People 's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

THIRD CYCLE LMD FORMATION

A Thesis submitted in partial execution of the requirements of

DOCTORATE DEGREE

Suggested by

Mohamed Khidher University Biskra

Presented by

Souad BENBRAIKA

Titled

Conditional Field for Image Classification

Supersivor: **Pr.Zouhir MOKHTARI**

Examination Committee :

Mr.Khaled MELKEMI	Professor	University of Batna 2	President
Mr. Zouhir MOKHTARI	Professor	University of Biskra	Supervisor
Mr. Mokhtar HAFAYEDH	Professor	University of Biskra	Examiner
Mr. Djabrane YAHIA	MCA	University of Biskra	Examiner
Mr. El Amir DJEFFAL	MCA	University of Batna 2	Examiner

In the memory of my little brother.

To my parents.

To my brothers and sisters

Acknowledgement

Words cannot thank them enough, I would just extend profound thanks to :

Particularly my supervisor Pr. MOKHTARI Zouhir for accompanying me throughout the academic course as well as the moral supports.

The honorable examination committee members for accepting the evaluation and discussion of this dissertation.

The Dr. MELGANI Farid from Trento University, for his guidance, valuable advice, constructive criticism and counsel. Besides to all DISI Lab members Trento University who have provided me with tremendous assistance during my internship.

I would like to express my special appreciation to my parents for giving me the strength to reach this accomplishment. At this moment, i am greatly indebted to both of them for everything. My

sincere thanks to my friends R. Benbrahim, K.Saibi, D. Amira and K.Hanane that helped me in variety of ways.

As well as anyone who helped the cause of this work in any possible way.

Abstract

We propose a novel multi label (ML) classification approach based on the Conditional Random fields (CRF) for the high resolution UAV images. The inderlying idea of the proposed model integrate 1) spatial information within the same class; jointly with 2) cross-correlation information between different class labels after . The experiments were done on two different UAV image datasets and the experimental results show that the new model outperforms conventional approaches.

Keywords: Conditional random fields (CRF), Markovian random fields(MRF), Image multilabeling classification, Spatial contextual information, UAV images.

Résumé

Nous avons proposé une nouvelle approche de classification multilabel (ML), basée sur les champs conditionnels pour les images connues par UAV avec une haute résolution. L'idée principale du modèle proposé est d'intègrer les deux types d'informations suivantes: 1) l'information spatiale au sein de la même classe, (conjointement avec) 2) Les informations d'intercorrélation entre différents labels de classes. Les implémentations sont réalisées sur deux ensembles d'images UAV différentes et les résultats expérimentaux montrent que le modèle proposé surpasse les approches conventionnelles.

Mots clés: Champs aleatoires conditionels, Champ markovien, Classification de l'image multilabeling, Information contextuelle spatiale, Les images UAV (Drone).

table of acronyms

AVG	Average Accuracy
CRFs	Conditional Random Fields
CRF-bias	standard conditional random fiels with bias feature in the potential functions
CRF-ML	Conditional Random Field for Multilabel data
FN	the False Negative
FP	the False Positive
Full-ML-CRF	reference for the proposed method which exploits also the interclass correlation
ICM	Iterated Conditional Mode
ML-CRF	Multi-Label Conditional Random Field
ML-Unary	reference to the thresholding of resulting maps obtained by bag-of-words strategy coupled with autoencoder network and multilayer perceptron classifier.
MRFs	Markov Random Fields
SEN	The Sensitivity
SPE	The Specificity
TN	the True Negative
TP	the True Positive
UAVs	Unmanned aerial vehicles

Table of symbols

X, X_k	The observed data
Y, Y_k	The labels of an observed data
$P(Y\|X)$	The conditional probability
$\mathcal{N}, \mathcal{N}_\rangle$	Neighborhood system and thier adjecent relationship
$\mathcal{C}$	set of all cliques possible
$\mathcal{S}$	set of sites in regular square lattice
$\mathcal{L}$	set of possible labels
$P(y\|\theta)$	Likelihood function
$E(.)$	Energy function
$\mathcal{G}$	the graph which is defined by a set of sites and neighboring system
$P(Y, X)$	the joint probability
λ_1, λ_2	the cross-correlation coefficients
$\widehat{y}, \widehat{y_k}$	The optimale solution

Contents

List of Figures

List of Tables

Introduction

The multilabel and the multi class classification are the most popular techniques to classify a query image, where multi-label outputs can hold more than one labels simultaneously unlike multi-class cases in which the samples belong to only one class. As a result, the space of multi-labeling problems exponentially grows as the number of classes. Besides, such datasets normally contain complicated structure and often the result of these kinds of complexity is the failure of typical ad-hoc approaches to make reasonable classification. Consequently, more synthetic approaches have been proposed in the past decade on the basis of machine learning tools such as:(Lu & Weng, 2007) in which the authors reviewed the pipelines of remote sensing imagery classification. After a tiles-representation of considered image, a deep learning tools(MacKay, 2003) are applied and a mutilabel solutions are resulting in (Zeggada, Melgani, & Bazi, 2017) (Zeggada & Melgani, 2016) (Zeggada & Melgani, 2017) that, most of the approaches addressed the problem of the high spatial resolution imagery regarding the high level of detail.

Among the machine learning tools, the structured random fields are known to be one of the most promising tools for encapsulating spatially neighbouring information in the classification model. It gained highly attention as application in the remotely sensed field such as (Tarabalka, Fauvel, Chanussot, & Benediktsson, 2010),(Schindler, 2012) (M. Li, Zang, Zhang, Li, & Wu, 2014) and(Zhang, Yan, Li, & Chen, 2016)˙

The Markov Random Fileds is graphical model that provides a flexible, powerful and theoretically well-established way to model relationships between random variables, and they have been applied with great success for different image types. For instance,

(Melgani & Serpico, 2003) considered a MRF approach for multitemporel image that the proposed model has been aimed at improving both the accuracy and the reliability of the classification process by means of a better exploitation of the temporal information beside the traditional spatial information. Dynamic MRFs were solved in order to perform real-time segmentation in video sequences (Kohli & Torr, 2005), that when we perform real-time segmentation, the MRF changes with every image frame. The authors in previous paper have proposed the solution through the minimization of energy function associate to the MRF using the graph cuts algorithm. Indeed, the graph cuts is well known as one of promising methods that the solution is a global one as described in (Kolmogorov & Zabin, 2004). (Lafferty, McCallum, & Pereira, 2001) has presented other class of random fields which modeling the posterior distribution as an MRF, this class of RFs is well known by Conditional Random Fields CRFs that the authors have considered a CRFs for segmenting and labeling problems of sequence data. Based on the concept of CRFs, (S. Kumar et al., 2003) have proposed a Disciminative random fields that permits interactions in both the observed data(inputs) and the labels for detection the man-made structure in natural scences. Moreover, the CRFs frameworks have been extended substantially according to the increasing complexity of remote sensing data as well (Volpi & Ferrari, 2015) address the problem of semantic segmentation in urban remote sensing into land cover maps. They proposed to embed geographic context potentials into a pairwise CRF coupled with unary potentials from a random forest (RF) classifier. On the otherhand, (Paisitkriangkrai, Sherrah, Janney, & van den Hengel, 2016) introduce a semantic pixel labeling framework of aerial and satellite imagery. They exploit different types of features in order to generate per-pixel class probabilities followed with a CRF as a post-processing step. Furthermore (Zhao, Zhong, Wu, Zhang, & Shu, 2015) present a sub-pixel mapping algorithm based on CRF for hyperspectral remote sensing imagery. The methodologies that are presented in this thesis address the problem of the multi-labeling classification image within a conditional random field (CRF) framework with taking into consideration the spatial contextual information and cross-correlation be-

tween labels simultaneously. We formulate the multilabeling problem applied at a tile level under a CRF perspective, with the aim to assign to each tile a vector of labels instead of just one class label. The main novelty is that integrate the above two properties in optimization process. To achieve the objectif of this thesis, a CRF model is applied for multilabeling data. The purpose of this step is to test the effectiveness of the information importing from adjecent classes side to side with traditional contextual information within the same class.

In this sense, the thesis is divided as follows:

- **Chapters 1 and 2:** First of all, we introduce preliminary concepts and the backgound of structured random fields ,in particular MRFs and CRFs, and theirs properties. Then we describe briefly the relationship between Gibbs distribution and MRFs. The second chapter reviews the classification literature using the structured random fields. Starting by briefly explaining the classification problem and how modelized the problem through probabilistic model. Then we review some methods for related problem : Parameters estimation and the related optimization procedure.

- **Chapters 3 and 4:** in the 3^{rd} chapter, we start with an explication of the problem which is dealt with in this thesis through two proposed models on basic of CRFs concept. The two models are detailing in two separately sections. The pipeline of the framework consists of two main phases. First, the considered input image is subdivided into a grid of tiles, which are processed thanks to an opportune representation and a multilayer perceptron classifier are providing tile-wise multilabel prediction probabilities. In the second phase, a multilabel CRF model is applied to integrate spatial correlation between adjacent tiles and the correlation between labels within the same tile, with the objective to improve iteratively the multilabel classification map associated with the considered input image. Chapter number 4 is devoted for experimental results which are achieved on two different UAV image datasets are reported and discussed.

CHAPTER 1

Preliminary on structured random fields MRFs and CRFs

Contents

1.1 Preliminary

1.1.1 graph

The graph is one of basic and vaste concepts in graph theory framework. for that, we introduce the notations and concepts juste that they are related in our study.

Neighborhood system

Let consider $\mathcal{S}$ a set of sites in regular square lattice. A neighborhood system for $\mathcal{S}$ is defined as:

$$\mathcal{N} = \{\mathcal{N}_i | i \in \mathcal{S}\} \tag{1.1}$$

where $\mathcal{N}_i$ is the set of sites adjacent i. this neighboring relationship is given as:

$$\mathcal{N}_i = \{j \in \mathcal{S} \mid \ [dist(site(i), site(j))]^2 \leq r, i \neq j\} \tag{1.2}$$

with $dist(a, b)$ refer to Euclidean distance between to elements a and b, and r takes integer value. Notice that, the sites at or near the boundaries have fewer neighbors. Besides, for each i, $\mathcal{N}_i$ is verified the two following properties:

- A site is not neighboring to itself i.e. $i \notin \mathcal{N}_i$.
- The neighboring relationship is mutual i.e. $i \in \mathcal{N}'_i \Leftrightarrow i' \in \mathcal{N}_i$ with $i \neq i'$.

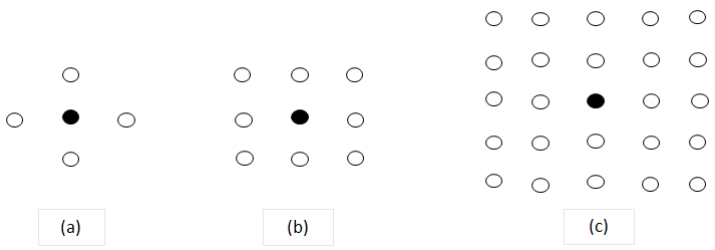


Figure 1.1: Neighborhood on a lattice of regular sites.

Indeed the value of r lead to many orders for a system. The first-order neighborhood system is defined for $r = 1$, also called the 4-neighborhood system, in which every interior site has four neighbors Fig.1.1(a) while the boundry sites have three and the sites at corners have two neighbors. We obtain the second-order neighborhood system, also called the 8-neighborhood system for $r = 2$. As shown in Fig.1.1 (b), the interior sites have eights neighbors, the boundries have five adjecents sites and the corners nodes have three. A widely neighboring relationships can defined for the square lattice in Fig.1.1(c) where we have the possibility to define 5 orders for the neighboring system (i.e. $r \in \{1, ..., 8\}$).

The couple $(\mathcal{S}, \mathcal{N})$ is called a Graph in the usual way. The Graph $\mathcal{G} = (\mathcal{S}, \mathcal{N})$ is constituted by the set of nodes $\mathcal{S}$ while neighborhood system $\mathcal{N}$ determines the edges between the nodes in which all the edges have no orientation.

Clique

One of the most concepts related to the considering graph $\mathcal{G} = (\mathcal{S}, \mathcal{N})$ is the clique. All subset c of sites in $\mathcal{S}$ is defind as clique (S. Z. Li, 2009). It consists from a union of a single-sites $\mathcal{C}_1 = \{i | i \in \mathcal{S}\}$, a pair of neighboring sites $\mathcal{C}_2 = \{(i, i') | i \in \mathcal{S}, i' \in \mathcal{N}_i\}$, triple of neighboring sites $\mathcal{C}_3 = \{(i, i', i'') | i, i', i'' \in \mathcal{S} are neighbors\}$ and so on. $\mathcal{C}$ denotes the set of all possibles cliques of the graph $(\mathcal{S}, \mathcal{N})$ is given by:

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3 ... \tag{1.3}$$

Cliques for $(\mathcal{S}, \mathcal{N})$ of a regular square lattice have a size, shape, and orientation, according the neighborhood system. The single-site, horizontal and vertical pair-site clique Fig.1.2(a) are all the possible cliques for the first-order neighborhood system. for the second-order neighboring system, the cliques types includes those one defining in (a) and the diagonal pari-site clique (b) and triple-site (c) and the quadruple-sites cliques(d) see Fig.1.2 Ṅoting that the number of cliques grows rapidly compared the order of neighborhood system.

