

NNT : 2015SACLC006

THÈSE DE DOCTORAT
DE
L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À
L'CENTRALESUPÉLEC

ÉCOLE DOCTORALE N°580
Sciences et Technologies de l'Information et de la Communication

Spécialité de doctorat : Traitement du Signal et des Images

Par

M Jorge PRENDES

New statistical modeling of multi-sensor images with
application to change detection

Thèse présentée et soutenue à Toulouse, le 22 octobre 2015

Composition du Jury :

M. Frédéric Pascal, *L2S – Supélec*, Directeur de thèse
Mme Marie Chabert, *INP – ENSEEIHT*, Co-directrice de thèse
M. Jean-Yves Tournet, *INP – ENSEEIHT*, Encadrant
M. Alain Giros, *CNES*, Encadrant
M. Pascal Bondon, *CNRS – Université Paris XI*, Président du jury
M. Lionel Moisan, *Université Paris Descartes*, Rapporteur
Mme Florence Tupin, *Institut Telecom – TelecomParisTech*, Rapporteur
Mme Josiane Zerubia, *INRIA Sophia-Antipolis*, Examinatrice

Titre : Nouvelle modélisation statistique des images multi-capteurs et son application à la détection des changements

Mots clés : Détection de changements, Images optiques, Images SAR, Modèle de mélange, Estimation Bayésienne nonparamétrique

Résumé : Les images de télédétection sont des images de la surface de la Terre acquises par des satellites ou des équipements embarqués. Ces images sont de plus en plus largement disponibles, à travers de nombreux services commerciaux et non commerciaux. La technologie des capteurs nécessaires pour capturer ce genre d'images est en pleine évolution. Non seulement les capteurs classiques s'améliorent en termes de résolution et de niveau de bruit, mais aussi de nouveaux types de capteurs commencent à être utilisés. Les capteurs d'images multispectraux sont la norme aujourd'hui, les images radar à ouverture synthétique (SAR) sont très populaires, et des capteurs hyperspectraux reçoivent de plus en plus d'attention dans de nombreuses applications.

L'une des principales applications des images de télédétection est la détection des changements dans les ensembles de données multitemporales, à savoir, la détection des changements dans les images de la même zone acquises à des moments différents. La détection des changements pour les images acquises par des capteurs homogènes a reçu un fort intérêt pendant une longue période. Toutefois, le large éventail de différents capteurs présents dans la télédétection fait de la détection des changements dans les images acquises par des capteurs hétérogènes un défi intéressant.

L'intérêt principal de cette thèse est d'étudier les approches statistiques pour détecter des changements dans les images acquises par des capteurs hétérogènes.

Title: New statistical modeling of multi-sensor images with application to change detection

Keywords: Change detection, Optical images, SAR images, Mixture model, Bayesian nonparametric estimation

Abstract: Remote sensing images are images of the Earth surface acquired from satellites or airborne equipment. These images are becoming widely available nowadays, with many commercial and non-commercial services providing them. The technology of the sensors required to capture this kind of images is evolving fast. Not only classical sensors are improving in terms of resolution and noise level, but also new kinds of sensors are proving to be useful. Multispectral image sensors are standard nowadays, synthetic aperture radar (SAR) images are very popular, and hyperspectral sensors are receiving more and more attention in many applications.

One of the main applications of remote sensing images is the detection of changes in multitemporal datasets, i.e., detecting changes in images of the same area acquired at different times. Change detection for images acquired by homogeneous sensors has been of interest for a long time. However the wide range of different sensors found in remote sensing makes the detection of changes in images acquired by heterogeneous sensors an interesting challenge.

The main interest of this thesis is to study statistical approaches to detect changes in images acquired by heterogeneous sensors.

“Choose a job you love, and you will
never have to work a day in your life.”

Confucius

Acknowledgment

It seems yesterday that I arrived to Toulouse to start working on this thesis and I met for the first time in person the people I would be working with for the following three years. It was the beginning of three wonderful years where I could learn and grow up a lot, not only from an academic perspective but also as a person.

These past three years I had the pleasure to work with Jean-Yves, Marie, Frédéric and Alain. All of them have contributed in different ways to make of these years a memorable period in my life, and I am very grateful for that. Jean-Yves and Frédéric have a special ability, like a super power, to always be present for their students. Even if they are busy with their several different activities, they always find time to discuss and come up with interesting ideas (I secretly think they have several clones). Marie has always been very helpful and supportive throughout the thesis work (and particularly throughout my Rubik's cube addiction!). Specially when writing seemed like an endless task, she was always very positive and encouraging. It is also thanks to her that I had the opportunity to teach in the university, which was a very interesting experience. Finally, Alain has always contributed with very stimulating ideas, providing a different point of view that never failed to enrich our discussions.

I also got to spend these three years at the TéSA laboratory. A simple "thank you" would not be enough to express my gratitude towards the people in the lab, who were like my step family in France. I have seen a few generations of people pass through the lab, and I keep very nice memories from all of them. Specially, I would like to thank Corinne, David, Jean-Adrien, Florian, Nil, Olivier, Philippe, Raoul, Romain, Tarik, Victor and Qi, for countless weekends and nights working at the office, trips to conferences, boardgames evenings, cinema nights and so many good moments and memories together.

Moreover, none of this would have been possible without the key support of a lot of people in Argentina, specially Daniel and Roxana, who encouraged me to apply for a PhD, my parents, who always pushed me to do what I love and made all they could to make that possible for me, and Christian, for receiving me in his family during all the years of engineering school.

Finally, I would like to thank Annemarie, for being there, for staying always positive and smiley and for being an excellent partner in life. Throughout these years, you have been the cherry on the icing on the cake. Thank you.

Toulouse, November 2015

Jorge Prendes



Abstract

Remote sensing images are images of the Earth surface acquired from satellites or airborne equipment. These images are becoming widely available nowadays, with many commercial and non-commercial services providing them. The technology of the sensors required to capture this kind of images is evolving fast. Not only classical sensors are improving in terms of resolution and noise level, but also new kinds of sensors are proving to be useful. Multispectral image sensors are standard nowadays, synthetic aperture radar (SAR) images are very popular, and hyperspectral sensors are receiving more and more attention in many applications.

The availability of different kind of sensors (also called modalities) is very advantageous since it allows us to capture a wide variety of properties of the objects contained in a scene as measured by each sensor. These properties can be exploited to extract richer information about the objects contained in the scene. One of the main applications of remote sensing images is the detection of changes in multitemporal datasets, i.e., detecting changes in images of the same area acquired at different times. Change detection for images acquired by homogeneous sensors has been of interest for a long time. However the wide range of different sensors found in remote sensing makes the detection of changes in images acquired by heterogeneous sensors an interesting challenge.

Accurate change detectors adapted to either homogeneous or heterogeneous sensors are needed for the management of natural disasters such as floods, volcano eruptions or earthquakes [UY12]. Databases of optical images are readily available for an extensive catalog of locations. However, good climate conditions and daylight are required to capture them. On the other hand, SAR images can be quickly captured, regardless of the weather conditions or the daytime. For these reasons, optical and SAR images are of specific interest for tracking natural disasters, by detecting the changes before and after the event. For this kind of applications an optical image is usually available from a database before the disaster and a SAR image can be quickly captured after the disaster.

The main interest of this thesis is to study statistical approaches to detect changes in images acquired by heterogeneous sensors.

Chapter 1 presents an introduction to remote sensing images. It also briefly reviews

the different change detection methods proposed in the literature. Additionally, this chapter presents the motivation to detect changes between heterogeneous sensors and its difficulties.

Chapter 2 studies the statistical properties of co-registered images in the absence of change, in particular for optical and SAR images. In this chapter a finite mixture model is proposed to describe the statistics of these images. The performance of classical statistical change detection methods is also studied by taking into account the proposed statistical model. In several situations it is found that these classical methods fail for change detection.

Chapter 3 studies the properties of the parameters associated with the proposed statistical mixture model. We assume that the model parameters belong to a manifold in the absence of change, which is then used to construct a new similarity measure overcoming the limitations of classic statistical approaches. Furthermore, an approach to estimate the proposed similarity measure is described. Finally, the proposed change detection strategy is validated on synthetic images and compared with previous strategies.

