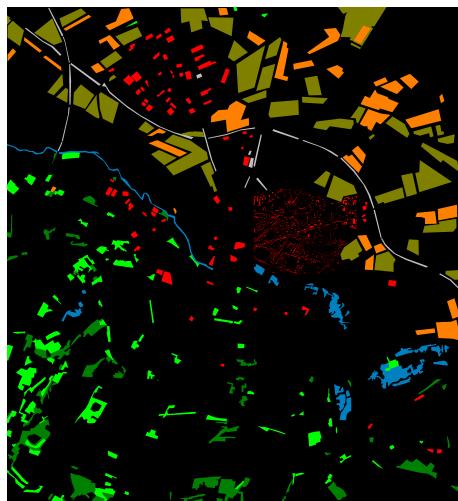
Class 1: 132005
Class 2: 76330
Class 3: 351949
Class 4: 933512
Class 5: 451999
Class 6: 363058
Class 7: 170921
Total : 2479774

TE:

Class 1: 60793
Class 2: 40123
Class 3: 282545
Class 4: 461562
Class 5: 194213
Class 6: 203286
Class 7: 123744
Total : 1366266



(a) Immagine telerilevata



(b) Mappa di *training*

Edifici	Terreno coltivabile
Strade e marciapiedi	Alberi
Aree Vegetate	Acqua
Suolo nudo	

(c) Legenda classi della mappa di *training*

Figura 5.4. *Dataset* con composizione RGB in falso colore (4001×4400 pixel) acquisito su Amiens (Francia) dal sensore SPOT5 HRG nel 2012

5.2 Applicazione del classificatore SVM

Il classificatore usato è un classificatore *soft-margin* SVM con *kernel* gaussiano. L'implementazione adottata è una variante *ad hoc* della LIBSVM [3] scritta in C++.

Il setup utilizzato per la fase di classificazione è composto da un MacBookPro Retina con un processore dual core Intel Core i5 2.8 GHz e 16 GB di memoria e un iMac con processore quad core Intel Core i7 3.4 GHz e 8 GB di memoria.

La fase di *training* della SVM, la quale ha complessità temporale $O(n^\alpha)$ (α tipicamente compreso tra 2 e 3) polinomialmente proporzionale al numero di vettori di *training*, ha impiegato, in media, 30 minuti per completare l'ottimizzazione dei parametri, mentre la fase di etichettatura (con complessità $O(n)$ dove n è il numero di vettori da etichettare) ha impiegato, in media, un'ora per esperimento.

L'ottimizzazione dei parametri C e σ del classificatore è stata effettuata mediante l'applicazione del metodo in [12] che minimizza un maggiorante sull'errore di generalizzazione del classificatore, detto "*span bound*", mediante l'algoritmo numerico di Powell (per maggiori approfondimenti si faccia riferimento anche a [11]).

5.3 Applicazione del metodo HOG

Qui di seguito verranno presentate le variazioni di accuratezza dell'algoritmo sviluppato al variare delle diverse combinazioni dei parametri in gioco, evidenziando le motivazioni alla base delle scelte progettuali effettuate. La calibrazione dei parametri è stata effettuata sul *dataset* di Amiens 2006-5m (Figura 5.1).

5.3.1 Riduzione del rumore

La riduzione del rumore è stata effettuata, come già illustrato in figura 2.3 nel Capitolo 2, tramite un filtraggio passa-basso attraverso un filtro gaussiano bi-dimensionale avente varianza σ pari a 2 pixel. Applicando questa gaussiana sia sull’immagine in ingresso all’algoritmo HOG sia sulle immagini HOG risultanti, si opera al fine di ottenere un notevole incremento nella *average accuracy*. In particolare, l’utilizzo di un filtro gaussiano in ingresso aumenta l’AA del 12%, mentre l’algoritmo di *noise cleaning* applicato prima e dopo l’estrazione delle *feature* ha fatto registrare un’ulteriore incremento di 2 punti percentuali, portando l’AA al 14%.

5.3.2 Calcolo dei gradienti

Per quanto riguarda la scelta della maschera da utilizzare per il calcolo del gradiente, sono state valutate diverse opzioni, tra cui la semplice maschera in $1D$ a differenze separate $[-1, 0, 1]$, la maschera cubica $[1, -8, 0, 8, -1]$ e filtri $2D$ più classici come quelli di Prewitt e Sobel.

I risultati migliori sono stati ottenuti con il *kernel* più semplice a $1D$ 3×1 . Variazioni sulla maschera utilizzata non hanno modificato significativamente i risultati per giustificare un aumento computazionale dovuto all’utilizzo di filtri più complessi. In particolare, con la maschera cubica 5×5 l’incremento della AA è stato di appena +1,5%; negli altri casi valutati il risultato è sempre stato peggiore (-7% con Prewitt e -3% con Sobel). Per questo motivo abbiamo deciso di utilizzare in tutti e tre i casi la soluzione più semplice e ottimale.

La direzione del gradiente è stata considerata tra 0 e π (ignorandone quindi il segno) in quanto le strutture geometriche di cui ci interessa avere informazioni (quali

strade, fiumi, ...) possono essere identificate dalla direzione e non è invece rilevante il verso del gradiente.