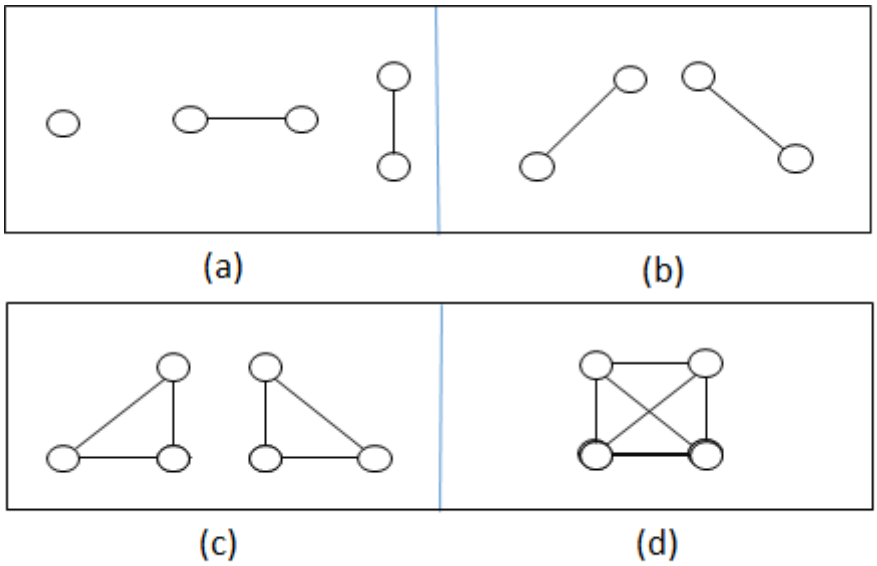


Figure 1.2: Cliques on a set of regular sites

1.1.2 Gibbs distribution and properties

Gibbs distribution

Lets consider a set of random variables $\mathcal{Y} = \{Y_1, ..., y_n\}$ defined on the graph $(\mathcal{S}, \mathcal{N})$ in which each variable Y_i and $i \in \{1, 2, ..., n\}$ takes a value y_i. All random variables obey the Gibbs distribution is namely *Gibbs Random Field* (GRF). The random variables obeys the a Gibbs distribution if and only if verified :

$$P(Y = y) = \frac{1}{Z} exp(-E(y)). \tag{1.4}$$

where

$$Z = \sum_{allpossibilityofy} exp(-E(y)) \tag{1.5}$$

Z is a normalizing constant called also the partition function, and the $E(y)$ is namely the energy function and it takes the following form :

$$E(y) = \sum_{c \in \mathcal{C}} V_c(y) \tag{1.6}$$

the energy function is defined as a sum of cliques function V_c over all possible cliques $\mathcal{C}$. According to the set of all possibles cliques of the graph $(\mathcal{S}, \mathcal{N})$ Eq.1.3, thus the most general energy function (S. Geman & Geman, 1987) expressed as following:

$$E(y) = \sum_{i \in \mathcal{C}_1} V_1(y_i) + \sum_{i,i' \in \mathcal{C}_2} V_2(y_i, y_i') + \sum_{i,i',i'' \in \mathcal{C}_3} V_3(y_i, y_i', y_i'') + ... \tag{1.7}$$

An important special case is the model of *Ising* (Ising, 1925), in which its has been defined when asuming only up to pairwise cliques, that is :

$$E(y) = \alpha \sum_{i \in \mathcal{C}_1} V_1(y_i) + \beta \sum_{i,i' \in \mathcal{C}_2} V_2(y_i, y_i') \tag{1.8}$$

α and β are the parameters of unary and parwise interaction respectivly. Notice that Ising model has been highly paid attention as an application in many different fields especially with the homogeneous and isotropic properties see (MacKay, 2003).

Properties of a GRF

We discuss here the two following properties: the homogeneity and isotropic of a GRF. Lets consider the GRF Y on $\mathcal{S}$ with regarding to the neighborhood system $\mathcal{N}$.

A GRF Y is homogeneous if the clique function V_c is independent of the ralative position of clique c in the set of sites $\mathcal{S}$.

A GRF Y is isotropic if the clique function V_c is independent of the orientation of clique c in the set of sites $\mathcal{S}$. It is considerably simpler to specify a GRF distribution if it is homogeneous or isotropic than one without such properties. The homogeneity is assumed in most Markov Random Field vision models for mathematical and computational convenience.

The evaluation of the normalizition constant(partition function) Z is necessary to calculate a Gibbs distribution. Recall that Z is defined as sum over all possible configuration of Y. even the size of random field and neighboring system are moderates, the evaluata-

tion is prohibitive. thus several approximation methods are considering for solving the problem see (S. Z. Li, 2009) (chapter number 8).

1.2 Structured random fields

The Structured random fields constitute a part of probability theory, in particular Markov Random Fields MRFs and Conditional Random Fields CRFs in which they provides a foundation for the characterization of contextual constraints.

In point of fact, the concept of MRF is extending of Markov process, commonly kown Markov chain, which is a sequence of random variables defind on the time indices (one-dimensional) as we can show in (Rozanov, 1982)the next sections reviews the MRF theory and properties.

1.2.1 Markov Random Fields

Lets consider the family of random variables $\mathcal{Y} = \{\mathcal{Y}_1, ..., \mathcal{Y}_n\}$ defined over the graph $\mathcal{G} = (\mathcal{S}, \mathcal{N})$, and its realization is $y = \{y_1, ..., y_n\}$ in which each value y_i is defined in the set $\mathcal{L} = \{0, 1\}$ where $\mathcal{Y}_i = y_i$ for every $i = \overline{1, n}$. with view of simplification, we abbreviate $\mathcal{Y} = y$ where the $y = \{y_1, ..., y_n\}$ is a configuration of the random field $\mathcal{Y}$. Aslo, the probability $P(\mathcal{Y} = y)$ is abbreviated $P(y)$.

$\mathcal{Y}$ is said to be a Markov random field(MRF), with respect to $(\mathcal{S}, \mathcal{N})$, if and only if the following two conditions are satisfied:

- $P(y) > 0 \qquad \forall y$ (positivity).
- $P(y_i|y_{\{\mathcal{S}\setminus i\}}) = P(y_i|y_{\mathcal{N}_i})$ (Markov Property).

where, $\mathcal{S}\setminus i$ refer to all sites of $\mathcal{S}$ except the site i. and $y_{\{\mathcal{S}\setminus i\}}$ signified to the configuration of the set of sites $\mathcal{S}\setminus i$ while $y_{\mathcal{N}_i}$ are the labels of adjecent sites for i i.e. $y_{\mathcal{N}_i} = \{y_{i'}|i' \in \mathcal{N}_i\}$. The essence of the markov property is that depicts the local characteristics. Since (Besag,

1974) assuming that if the positivity condition is satisfied, the probability (joint probability)of any random field is uniquely determined by its local probabilities.

The homogeneity and isotropy are additional properties. An MRF is said to be homogeneous if the local probability (for the $\mathcal{Y}$ is an MRF, the local probability is $P(y_i|y_{\mathcal{N}_i})$), is independent of the relative location of the site i in $\mathcal{S}$. we will discuss the isotropy property of an MRF following subsection because it related with clique clique function.

Besides the above properties of a MRF, we refer to two other variants of markov properties which are:

- **Pairwise Markovianity**: is the independent of two labels non-adjecent ,y_i and y_j, given the labels of the other sites i.e. for any non-adjecent sites i and j we have : $P(y_i|y_j) = P(y_i|y_{\mathcal{S}\setminus i\setminus j})$.
- **Global Markovianity**: for any disjoint subsets $\mathcal{A}$,$\mathcal{B}$ and $\mathcal{C}$ where $\mathcal{C}$ separating $\mathcal{A}$ from $\mathcal{B}$. we said that a Markovianity is global if $P(y_{\mathcal{A}}|y_{\mathcal{B}}) = P(y_{\mathcal{A}}|y_{\mathcal{C}})$. This meas that, given a set of sites, the labels of any two sparated subsets are independent.

Markov-Gibbs equivalence

An MRF is specify through two terms which are : the joint probability $P(y)$ and the local conditiona probability $P(y_i|y_{\mathcal{N}_i})$. under the disadvantages of the conditional probability approach, (Besag, 1974) debated for the joint probability $P(y)$. That no obvious method is available to deduce the joint probability from the conditional probabilities while the conditional probabilities themselves are subject to some highly restrictive consistency conditions. The Hammersley-Clifford theorem (Hammersley and Clifford 1971)establishes the equivalence between an MRF and GRF.

The theorem is assuming that: $\mathcal{Y}$ is an MRF on $\mathcal{S}$ w.r.t $\mathcal{N}$ if only if it have a Gibbs distribution, and the proof of the equivlance is showed with details in this work (Besag, 1974). for now, we ll proof that if $\mathcal{Y}$ is a GRF so it is an MRF.

Lets consider that $\mathcal{Y}$ have Gibbs distribution w.r.t the graph $\mathcal{G}$. According to Bayes theorem we have:

$$\begin{aligned} P(y_i|y_{\{\mathcal{S}\setminus i\}}) &= \frac{P(y_{\{\mathcal{S}\setminus i\}}|y_i)P(y_i)}{P(y_{\{\mathcal{S}\setminus i\}})} \\ &= \frac{P(y_{\{\mathcal{S}\setminus i\}}, y_i)}{P(y_{\{\mathcal{S}\setminus i\}})} \\ &= \frac{P(f)}{P(y_{\{\mathcal{S}\setminus i\}})} \end{aligned} \tag{1.9}$$

Recall that y is a GRF that means it can expressed in terms of energy function. Noting that $y' = y_{\{\mathcal{S}\setminus i\}}$ is the configuration of y without the variable y_i and $y'_i \in y'$. the equation 1.9 rewritting as :

$$\begin{aligned} P(y_i|y_{\{\mathcal{S}\setminus i\}}) &= \frac{exp\{-\sum_{c\in\mathcal{C}} V_c(y)\}}{\sum_{y'_i} P(y')} \\ &= \frac{exp\{-\sum_{c\in\mathcal{C}} V_c(y)\}}{\sum_{y'_i} exp\{-\sum_{c\in\mathcal{C}} V_c(y')\}} \end{aligned} \tag{1.10}$$

Divide $\mathcal{C}$ into two sets A and B with A consisting of cliques containing i and B cliques not containing i, Then 1.10 can be written as:

$$P(y_i|y_{\{\mathcal{S}\setminus i\}}) = \frac{[exp\{-\sum_{c\in A} V_c(y)\}][exp\{-\sum_{c\in B} V_c(y)\}]}{\sum_{y'_i}\{[exp\{-\sum_{c\in A} V_c(y')\}][exp\{-\sum_{c\in B} V_c(y')\}]\}} \tag{1.11}$$

Because $V_c(y) = V_c(y')$ for any clique c that does not contain i, so $exp\{-\sum_{c\in B} V_c(y)\}$ cancels from the both the numerator and denominator giving the following equation :

$$P(y_i|y_{\{\mathcal{S}\setminus i\}}) = \frac{exp\{-\sum_{c\in A} V_c(y)\}}{\sum_{y'_i} exp\{-\sum_{c\in A} V_c(y')\}} \tag{1.12}$$

Consequently, the equation 1.12 depends on labels at i's neighbors, in other words y is an MRF.

The essence of the theorem results is that, the joint probability can be writting in terms of energy function by specifying the cliquespotentials functions :

$$P(y) = \frac{1}{Z} exp\{-\sum_{c\in\mathcal{C}} V_c(y)\}. \tag{1.13}$$

Taking into account the definition proposed in 1.8 it is defined for pair-site clique functions. The joint probability of an MRF take the form :

$$P(y) = \frac{1}{Z} exp\{\sum_{i \in \mathcal{C}_1} V_1(y_i) + \sum_{i,i' \in \mathcal{C}_2} V_2(y_i, y_i')\}. \tag{1.14}$$

Such models are called Auto-models (Besag, 1974) for some definitions for V_1 and V_2 as well the *auto-logisitc model*, *auto-binomial model*, and *auto-normal model*, also called *caussian MRF*. those auto model are the popular auato model generally used, (S. Z. Li, 2009) is discussed and detailed those models and others. below,we introduce briefly the auto-logistic model and Gaussian-MRF.

Auto model

Under the assumption that y_i have only two possible value, those value defining by the discrete label set $\mathcal{L} = \{0, 1\}$ (for other references $\mathcal{L} = L = \{-1, 1\}$). Assuming only up the paiwise cliques functions to be nonzero, (Besag, 1986) was proposed the joint probability $P(y)$ as a pairwise interaction MRF and expressed by :

$$P(y) \propto exp\{\sum_{i \in \mathcal{S}} G_i(y_i) + \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{N}_i} G_{ij}(y_i, y_j)\} \tag{1.15}$$

where the G-functions are arbitrary [(Ganan & McClure, 1985), (Besag, 1986)].