Chapter 4 studies Bayesian non parametric algorithm to improve the estimation of the proposed similarity measure. This algorithm is based on a Chinese restaurant process and a Markov random field taking advantage of the spatial correlations between adjacent pixels of the image. This chapter also defines a new Jeffreys prior for the concentration parameter of this Chinese restaurant process. The estimation of the different model parameters is conducted using a collapsed Gibbs sampler. The proposed strategy is validated on synthetic images and compared with the previously proposed strategy.

Finally, Chapter 5 is dedicated to the validation of the proposed change detection framework on real datasets, where encouraging results are obtained in all cases. Including the Bayesian non parametric model into the change detection strategy improves change detection performance at the expenses of an increased computational cost.

Résumé

Les images de télédétection sont des images de la surface de la Terre acquises par des satellites ou des avions. De telles images sont de plus en plus disponibles et fournies par des services commerciaux ou non-commerciaux. Les technologies des capteurs utilisés pour faire l'acquisition de ces images évoluent rapidement. On peut observer une amélioration significative des capteurs existants en termes de résolution et de rapport signal sur bruit. Mais de nouveaux types de capteurs ont également vu le jour et ont montré des propriétés intéressantes pour le traitement d'images. Ainsi, les images multispectrales et radar sont devenues très classiques et les capteurs hyperspectraux reçoivent de plus en plus d'attention dans de nombreuses applications.

La disponibilité de différents types de capteurs (donnant lieu à différentes modalités) est très intéressante car elle permet de capturer une grande variété de propriétés des objets contenus dans une scène et mesurés par chaque capteur. Ces propriétés peuvent être exploitées pour extraire des informations plus riches sur les objets contenues dans une scène. Une des applications majeures de la télédétection est la détection de changements entre des images multi-temporelles, i.e., détecter les changements dans des images de la même scène acquise à des instants différents. Détecter des changements entre des images acquises à l'aide de capteurs homogènes est un problème classique qui a reçu un grand intérêt dans la littérature. Mais le problème de la détection de changements entre images acquises par des capteurs hétérogènes est un problème beaucoup plus difficile

Avoir des méthodes de détection de changements adaptées aux images issues de capteurs homogènes et hétérogènes est nécessaire pour le traitement de catastrophes naturelles comme les inondations, les éruptions volcaniques ou les tremblements de terre [UY12]. Des bases de données constituées d'images optiques sont déjà disponibles avec une liste conséquente de lieux géographiques. Mais il est nécessaire d'avoir de bonnes conditions climatiques et une lumière suffisante pour faire l'acquisition de nouvelles images optiques associées à la même scène. En revanche, les images radar sont accessibles rapidement quelles que soient les conditions climatiques et peuvent même être acquises de nuit. Pour ces raisons, détecter des changements entre des images optiques et radar (acquises avant et après une éventuelle catastrophe) est un problème d'un grand intérêt

en télédétection.

L'intérêt principal de cette thèse est d'étudier des méthodes statistiques de détection de changements adaptés aux images issues de capteurs hétérogènes.

Chapitre 1 rappelle ce qu'on entend par une image de télédétection et résume rapidement quelques méthodes de détection de changements disponibles dans la littérature. Le chapitre présente également les motivations à développer des méthodes de détection de changements adaptées aux images hétérogènes et les difficultés associées.

Chapitre 2 étudie les propriétés statistiques des images de télédétection en l'absence de changements, en particulier, pour les images optiques et radar. Ce chapitre introduit un modèle de mélange de lois adapté aux propriétés statistiques de ces images. La performance des méthodes classiques de détection de changements est également étudiée en considérant ce modèle statistique. Dans plusieurs cas d'études, ce modèle permet d'expliquer certains défauts de certaines méthodes de la littérature.

Chapitre 3 étudie les propriétés des paramètres du modèle de mélange introduit au chapitre précédent. Ce chapitre fait l'hypothèse que les paramètres de ce modèle appartiennent à une variété en l'absence de changements entre les différentes images. Cette hypothèse est utilisée pour définir une nouvelle mesure de similarité entre images qui permet d'éviter les défauts des approches statistiques classiques utilisées pour la détection de changements. Une méthode permettant d'estimer cette mesure de similarité est également présentée. Enfin, la stratégie de détection de changements basée sur cette mesure de similarité est validée à l'aide d'images synthétiques et comparées avec plusieurs méthodes de référence.

Chapitre 4 étudie un algorithme Bayésien non-paramétrique qui permet d'améliorer l'estimation de la variété introduite au chapitre précédent. Cet algorithme est basé sur un processus de restaurant Chinois et un champ de Markov qui exploite la corrélation spatiale entre des pixels voisins de l'image. Ce chapitre définit également une nouvelle loi a priori de Jeffrey pour le paramètre de concentration de ce champ de Markov. L'estimation des paramètres de ce nouveau modèle est effectuée à l'aide d'un échantillonneur de Gibbs de type "collapsed Gibbs sampler". La stratégie de détection de changement issue de ce modèle non-paramétrique est validée à l'aide d'images synthétiques et comparée avec la stratégie proposée au chapitre précédent.

Le dernier chapitre de cette thèse est destiné à la validation des algorithmes de

détection de changements développés dans cette thèse sur des jeux d'images réelles montrant des résultats encourageant pour tous les cas d'étude. Le modèle Bayésien non-paramétrique permet d'obtenir de meilleurs performances que le modèle paramétrique, mais ceci se fait au prix d'une complexité calculatoire plus importante.

Acronyms and notations

Acronyms

The acronyms used throughout this thesis can be found in the following table.

BNP	Bayesian non parametric
DMM	Dirichlet mixture model
DPMM	Dirichlet process mixture model
EM	Expectation maximization
MCMC	Markov chain Monte Carlo
MRF	Markov random field
pdf	Probability density function
pdfs	Probability density functions
ROC	Receiver operating characteristic
SAR	Synthetic aperture radar

Notations

The notations used throughout this thesis can be found in the following table. Lowercase bold letters denote column vectors, while uppercase bold letters denote matrices. Please note that sub-indexes are denoted using a lowercase letter, and their upper limit is denoted by the capital letter, e.g., the sub-index k can take the values $1 \leq k \leq K$. The notation $\mathbf{z}_{\setminus n}$ denotes all the elements of the vector \mathbf{z} , except the n -th element.

D	Number of images in a dataset.
W	Analysis window.
N	Number of pixels in W .
S	Sensor used to acquire an image.
S_d	Sensor used to acquire the d -th image.
$i_{n,S}$	Intensity measured by S for the n -th pixel.
\mathbf{i}_n	D -dimensional vector of pixel intensities acquired by all sensors for the n -th pixel.

\mathbf{I}	Matrix containing all the intensities \mathbf{i}_n .
K	Number of different objects in W .
P	Physical properties of an object.
P_k	Physical properties of the k -th object.
w_k	Proportion of W covered by the k -th object.
\mathbf{w}	K -dimensional vector containing all w_k .
$T_S(P)$	How S images an object with properties P .
η_S	Acquisition noise of S .
$\mathcal{F}(\mathbf{v})$	Application dependent distribution family with parameter vector \mathbf{v} .
\mathcal{V}_0	Prior distribution for \mathbf{v} .
\mathbf{v}_n	Parameter vector that identifies a distribution from \mathcal{F} for the n -th pixel.
\mathbf{V}	Matrix containing all the \mathbf{v}_n .
\mathbf{v}'_k	Parameter vector that identifies a distribution from \mathcal{F} for the k -th object.
\mathbf{V}'	Matrix containing all the vectors \mathbf{v}'_k .
\mathcal{M}	“No-change” Manifold.
\hat{T}_{k,S_d}	Estimated parameters for the k -th component of the d -th sensor.
\hat{v}_k	Estimated parameters for the k -th component.
d_W	Similarity measure for the analysis window W .
$d_{\mathcal{M}}(\hat{v}_n)$	Similarity measure for the n -th pixel.
$DP(\alpha, \mathcal{V}_0)$	Dirichlet process with concentration parameter α and base pdf \mathcal{V}_0 .
$SBP(\alpha)$	Stick breaking process with concentration parameter α .
$CRP(\alpha)$	Chinese restaurant process with concentration parameter α .
$\psi(\cdot)$	Digamma function.
α	Concentration parameter.
z_n	Class label for the n -th pixel.
\mathbf{z}	Vector containing the class labels of all pixels.
$\rho_{X,Y}$	Population correlation coefficient between X and Y .
$r_{X,Y}$	Sample correlation coefficient between samples of X and Y .
$MI(X; Y)$	Mutual information between X and Y .
$D_{KL}(P Q)$	Kullback-Liebler divergence of P respect to Q .
$D_{KL}(P, Q)$	Symmetric Kullback-Liebler divergence.
G	A simple weighted graph.
ω_{nm}	Weight of the edge connecting the nodes n and m in the graph G .