5.3.3 Numero di componenti dei vettori delle *feature*

Per quanto riguarda il numero di canali utilizzati per l'istogramma, la scelta che ha fornito prestazioni migliori è risultata essere quella con numero di bande pari a 4. Quantitativamente parlando, i risultati ottenuti con 9 *orientation bins* hanno portato ad un decremento nella AA di -11% . Inoltre, un aumento del numero di *bins* comporta un aumento della risoluzione angolare e, pur apportando complessivamente poca informazione aggiuntiva, aumenta la dimensionalità dello spazio delle *feature* \mathbb{R}^d . Infatti, in primo luogo, all'aumentare di d , cresce la complessità computazionale del classificatore, che si traduce in un allungamento dei tempi di calcolo e in una maggiore occupazione di memoria.

5.3.4 Dimensione delle celle e dei blocchi

La scelta di utilizzare celle di elevate dimensioni introduce un'alta correlazione tra i vettori delle *feature*, mentre un'eccessiva riduzione comporta marcata sensibilità al rumore. Un compromesso è stato trovato empiricamente attraverso diverse sperimentazioni ed è risultato essere diverso (come ci si può aspettare) a seconda della risoluzione spaziale con cui si operava:

- cella da 4×4 pixel, nel caso di pixel di dimensione spaziale pari a 5 m (Figura 5.1)
- cella da 2×2 pixel, nel caso di pixel di dimensione spaziale pari a 2.5 m (Figure 5.3 e 5.4)

Si è deciso di mantenere costante il numero di pixel usati per la normalizzazione dei blocchi ad un valore di 16×16 , dal momento che questo valore non inficia la quantità di informazione a disposizione, ma semplicemente normalizza i valori già calcolati, limitandone l'escursione entro un certo intervallo predefinito.

5.4 Discussione dei risultati sperimentali

5.4.1 Amiens 2006 - 5m - 10 classi

È riportata in Figura 5.5 la mappa di classificazione con *feature* HOG aggiuntive in configurazione di 4 *bin*, celle di dimensione 4×4 pixel e blocchi di normalizzazione con 16×16 pixel, con filtraggio gaussiano sia per l’immagine in ingresso che per le immagini HOG.

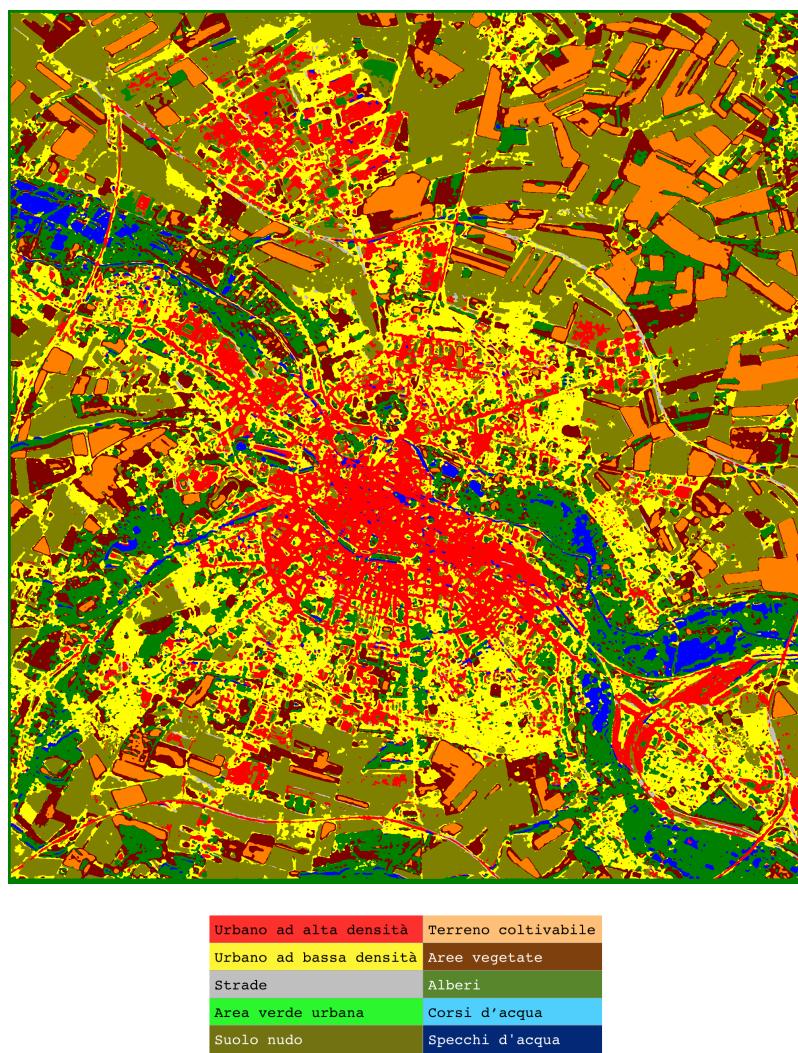


Figura 5.5. Mappa di classificazione ottenuta applicando SVM a *feature* spettrali e HOG per il dataset *Amiens 2006 - 5m - 10 classi*

Da una prima analisi visiva, si può chiaramente constatare come nel complesso la mappa di classificazione sia soddisfacente, sebbene si noti già adesso difficoltà nella classificazione delle strade (classe 3) soprattutto all'interno dell'area urbana. Inoltre, è evidente che la classe relativa ai corsi d'acqua (classe 9) non è presente (o almeno non apprezzabile).