- **Auto-logistic model:** An auto-model is said to be an auto-logistic model if the corresponding energy function is of the form:

$$E(y) = \sum_{i \in \mathcal{S}} \alpha_i y_i + \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{N}_i} \beta_{ij} y_i y_j \tag{1.16}$$

  with β_{ij} is the interaction coefficients. The conditional probability for this model is

:

$$P(y_i|y_{\mathcal{N}_i}) = \frac{exp\{\alpha_i y_i + \sum_{j\in\mathcal{N}_i}\beta_{ij}y_i y_j\}}{\sum_{y_i\in\mathcal{L}} exp\{\alpha_i y_i + \sum_{j\in\mathcal{N}_i}\beta_{ij}y_i y_j\}} = \frac{exp\{\alpha_i y_i + \sum_{j\in\mathcal{N}_i}\beta_{ij}y_i y_j\}}{1 + exp\{\alpha_i y_i + \sum_{j\in\mathcal{N}_i}\beta_{ij}y_i y_j\}} \tag{1.17}$$

- **Gaussian-MRF model:** An auto-model is said to be an Gaussian-MRF model if the set label $\mathcal{L}$ is a real line (compact of $\mathbb{R}$) and The joint probability is a Gibbs distribution:

$$P(y) = \frac{\sqrt{det(M)}}{\sqrt{(2\pi\sigma^2)^m}} exp\{\frac{(y-\mu)^t M(y-\mu)}{2\sigma^2}\} \tag{1.18}$$

where: y reconstruct as a vector, μ is the $m\times 1$ vector of the conditional means, M is the interaction matrix of $m\times m$ dimension, whose diagonal elemeants are unity and the off-elements are $-\beta_i j$. since, we expressed the MRF by his energy function, so needed to define the single-site an pair-site clique functions V_1 and V_2 respectively. the energy function of an auto-normal model is:

$$E(y) = \sum_{i\in\mathcal{S}} \frac{(y_i-\mu_i)^2}{2\sigma^2} + \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{N}_i} \beta_{ij}\frac{(y_i-\mu_i)(y_j-\mu_j)}{2\sigma^2} \tag{1.19}$$

Its conditional probability density function is :

$$P(y_i|y_{\mathcal{N}_i}) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{1}{2\sigma^2}[y_i - \mu_i - \sum_{j\in\mathcal{N}_i}\beta_i j(y_j-\mu_j)]^2\} \tag{1.20}$$

given the conditional mean and conditional variance :

$$E(y_i|y_{\mathcal{N}_i}) = \mu_i - \sum_{j\in\mathcal{N}_i}\beta_i j(y_j-\mu_j) \tag{1.21}$$

$$var(y_i|y_{\mathcal{N}_i}) = \sigma^2 \tag{1.22}$$

1.2.2 Conditional Random Fields

CRFs are a probabilistic framework for labeling and segmenting data. the first proposed of the CRFs addressed the segmenting and labeling a text sequences data (Lafferty et al., 2001).

The underlying main of CRFs is modeling the Posterior probability directly as an MRF. If we consider an observed data $x = \{x_1, ..., x_n\}$, $Y = y$ the random field that is already defined previously w.r.t the graph $\mathcal{G}$. the couple (y, x) is said to be condional random field if, when every random variable y_i conditioned on x obey the markov property. That is :

$$P(y_i|x, y_{\{\mathcal{S}\setminus i\}}) = P(y_i|x, y_{\mathcal{N}_i}) \tag{1.23}$$

According the Gibbs-Markov equivalance, the Posterior is given by:

$$P(y|x) = \frac{1}{Z} exp\{-E(y, x)\}. \tag{1.24}$$

Assuming the only up to pairwise clique functionsare nonzero, the posterior distribution has the form :

$$P(y|x) = \frac{1}{Z} exp\{-\sum_{i\in\mathcal{S}} V_1(y_i, x) - \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{N}_i} V_2(y_i, y_j, x)\} \tag{1.25}$$

in the CRF literature (Lafferty et al., 2001), $-V_1$ and $-V_2$ are called the association and interaction Potentials respectively. Generally, these potentials are computed as a linear combination of some feature attributes extracted from the observation.

MRF Vs CRF

- MRF model is a generative model that the best configuration is passing by two steps
  - Infer likelihood $P(x|y)$ and the prior $P(y)$
  - The posterior distribution $P(y|x)$ is determined using the Bayes theorem.
- In accordance with the proposed model in (Lafferty et al., 2001) and (S. Kumar et al., 2003), CRF model is a discriminatice model :
  - Directly infer the Posterior $P(y|x)$.
  - The likelihood is due to the observation model $P(x|y)$. Usually, for tractability reasons (Besag, 1974) is assumed to have the factorized form:

$$P(x|y) = \prod_{i \in \mathcal{S}} P(x_i|y_i). \tag{1.26}$$

- CRF may be suitable for dealing with situations where the likelihood of an MRF is not of a factorized form such that all the $x_i(\forall i \in \mathcal{S})$ can explicitly exist in both unary and pairwise potentials.Where, in CRF, both Association and Interaction Potentials are functions of all the observation data as well as that of the labels :

– For a CRF

$$P(y|x) = \frac{1}{Z}\exp(\sum_{i\in\mathcal{S}} V_1(y_i, x) + \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{N}_i} V_2(y_i, y_j, x))$$

$\text{info}(y_i, x_{\mathcal{N}_i})$ $\text{info}(y_i, y_{\mathcal{N}_i})$ $\text{info}(y_i, x_{\mathcal{N}_i})$

– For a MRF

$$P(y) = \frac{1}{Z}\exp(\sum_{i\in\mathcal{S}} V_1(y_i) + \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{N}_i} V_2(y_i, y_j))$$

$\text{info}(y_i)$ $\text{info}(y_i, y_{N_i})$

The main difference between the CRF and MRF is that, in the CRF the unary potential is expressed in term of the observed data x i.e.info$(y_i, x_{\mathcal{N}_i})$ as well as of the label y_i, in an MRF i.e. info(y_i). Also, the pairwise potential for an CRF is expressed in the function of the observed data Thus we have two different information embeded in the pairwise term V_2 that one of them is the standard information for an MRF info$(y_i, x_{\mathcal{N}_i})$, while the second one is an information which changes with the change of the observed data info$(y_i, x_{\mathcal{N}_i})$. Consequently, unlike in an MRF, where $x_{i'}$ can influence y_i ($i \neq i'$) indirectly through the neighborhood system, in a CRF, this is done directly by the link between $x_{i'}$ and i (S. Z. Li, 2009).

CHAPTER 2

Classification and Prediction using MRFs

Contents

2.1 Introduction

As we reviewed in the first chapter, the both MRFs and CRFs represent an effective and theoretically well-established mathematical tool. The practical use of the MRFs and CRFs models is largely ascribed to the Hammersley–Clifford theorem which states the Gibbs and Markov random fields equivalence and further developed by (Besag, 1974). that model a given set of sites by expressing contextual information through an adequate energy functions. They allow passing from a global model to a model of local properties, defined in terms of both the potential function of single sites and the interactions potentials functions. For the above reasons and many others, the interest in MRF modeling in image and vision analysis and synthesis is increasing, as reflected by books as well as journal and conference papers where its have been widely employed to solve different problems on image processing such as image restoration (S. Geman & Geman, 1987) (Greig, Porteous, & Seheult, 1989) (Berthod, Kato, Yu, & Zerubia, 1996), also The segmentation was treated in this research papers (Derin, Elliott, Cristi, & Geman, 1984), (S. Geman & Graffigne, 1986) and (Kohli & Torr, 2005). while some researchs a represent different applications of MRF in computer vision(D. Geman & Gidas, 1991) and (S. Z. Li, 1994).

So, as our topic is the classification of imagery, we need to introduce the classification framework. the classification problem is a set of labels assigned to image pixels or features. In the rest of section, some definitions and notations will be used throughout the thesis.

2.1.1 Labeling Problem

Let $\mathcal{S}$ index a discrete set of n sites and their elements are indices: $\mathcal{S} = \{1, ..., n\}$. Usually, a set reptresent a region or a points in euclidean space such as an image pixel or image feature (corner point, a line segment, object detection). the definition of the set $\mathcal{S}$ is

changing by the space dimension that the set in 1-dimension space different totally with a set in 2-dimension space :

- For a set defined in 1-dimensional space :

$$\mathcal{S} = \{1, ..., n\}. \tag{2.1}$$

- For a set defined in 2-dimensional space :

$$\mathcal{S} = \{(i, j)|1 \leq i, j \leq n\} \tag{2.2}$$

For an $N \times N$ image, the pixel in the site (i, j) is conveniently reindexed by a single index m (double indicator) where m take a values in $\{1, ..., n\}$ with $n = N \times N$. In this work, We will adopt this notation for indexing an image.

Let's define the set of labels that denoted by $\mathcal{L}$ where avery label (element) from $\mathcal{L}$ represent an event that may happen to a site. The set of labels consists from a continuous or discrete labels. In the continuous case, the label set is defining as real line $\mathbb{R}$ or a part of it (compact interval) Eq.2.3 . In the discrete case, a label takes a discrete value $\mathcal{L} = \{l_1, ..., l_k\}$ with all og the labels are element of $\mathbb{N}$ Eq.2.5 :

$$\mathcal{L}_c = [a, b] \subset \mathbb{R} \tag{2.3}$$

$$\mathcal{L}_d = \{l_1, ..., l_k\} \tag{2.4}$$

Or

$$\mathcal{L}_d = \{1, ..., K\} \subset \mathbb{N} \tag{2.5}$$

Now, we called a labeling of $\mathcal{S}$ the following set $y = \{y_1, ..., y_n\}$ in terms of the labels in $\mathcal{L}$. The terminology 'labeling' is equivalance to configuration especially for random fields. We sammurize the problem of image labeling as follow :

- Given
  - An observed data from input image $x = \{x_1, ..., x_n\}$

  - Set of label $\mathcal{L}$ which is already defined.

- The best corresponding configuration or labeling to input image $y = \{y_1, ..., y_n\}$ is finding by maximise the conditional probability, namely also the posterior distribution :

$$\widehat{y} = \underset{y}{argmax} P(y|x) \tag{2.6}$$

The problem of maximizating the posterior is known as the MAP estimate. The MAP framework (MAP-MRF) is supported through many researchs papers (Ganan & McClure, 1985), (S. Geman & Geman, 1987),(Greig et al., 1989),(S. Kumar et al., 2003) and (M. P. Kumar, Kolmogorov, & Torr, 2009).

An equivalently solution to MAP solution is the minimizing the energy function of an MRF i.e. Maximizing a posterior probability $P(y|x)$ is equivalent to minimizing the posterior energy function $E(y|x)$ according to the relationship between them :

$$\widehat{y} = \underset{y}{argmax} E(y|x) \tag{2.7}$$

2.1.2 Summary of MAP estimate

The piplines of MAP solution are sammurized as follow :

(1) Defin the neighborhood system $\mathcal{N}$.

(2) Define the Clique function V_c

(3) Define the prior probability (Prior energy $E(y)$ if we are dealling with energy function).

(4) Derive the likelihood (or the likelihood energy $E(x|y)$)

(5) Maximize the posterior probability (or the mininizing the posterior energy $E(y|x)$).

The derive form posterior distribution and detemine the parameters of the model are two essential steps To define completely the posterior distribution. And finally we need to adopt an optimization algorithms for finding the maximum of the posterior distribution .

2.2 Parameters estimation

In this section, we will show how statistical criteria are intertwined with our topic exactly in the parameters estimation. The parameters estimation have an important performance because any probabilistic distribution are recognised through two elements which are : the form of the function and the parametrs involved. Historically, the existing techniques usually used are : Maximum Likelihood (ML), Pseudo-Likelihood (PL), Expectation- maximization that the estimating parameters has a much shorter history. The next subsections review the techniques ML and PL to estimate the parameters.

2.2.1 Maximum Likelihood

Denote the parameters set by θ, and a realization y of an MRF (true labeling or ground truth). The underlying idea of maximum likelihood is estimating the maximum of the conditional probability $P(y|\theta)$ which is the likelihood function of θ, i.e.

$$\widehat{\theta} = \underset{\theta}{argmax} P(y|\theta). \tag{2.8}$$

According to the bayes rule, the posterior density over the parameters is given by :

$$P(\theta|y) \propto P(y|\theta)p(\theta). \tag{2.9}$$

In this step, we distinguish two cases :

- If the prior probability function of the parameters $p(\theta)$ is known, the MAP estimation that maximizes the posterior Eq.2.9 can be sought.
- If the prior information is totally unavailable, the prior is assumed to be flat, thus the MAP estimation is equivalent to the maximum likelihood.