Contents

Acknowledgment	v
Abstract	vii
Résumé	ix
Acronyms and notations	xiii
Contents	xv
Introduction	1
1 Remote sensing and change detection	7
1.1 Remote sensing	7
1.1.1 Optical sensors	7
1.1.2 SAR sensors	10
1.1.3 Heterogeneous sensors	10
1.2 Change detection	12
1.2.1 Change detection for optical images	12
1.2.2 Change detection for SAR images	13
1.2.3 Change detection based on a similarity measure	13
1.3 Conclusions	17
2 Statistical model for remote sensing images	19
2.1 Image generation	20
2.1.1 Optical images	21
2.1.2 Synthetic aperture radar images	23
2.2 Joint behavior of the images	24
2.2.1 Uniform regions	25
2.2.2 Non uniform regions	26
2.2.3 Model validation	27
2.3 Dependency based measures	28

2.3.1	Unchanged optical images	28
2.3.2	Changed optical images	30
2.3.3	Uniform regions	31
2.3.4	Non linearly related sensors	33
2.3.5	Real images	34
2.4	Conclusions	36
3	A new manifold based similarity measure	37
3.1	Manifold definition	38
3.1.1	Unchanged areas	39
3.1.2	Changed areas	40
3.2	A new similarity measure	41
3.2.1	Distance to the manifold	42
3.2.2	Estimation of the distance to the manifold	45
3.3	Parameter estimation	47
3.4	Performance on synthetic images	49
3.5	Conclusions	56
4	A BNP model for change detection	59
4.1	Bayesian modeling	60
4.1.1	Finite mixture distributions	60
4.1.2	Non parametric mixture distribution	62
4.1.3	Parameter estimation	66
4.1.4	Concentration parameter	68
4.2	Spatial correlation	73
4.2.1	Markov random field	75
4.2.2	Markov random field as a prior in a DPMM	76
4.3	Performance on synthetic images	79
4.4	Conclusion	82
5	Validation on real images	87
5.1	Homogeneous optical images	88
5.2	Heterogeneous optical images	92
5.3	Heterogeneous optical and SAR images	96
5.4	Conclusions	100

CONTENTS

xvii

6 Conclusion and perspectives

101

Bibliography

105

List of Figures

1.1	Spectral bands of the Pléiades sensors.	8
1.2	Pansharpening process on a multispectral Pléiades image.	9
1.3	Classic framework for statistical change detection based on an analysis window.	14
2.1	Spectral bands of the Pléiades sensors.	23
2.2	(a) Optical image from the south of Toulouse before the construction of a new road, (b) after its construction, and (c) the mask indicating the changed areas. (d) Shows the p-value distribution for the proposed mixture model and a histogram description.	27
2.3	Images of an unchanged area in south of Toulouse and the corresponding joint distribution estimated from a histogram.	29
2.4	Images before and after the construction of a road in south of Toulouse and the corresponding joint distribution estimated from a histogram.	30
2.5	Old image with a group of trees, new image preparing the ground for a new construction, and their corresponding joint distribution estimated from a histogram.	31
2.6	(a) and (b) Optical images for an unchanged homogeneous area with different brightnesses and contrasts, and (c) the corresponding joint distribution (estimated using a 2D histogram).	32
2.7	(a) Optical image with two different objects, (b) the unchanged image corresponding to a sensor with $T_S(P) = T_{Opt}(P)[1 - T_{Opt}(P)]$ and (c) the corresponding joint distribution (estimated using a 2D histogram).	33
2.8	Similarity of the images in Figs. 2.2(a) and 2.2(b) as measured by (a) the correlation coefficient and (b) the mutual information.	34

2.9	Even if the proposed model results in an improved model of the joint pixel intensities, its use does not reflect in any measurable improvement in the change detection performance compared to a histogram description when the mutual information is used as a similarity measure, as shown by these ROC curves.	35
3.1	Manifolds \mathcal{M} obtained with simple sensor combinations (Red curve: two identical sensors, Blue curve: sensors with different brightness and contrast with a saturation).	39
3.2	(a) and (b) are two (different) synthetic images obtained for the same sensor on a unchanged area. The $\mathbf{v}(P)$ of every each region is shown in (c).	40
3.3	(a) and (b) are two synthetic images obtained for the same sensor with different brightness and contrast settings on an unchanged area. The $\mathbf{v}(P)$ of every each region is shown in (c).	41
3.4	(a) and (b) are two synthetic images obtained for the same sensor on a changed area. The $\mathbf{v}(P_1, P_2)$ of every each region is shown in (c).	42
3.5	(a) and (b) are two synthetic images obtained for the same sensor with different brightness and contrast settings on a changed area. The $\mathbf{v}(P_1, P_2)$ of every each region is shown in (c).	43
3.6	Comparison between the (left) ground truth vectors $\mathbf{v}(P)$ or $\mathbf{v}(P_1, P_2)$ with the (right) vectors $\hat{\mathbf{v}}$ obtained from noisy data.	44
3.7	Relationship between Fig. 3.7(a) the distance to the manifold \mathcal{M} and Fig. 3.7(b) the density $p(\hat{\mathbf{v}})$. Black (blue on the right) color represent a value near zero while white (red on the right) represents a high value.	46
3.8	Figures (a) and (b) show two images captured by sensors S_1 and S_2 respectively. Figure (c) shows the mixture distribution of their pixel intensity, and their estimated parameters represented by a circle with a cross. Figure (d) presents the resulting probability of each pixel belonging to each component.	49
3.9	Example of synthetic images with changed and unchanged areas, where $T_{\text{opt}}(P) = P$ and $T_{\text{SAR}}(P) = P(1 - P)$	52
3.10	Estimated change maps for the images of Fig. 3.9. Red areas indicate high similarity, while blue areas indicate low similarity.	53
3.11	ROC curves for synthetic images (a) for different methods, (b) for the proposed method with different SNRs.	54

3.12	Impact of overestimating the number of components K in the mixture distribution.	55
3.13	Process to estimate the manifold \mathcal{M} from training data. A pair of unchanged optical and SAR images are used as an example.	57
4.1	Graphical representation of a Chinese restaurant process.	65
4.2	Evolution of the parameter α with a gamma prior for $\alpha_0 = 1$. The lines indicate the values of α for different iterations and the shaded area corresponds to 1 standard deviation respect to the mean.	70
4.3	Evolution of the parameter α with a Jeffreys prior for $\alpha_0 = 1$. The red line indicates the values of α for different iterations and the red shaded area corresponds to 1 standard deviation respect to the mean.	74
4.4	Two pairs of unchanged synthetic images. In (a) the images present spatial correlation. In (b) the spatial correlation had been broken by randomly shuffling the pixels jointly in both images.	75
4.5	Example of two neighborhood weights, for $\sigma = 1$ (left) and $\sigma = 2$ (right). The size of the circles represents the weight of that pixel with respect to the pixel in $(0, 0)$	78
4.6	Figures (a) and (b) show the same images as in Fig. 3.8 captured by sensors S_1 and S_2 respectively. Figure (c) and (e) show the mixture distribution of their pixel intensity, and their estimated parameters represented by a circle with a cross for the EM and the BNP-MRF estimations respectively. Figure (d) and (f) presents the resulting classification of each pixel into each mixture component for the EM and the BNP-MRF estimations respectively.	81
4.7	Results obtained by applying the proposed method into a dataset consisting of two synthetic heterogeneous optical and SAR images.	85
5.1	Results obtained by applying the proposed method into a dataset consisting of two real homogeneous optical images.	90
5.2	Change detection performance for the images of Fig. 5.1.	91
5.3	Results obtained by applying the proposed method into a dataset consisting of two real heterogeneous optical images.	94
5.4	Change detection performance for the images of Fig. 5.3.	95