Sono stati evidenziati, principalmente, due problemi:

- Il primo problema è legato alla risoluzione spaziale dell'immagine in ingresso che permette l'identificazione di strade abbastanza larghe, ma rende difficoltosa l'identificazione di strade più strette.
- Il secondo problema si correla col fatto che il *dataset* include due classi associate a corpi idrici: pur rappresentando essi usi del suolo differenti, le loro coperture del suolo sono effettivamente analoghe, il che ne rende difficile la discriminazione mediante dati satellitari, soprattutto caratterizzati da pochi canali spettrali (solo tre nel nostro caso).

Queste prime impressioni vengono confermate anche dall'analisi numerica della matrice di confusione (Figura 5.6).

Analizzando la matrice di confusione saltano subito all'occhio alcuni comportamenti interessanti nella classificazione. La zona urbana (classe 1 e 2) viene discriminata in modo soddisfacente (con accuratezze nell'ordine del 75%); ciò che si perde nell'accuratezza deriva in massima parte dalla difficoltà del classificatore di distinguere le aree urbane di diversa densità.

L'analisi quantitativa della matrice di confusione conferma quanto già ipotizzato

Mappa di test	Mappa di classificazione									
	Urbano ad alta densità	Urbano ad bassa densità	Strade	Area urbana verde	Suolo nudo	Terreno coltivabile	Area vegetata	Alberi	Fiumi	Bacini
Area urbana ad alta densità	56516	9388	578	0	1491	0	127	1273	0	926
Area urbana ad bassa densità	6129	49897	261	0	3819	12	4603	3680	0	34
Strade	4697	2492	1184	0	1367	0	59	220	0	0
Area urbana verde	80	2441	16	0	175	956	5865	6121	0	358
Suolo nudo	2631	8719	780	0	99772	0	2050	1615	0	3
Terreno coltivabile	0	4	0	0	3	44228	4234	146	0	0
Area vegetata	72	6670	145	0	1659	3884	32972	8826	0	354
Alberi	38	2299	31	0	285	968	9554	37326	0	200
Fiumi	384	423	26	0	0	0	100	567	0	1512
Bacini	315	820	0	0	0	7	858	7174	0	18750
	80.39%	72.91%	11.82%	0.00%	86.33%	90.98%	60.41%	73.62%	0.00%	67.15%

Figura 5.6. Matrice di confusione associata alla classificazione del dataset *Amiens 2006 - 5m - 10 classi* mediante l'applicazione di SVM a feature spettrali e HOG

precedentemente, ovvero che le strade sono quasi del tutto perse registrando un'accuratezza di circa 11%, a scapito soprattutto dell'area urbana.

Si è registrata una buona accuratezza nelle classificazioni di aree più uniformi quali suolo nudo (classe 5), terreni coltivabili e non coltivabile (classi 6 e 7): pochissimi sono i pixel etichettati in modo errato per queste regioni, analogamente ai terreni non coltivabili, per i quali tuttavia la confusione con la classe "alberi" causa accuratezza inferiore.

Tuttavia, il classificatore ha avuto serie difficoltà in quelle classi minoritarie per le quali il numero di pixel di training era inferiore. In particolare, vengono completamente perse l'area verde urbana e i corsi d'acqua (0%), le cui etichette sono state

assegnate in maggioranza alla classe alberi e alla classe bacini d’acqua, rispettivamente.

Come osservato prima, si tratta di un errore di classificazione dovuto alla difficoltà di distinguere classi associate a distinti usi del suolo ma sostanzialmente a coperture del suolo molto simili. In quest’ottica, il risultato ottenuto si ritiene già soddisfacente.

La difficoltà di classificazione di questo *dataset* risiede soprattutto nella sovrabbondanza del numero di classi da distinguere, in quanto molti errori fatti nella mappa di classificazione sono proprio causati dalla confusione di coperture di suolo simili tra loro.

In termini generali, l’*Overall Accuracy* (OA) complessivo di questo primo *dataset* è stato di 73.23%, mentre l’AA è risultato essere 54.36%. Il più basso valore di AA è legato al fatto che accuratezze più elevate si siano ottenute per classi minoritarie in termini di pixel di test.

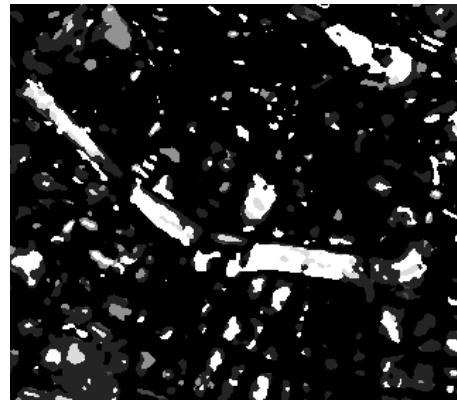
Per concludere questa prima sessione di discussione, è interessante confrontare questi risultati con quelli ottenuti senza estrazione delle *feature* HOG. La Figura 5.7 sintetizza i miglioramenti/peggioramenti classe per classe.