In the general way, the parameters are deducing through a necessary condition which is $\frac{\partial P(y|\theta)}{\partial \theta} = 0$ where maximizing the likelihood and its log-likelihood are equivalents. For a gibbs distribution, the likelihood is :

$$P(y|\theta) = \frac{1}{Z_\theta} exp\{-E(y,\theta)\} \tag{2.10}$$

Since, the partition functio, Z_θ is calculated by summing over all the possible configurations of y Eq.1.5 that, maximizing $P(y|\theta)$, generally, becomes intractable even for small problems because there are a combinatorial number of elements in the configuration space and this is the main difficult in the estimation of parameters.

2.2.2 Pseudo-Likelihood

The most advantage of this technique is that the partition function Z (Z_θ) does not appear within maximization of posterior density $P(y|\theta)$. The pseudo-likelihood is expressed as a product of the conditional likelihood functions :

$$\begin{aligned} PL(y|\theta) &= \prod_{i \in \mathcal{S}} P(y_i|y_{\mathcal{N}_i}, \theta) \\ &= \prod_{i \in \mathcal{S}} \frac{exp\{-E_i(y_i, y_{\mathcal{N}_i})\}}{\sum_{y_i} exp\{-E_i(y_i, y_{\mathcal{N}_i})\}} \end{aligned} \tag{2.11}$$

That $E_i(y_i, y_{\mathcal{N}_i})$ is defined by

$$E_i(y_i, y_{\mathcal{N}_i}) = \alpha V_1(y_i) + \sum_{j \in \mathcal{N}_i} \beta V_2(y_i, y_j) \tag{2.12}$$

Consequently, we can obtain $\theta = \{\alpha, \beta\}$ by solving the

$$\frac{\partial P(y|\theta)}{\partial \alpha} = 0 \qquad\qquad \frac{\partial P(y|\theta)}{\partial \beta} = 0 \tag{2.13}$$

2.3 The Maximum A posteriori-MAP- Framework

As aforemebtioned in previous chapter, the optimization problem related on MRF can expressed as minimization of an energy function associeted with the MRF (Szeliski et al., 2008). That many works proposed differents methods or algorithms, where some algorithms resulting a global solution as Graph cut (Greig et al., 1989),(Roy & Cox, 1998), and (Boykov & Kolmogorov, 2001), while others obtaining the local solution as Iterated Conditional Mode (ICM) (Besag, 1986), noting that the ICM is maximazing the local conditional probability.

In the next, we introduce briefly those two above methods.

2.3.1 Iterated Conditional Mode (ICM)

Given an initial solution, the ICM maximizes the local conditional probabilities iteratively :

$$\widehat{y_i} \leftarrow \underset{y_{ki}}{argmax} P(y_{ki}|y_{\mathcal{N}_i}, x) \tag{2.14}$$

This algorithms is adopted from several works thanks to the reasonably results obtained in (S. Geman & Geman, 1987) , (S. Kumar et al., 2003) , (Lempitsky, Rother, Roth, & Blake, 2010) and (Zeggada, Benbraika, Melgani, & Mokhtari, 2018)

Algorithm

1.Input : NodePot, EgdePot, done=0;

2. $y \leftarrow max(NodePot)$ % Initialization of solution

3. **while** $\sim$ done **do**

4. done $\leftarrow 1$;

5. **for** each node i **do**

6. $pot \leftarrow \psi_i$;

7. **for** each edge related to i **do**

8. pot $\leftarrow pot \times \psi_{ij}$;

9. **end for**

10. $(new - y)_i \leftarrow max(pot)$;

11. **if** $(new - y)_i \sim= y_i$ **do**

12. $y_i \leftarrow (new - y)_i$; $done \leftarrow 0$;

13. **end if**

14. **end for**

15. **end**

2.3.2 Graph cuts

Graph cuts is algorithms which obtain the global minimum of a binary MRF model expressed as a discrete energy minimization problems. this algorithm is based totally on the theorem of equivalence between the value of maximum flow and the capacity of the minimum cut in which this theorem was proven by Ford and Fulkerson in 1954 for undirected graphs and 1955 for directed graphs. The first proposed of this algorithm is obtained by (Greig et al., 1989) for a global minimum solution of a two-label MRF model (Ising model). In order to solve convex multilabel problems, the algorithm proposed have been extended by (Roy & Cox, 1998) and an approximate global solution have been proposed for general multilabel MRF problems by (Boykov, Veksler, & Zabih, 2001).For an binary image, the main idea of The min cut(maxflow) algorithms is divide the set pixels of the considered image into two subsets. that all the nodes or pixels belongs to one of those two subsets having the same label.

CHAPTER 3

Multilabel CRFs model for classification

Contents

3.1 The Problematic

The multilabel classification is a challenging task in machine learning. that the multilabel image is represented as a series of binary images namely classes, in which each class reflect the state of an predefined object in considered image. In order to classify the input image we need to re-indexed the image taking into consideration the successive classes maps Fig. 3.1.

Let' condider that the number of classes is c indexing by k that means, if we denote the considered image by x we have :

$$X = \{X_1, ..., X_k, ..., X_c\} \tag{3.1}$$

where X_k refer to k^{th} class map. Every class of them indexed by the set of site $\mathcal{S}$. By other words, every node (pixel) is expressed by a multilabel description vector i.e. $X_i = \{X_{1i}, ..., X_{ki}, ..., X_{ci}\}$

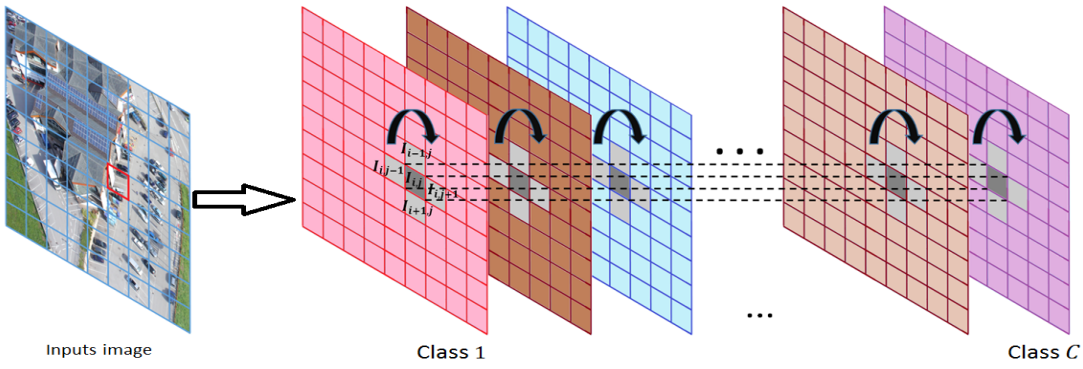


Figure 3.1: Multilabel representation ok an image

Under the multilabeling representaion lead us to test the hypothesis of the correlation between classes maps considering a CRF model .

3.2 CRF model for multilabeling inputs

Unlike MRFs that are generative frameworks which regularize the classification output of an image by enforcing priors assumption between neighboring sites, discriminative CRFs are globally conditioned on the observations X, where they directly model the posterior distribution as a Gibbs model. This means that the potential functions in CRFs are more flexible in capturing complex spatial dependencies between labels (Lafferty et al., 2001). The posterior distribution of Y_k conditioned on X_k with $k \in \mathcal{C} = \{1, ..., C\}$ is defined as:

$$P(Y_k|X_k) = \frac{1}{Z} exp\{-E(Y, X)\}. \tag{3.2}$$

According to 2.1.1 and 2.1.2 , maximizing the a posteriori probability (MAP) of the whole image is equivalent to minimizing its corresponding energy function $E(Y, X)$. This latter is expressed as sum of unary and pairwise terms:

$$E(Y_k, X_k) = E_{data}(Y_k, X_k) + E_{spatial}(Y_k, X_k) \tag{3.3}$$

where

- E_{data} represents the local decision term. It computes the cost of associating a given tile to a certain class without including its neighbors.
- The second term $E_spatial$ is a smoothness term. It imposes spatial smoothness by penalizing dissimilarities between adjecent labels.

In dealing with multilabeling imagery 3.1 , the spatial information integrate two types of information 1) the traditional spatial information over each classification map X_k jointly with 2) Latent spatial information between neighboring classes within same site i. Tat by changing the interaction potential term from a pairwise term into a ternary one. According to the Hammersley-Clifford theorem (Besag, 1974) , for each map k, we define a CRF over the outputs$Y_k$given the inputsX_k, through the following posterior distribution:

$$P(Y_k|X_k) = \frac{1}{Z} exp\{\sum_{i=1}^{n} V_1(y_{ki}, X_k) + \sum_{i=1}^{n} \sum_{(l,j)\in\mathcal{N}_{k,i}} V_2(y_{ki}, y_{kj}, y_{li}X_k)\} \tag{3.4}$$

where V_1 is the unary potential and V_2 is the interaction term expressed as a trinary-wise potential that depends on 2 types of information. The first one is the traditional spatial information of neighboring tiles within the same class map k . The second one is the information obtained from the neighborhood classes of k^{th} class. We are integrating the aboves spatial information using an neighborhood system defining by:

$$\mathcal{N}_{k,i} = \{(l, j) \in \mathcal{C} \times \mathcal{S} || i - j| = 1, |l - k| = 1\} \tag{3.5}$$

Its neighboring system dependence on two indicators, the first one k refers to adjecents classes within same sites i while the second one i refers the neighboring sites within same class. thus, we can rewrite it as union of two system : $\mathcal{N}_{k,i} = \mathcal{N}_k^i \cup \mathcal{N}_i^k$. consequently, the Eq.3.4 becomes :

$$P(Y_k|X_k) = \frac{1}{Z} exp\{\sum_{i=1}^{n} V_1(y_{ki}, X_k) + \sum_{i=1}^{n} \sum_{j\in\mathcal{N}_i^k} \sum_{l\in\mathcal{N}_k^i} V_2(y_{ki}, y_{kj}, y_{li}X_k)\} \tag{3.6}$$

where

- V_1 is the unary potential andV_2 is trinary-wise potential (interaction potential) that depends on 2 types of informations mentioned earlier. The first one is the traditional spatial information of neighboring sites within the same class map k using the neighboring system $\mathcal{N}_i^k$. The second one is the information obtained from the neighborhood classes within site i as shown on the multilabeling image representation Fig.3.1.
- As we explained previously, we consider the first order neighborhood system :$N_i^k = \{j, |i - j| = 1\}$ is the neighboring system defining over each class map, and $N_k^i = \{l, |k - l| = 1\}$ is the neighborhood system defining for capturing the adjecent classes. That, the neighborhood system N_i^k have no directly relationship with the system (N_k^i).

Therefore, we tackle each of these information sources separately becauce the separability of neighborhood systems over classes(N_k^i) and on level of each class N_i^k. We thus quantify V_2 as a sum of two interaction terms, the first one over each map (spatial information) and the second one across the maps to encode the interaction between the neighboring classes maps within the same site. Under these simplifying assumptions, the interaction potentialV_2 in the posterior $P(Y_k|X_k)$ can be written as:

$$P(Y_k|X_k) = \frac{1}{Z} exp\{\sum_{i=1}^{n} V_1(y_{ki}, X_k) + \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i^k} V_2(y_{ki}, y_{kj}, X_k) + \sum_{i=1}^{n} \sum_{l \in \mathcal{N}_k^i} V_2(y_{ki}, y_{li} X_k)\} \tag{3.7}$$

Now, for simplify the proposed model, we adopt the CRF representation in (Schmidt, Murphy, Fung, & Rosales, 2008) and after adding the cross-correlation term the Eq.3.7 becomes :

$$P(Y_k|X_k) = \frac{1}{Z} \prod_{i} \psi_i(y_{ki}, X_k) \prod_{(i,j)} \psi_{ij}(y_{ki}, y_{kj}, X_k) \prod_{(i,l)} \psi_{il}(y_{ki}, y_{li}, X_k) \tag{3.8}$$

where, $\psi_i, \psi_{ij}, \psi_{il}$ are the functions of node potential, edge potential and the cross-correlation potential respectively. Given $y_{ki} \in \mathcal{L} = \{0, 1\}$ that '0' refer to absence of object and '1' refer to the presence of object, the potential functions are defining as :

- The node potenital :

$$\psi_i(:, X_k) = (exp(v_{ki,1} x_{ki}, exp(v_{ki,2} x_{ki}) \tag{3.9}$$

  where x_{ki} is the node feature in the k^{th} class and the i^{th} site. $v_{ki} = \{v_{ki,1}, v_{ki,2}\}$ are the associated weights in the different label state $\{0$ or $1\}$.

- The edge potential :

$$\psi_{ij}(:, :, X_k) = \begin{pmatrix} exp(w_{kij,11} h_{kij}) & exp(w_{kij,12} h_{kij}) \\ exp(w_{kij,21} h_{kij}) & exp(w_{kij,22} h_{kij}) \end{pmatrix}$$

where where h_{kij} being an arbitrary decreasing function that takes values in $[0, 1]$, in this purpose, tis takes the form $h(x_{ki}, x_{kj}) = \frac{1}{1 + |x_{ki} + x_{kj}|}$. $w_{kij} = \{w_{ki,11}, w_{ki,12}, w_{ki,21}, w_{ki,22}\}$ are the associated edge weights.