5.5	Results obtained by applying the proposed method into a dataset consisting of two real heterogeneous optical and SAR images.	98
5.6	Change detection performance for the images of Fig. 5.5.	99

List of Tables

3.1	Performance of the different methods for detecting changes on the images of Fig. 3.11(a).	51
4.1	Change detector performance for different methods for the images in Fig. 3.11(a).	83
5.1	Performance of the different methods for detecting changes on the images of Fig. 5.1.	89
5.2	Performance of the different methods for detecting changes on the images of Fig. 5.3.	93
5.3	Performance of the different methods for detecting changes on the images of Fig. 5.5.	97

Introduction

Remote sensing images are images of the Earth surface acquired from satellites or airborne equipment. These images are becoming widely available nowadays, with many commercial and non-commercial services providing them. The technology of the sensors required to capture this kind of images is evolving fast. Not only classical sensors are improving in terms of resolution and noise level, but also new kinds of sensors are proving to be useful. Multispectral image sensors are standard nowadays, synthetic aperture radar (SAR) images [CM91; CGM95] are very popular, and hyperspectral sensors [Sch06] are receiving more and more attention in many applications.

One of the main applications of remote sensing images is the detection of changes in multitemporal datasets, i.e., detecting changes in images of the same area acquired at different times. Change detection for images acquired by homogeneous sensors has been of interest for a long time. However the wide range of different sensors found in remote sensing makes the detection of changes in images acquired by heterogeneous sensors an interesting challenge.

The main interest of this thesis is to study statistical approaches to detect changes in images acquired by heterogeneous sensors. The chapters composing this thesis are detailed below.

Organization of the thesis

Chapter 1

This chapter presents an introduction to remote sensing images, the different existing sensors and their properties, with a special focus on optical and SAR images: how the sensors work, different sensor variants, what are their advantages, weaknesses, uses, and what properties of a given scene is observed with them. It also presents the motivation to detect changes between heterogeneous sensors and its difficulties. This chapter also

explores the different change detection methods proposed in the literature, their strengths and limitations. First the methods proposed for homogeneous sensors are presented, most notably for two optical images or two SAR images. This chapter also presents the methods available to deal with heterogeneous images, especially the method proposed in [MMS08].

Chapter 2

This chapter studies the statistical properties of co-registered images in the absence of change. To achieve that, we introduce a general model for the statistics of the pixel intensity of independent sensors. The particular cases of optical and SAR images is developed. The joint distribution of these intensities on uniform areas (i.e., an areas of the image where all the pixels present the same statistics) is analyzed. To do this, the intensities associated with a given pixel (i.e., a given location) on all the images are collected into a multidimensional random vector. The model is extended to non-uniform areas, resulting in a finite mixture model. The goodness of fit (as measured using a Kolmogorov–Smirnov test) is compared between the proposed model and descriptions of the joint probability density function using histograms. The performance of classical statistical change detection methods are studied by taking into account the proposed statistical model in several different situations. Based on the proposed model, several situations where these measures fail are analyzed. The reason why measures based on statistical dependency fail as change detectors in these situations is also analyzed.

Chapter 3

This chapter studies the properties of the parameters associated with the model presented in Chapter 2. The behavior of these parameters is analyzed on changed and unchanged areas of the image. Based on this, it is proposed that the model parameters belong to a manifold in the absence of change [P3,P4]. This manifold is then used to construct a new similarity measure overcoming the limitations of classic statistical approaches as noted in Chapter 2. The similarity measure is constructed as the distance between the estimated mixture parameters (computed using the pixel intensities belonging to the estimation window) and this manifold. An approach to estimate a distance measure between the mixture parameters and the manifold is finally described. The estimation of the mixture parameters is addressed, based on a modified expectation maximization (EM) algorithm [FJ02]. The proposed similarity measure is validated on synthetic

images, and its performance for change detection is compared with that of other methods.

Chapter 4

The parameter estimation method considered in Chapter 3 presents two main limitations: the number of components in the mixture must be known, and the spatial correlation between neighboring pixels has been ignored [P1,P2]. A non parametric Bayesian framework is proposed in this chapter to deal with these limitations. To overcome the first limitation, the number of components in the mixture distribution is considered as a random variable to be estimated using a Bayesian nonparametric model [GB12] (BNP), namely a Dirichlet process mixture model (DPMM). A hierarchical Bayesian model is proposed to estimate some of the parameters required by the DPMM by means of a Jeffreys prior for which its density is derived. The second limitation is addressed by the introduction of a Markov random field [BKR11] (MRF) for the pixel labels associating each pixel to a component of the mixture distribution. To estimate the parameters of this new Bayesian non parametric model, a Markov chain Monte Carlo (MCMC) approach is considered. A collapsed Gibbs sampler algorithm is derived to improve the convergence of the MCMC based estimation.

Chapter 5

The proposed model was validated on synthetic images generated according to the model presented in Chapter 2. This validation favors the proposed method since they are based on the same assumptions. This chapter is dedicated to the validation of the proposed change detection framework on real datasets. Three study cases considered of relevance are included, namely a pair of homogeneous Pléiades images, a pair of heterogeneous optical images, and a pair of heterogeneous optical and SAR images [P5]. The proposed EM based method is found to outperform the classical methods in all the studied cases. Moreover, the BNP-MRF based approach outperforms the EM approach in all cases, at the cost of an increased computational time.

Main contributions

In what follows the contributions of the different chapters of this thesis are summarized.

Chapter 1

In this chapter an overview about the properties of remote sensing images, most notably optical and SAR images is detailed. An overview of change detection approaches for homogeneous and heterogeneous is also presented.

Chapter 2

This chapter studies the statistical properties of co-registered images in the absence of change. We propose a statistical model characterized by a mixture distribution to describe the joint distribution of the pixel intensities. The performances of classic statistical change detection methods are studied by taking into account the proposed statistical model. In particular, we explain why some methods based on statistical dependency measures can fail for detecting changes in remote sensing images.

Chapter 3

This chapter analyzes the behavior of the parameters of the statistical model presented in Chapter 2. Based on this, we propose a new similarity measure overcoming the limitations highlighted in the previous chapter. This similarity measure is data based, which makes it very flexible and can be built based on the application needs. Moreover, the proposed statistical model is validated on synthetic images.

Chapter 4

This chapter presents a parameter estimation approach that improves the one presented in Chapter 3. This is achieved by introducing a non parametric Bayesian framework into the parameter estimation. More precisely, a Dirichlet process mixture model [GB12] is studied to deal with the variable number of objects within each analysis window. A Jeffreys prior is proposed and derived to deal with the uncertainty on this number of objects. A Markov random field [BKR11] (MRF) prior is also assigned to the pixel labels associating each pixel to a component of the mixture distribution. Finally we propose a collapsed Gibbs sampler [VP08] to estimate the model parameters of this new Bayesian non parametric model.

Chapter 5

This chapter validates the proposed change detection framework on real datasets, where the proposed model outperforms the other methods studied [P5]. The EM based method is found to outperform the previous methods by an important margin. Moreover, the BNP-MRF based method outperforms all studied methods, included the EM based one. However, this improvement comes at the cost of an incremented computational complexity. The chapter concludes that each method (BNP-MRF or EM) is suitable for different applications depending whether the computational cost is an important factor.

List of publications

Submitted to journal

- [P1] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tournet, “A Bayesian nonparametric model coupled with a Markov random field for change detection in heterogeneous remote sensing images”.