Si può chiaramente osservare che l’introduzione delle *feature* HOG ha quasi sempre apportato miglioramenti, con benefici notevoli soprattutto per le classi di suolo urbano. La Figura 5.7 dimostra come effettivamente il risultato della classificazione dell’immagine a tre canali mostri in generale più dettaglio; così facendo, però, ciò

Mappa di test									
Urbano ad alta densità	Urbano ad bassa densità	Strade	Area urbana verde	Suolo nudo	Terreno coltivabile	Area vegetata	Alberi	Fiumi	Bacini
7'642	12'419	1'183	0	-259	858	3'802	1'476	0	345

Figura 5.7. Differenze fra il numero dei campioni di test classificati correttamente classe per classe con e senza estrazione di *feature* HOG per il data set *Amiens 2006 - 5m - 10 classi*

che si guadagna in risoluzione si perde in accuratezza nella classificazione, in quanto molti pixel urbani vengono etichettati come appartenenti ad altre classi (in particolare la classe 10, gli specchi d'acqua, rappresentata in bianco nelle immagini). Il miglioramento ottenuto per le classi urbane è coerente con l'informazione direzionale estratta dalle *feature* HOG che risulta utile soprattutto ai fini della discriminazione di classi caratterizzate da struttura geometrica.



(a) Mappa di classificazione con HOG



(b) Mappa di classificazione senza HOG

Figura 5.8. Confronto su una porzione dell'area urbana di 250×250 pixel della mappa di classificazione con e senza estrazione di *feature* HOG sul *dataset* Amiens 2006 a 5 m (i livelli di grigio distinti rappresentano etichette di classi differenti)

5.4.2 Amiens 2012 - 2.5m - 7 classi

È riportata in Figura 5.9 la mappa di classificazione con *feature* HOG, in aggiunta alle tre *feature* spettrali, in configurazione di 4 *bin*, celle di dimensione 2×2 pixel e blocchi di normalizzazione con 16×16 pixel, con filtraggio gaussiano sia per l'immagine in ingresso che per le immagini HOG.

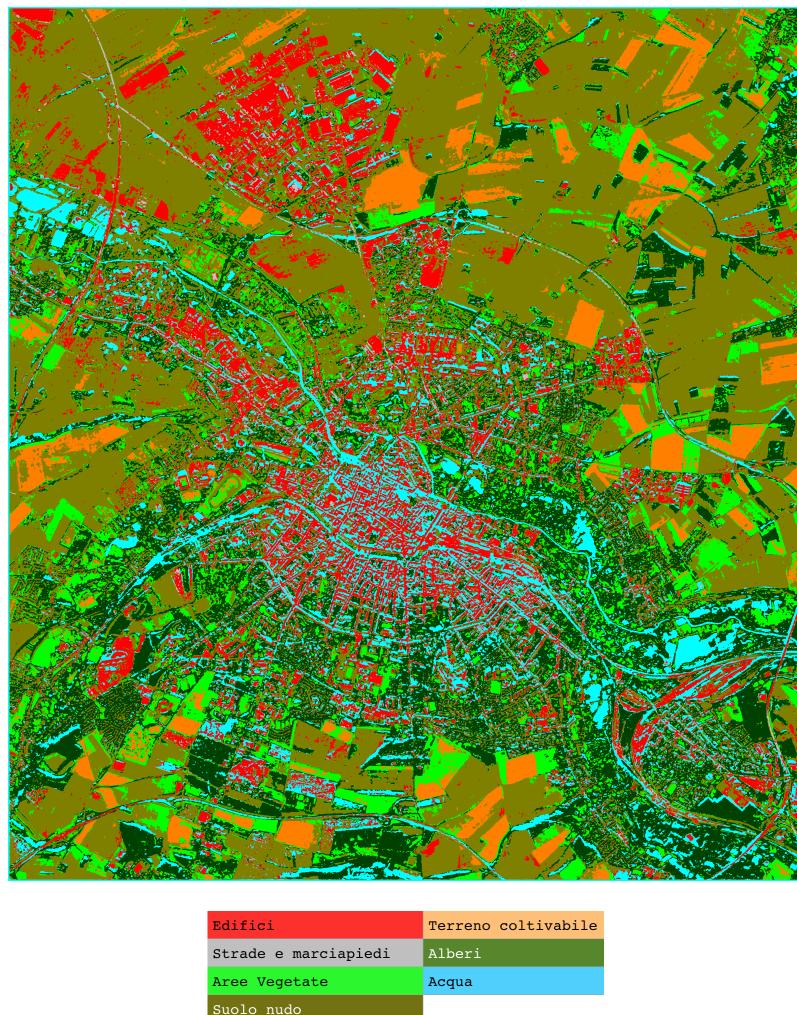


Figura 5.9. Mappa di classificazione ottenuta applicando SVM a feature spettrali e HOG per il dataset *Amiens 2012 - 2.5m - 7 classi*

È importante far notare fin da subito che l'immagine telerilevata che fa parte di questo *dataset* è stata acquisita in situazioni ambientali e in periodi di tempo diversi da quelle precedenti (nell'inverno 2012). Questo costituisce un evidente deficit nell'informazione proveniente dal suolo, in quanto la stagione di acquisizione ha reso altamente più complesso la discriminazione delle 7 diverse coperture di suolo. Le classi vegetate, in particolare, corrispondono qui a coperture del suolo che presentano similarità col suolo nudo, avendo l'acquisizione avuto luogo in un periodo dell'anno in cui gli alberi avevano perso le foglie e le aree agricole non presentavano già più vegetazione coltivata evidente.