- The cross-correlation potential:

$$\psi_{il}(:, :, X_k) = \begin{pmatrix} exp(\beta_1 h_{kil}) & exp(\beta_0 h_{kil}) \\ exp(\beta_0 h_{kil}) & exp(\beta_1 h_{kil}) \end{pmatrix}$$

where $\{\beta_0, \beta_1\}$ are the interaction coefficient between classes that β_1 is defining for the sites which have same lebel while β_0 is nulle . h function is the same one using for edge features.

In order to complete the posterior propability definition we need to estimate th node and adge weights. The Loopy Belief Propagation (LBP) method (Pearl, 1982) which discrebed later.

for the optimization related problem, we adopt the iterated conditional modes (ICM) algorithm. Given an initial solution, this algorithm maximizes the local conditional probabilities iteratively. The label that maximizes the local conditional probability is chosen as an optimal local solution (Besag, 1986). In this model, the local probability of each site is conditioned by the labels of adjecent classes over the same site, in addition to the corresponding neighboring sites over the same label map level. Starting from an initial multilabel combination generated, at each iteration, the ICM maximizes the conditional MAP estimation:

$$widehaty_{ki} \leftarrow \underset{y_{ki}}{argmax} P(y_{ki} | y_{k\mathcal{N}_i^k}, y_{k\mathcal{N}_k^i}, X_k) \tag{3.10}$$

3.3 Multilabel CRF Description

as aforementioned, our purpose is integrate the cross-correlation In wich this potential between labels can serve as an additional important source of information to build a ro-

bust classification framework with a strong capability of penalizing the co-occurrence of uncorrelated labels. According to the Hammersley-Clifford theorem (Besag, 1974) , for each map k, we define a CRF over the outputs $Y_k$given the inputs X_k, through the following posterior distribution:

$$P(Y_k|X_k) = \frac{1}{Z} exp\{\sum_{i=1}^{n} V_1(y_{ki}, X_k) + \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} \sum_{l \in c_{ki}} V_2(y_{ki}, y_{kj}, y_{li} X_k)\} \tag{3.11}$$

where

- V_1 is the unary potential, which is the probabilistic output of our discriminative classifier when considering each tile in isolation
- and V_2 is the interaction term expressed as a trinary-wise potential that depends on 2 types of information. The first one is the traditional spatial information of neighboring sites within the same class map $k$using the$\mathcal{N}_i$ neighborhood system for site i. The second one is the information obtained from the neighborhood c_{ki} of i, namely, the multilabel components of the binary vector of site i). That we are defining the both of $\mathcal{N}_i$ and c_{ki} as follow :
  - $\mathcal{N}_i$ is the neighboring system which encapsulate the adjecent nodes of the site i without takinf in consideration the class under process.
  - $c_{ki} = \{(l, i), l \in \mathcal{C} - k, i = \overline{1, n}\}$

In this work, for simplicity, we consider the first order neighborhood system for the traditional spatial information and the labels belonging to the same site (i.e., multilabel vector descriptor of one site) for the cross-correlation information. Moreover, we tackle each of these information sources separately. We thus quantify V_2 as a sum of two interaction terms, the first one over each class map (spatial information) and the second one

across the classes maps to encode the cross-correlation between the multiple labels lying within the same site. Under these simplifying assumptions, the interaction potential V_2 in the posterior $P(Y_k|X_k)$ can be written as:

$$P(Y_k|X_k) = \frac{1}{Z} exp\{\sum_{i=1}^{n} V_1(y_{ki}, X_k) + \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} I_1(y_{ki}, y_{kj}, X_k) + \sum_{i=1}^{n} \sum_{l \in \mathcal{C}_{ki}} I_2(y_{ki}, y_{li} X_k)\} \tag{3.12}$$

where I_1 is the interaction function at the level of each class map (spatial term), and I_2 is the cross-correlation function, $c_k i$. According to the CRF representation in (Schmidt et al., 2008) , and after adding the cross-correlation term, 3.12 becomes:

$$P(Y_k|X_k) = \frac{1}{Z} \prod_{i} \psi_i(y_{ki}, X_k) \prod_{(i,j)} \psi_{ij}(y_{ki}, y_{kj}, X_k) \prod_{(i,l)} \psi_{il}(y_{ki}, y_{li}, X_k) \tag{3.13}$$

The terms $\psi_i$and ψ_{ij} are respectively the node and the edge potential functions over the class map k, whileψ_{il} is the cross-correlation function (i.e., edge potential through different labels). In the following, we define each of the potential functions, given thaty_{ki} takes a binary state, $y_{ki} \in \mathcal{L} = \{0, 1\}$. The node potential takes the following form:

$$\psi_i(:, X_k) = (exp(v_{ki,1}^t f_{ki}), exp(v_{ki,2}^t f_{ki}) \tag{3.14}$$

where $f_{ki} = [1, x_{ki}]$ and $v_{ki} = v_{ki,1}, v_(ki, 2)$ are respectively, the node features and their associated weights. The traditional edge potentials are defined as:

$$\psi_{ij}(:, :, X_k) = \begin{pmatrix} exp(w_{kij,11}^t f_{kij}) & exp(w_{kij,12}^t f_{kij}) \\ exp(w_{kij,21}^t f_{kij}) & exp(w_{kij,22}^t f_{kij}) \end{pmatrix}$$

where f_{kij} are the edge features, defined as: $f_{kij} = [1, h(x_{ki}, x_{kj})]$, and $w_{kij} = \{w_{kij,11}, w_{kij,12}, w_{kij,21}, w_{kij,2}$ being the weights associated with edges. while with h being an arbitrary decreasing function that takes values in $[0, 1]$, that in our model it is taking the following form :

$$h(x_{ki}, x_{kj}) = \frac{1}{1 + |x_{ki} - x_{kj}|} \tag{3.15}$$

Regarding the correlation potentials ψ_{il} , we define first an auxiliary function $g_{kl} : \mathcal{L}^n \times \mathcal{L} \rightarrow [0, 1]$ that measures the co-occurrence probability of two labels y_{li} and y_{ki} as:

$$g_{kl}(y_{l.}, y_{ki}) = \frac{1}{|\{y_{k.} = y_{ki}\}|} \sum_{i=1}^{n} \mathbf{1}_{y_{ki}}(y_{li}) \tag{3.16}$$

the notation $|\{y_{k.} = y_{ki}\}|$ means the cardinal of the set of all nodes in the class map k that having the same label y_{ki} with :

$$l \in \{1, ..., C | l \neq k\} = c_{ki} \tag{3.17}$$

Note that g_{kl} is defined when y_{li} and y_{ki} holds the same label value (i.e., 0 or 1), for the case where y_{li} and y_{ki} have different labels, we set $\overline{g_{kl}}$ as:

$$\begin{aligned} \overline{g_{kl}} &= \overline{g_{kl}}(y_{l.}, y_{ki}) \\ &= 1 - g_{kl}(y_{l.}, y_{ki}) \end{aligned} \tag{3.18}$$

The values of g_{kl} and $\overline{g_{kl}}$ are estimated from the training data to assess the correlation degree between label classes.

Once g_{kl} and $\overline{g_{kl}}$ are computed, the cross-correlation weights for a given test data at location$i$, are defined as:

$$\mu_{ki}(y_{.i}, y_{ki}) = \sum_{l \in c_{ki}} \mathbf{1}_{y_{ki}}(y_{li}) g_{kl} + [1 - \mathbf{1}_{y_{ki}}(y_{li})] \overline{g_{kl}} \qquad s.t \quad y_{ki} = 1. \tag{3.19}$$

$$\tau_{ki}(y_{.i}, y_{ki}) = \sum_{l \in c_{ki}} \mathbf{1}_{y_{ki}}(y_{li}) g_{kl} + [1 - \mathbf{1}_{y_{ki}}(y_{li})] \overline{g_{kl}} \qquad s.t \quad y_{ki} = 0. \tag{3.20}$$

The functionsμ_{ki} andτ_{ki}define the cross-correlation weights for the two possible states of labely_{ki}, namely, presence and absence respectively by using the sum of the label pairs scores to evaluate the degree of the relationship of a given test site. Subsequently, the correlation potential is given as:

$$\psi_{il}(:, y_{c_k i}) = (exp(\lambda_1 \mu_{ki}(1, y_{.i})), exp(\lambda_1 \tau_{ki}(0, y_{.i}))) \tag{3.21}$$

where $\lambda_1$and λ_2 are coefficients used to ponder the importance between the two states that a given label may take i.e., presence and absence, respectively.

In order to obtain an optimum multilabeling over a given test image, the use of an inference algorithm is necessary in order to get an estimation of the MAP solution. For such purpose, we make use of the loopy belief propagation (LBP) method (Pearl, 1982), which is a widely used learning a graphical model in approximation algorithms (Lempitsky et al., 2010). As mentioned earlier, we adopt the iterated conditional modes (ICM) algorithm. This algorithm maximizes the local conditional probabilities iteratively, given an initial labeling. The label that maximizes the local conditional probability is chosen as an optimal local solution.

In this work, the local probabilities of each tile are conditioned by its multilabel vector descriptor in addition to the corresponding neighboring tiles over the same label map level. Starting from an initial multilabel combination generated, at each iteration, the ICM maximizes the conditional MAP estimation:

$$\widehat{y_{ki}} \leftarrow \underset{y_{ki}}{argmax} P(y_{ki} | y_{k\mathcal{N}_i}, y_{c_k i}, X_k) \tag{3.22}$$

passing through all label maps of the test image repeatedly up to a convergence is reached, producing thus a final multilabel classification map of the considered image.

CHAPTER 4

Experimental Results

Contents

4.1 Introduction

The Unamend Aerial Vehicles abbreviate UAVs are commonly known as Drones. Mainly, they were created for military usage, but the increasing interest on it in the last decade has paved the way to be a very promising and effective technology through their exploitation in numerous civilian applications, such as emergecy and surveillance and urban monitoring. Besides that their advantages exhibited through the acquired images that are characterized by the Extremely High Resolution (EHR). So, the wide range of uses out in light many studies have been proposed in the past few years on the basis of machine learning tools such as deep neural networks and pattern matching. It is well known that EHR imagery contains complex structure. Since typical ad-hoc approaches often fail to make reasonable classification because of these kinds of complexity, for that we propose to evalute the effectiveness of the proposed models were described earlier. As the multilabeling image is consisting of finite number of classes that every classses reflect the state of one object see the Fig.4.1 . Since, in the dealing with multilabel images, one may observe that: some objects are more likely to coincide with others in many image, and some objects are rarely to appear with one another. For instance, “solar panel” labels are frequently correlated with “roof” labels compared by “tree” labels, because obviously solar panels require flat places to absorb much sunshine or, the label ‘car’ is strongly correlated with the label ‘asphalt’, on the contrary, it is seldom that the label ‘car’ appears together with label ‘railway’ in the same image tile. Consequently, the cross-correlation between labels (objects) can serve as an additional important source of information that makes the result smuch more reasonable and fruitful. for this purpose, the numerical experiments performed on two differents UAV data sets are represented and descibed in the nexet section.