Conference papers

- [P2] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tournet, “Change detection for optical and radar images using a Bayesian nonparametric model coupled with a Markov random field”, in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Brisbane, Australia, April 2015.
- [P3] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tournet, “A multivariate statistical model for multiple images acquired by homogeneous or heterogeneous sensors”, in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Florence, Italy, May 2014.

Journal papers

- [P4] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tournet, “A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors”, *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, March 2015.

- [P5] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, “Performance assessment of a recent change detection method for homogeneous and heterogeneous images”, *Revue Française de Photogrammétrie et de Télédétection*, vol. 209, pp. 23–29, January 2015.

Other conference papers

- [P6] M. Ressler, J. Prendes and R. Saint-Nom, “Undergraduate speech processing awareness”, in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Kyoto, Japan, March 2012.
- [P7] J. Prendes, M. Ressler, and R. Saint-Nom, design and execution”, “Speaker verification: A flexible platform architecture for experiment in *Proc. IASTED Int. Conf. Artificial Intell. and Appl.*, Innsbruck, Austria, February 2011.

CHAPTER 1

Remote sensing and change detection

Contents

1.1 Remote sensing	7
1.1.1 Optical sensors	7
1.1.2 SAR sensors	10
1.1.3 Heterogeneous sensors	10
1.2 Change detection	12
1.2.1 Change detection for optical images	12
1.2.2 Change detection for SAR images	13
1.2.3 Change detection based on a similarity measure	13
1.3 Conclusions	17

1.1 Remote sensing

For a long time, airborne or satellite remote sensing imagery has been used to track changes on the Earth surface for applications including urban growth tracking [SSS12; Tis+04], plantation monitoring [Sun+11; Sen+13], and urban database updating [Pou+11]. For this purpose, different sensors have been investigated including optical [Sch06; LFT12; Alp+07], synthetic aperture radars (SAR) [Sch06; CM91; CGM95; Fro+82] or multi-spectral sensors [Sch06; QP07].

1.1.1 Optical sensors

Due to the involved wavelengths, optical sensors provide high resolution images, and are generally easy to interpret, since they directly correlate to what a human would see. Because of this, huge databases of optical images are currently available. Optical sensors are usually composed of at least three different “sub-sensors” measuring different spectral

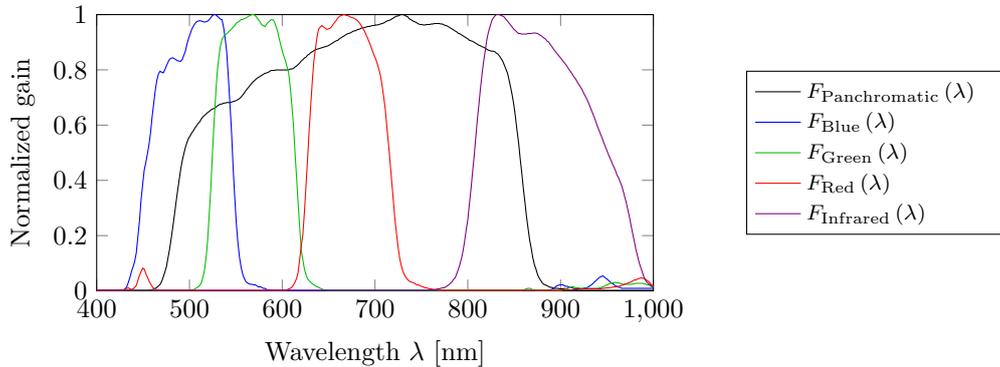


Figure 1.1: Spectral bands of the Pléiades sensors.

bands, one sensible to red light wavelengths, another sensible to green light wavelengths, and a third one sensible to blue light wavelengths, being also common to find another sensor, sensible to infrared light wavelengths. When more than three spectral bands are acquired, the resulting image is called a multispectral image or hyperspectral image depending on the number of acquired spectral bands. A sensing element for each spectral band is repeated in a grid, once for each pixel. Having several elements per pixel reduces the number of pixels we can measure at the same time, and thus it limits the image resolution. In other words, the more spectral bands to measure the lowest the resolution. This problem is usually addressed by including another sensor called panchromatic, which is sensible to a wide range of wavelengths (trying to include all the measured wavelengths), which is in its own grid and thus is not affected by the presence of the other sensors, allowing it to produce higher resolution images. These images are then combined through a pan-sharpening process [LFT12] to generate an image with the spectral characteristics of the multispectral image and the spatial resolution of the panchromatic image. One particular kind of optical images we are interested in results from the Pléiades satellites, which are two satellites capturing high resolution multispectral images. The spectral sensitivity of the sensors associated with these satellites [Bar+14] is depicted in Fig. 1.1. The multispectral image obtained with these satellites consists of 4 bands (blue, green, red and infra-red) with a spatial resolution of 2m, while the panchromatic image has a resolution of 70cm, although it is distributed as a resampled image with 50cm of spatial resolution. This pansharpening process for a Pléiades image can be observed in Fig. 1.2, where the multispectral image has been pansharpened using the corresponding panchromatic image to obtain a result combining the advantages of both.



(a) Multispectral image, 2m resolution, 4 bands



(b) Panchromatic image, 50cm resolution, 1 band



(c) Pansharpened image, 50cm resolution, 4 bands

Figure 1.2: Pansharpening process on a multispectral Pléiades image.

1.1.2 SAR sensors

On the other hand, SAR images can be acquired even at night or under bad weather conditions and thus are more rapidly available in emergency situations [IM09]. The physical principle behind SAR sensors is completely different from optical images. One key difference is that while optical sensors capture the reflected sunlight of the Sun on the different objects, SAR sensors are active, in that they generate the “light” to be reflected. This allows SAR sensors to acquire images during the day as well as during the night. To form the image a pulse of a coherent radio wave is emitted through an antenna, and the echo produced by the wave being reflected on the scene is recorded in amplitude and phase. This process is repeated in different locations, which is easily achieved if the antenna is mounted on a moving object, like an airplane or a satellite. All the echoes at different positions are then combined to focus the image. This process allows one to synthetically increase the aperture of the antenna, obtaining a result equivalent to that of using an antenna as long as the distance traveled while emitting and recording the radio waves. Wavelengths emitted by the SAR antenna can usually range anywhere from meters to millimeters, with smaller wavelengths allowing higher resolution images. The wavelength of the radio signal plays an important role on how the signal interacts and is reflected by the objects being imaged. Not only the material of the object is important but also its shape and size, since radio signals interact with objects of sizes similar to their wavelength. Because of this, clouds do not interact with the generated radio wave, allowing images to be obtained even with bad weather conditions. When focused, these images present a high variance multiplicative noise (referred to as “speckle noise”) due to the constructive and destructive interferences on each pixel of the reflections of the emitted coherent radio wave. Consequently, several beams are emitted in order to generate several images which are then averaged in order to reduce the noise, resulting in the so-called multi-look SAR images, where the number of looks is the number of images being averaged. Nevertheless, SAR images are generally more difficult to interpret because of their lower signal to noise ratio and because the observation is not linked to what a human would see.

1.1.3 Heterogeneous sensors

The availability of different kind of sensors (also called modalities) is very advantageous. This allows us to capture a wide variety of properties of the objects contained in a scene as measured by each sensor, which can be exploited to extract richer information about the

objects in the scene. This information can be used to classify the content of the image in different groups, such as: urban areas, rural areas, forestation, water surfaces, etc. Note that hyperspectral images are particularly useful for this classification, although they present much lower resolution than multispectral or SAR images. Combining information from different sensors requires the different images to be coregistered so that pixels corresponding to the same spatial location are identified as such. Remote sensing images usually include meta-information about the location of the image, with different degrees of accuracy, which helps their coregistration. When the images are not coregistered, different image processing techniques can be used to coregister them. Another advantage of having images from different modalities is that their complementary properties can be combined by image processing techniques, as with the pansharpening process that combines the spatial resolution of the panchromatic image and the spectral resolution of the multispectral image.