Già con un'analisi della mappa, si osserva che qualitativamente il risultato sembra peggiorare rispetto all'esperimento precedente. Due sono i punti critici già a prima vista apprezzabili:

- i pixel etichettati come appartenenti al suolo nudo (classe 4) sembrano essere in misura sovrabbondante rispetto ai risultati precedenti: ciò può essere giustificato dal commento precedente;
- la quasi totale scomparsa delle strade (classe 2) nel centro, che talora vengono classificate come pixel appartenenti alla classe acqua (classe 7). Ciò si spiega non solo in relazione alla risoluzione spaziale dell'immagine considerata ma anche col fatto che molte strade presentino ombre di alcuni edifici; soprattutto quando osservata con tre soli canali, la risposta spettrale delle zone in ombra si presenta molto simile a quella dei corpi idrici.

Queste considerazioni trovano conferma negli indici di accuratezza che raggiungono, in assoluto, valori meno soddisfacenti rispetto ai casi esaminati in precedenza.

L'OA per questo esperimento, infatti, si è fermata al 56.63%, mentre l'AA ha raggiunto il valore di 57.33%.

In Figura 5.10 viene presentata la matrice di confusione di questo esperimento.

Mappa di test	Mappa di classificazione						
	Edifici	Strade	Area vegetata	Suolo nudo	Terreno coltivabile	Alberi	Acqua
Edifici	57441	1306	762	36352	17	3439	9523
Strade	11674	4272	1080	12983	26	2763	7325
Area vegetata	2970	284	74832	116879	10120	65691	11769
Suolo nudo	6518	465	1557	120127	10	16806	2379
Terreno coltivabile	0	0	4315	2810	80915	0	120
Alberi	136	97	16686	39398	163	138850	7956
Acqua	4503	1769	3232	15321	22	12145	86744
	52.78%	10.65%	26.48%	81.24%	91.78%	68.30%	70.10%

Figura 5.10. Matrice di confusione associata alla classificazione del dataset *Amiens 2012 - 2.5m - 7 classi* mediante l'applicazione di SVM a feature spettrali e HOG

Si può osservare che la classe per la quale la classificazione ha ottenuto i risultati migliori è stata quella relativa al terreno coltivabile (classe 5), che ha registrato un accuratezza di oltre il 90%. Prestazioni soddisfacenti sono state raggiunte anche per le classi di suolo nudo e acqua (classi 4 e 7), in cui una buona percentuale di pixel della mappa di classificazione coincide con la verità al suolo.

Mappa di test						
Edifici	Strade	Area vegetata	Suolo nudo	Terreno coltivabile	Alberi	Acqua
-11264	4110	18829	-1450	395	3824	3611

Figura 5.11. Differenze tra il numero di campioni di test classificati correttamente classe per classe con e senza estrazione di *feature* per il *dataset* Amiens 2012 a 2.5 m

Confrontando i risultati ottenuti dalla classificazione con HOG e da quella avvenuta come *feature* le sole tre bande telerilevate, si osserva come, anche questa volta, l'introduzione degli HOG porti, in generale, ad un incremento della percentuale di pixel classificati correttamente: in particolare in Figura 5.11 si può notare un netto miglioramento (+2537%) per le strade, benché in termini assoluti restino ancora ben poco classificate (10%) probabilmente a causa della presenza di ombre . Si è rilevato, tuttavia, un leggero peggioramento (-16%) nell'individuazione degli edifici. In conclusione, la presenza degli HOG migliora anche OA e AA seppure in maniera limitata.

Vista la non ancora soddisfacente percentuale di pixel di strada classificati in maniera adeguata, si è deciso di inserire una *feature* aggiuntiva della struttura cartografica delle strade di Amiens. L'immagine binaria in Figura 5.12 è stata inserita nel vettore delle *feature* in coda all'immagine precedente con vettori HOG.

Quest'immagine è il risultato della fusione di tre mappe tematiche (strade principali, strade secondarie e binari ferroviari, riportati unitamente a parte del territorio circostante quando tale territorio non presenta altro uso che quello associato ai trasporti), ottenuta dal servizio di mappatura stradale della European Urban Atlas. In particolare, l'immagine è aggiornata all'anno 2006 e possiede una risoluzione spaziale

di 5 m/pixel.



Figura 5.12. Immagine binaria della struttura cartografica di strade ed autostrade della regione di Amiens: strade, autostrade e terreni ad esse collegati sono rappresentati con il bianco e lo sfondo in nero.

In questo modo, si è riusciti a limitare la degradazione della classe strade (classe 2) portandola ad un più che accettabile valore di accuratezza (64.61%); dal momento che il numero di pixel classificati come strade non è molto significativo nella totalità dei pixel dell'immagine, non si ha un'apprezzabile incremento dell'OA, al contrario della AA che, pesando gli errori uniformemente, registra un significativo miglioramento (63.54%).