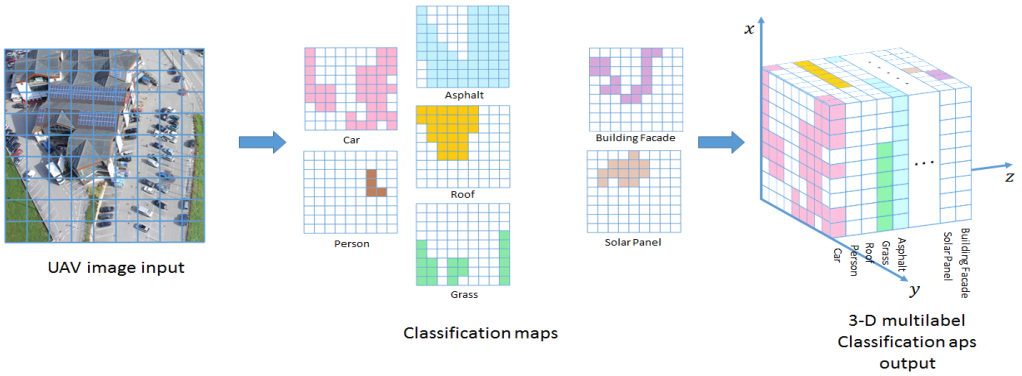


Figure 4.1: An example of inpouts/outputs of the proposed modes

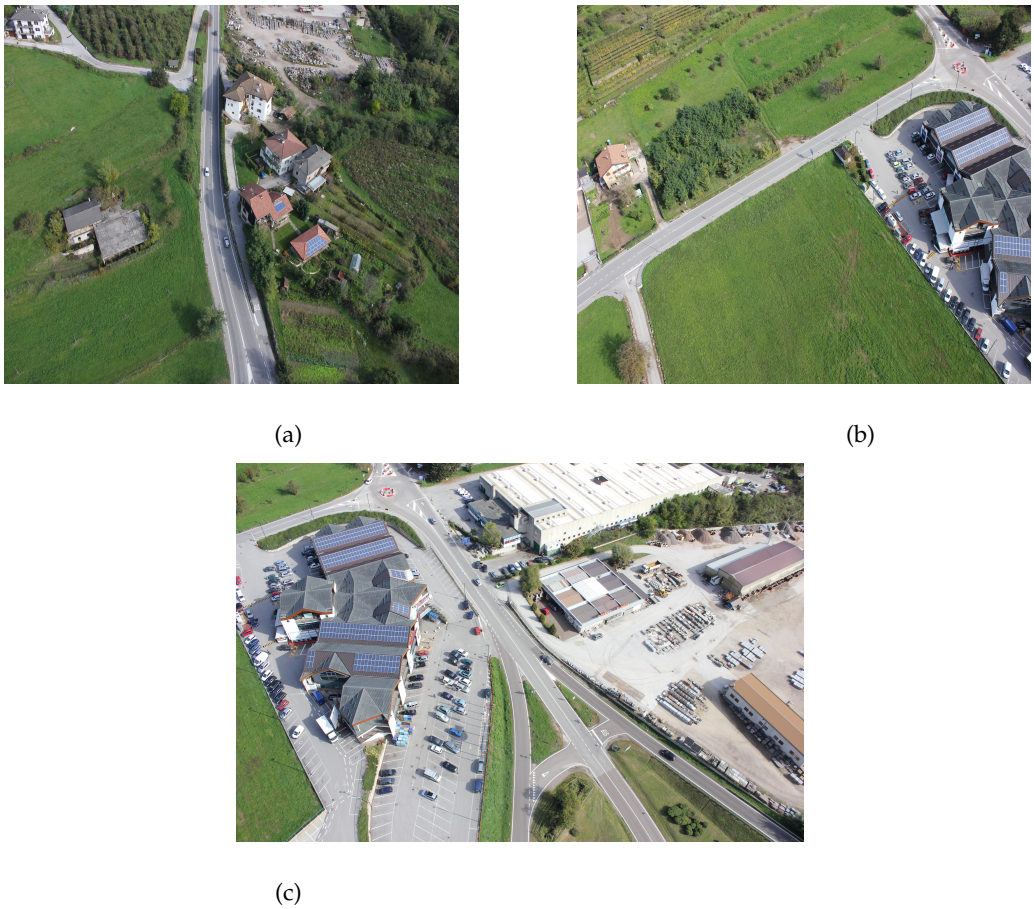
4.2 Datasets Description

In order to evaluate the performance of the proposed classification method, we exploited two real datasets of UAV images acquired over two different locations. The first set of images was taken over the Faculty of Science of the University of Trento (Italy) Nadir acquisition on the 3rd October 2011. The second set of images was acquired near the city of Civezzano (Italy) at different off-nadir angles, on the 17th October 2012. Both acquisitions were performed with a picture camera Canon EOS 550D characterized by a CMOS APS-C sensor with 18 megapixels. The UAV images are characterized by three channels (RGB) with a spatial resolution of around 2 cm. The image size is 5184×3456 pixels and the radiometric resolution is 8 bits for both datasets.

The both of datasets are composed of 10 images, subdivided into two groups:

- **Training set**: Three images are selected as training 4.2 which chosen from overall set in such a way they contain all predefined classes of objects. The training data is used to learn the parameters of the proposed models.
- **Test set**: we have seven images in this set for testing the validation of proposed model 4.3.

For the images of first data, they contain 13 classes of objects, namely, 'Asphalt',



(a) (b)

(c)

Figure 4.2: Training images

'Grass', 'Tree', 'Vineyard', 'Pedestrian Crossing', 'Person', 'Car', 'Roof 1', 'Roof 2', 'Solar Panel', 'Building Facade', 'Soil' and 'Shadow'.

While the second dataset includes 14 classes, which are 'Asphalt', 'Grass', 'Tree', 'Vineyard', 'Low Vegetation', 'Car', 'Roof 1', 'Roof 2', 'Roof 3, 'Solar Panel', 'Building Facade', ' Soil', 'Gravel', and 'Rocks'.

4.2.1 Inputs Construction

The pipeline of the framework consists of two main phases. First, the considered input UAV image is subdivided into a grid of tiles, which are processed thanks to an opportune representation thus tile-wise multilabel prediction probabilities. In the second phase, a multilabel CRF model is applied to integrate spatial correlation between adjacent tiles

Figure 4.3: test

and the correlation between labels within the same tile.

In the first phase, starting by subdivide the query image (RGB) EHR image (I) acquired by means of a UAV into a grid of tiles whose size depends on the image resolution and the expected sizes of objects to recognize. For the both datasets, all the tiles have size 50×50 pixels. Then, the multilabel tile-based approach has started with the extraction of features, that the multilabel tile-based classification described in (?, ?) has been highly paid attention especially with interesting result obtained by (Zeggada & Melgani, 2016). Due to the EHR imagery, each tile is characterized by a high level of detail and thus rich information content. In order to extract a compact signature that describes efficiently each tile, the well-known strategy, namely, the Bag of Visual Words (BOW) representation is adopted. Then, an Auto Encoder neural network (AE) is applied to enhance the features discrimination capability(Zeggada & Melgani, 2017). We feed the extracted features to an AE network which constructs new learned features. The next step consists in adding a multilayer perceptron (MLP) network(MacKay, 2003), as a classifier at the end of the encoding part in order to classify the resulting features. This classifier fits the mul-

tilabeling requirements, and thereafter can be used on the test tiles to infer their object lists. Indeed, the MLP can handle simultaneously multiple outputs which may characterize each tile of the query image. The number of the MLP outputs C corresponds to the number of predefined object classes. The resulting MLP outputs generate what is called the classification maps X_k for $k \in 1, \ldots, C$, such that $X_k = x_{ki} | i = 1, \ldots, n$. These maps are inferred from the posterior probability distribution provided by the MLP outputs. In each classification map X_k, the presence/absence of object k is indicated for each tile x_{ki}. In other words, each tile x_i is associated with a multilabel descriptor vector of size C, defined as $x_i = (x_{1i}, x_{2i}, \ldots x_{Ci})$.

The following sections reports the numerical results of the two proposed methods, descibed in previous chapter, with consideration our intputs the classification maps X_k.

4.2.2 Accuracy Measures

The effectiveness of the proposed framework is evaluated by the two well kwon accuracy measures, namely, the sensitivity, the specificity and their average defined as follows:

- **the sensitivity**: $SEN = \dfrac{TP}{TP + FN}$
- **the specificity**: $SPE = \dfrac{TN}{TN + FP}$
- **the average**: $AVG = \dfrac{SEN + SPE}{2}$

Clarifying the annotations mentioned in the two first equations are given in the Table.4.1

		Estimated Label	
		0	1
True	0	TN	FP
label	1	FN	TP

Table 4.1: Confusion Matrix for the computation of the SEN, SPE and AVG accuraries

4.3 Numerical Results

In order to complete our experimental assessment, we compered our method with others. The first one is that thresholding of resulting maps obtained by bag-of-words strategy coupled with autoencoder network and MLP classifier as described earlier (termed as ML-Unary). The second one is the traditional monolabel CRF reference method (Eq.1.25) which was run on each binary map independently from the others, without taking into account the multilabel context and we refer to this method by CRF. we added an bias as feature beside to G-functions(Schmidt et al., 2008) on the standard CRF and its refered by CRF-bias. The CRF-ML refers to the proposed method which exploit the importing information from neighbouring classes as described previously. The Full-ML-CRF stands for the proposed method which exploits cross-correlation between all defined classes for multilabeling image proposed by (Zeggada et al., 2018). The next paragraphs discuss the results, which are summarized in Table.4.2 and Table.4.3, for each data set separately.

	Accuracies (%)		
Methods	SPE	SEN	AVR
ML-Unary	97.1	47.9	72.5
CRF	79.9	64.5	72.22
CRF-bias	86.8	53.4	70.1
CRF-ML	**75.04**	**75.03**	75.03
Full-ML-CRF	82.2	70.3	**76.2**

Table 4.2: SEN, SPE and AVG accuracies in percent obtained by the different classification methods on first datasets.

	Accuracies (%)		
Methods	SPE	SEN	AVR
ML-Unary	98.2	62.6	80.4
CRF	84.42	81.27	82.8
CRF-bias	92.5	70.6	81.5
CRF-ML	**82.32**	**82.35**	82.34
Full-ML-CRF	90.8	75.9	**83.4**

Table 4.3: SEN, SPE and AVG accuracies obtained by the different classification methods on second datasets.

In the following subsections, the results of our approachs are reported and discussed.

4.3.1 the results of CRF approach for Multilabel classification

Recall the underlying idea of this approach. According to the multilabeling representation, we develop the monolabel CRF reference method to express the neighboring labels information beside with tradition contextual information. In other words the proposed model integrates: 1) the spatially neighbouring information between adjacent tiles within the same class, jointly with, 2) cross-correlation information between neighbouring class labels within the same tile as shown in the Fig.4.4. Subsequently, applying the Iterated Conditional Modes (ICM) (Besag, 1986) algorithm to solve the related optimization problem.

In the following paragraphs, the results of our approach are reported and discussed. In particular, we will first look at the different accuracy values for $\beta \in [0, 4]$ in which they are reporting in Fig.4.5 and Fig.4.6 for dataset 1 and 2 respectively that the both graphs reported using the MATLAB platform.

Since a certain point $\widehat{\beta}$, the SEN and the SPE will be balanced and stabilized with outper-

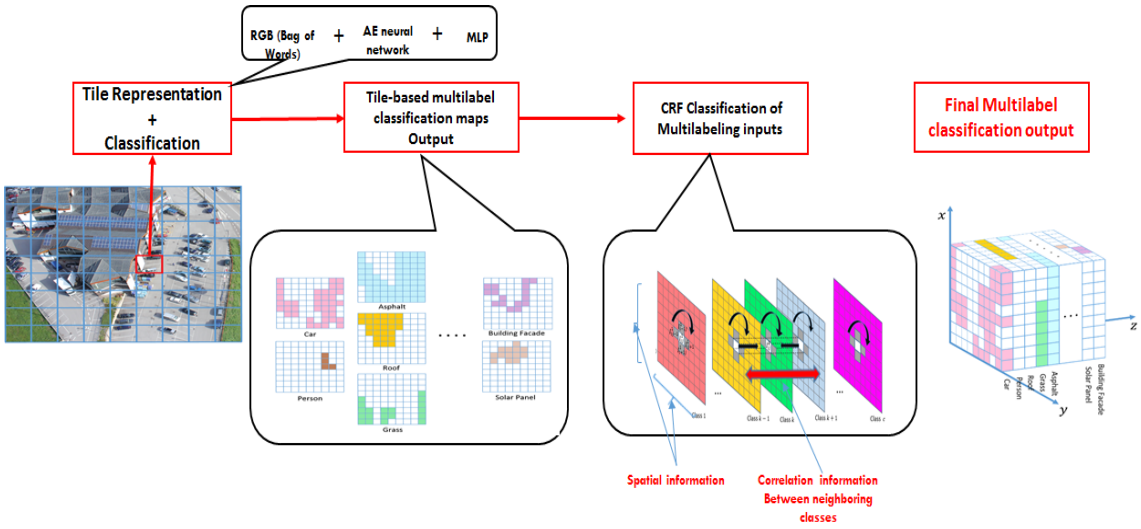


Figure 4.4: Flowchart of the proposed CRF clasification method for ML inputs

form of SEN i.e. we have the advantage of correct labeling after this point $\widehat{\beta}$, in which $\widehat{\beta} = 1.6$ for the first data set Fig.4.5 and $\widehat{\beta} = 2.2$ for the second data set Fig.4.6Unlikethe average accuracy concerned second data, which recorded a smal rate of decreasing while β is increasing, the average accuracy of first data is increasing simultaneously with β. but for the both data we note that this divergence turn into stability over all the accuracy measures after specifying the $\widehat{\beta}$ value.

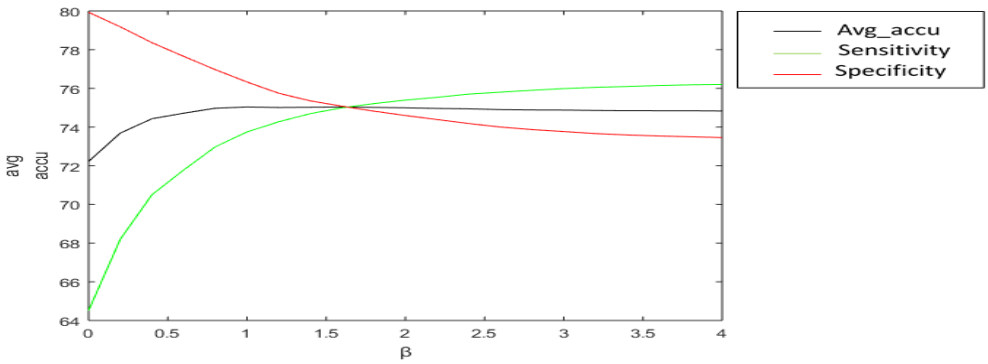


Figure 4.5: AVR, SEN, SPE graphs data 1

Now, we discuss the results of dataset 1 and 2 reporting in the Tab.4.2 and Tab.4.3 respectively.