Image processing techniques can also be used on remote sensing images to find objects such as planes, buildings, forestation and cars. In a sequence of multitemporal images, these objects can be tracked from one image to another, where the modality of the image can change according to its availability (e.g., tracking objects at night is more easily achieved by SAR images). Another application of multitemporal images is the detection of changes, which consists in finding the differences between two images captured at different times. Since the different image modalities have advantages and disadvantages in different conditions, these multitemporal images are likely to be obtained by different sensors. This situation is particularly challenging since the images to be compared do not reflect the same characteristics of the objects present in the scene, so that they cannot be directly compared. In this thesis we will focus on the detection of changes in coregistered images captured by heterogeneous sensors. More precisely this work will pay a specific attention to detecting changes between optical and SAR images since. As previously discussed, combining the advantages of these sensors can be very interesting for several image processing applications.

Accurate change detectors adapted to either homogeneous or heterogeneous sensors are needed for the management of natural disasters such as floods, volcano eruptions or earthquakes [UY12]. Databases of optical images are readily available for a extensive catalog of locations. However, good climate conditions and daylight are required to capture them. On the other hand, SAR images can be quickly captured, regardless of the weather conditions or the daytime. For these reasons, optical and SAR images are

of specific interest for tracking natural disasters, by detecting the changes before and after the event. For this kind of applications an optical image is usually available from a database before the disaster and a SAR image can be quickly captured after the disaster.

To finish this part, we would like to point out the difference between sensors with a different working principle and homogeneous or heterogeneous sensors. We consider two sensors to be heterogeneous when they observe different physical properties of the captured scene, regardless of their working principle being the same or not. As an example, optical and SAR sensors have a different working principle, i.e., the physical phenomena on which they function is different. These are clearly heterogeneous sensors. Let now consider two optical sensors, with the same working principle, where one of these sensors is sensible to the visible spectra while the other is sensible to the infrared spectra. We also consider these sensors heterogeneous.

1.2 Change detection

Change detection is a classification problem between two different classes representing change and no change between two or more images. Generally the detection of changes becomes more challenging when the images to be compared are not coregistered [BS97]. The focus of this thesis is the detection of changes between multitemporal coregistered remote sensing images. The features considered by change detection methods are generally chosen according to the kind of sensor. Many techniques have been developed in the literature to detect changes between homogeneous images, especially in the case of two optical images. As a consequence many different approaches have been developed for optical and SAR images separately. In this thesis we propose a new flexible change detection strategy capable of dealing with homogeneous and heterogeneous sensors. More precisely this work will pay a specific attention to detecting changes between optical and SAR images, since the combination of these sensors is interesting in several practical applications.

1.2.1 Change detection for optical images

Well illuminated optical images are usually affected by an additive random Gaussian noise due to the thermal noise present in the sensor components (see Section 2.1.1 for more details). For this reason, many detection methods are based on the difference between intensities or on the difference between spectral bands in the case of multispectral

images leading to the so-called spectral change vector [BB07]. The difference image was initially derived pixel-wise [Sin89; Fun90; BF00b; BF02; Cel09b; MMS02]. However a parcel-wise derivation, using local averaging increases robustness with respect to noise, misregistration, miscalibration and other artifacts [BF00a; PRC05]. The difference image can also be derived in a transformed domain related to the wavelet transform for instance [Cel09a; CK10; MAS11], or the histogram of oriented gradients [SKZ14]. Some of these approaches also exploit the spatial correlation present in images by combining different image features with Markov random fields [Ben+15a; BS09; SKZ14; Ben+15b; MAS11]. These methods consider that in absence of change all sensors capture similar intensities on all images. Unfortunately this assumption does not hold in the case of heterogeneous sensors and thus these methods cannot be directly applied in such case.

Some interesting change detection methods adapted to optical images are based on neural networks [Gho+07; GSB13; Pac+07; Paj06; Lon+12; Hus+13] and support vector machines [PF10; NC06; Lon+12; Hus+13]. However, these “black-box” methods do not exploit the properties of these images. In this chapter we are going to develop the techniques of interest for this thesis. However, a survey of many popular change detection methods has been done in [Rad+05] and the reader can refer to it for more details.

1.2.2 Change detection for SAR images

As previously discussed, SAR images are affected by a multiplicative speckle noise due to the constructive and destructive interference on each pixel of the reflections of the emitted coherent radio wave. Consequently, many change detection methods are based on the ratio of the image intensities rather than the difference [TLB88; RV93; VFH93; Fjo+88; BBM05; CDB06; MS06]. In this case the difference image is usually computed as the difference between the logarithm of the images, which is referred to as the log-ratio. Moreover, some techniques have been proposed to merge the information obtained from different channels in multi-channel SAR images [MS09; MSV07].

As in the case of optical images, some change detection methods are based on neural networks [BF97; Pra+13] or on the joint distribution of the two images [Cha+07; Qui+14; Giu+13].

1.2.3 Change detection based on a similarity measure

There exist other groups of change detection methods that are based on some statistical measures obtained from the target images. These methods are generally based on a

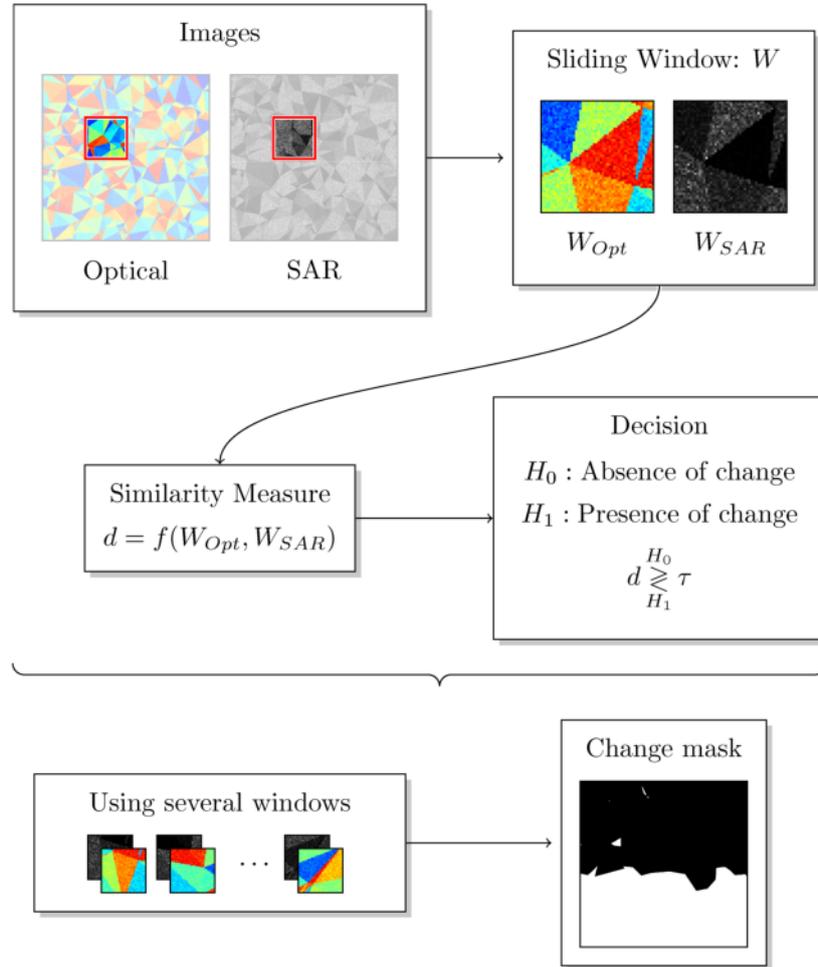


Figure 1.3: Classic framework for statistical change detection based on an analysis window.

so-called moving or analysis window. The strategy consists in decomposing the images in several small (coregistered) windows. A function assigning a scalar number to the analysis window is defined so that it assigns a low score to images which are different, and a high score to images which are similar. Such function is the so-called similarity measure. The similarity measure is evaluated on all the analysis windows until the whole image is analyzed, to obtain a similarity map.

A similarity measure can generally be associated with a binary hypothesis test, where the null hypothesis H_0 reflects the absence of change, and H_1 corresponds to the presence of change. The decision on whether there was a change or not is obtained by thresholding the similarity measure. When a theoretical distribution of the similarity measure in

presence and absence of change is available, such threshold can be adjusted for a given probability of false alarm.