In Figura 5.13 viene mostrato l'effetto di questa *feature* aggiuntiva per la classificazione delle strade.

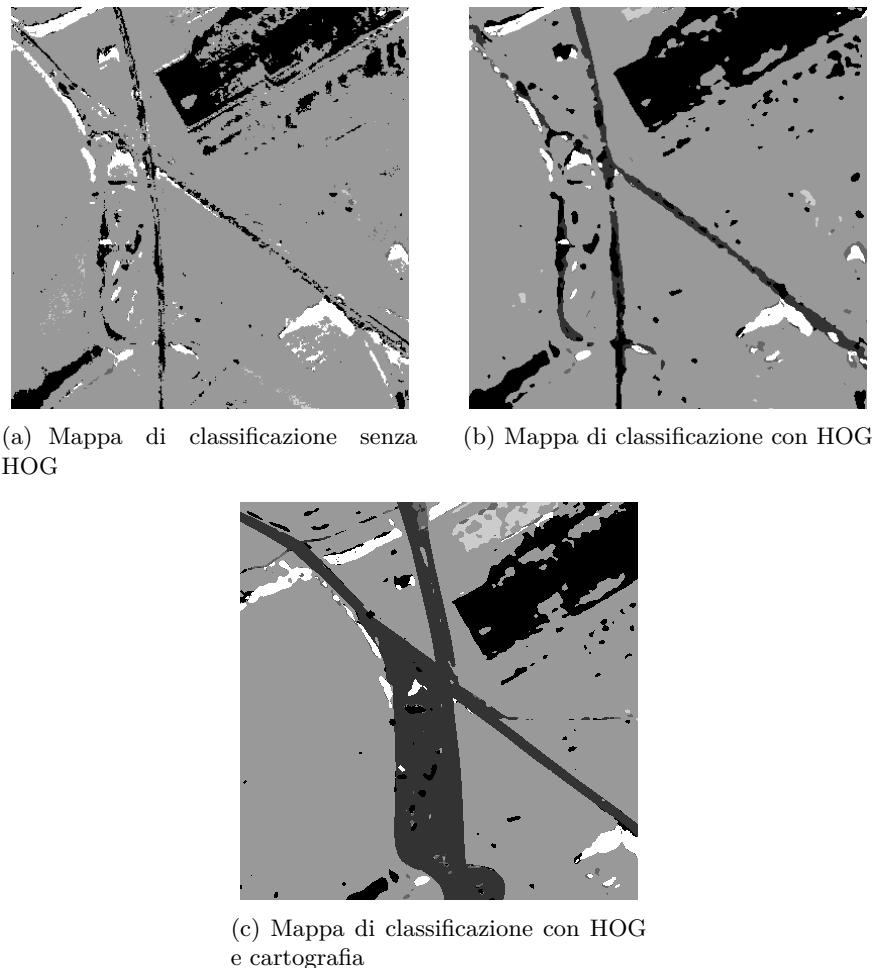


Figura 5.13. Confronto su una porzione della periferia di Amiens di 250×250 pixel della mappa di classificazione con e senza estrazione di *feature* HOG e con l'aggiunta della cartografia stradale sul *dataset* Amiens 2012 a 2.5 m (i livelli di grigio distinti rappresentano etichette di classi differenti)

5.4.3 Amiens 2006 - 2.5m - 7 classi

È riportata in Figura 5.14 la mappa di classificazione con *feature* HOG aggiuntive in configurazione di 4 *bin*, celle di dimensione 2×2 pixel e blocchi di normalizzazione con 16×16 pixel, con filtraggio gaussiano sia per l'immagine in ingresso che per le immagini HOG e con l'aggiunta della *feature* cartografica delle strade di Amiens.