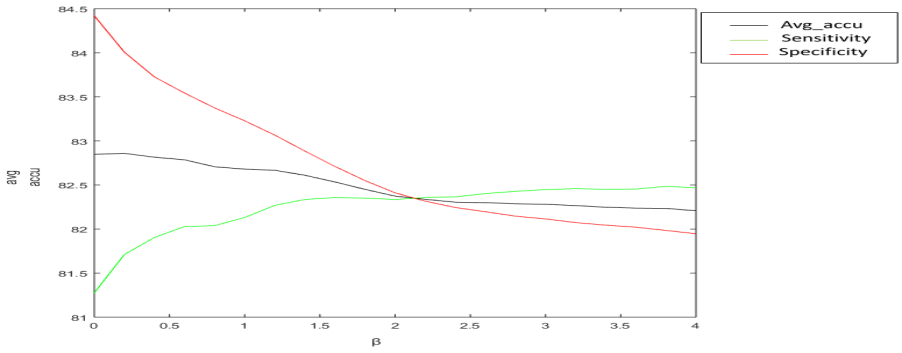


Figure 4.6: AVR, SEN, SPE graphs data 2

First dataset: the results of our model (CRF-ML) outperforms both the ML-Unary, CRF and CRF-bias methods respectively. As can be seen, there is a clear average accuracy improvement scoring 75.04% where the approach records an increment of around 2.5% , 2.7% and 5% over ML-Unary, CRF and ML-CRF-bias methods, successively. An interesting fact to point out is that the significant increase in sensitivity value of roughly 27% from 47.9% to 75.03% when compare the ML-Unary and CRF-ML whereas the rate of increase is 22% compared by CRF-bias. the score of increase is 10% when we compare the CRF and CRF-ML methods (from 64.5% to 75.03%). Simultaneously with this higher gain in the sensitivity, there is a decreasing in specificity value, from 97.1% to 75.06% compared with ML-Unary, from 86.8% to 75.06 % compared with CRF-bias while it decrease from 79.9% to 75.06% compared with CRF. This dissimilarity between SEN and SPE can interpreted by the fact that our model has allowed to recover many lost objects (true positives i.e. SEN) but at the expense of a higher number of errors on absent objects (true negatives i.e. SPE).

Second dataset: Unlike the results for the first data set, the CRF-ML outperform the ML-Unary and the CRF-bias methods, in terms of average accuracy, scoring 82.9% with rate of increase around 2.5% for ML-Unary, and around 1.4% for CRF-bias, while a negligible improvement is recording compared with CRF (around 0.1%). Regarding the

numerical results, one also can notice the similarity between CRF-ML and CRF methods in terms of SEN, SPE and average accuracy (AVG). Indeed, a lowest rate of increasing is observed scoring 0.5% for the SEN and a lowest rate of decrease 0.4% for the SPE. However, our strategy (CRF-ML) records significant results compared with ML-Unary and CRF-bias regarding the SEN and SPE values. For the SEN is increased from 62.6 to 81.7 (around 19% rate of increase) compared with ML-Unary and from 70.6% to 81.7% compared with CRF-bias, while the SPE is decreased from 98.2% to 84% with rate of decline 14% comparing with ML-Unary and from 92.8% to 84.01% comparing with CRF-bias. According to this data set results, we can confirmed that the exploitation of spatial contextual information has led to a further substantial boost in the correct detection of true positives.

4.3.2 The results of Multilabel CRF

Firstly, we remind the main idea of the proposed model. We formulate the multilabeling problem applied at a tile level under a CRF perspective, with the aim to assign to each tile a vector of labels instead of just one class label. The main novelty is that the proposed CRF integrates the cross-correlation information between different class labels jointly with spatial information within the same class as shown in the Fig.4.7The parameters of the proposed multilabel CRF model i.e.,$v_k i, w_k ij$ are learned from the training data using LBP inference. We set $v_{ki2} = 0$ and $w_{kij12} = w_{kij21} = 0$ in order to avoid overparametrization of the model(Schmidt et al., 2008). The correlation parameters λ_1, λ_2 are fixed to $\lambda_1 = \lambda_2 = 1$. The effect of varying the correlation parameters on the obtained results is analyzed.

Our proposed strategy (Full-ML-CRF) outperforms both the ML-Unary and ML-CRF methods in terms of average accuracy scoring, 76.2% and 83.4% for datasets 1 and 2, respectively. It records an increment of around 4% and 6% in dataset 1, and 3% and 2% in dataset 2 over ML-Unary CRF-bias methods respectively. Moreover, our strategy of-

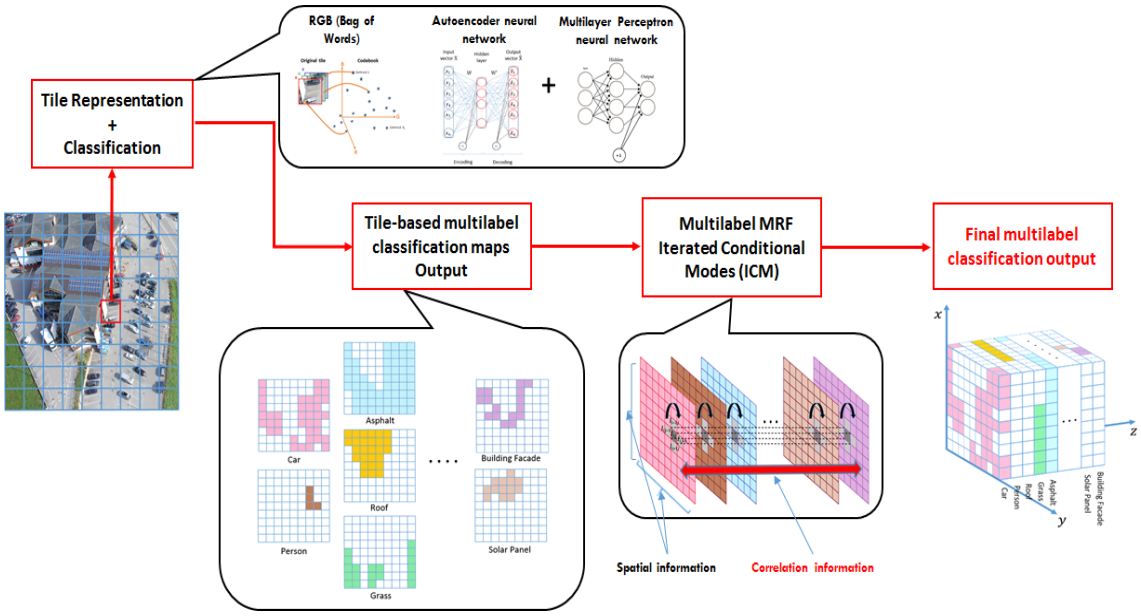
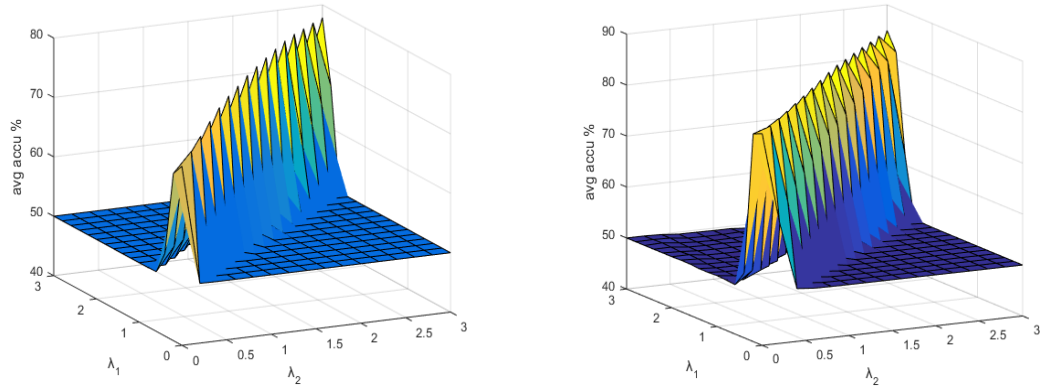


Figure 4.7: Flowchart of the proposed multilabel clasification method

fers the advantage of yielding higher sensitivity in both datasets while maintaining an appropriate rate of specificity. An interesting fact to point out is that a lowest average accuracy in dataset 1 is observed for the ML-CRF method scoring 70.1%. It has failed to outperform the ML-unary classification of independent tiles, which has scored 72.5%. This can be interpreted by the fact that spatial information has allowed to recover some lost objects (true positives) but at the expense of a higher number of errors on absent objects (true negatives). The exploitation of the cross-correlation information by Full-ML-CRF has led to a further substantial boost in the correct detection of true positives (and thus a higher SEN). In general, the exploitation of spatial information incurs in a loss of true negatives (confirmed for both datasets) which are typically dominant in multilabel maps. This is however accompanied by an increase of the true positives, which can be not always sufficient to compensate the above loss (case of ML-CRF on dataset 1) or very substantial (case of Full-ML-CRF on both datasets).

Another element to discuss is the influence of varying the correlation parameter values i.e.λ_1, λ_2 on the classification outcomes. Fig.4.8 describes the behavior of the average accuracy against λ_1 and λ_2. By analyzing Fig.4.9, one can notice that the best results

are obtained when λ_1 and λ_2 take similar values. Indeed, balancing between λ_1 and λ_2 which represent respectively the presence and the absence state of a given label leads to a better balance between sensitivity and specificity. By contrast, magnifying one parameter at the expense of the other, leads to a very high sensitivity (i.e., 100%) and very low specificity (i.e., 0%) or vice-versa, which results in 50% of average accuracy.



(a) (b)

Figure 4.8: Average accuracy versus spatial parameters achieved by the Full-ML-CRF method on (a) dataset 1 and (b) dataset 2

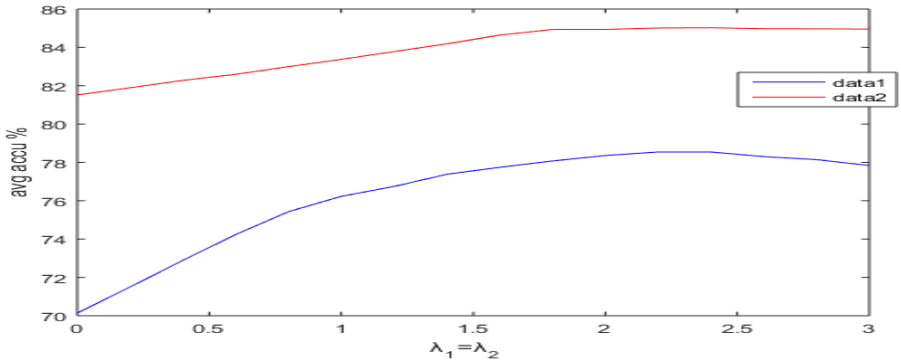
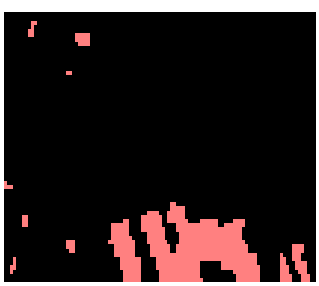


Figure 4.9: Average accuracy versus spatial parameters taking same value for Fll-ML-CRF model

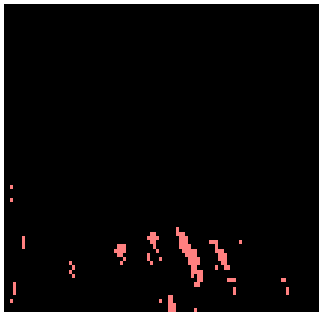
4.4 Comments and Comparaison

Already we discussed and analyzed the numerical reslts in terms of accuracies SEN, SPE and AVG. Now, compared the results over the resulting classification maps.