The principle of change detection using a similarity measure is depicted in Fig. 1.3, where two synthetic optical and SAR images are considered as an example. An analysis window is highlighted in red in each image. The region within this window is then isolated from the rest of the image, since the different analysis windows will be analyzed independently. A similarity measure is then applied to them to obtain the scalar value d , which is the similarity measure between the two images. This similarity measure d is then thresholded according to a binary hypothesis test to decide whether a change has been found within the analysis windows or not. Finally, this process is repeated for all the analysis windows within the target images to obtain a change mask indicating the regions of the image which have been affected by changes between the two acquisitions. Obviously, the key component of this process is the definition of the similarity measure. Such statistical approaches have been extensively developed in the literature [SS14; Alb+07; IG04; Alb09; PV13; Hus+13]. In this work we are going to focus on several similarity measures appropriate to the application under consideration, with certain limitations as detailed hereafter.

The correlation coefficient is a very popular similarity measure to detect changes between images acquired by homogeneous sensors [IG04]. This coefficient measures the linear correlation between two random variables. In this case, it is assumed that, in the absence of change, the pixel intensities in the two images are linearly correlated. The correlation coefficient ρ for two random variables X and Y with a joint distribution $p(x, y)$ is given by

$$\rho_{X,Y} = \frac{\int_Y \int_X (x - \mu_X)(y - \mu_Y)p(x, y)dx dy}{\sqrt{\int_X (x - \mu_X)^2 p(x)dx} \sqrt{\int_Y (y - \mu_Y)^2 p(y)dy}} \quad (1.1)$$

where $\mu_X = E[X]$ and $\mu_Y = E[Y]$ are the population means of X and Y respectively. Moreover, a maximum likelihood estimation of this coefficient, denoted $r_{X,Y}$, can be easily obtained from a group of samples $[x_n, y_n]$ with $1 \leq n \leq N$ as

$$r_{X,Y} = \frac{\sum_{n=1}^N (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_{n=1}^N (x_n - \bar{x})^2} \sqrt{\sum_{n=1}^N (y_n - \bar{y})^2}} \quad (1.2)$$

where \bar{x} and \bar{y} are the sample means of $[x_1, \dots, x_N]$ and $[y_1, \dots, y_N]$ respectively. However, the assumption of linear correlation is generally not valid in the case of heterogeneous sensors [Ing02; IG04; IM09; CVA03], which makes the correlation coefficient inefficient for random variables X and Y associated with heterogeneous images.

To address this limitation other techniques have been developed, including the mutual information, which is defined as

$$\text{MI}(X; Y) = \int_Y \int_X p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) dx dy \quad (1.3)$$

The mutual information allows the dependency between two non linearly dependent images to be measured, yielding $\text{MI}(X; Y) = 0$ when X and Y are statistically independent, and a non-zero value when they are dependent (regardless the dependency being linear or not). Of course, the mutual information is more difficult to estimate than the correlation coefficient since it requires to estimate the joint distribution of the pixel intensities. This estimation can be achieved using a joint histogram or methods based on Parzen windows [KC02]. Unfortunately, the resulting mutual information is strongly dependent on the bins used to generate the histogram [FS86] or on the Parzen window size [KC02].

One alternative considered by several authors in the literature and the first one attempted in this thesis is to consider parametric distributions and to estimate their parameters using pixels located within a sliding window. Distributions that have been recently considered include bivariate gamma distributions for two SAR images, where a specific definition of the bivariate gamma distribution was considered in [Cha+07]. The parameters of this bivariate gamma distribution are then estimated based on the observed pixel intensities. The mutual information is then obtained from the analytic expression of the bivariate gamma distribution, and used as a similarity measure. Extensions to heterogeneous sensors, where the statistics of the two marginal distributions are not necessarily the same have also been proposed in [IG04; CT11]. In particular, a multivariate Pearson distribution was proposed in [CT11]. The family of univariate Pearson distributions has the interesting property of comprising both, the gamma and the normal distribution. Their multivariate extensions can be adopted to fit different image modalities on each marginal distribution, which can be interesting for SAR and optical images. Again, the mutual information could be obtained from the analytic expression of the multivariate Pearson distribution, and used as a similarity measure.

However, change detection between images acquired by heterogeneous sensors has received much less attention in the literature than the optical/optical or radar/radar cases. One can cite the recent approach developed in [MMS08]. This method is based on the idea that the Kullback–Leibler divergence D_{KL} can be used as a similarity measure between two analysis windows when the two images are captured by the same sensor. This divergence is asymmetric and computes the dissimilarity between two distributions

P and Q as follows,

$$D_{KL}(P|Q) = \int_{-\infty}^{+\infty} p_P(x) \log \left(\frac{p_P(x)}{p_Q(x)} \right) dx \quad (1.4)$$

where $p_P(x)$ and $p_Q(x)$ are the probability density functions (pdfs) of the distributions P and Q respectively. This divergence is asymmetric in the sense that $D_{KL}(P|Q)$ is not necessarily equal to $D_{KL}(Q|P)$. To address this issue, different approaches have been proposed in the literature. In [MMS08], the following symmetric version is used

$$D_{KL}(P, Q) = D_{KL}(Q, P) = D_{KL}(P|Q) + D_{KL}(Q|P). \quad (1.5)$$

The use of this divergence as a similarity measure is based on the assumption that for two homogeneous sensors, the distribution of the pixel intensities within an unchanged area should be similar, yielding a low value for D_{KL} . However, the pdfs of pixel intensities are generally different when heterogeneous sensors are considered. In [MMS08] it is proposed to transform the probability distribution associated with one sensor into the distribution of the other before computing the D_{KL} . In order to ensure symmetry into the process, this transformation is performed twice, i.e., from sensor A to sensor B and from sensor B to sensor A. The method investigated in [MMS08] to learn these transformations was based on the estimation of appropriate copulas from training samples belonging to a “no change” area. Generally this method is not straight forwards for the case of more than two images. To the best of our knowledge the method presented in [MMS08] is one of the latest change detection approaches proposed for heterogeneous sensors.

1.3 Conclusions

Section 1.1 introduced remote sensing images, with a specific attention to optical and SAR remote sensing images. The functioning of the sensors capturing these images was summarized, with a particular interest for the physical properties of the different objects that are captured by each sensor. Their advantages and weaknesses were highlighted. We also introduced the concept of homogeneous and heterogeneous sensors that will be of importance in this thesis. Section 1.1.3 analyzed the advantages of having heterogeneous sensors, and motivated the change detection problem for these images. One of the main reasons supporting this analysis is that the available images do not always arise from the same modality, especially considering the strengths of each sensor.

The different change detection methods proposed in the literature, their strengths and limitations were explored in Section 1.2. First the methods proposed for homogeneous

sensors were briefly summarized, most notably for two optical images in Section 1.2.1 and for two SAR images in Section 1.2.2. Statistical methods dealing with these heterogeneous images were presented in Section 1.2.3, with a specific interest for the method proposed in [MMS08].

CHAPTER 2

Statistical model for remote sensing images

Contents

2.1	Image generation	20
2.1.1	Optical images	21
2.1.2	Synthetic aperture radar images	23
2.2	Joint behavior of the images	24
2.2.1	Uniform regions	25
2.2.2	Non uniform regions	26
2.2.3	Model validation	27
2.3	Dependency based measures	28
2.3.1	Unchanged optical images	28
2.3.2	Changed optical images	30
2.3.3	Uniform regions	31
2.3.4	Non linearly related sensors	33
2.3.5	Real images	34
2.4	Conclusions	36

As discussed in Chapter 1, various statistical strategies were proposed for change detection that rely on the estimation of the joint probability distribution of the pixel intensities. These strategies usually derive a similarity measure from a statistical measure obtained from this joint distribution. In most cases, the correct estimation of the joint distribution is of great importance, and thus several joint distribution models have been proposed. In [Cha+07] a bivariate version of the gamma distribution was proposed, while [CT11] proposed a bivariate generalization of the Pearson distribution. The goal of this chapter is to investigate the joint behavior of the pixel intensities and derive a model for it.