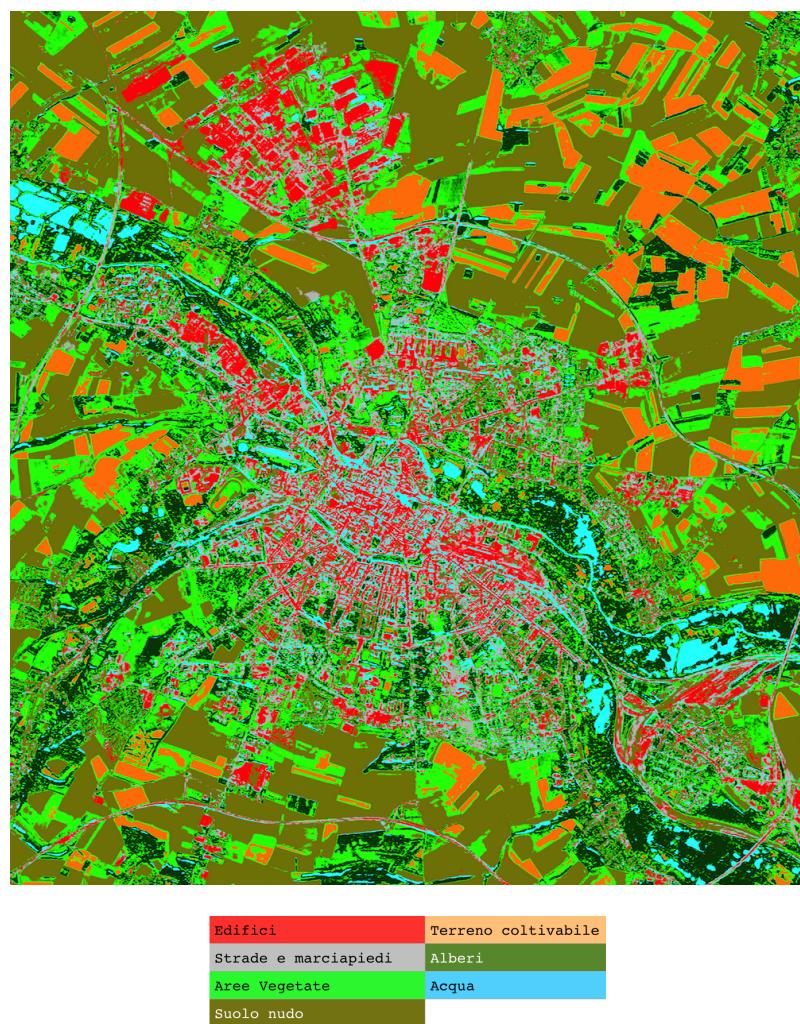


Figura 5.14. Mappa di classificazione ottenuta applicando SVM a feature spettrali e HOG per il dataset *Amiens 2006 - 2.5m - 7 classi*

Su questo *dataset* i risultati sono stati interessanti, poiché si differenziano notevolmente dai due casi studiati in precedenza.

E’ innanzitutto importante far notare che questo esperimento ha registrato il valore più alto a livello di OA (74.88%) e AA (70.61%).

L’analisi della matrice di confusione (riportata in Figura 5.14) conferma il fatto che il classificatore, per questo *dataset*, si comporti mediamente bene su tutte le classi, presentando accuratezze abbastanza elevate, soprattutto in rapporto alla difficoltà del problema di classificazione affrontato, per tutte le 7 classi presenti.

L’unico caso in cui i risultati non sono del tutto accettabili riguarda la distinzione tra area vegetata (classe 3) e alberi (classe 6), i cui pixel vengono spesso confusi tra le due categorie.

In Figura 5.16 vengono proposte le variazioni di prestazioni classe per classe nel confronto fra questo risultato e quello ottenuto senza *feature HOG*.

Si osserva come l’introduzione delle *feature HOG* peggiora quasi tutti i valori di precisione del classificatore.

Tuttavia, come era possibile ipotizzare già dopo l’analisi dei due precedenti esperimenti, l’uso di *HOG*, unito all’aggiunta della mappa cartografica, permette di far registrare un incremento notevole alla classe deputata all’etichettatura delle strade, che senza *HOG* è sempre la classe che registra i livelli di accuratezza più bassi, spesso insufficienti a dare una buona rappresentazione della verità al suolo. Tale risultato, insieme con quelli osservati in riferimento agli altri due data set, suggerisce come

Mappa di classificazione							
Mappa di test	Edifici	Strade	Area vegetata	Suolo nudo	Terreno coltivabile	Alberi	Acqua
Edifici	37195	8159	2095	10887	232	137	2088
Strade	1043	26389	3351	5186	909	1924	1321
Area vegetata	1995	6526	168827	22346	30447	45349	7055
Suolo nudo	9783	13535	21025	412260	2351	2224	384
Terreno coltivabile	2	1	17631	6629	169554	339	57
Alberi	182	1187	71698	3242	4062	119559	3356
Acqua	252	1336	16165	549	196	15974	89272
	61.18%	65.77%	59.75%	89.32%	87.30%	58.81%	72.14%

Figura 5.15. Matrice di confusione associata alla classificazione del dataset *Amiens 2006 - 2.5m - 7 classi* mediante l'applicazione di SVM a feature spettrali e HOG

Mappa di test						
Edifici	Strade	Area vegetata	Suolo nudo	Terreno coltivabile	Alberi	Acqua
-408	8'867	-5'568	-24'671	-4'589	-27'370	-1'456

Figura 5.16. Differenze tra il numero di campioni di test classificati correttamente classe per classe con e senza estrazione di *feature* per il dataset Amiens 2006 a 2.5 m

l'uso di HOG in un problema di classificazione a risoluzione spaziale abbastanza elevata in aree urbane presenti benefici in termini di capacità di discriminazione di classi con struttura geometrica ben definita, ma anche potenziali svantaggi, soprattutto in riferimento all'identificazione di classi prive di tale struttura a causa della loro natura fisica (ad es., alberi) o delle condizioni di acquisizione (ad es., sfocatura nei dati o presenza di ombre).

Capitolo 6

Conclusioni

In questa tesi si è affrontato il problema della classificazione di immagini telerilevate ad alta risoluzione (VHR), con particolare attenzione al contributo che l'estrazione di *feature* aggiuntive può apportare all'accuratezza dei risultati ottenuti sulle mappe di classificazione. In particolare è stata esplorata l'opportunità di usare l'approccio HOG proposto inizialmente in altre applicazioni di *computer vision*. Tale approccio è stato analizzato e implementato, testandone l'applicabilità al problema della mappatura del suolo in aree urbane tramite immagini multispettrali telerilevate.