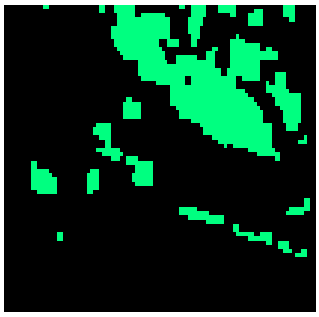


(a) Full-ML-CRF

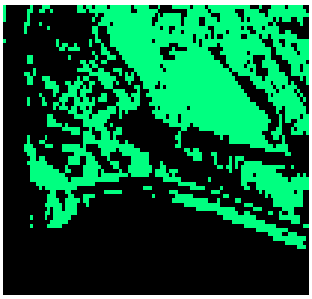
(b) CRF-ML



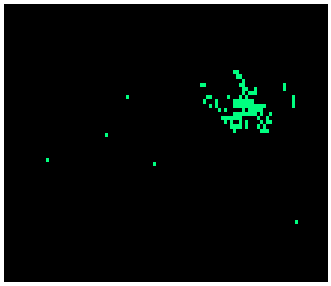
(c) ML-unary



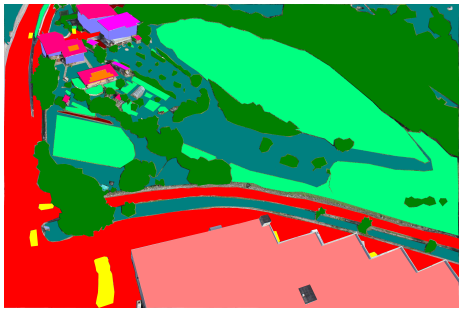
(d) Ful-lML-CRF



(e) CRF-ML



(f) ML-unary



(g) Ground Truth

(h) Original Image

Figure 4.10: Example of multilabel classification maps obtained by the three reference methods (ML-Unary, CRF-ML, Full-ML-CRF) on a test image, along with their related ground truth and original image.

That Fig.4.10 illustrates examples of multilabel classification maps obtained with the proposed Full-ML-CRF along with the reported methods ML-Unary, CRF-ML and CRF-bias, on one example test image from dataset 2. Where (a),(b)and (c) represent the class of Roof 3(bright), while (d),(e) and (f) represent the low vegetation class. In a comparison between the three methods of Full-ML-CRF, CRF-ML and ML-Unary, we have noticed the following results : it has been confirmed that the ML-Unary is not accurate to classify an image because, on the basis of ground thruth Fig.4.10 (g), (c) and (f) show many losses in terms of the main conponents, whereas Full-ML-CRF and CRF-ML have showed more accuracy. However, between the Full-ML-CRF and CRF-ML, the Full-ML-CRF appears to be more reliable as a method of classification. For that, (b) and (e) show false classification while (a) and (d) show true prediction. Hence, we can say that CRF-ML needs to be enhanced to get more accurate resuls and more credible prediction.

- We need to explain the difference between this proposed model and the previous one. the first one (refered as CRF-ML) based in the principe of CRF (encapsulate the neighboring information 4.4) while this model (refered Full-ML-CRF) express the cross-correlation in terms of all labels beside the traditional spatial information see 4.7. In the theoretical side, the CRF-ML expressed by use the observed data, but the Full-ML-CRF expressed in terms of true labeling which lead us to extract the specific term (cross-correlation term) in the training phase as descibed in (Zeggada et al., 2018).

- It is worth to mention that it is possible to recover an MRF model (i.e., the joint probability $P(Y_k, X_k) \simeq P(X_k)$) by setting the edge features $f_{kij=1}$, so that w_{kij} will represent the unconditional potential edge between nodes i and j

- once we set $\lambda_1 = \lambda_2 = 0$, we recover the traditional monolabel CRF (CRF-bias) that works at the level of each class map separately. For the CRF-ML approach, if we set $\beta = 0$, we recover also the standard CRF reference model (CRF).

- As future development, expanding the action field of the cross-correlation term by

considering wider neighborhood systems could be worth investigating, though this would lead to an increase of the method complexity.

- Moreover, an automatic way for estimating the optimal value of $\lambda_1 = \lambda_2$ (here fixed to 1) could be interesting to improve further the results

In the last of this chapter, we atteched a tables that they report the results over all the classes for both datasets considering Full-ML-CRF and CRF-ML methods.

	Accuracies (%)		
classes	SPE	SEN	AVR
1	89.2374	83.8630	86.5502
2	87.5804	86.1087	86.8446
3	81.9942	87.0307	84.5125
4	84.8292	72.9642	78.8967
5	92.5366	47.2209	69.8788
6	83.3323	85.939	84.6360
7	91.5183	81.134	86.3263
8	88.8875	82.3764	85.6320
9	91.0585	62.3367	76.6976
10	93.2424	91.8860	92.5642
11	82.3575	83.7418	83.0497
12	83.3173	85.028	84.1727
13	83.3173	85.028	84.1727
14	80.548	84.268	82.4080

Table 4.4: CRF-ML results of all classes- Data 2

	Accuracies (%)		
classes	SPE	SEN	AVR
1	93.7	75.96	84.83
2	72.02	75.33	73.68
3	66.47	82.35	74.41
4	52.68	26.26	39.47
5	71.45	88.02	79.74
6	63.36	61.79	62.58
7	91.41	67.08	79.25
8	87.49	65.32	76.41
9	63.13	77.15	70.14
10	81.64	72.26	76.95
11	66.90	100	83.45
12	85.04	76.96	81.00
13	84.40	81.44	82.92

Table 4.5: CRF-ML results of all classes- Data 1

	Accuracies (%)		
classes	SPE	SEN	AVR
1	95.76	59.97	77.87
2	68.10	80.06	74.08
3	66.52	76.32	71.41
4	55.42	71.04	63.23
5	87.15	43.26	65.21
6	100	00	50
7	91.22	66.71	78.96
8	92.51	57.52	75.02
9	68.22	65.95	67.08
10	78.34	74.81	76.59
11	75.17	100	87.59
12	94.88	55.01	74.94
13	92.38	67.15	79.77

Table 4.6: Full-ML-CRF results of all classes- Data 1

	Accuracies (%)		
classes	SPE	SEN	AVR
1	91.0993	81.3780	86.2387
2	83.7681	88.8535	86.3108
3	89.7629	78.8965	84.3297
4	98.8397	5.2932	52.0664
5	92.5366	47.2209	69.8788
6	92.5366	47.2209	69.8788
7	96.4953	74.6528	85.5740
8	95.3088	70.0159	82.6624
9	90.7345	65.7960	78.2652
10	98.9309	82.8947	90.9128
11	95.2746	50.4902	72.8824
12	92.9612	69.4430	81.2021
13	52.1844	43.3809	47.7826
14	94.6856	57.0451	75.8654

Table 4.7: Full-ML-CRF results of all classes- Data 2

Conclusion

With the objectives of improving the performance of multilabel classification task, this thesis has treated the condotional random fields (CRFs) framework for binary classification of images and thier persepectives in this tendency of image treatments in the first two chapters. That, the CRFs have been related substantially with graph theory where the graph in our context is made up of nodes (pixels), and edges in which each edge operates to connect a pair of nodes. In addition to preliminary and other basics concepts such as cliques, neighborhood system and so on, are introduced in the first chapter that we represente the theoretical part for this type of random fields. In the second chapter, we have started with explaining the problem of classification and the underlying idea of CRF model as a classifier. Then, we described the essentiel elements of a CRF model which are Parameters estimation and the related optimization problem. The subject of the third chapter is the corp of the research paper (Zeggada et al., 2018) starting with a simple approach for multilabeling data then we have proposed the model of this paper. Where the main idea of first proposed approach is applying a Conditional Random Field over a multilabeling image with subjecting differents classes to the neighboring system adopted while second approach integrate the cross-correlation between all the compenants of multilabel description vector of on site jointly with the traditional spatial information under the CRF perpective. The efficiency of this approaches were illustrated by numerical results reported in the fourth chapter advocated by some graphs and visual results.

As a perspectives, we can re-struct the space of labels(multilabeling inputs) then we will

apply one of the above approaches. Also we will try to innovate the ML CRF by finding another approximate for cross correlation term with subjecting their parameters to estimation procedure.

Bibliography

References

Berthod, M., Kato, Z., Yu, S., & Zerubia, J. (1996). Bayesian image classification using markov random fields. *Image and vision computing*, *14*(4), 285–295.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 259–302.

Boykov, Y., & Kolmogorov, V. (2001). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *International workshop on energy minimization methods in computer vision and pattern recognition* (pp. 359–374).

Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, *23*(11), 1222–1239.

Derin, H., Elliott, H., Cristi, R., & Geman, D. (1984). Bayes smoothing algorithms for segmentation of binary images modeled by markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*(6), 707–720.

Ganan, S., & McClure, D. (1985). Bayesian image analysis: An application to single photon emission tomography. In *Proceedings of american statistical association* (pp. 12–18).

Geman, D., & Gidas, B. (1991). *Image analysis and computer vision*. University of Massachusetts.

Geman, S., & Geman, D. (1987). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In *Readings in computer vision* (pp. 564–584). Elsevier.

Geman, S., & Graffigne, C. (1986). Markov random field image models and their applications to computer vision. In *Proceedings of the international congress of mathematicians* (Vol. 1, p. 2).

Greig, D. M., Porteous, B. T., & Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, 271–279.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, *31*(1), 253–258.

Kohli, P., & Torr, P. H. (2005). Efficiently solving dynamic markov random fields using graph cuts. In *Computer vision, 2005. iccv 2005. tenth ieee international conference on* (Vol. 2, pp. 922–929).

Kolmogorov, V., & Zabin, R. (2004). What energy functions can be minimized via graph cuts? *IEEE transactions on pattern analysis and machine intelligence*, *26*(2), 147–159.

Kumar, M. P., Kolmogorov, V., & Torr, P. H. (2009). An analysis of convex relaxations for map estimation of discrete mrfs. *Journal of Machine Learning Research*, *10*(Jan), 71–106.

Kumar, S., et al. (2003). Discriminative random fields: A discriminative framework for contextual interaction in classification. In *Computer vision, 2003. proceedings. ninth ieee international conference on* (pp. 1150–1157).

Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Lempitsky, V., Rother, C., Roth, S., & Blake, A. (2010). Fusion moves for markov random field optimization. *IEEE transactions on pattern analysis and machine intelligence*, *32*(8), 1392–1405.

Li, M., Zang, S., Zhang, B., Li, S., & Wu, C. (2014). A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing*, *47*(1), 389–411.

Li, S. Z. (1994). Markov random field models in computer vision. In *European conference on computer vision* (pp. 361–370).

Li, S. Z. (2009). *Markov random field modeling in image analysis*. Springer Science & Business Media.

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, *28*(5), 823–870.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Melgani, F., & Serpico, S. B. (2003). A markov random field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(11), 2478–2487.

Paisitkriangkrai, S., Sherrah, J., Janney, P., & van den Hengel, A. (2016). Semantic labeling of aerial and satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(7), 2868–2881.

Pearl, J. (1982). Reverend bayes on inference engines: a distributed hierarchical approach. In *in proceedings of the national conference on artificial intelligence* (pp. 133–136).

Roy, S., & Cox, I. J. (1998). A maximum-flow formulation of the n-camera stereo correspondence problem. In *Computer vision, 1998. sixth international conference on* (pp. 492–499).

Rozanov, Y. A. (1982). Markov random fields. In *Markov random fields* (pp. 55–102). Springer.

Schindler, K. (2012). An overview and comparison of smooth labeling methods for land-cover classification. *IEEE transactions on geoscience and remote sensing*, *50*(11), 4534–4545.

Schmidt, M. W., Murphy, K. P., Fung, G., & Rosales, R. (2008). Structure learning in random fields for heart motion abnormality detection. In *Cvpr* (Vol. 1, p. 2).

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., … Rother, C. (2008). A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE transactions on pattern analysis and machine intelligence*, *30*(6), 1068–1080.

Tarabalka, Y., Fauvel, M., Chanussot, J., & Benediktsson, J. A. (2010). Svm-and mrf-based method for accurate classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, *7*(4), 736–740.

Volpi, M., & Ferrari, V. (2015). Structured prediction for urban scene semantic segmentation with geographic context. In *Urban remote sensing event (jurse), 2015 joint* (pp. 1–4).

Zeggada, A., Benbraika, S., Melgani, F., & Mokhtari, Z. (2018). Multilabel conditional random field classification for uav images. *IEEE Geoscience and Remote Sensing Letters*, *15*(3), 399–403.

Zeggada, A., & Melgani, F. (2016). Multilabel classification of uav images with convolutional neural networks. In *Geoscience and remote sensing symposium (igarss), 2016 ieee international* (pp. 5083–5086).

Zeggada, A., & Melgani, F. (2017). Multilabeling uav images with autoencoder networks. In *Urban remote sensing event (jurse), 2017 joint* (pp. 1–4).

Zeggada, A., Melgani, F., & Bazi, Y. (2017). A deep learning approach to uav image multilabeling. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), 694–698.

Zhang, T., Yan, W., Li, J., & Chen, J. (2016). Multiclass labeling of very high-resolution remote sensing imagery by enforcing nonlocal shared constraints in multilevel conditional random fields model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(7), 2854–2867.

Zhao, J., Zhong, Y., Wu, Y., Zhang, L., & Shu, H. (2015). Sub-pixel mapping based on conditional random fields for hyperspectral remote sensing imagery. *IEEE Journal of Selected Topics in Signal Processing*, *9*(6), 1049–1060.