Il metodo HOG di estrazione di *feature* è stato combinato con un classificatore non parametrico SVM e con un metodo di ottimizzazione automatica dei parametri di tale classificatore. Tale approccio è stato sperimentato con tre casi di studio, in cui l'eterogeneità spaziale e spettrale delle immagini coinvolte rende complesso il problema di classificazione. In particolare, le immagini utilizzate sono state caratterizzate da sfuocatura dovuta a risoluzione spaziale meno fine della dimensione del pixel, solo tre bande spettrali e presenza di occlusioni (ombre) nelle aree urbane coinvolte. I valori di accuratezza ottenuti non sono risultati elevati, tuttavia si sono dimostrati accettabili e soddisfacenti, soprattutto in relazione alla difficoltà nel discriminare

classi di uso del suolo coi dati disponibili. Infatti, confrontando i parametri di accuratezza ottenuti con e senza estrazione di *feature*, si è potuto constatare come l'utilizzo dell'algoritmo HOG abbia spesso portato ad un incremento di accuratezza, soprattutto per quelle classi per le quali è marcata la presenza di strutture geometriche regolari. Quest'approccio ha, invece, mostrato limitazioni nella mappatura di aree spazialmente omogenee.

Per queste ragioni, nei casi di studio in ambito urbano considerati, le *feature* di tessitura HOG si possono ritenere uno strumento utile per zone caratterizzate da un elevato livello di dettaglio geometrico, quali zone urbane e strade, mentre è consigliabile affiancare ad esse altri metodi di *feature extraction* per classificare con affidabilità porzioni di suolo strutturalmente omogenee.

Esempi di futuri sviluppi in tale ambito potrebbero riguardare l'utilizzo di HOG insieme non solo ai canali spettrali, ma anche ad altre tipologie di *feature* quali i parametri morfologici, il semivariogramma oppure la associazione di tale algoritmo all'informazione di segmentazione, in modo da integrare nel processo di classificazione informazione complementare a quella fornita dai descrittori HOG qui esposti.

Bibliografia

- [1] A.V. OPPENHEIM, R. S. *Elaborazione Numerica dei Segnali*. F. Angeli, 1990.
- [2] BURGES, C. J. C. A tutorial on support vector machine for pattern recognition. *Data Mining and Knowledge Discovery* 2 (June 1998), 121–167.
- [3] CHANG, C.-C., AND LIN, C.-J. Working set selection using second order information for training svm. *Journal of Machine Learning Research* 6 (May 2005), 1889–1918.
- [4] CHEN, Q., AND GONG, P. Automatic variogram parameter extraction for textural classification of the panchromatic ikonos imagery. *IEEE Transactions on Geoscience and Remote Sensing* 42, 5 (May 2004), 1106–1115.
- [5] CRISTIANINI, N., AND TAYLOR, J. S. *An introduction to Support Vector Machines*. Cambridge University Press, 2000.
- [6] DALAL, N., AND TRIGGS, B. Histograms of oriented gradientes for human detection. In *Computer Vision and Pattern Recognition* (June 2005), vol. 1, pp. 886–893.

- [7] GHAMINI, P., MURA, M. D., AND BENEDIKTSSON, J. A. A survey on spectral-spatial classification techniques based on attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing* 53, 5 (May 2015), 2335–2353.
- [8] KUO, B.-C., AND LANDGREBE, D. A. Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing* 42, 5 (May 2004), 1096–1105.
- [9] LANDGREBE, D. A. *Signal theory methods in multispectral remote sensing*. Wiley-InterScience, 2003.
- [10] MOSER, G. *Analisi di immagini telerilevate per osservazione della terra*. Ecig Universitas, 2007.
- [11] MOSER, G., AND SERPICO, S. Automatic parameter optimization for support vector regression for land and sea surface temperature estimation from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing* 47, 3 (March 2009), 909–921.
- [12] MOSER, G., AND SERPICO, S. Combining support vector machines and markov random fields in an integrated framework for contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing* 51, 5 (May 2013), 2734–2752.
- [13] MOSER, G., SERPICO, S., AND BENEDIKTSSON, J. A. Land-cover mapping by markov modeling of spatial-contextual information in very-high-resolution remote sensing images. *Proceedings of the IEEE* 101, 3 (March 2013), 631–651.

BIBLIOGRAFIA

- [14] PESARESI, M., AND BENEDIKTSSON, J. A. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing* 39, 2 (February 2001), 309–320.
- [15] RICHARDS, J., AND JIA, X. *Remote sensing digital image analysis*. Springer, 2005.
- [16] SMITH, B. T. *Lagrange Multipliers Tutorial in the Context of Support Vector Machines*. PhD thesis, Faculty of Engineering and Applied Science Memorial University of Newfoundland, June 2004.
- [17] TORRIONE, P. A., MORTON, K. D., SAKAGUCHI, R., AND COLLINS, L. M. Histograms of oriented gradientes for landmine detection in ground-penetrating radar data. *IEEE Transactions on Geoscience and Remote Sensing* 52, 3 (March 2014), 1539–1550.
- [18] VAPNIK, V. *The Nature of Statistical Learning Theory*, 2 ed. Springer-Verlag New York, 2000.