More precisely, this chapter introduces a flexible statistical model for the pixel intensities associated with several images acquired by different sensors. The approach to obtain this statistical model differs from previous approaches since it begins by studying the physical processes that generate the heterogeneous images. To achieve this, the marginal statistical properties of the pixel intensity for different kinds of sensors (mainly optical and radar) is studied in Section 2.1. The joint behavior of several pixels belonging to an analysis window and acquired by heterogeneous sensors is studied in Section 2.2. We first consider uniform areas in the image, and later extend this model to the more general case of non-uniform areas which are commonly observed in analysis windows.

The implications of this model on different classical change detection methods are analyzed in Section 2.3, with a specific interest on how this model allows us to predict the behavior of change detectors based on statistical dependency measures, addressing strengths and limitations. Finally, conclusions and implications of this chapter are discussed in Section 2.4.

2.1 Image generation

A uniform area of an image is a region where the pixels have the same physical properties (denoted as P). For example, this could be the case for a region of the image covered by a corn crop, or water, or grass. Each sensor measures some particular property of the imaged object, which depends on the physical properties P of this object. For instance, an optical sensor measures the object reflectivity to the wavelengths in the optical range, while a LiDAR sensor [Wei06] measures the height of the object. If we assume that P provides a full description of the object being imaged, we can denote the measurement obtained by the sensor S as a function of P , i.e.,

$$i_S = T_S(P) \tag{2.1}$$

where i_S represents the pixel intensity measured by the sensor S , and $T_S(\cdot)$ is a transformation depending on the sensor, which represents how it captures an object with physical properties P . For instance, for a LiDAR sensor, the transformation $T_{\text{LiDAR}}(P)$ extracts the height information from P .

However, the measurements of any sensor S are corrupted by noise, which can change

in behavior and nature depending on the sensor. Because of this, i_S can be more accurately expressed as

$$i_S = f_S[T_S(P), \eta_S] \quad (2.2)$$

where η_S is a random variable representing the noise produced by sensor S , and $f_S[\cdot, \cdot]$ is a function indicating how such noise interacts with the measured intensity. Consequently, i_S is a random variable. The distribution family of this variable depends on $f_S[\cdot, \cdot]$ as well as on the distribution of η_S , while the parameters identifying one particular distribution of the family depend on $T_S(P)$, i.e., we can write

$$i_S|T_S(P) \sim \mathcal{F}_S[T_S(P)] \quad (2.3)$$

where $\mathcal{F}_S(\cdot)$ is a distribution family computed from η_S and $f_S(\cdot, \cdot)$ is parameterized by $T_S(P)$, which for simplicity and without any loss of generality, can be expressed as

$$i_S|P \sim \mathcal{F}_S(P) \quad (2.4)$$

In what follows, we study $\mathcal{F}_S(P)$ through the analysis of η_S and $f_S[\cdot, \cdot]$ for the case of optical and SAR sensors, while $T_S(P)$ is considered to be unknown. It is important to note that the focus of this thesis is on the study of optical and SAR remote sensing images. However, the presented model is general and could be applied to other kind of sensors. In Chapter 3 the relationship between the transformations $T_S(P)$ of different sensors S_1, \dots, S_D will be studied, and a method to estimate it from training data will be proposed.

2.1.1 Optical images

Optical images have been thoroughly studied in the literature (see [Cas96; LJW08]). Several kinds of noises are identified, and their predominance depends on the conditions on which the image has been taken.

For instance, under low light conditions, a few number of photons arrives to the optical sensor, producing the so called “shot noise” [HMC03]. In this case, the detection of photons is considered a Poisson process, and thus the measured intensities thus follow a Poisson distribution.

$$i_{\text{opt}}|P \sim k_{\text{opt}} \text{Poisson}[k_{\text{opt}}^{-1} T_{\text{Opt}}(P)] \quad (2.5)$$

where k_{opt} is a constant defined by the sensor amplification, i.e., a high value of k_{opt} represents a high amplification. Under low light conditions optical sensors need to increase their sensitivity in order to obtain a bright enough image, which is achieved by incrementing k_{opt} . In digital cameras this amplification is associated with the ISO parameter [Sta06]. For k_{opt} small enough, the Poisson distribution can be approximated by a Normal distribution as follows

$$\begin{aligned} k_{\text{opt}}\text{Poisson}[k_{\text{opt}}^{-1} T_{\text{Opt}}(P)] \\ \approx k_{\text{opt}} \mathcal{N}[k_{\text{opt}}^{-1} T_{\text{Opt}}(P), k_{\text{opt}}^{-1} T_{\text{Opt}}(P)] \end{aligned} \quad (2.6)$$

$$\approx \mathcal{N}[T_{\text{Opt}}(P), k_{\text{opt}} T_{\text{Opt}}(P)] \quad (2.7)$$

Under good light conditions, k_{opt} is small. Since the variance of the shot noise is proportional to k_{opt} , for small k_{opt} the shot noise can be omitted. The most common noise observed under this conditions is the so called ‘‘thermal noise’’, caused by the thermal agitation of electrons in the components of the sensor. This noise is usually modeled as a zero-mean Gaussian noise, whose variance is proportional to the operation temperature of the sensor. In this situations, the following model is more adequate,

$$i_{\text{opt}} = T_{\text{opt}}(P) + \eta_{\text{opt}} \quad (2.8)$$

$$\eta_{\text{opt}} \sim \mathcal{N}(0, \sigma_{\text{opt}}^2) \quad (2.9)$$

where σ_{opt} is a constant that depends on the sensor and its operation temperature. This results in the following expression for i_{opt}

$$i_{\text{opt}}|P \sim \mathcal{N}[T_{\text{opt}}(P), \sigma_{\text{opt}}^2]. \quad (2.10)$$

These two noise sources could be combined in a single more complex model. However, this would increase the number of model parameters and certainly make their estimation more difficult. Moreover, most optical satellite images are obtained during the day, when the illumination conditions are good. Thus, a usual factor considered when capturing these images is the saturation due to reflections on particular objects of the scene. For the stated reasons, we assume that the statistical behavior of optical remote sensing images, is described by the following family of distributions

$$\mathcal{F}_{\text{opt}}(P) = \mathcal{N}[T_{\text{opt}}(P), \sigma_{\text{opt}}^2]. \quad (2.11)$$

Note that different optical sensors will capture slightly different features from the physical properties P . Consider for instance the sensor used by the Pléiades satellite

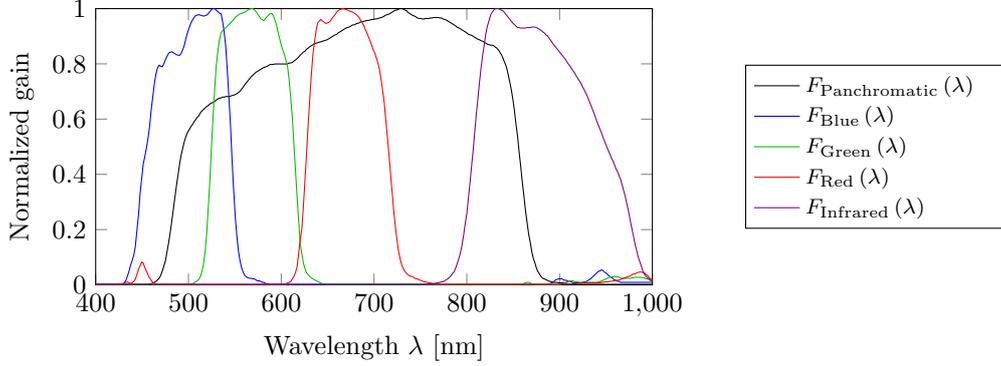


Figure 2.1: Spectral bands of the Pléiades sensors.

constellation. These satellites capture multispectral images in 4 bands: blue, green, red and near infrared, with a resolution of 2 meters per pixel, and a panchromatic image with a resolution of 0.5 meters per pixel. Usually these images undergo a pansharping process to generate a 4 band multispectral image with high spatial resolution. The 4 resulting images can be considered as 4 independent images acquired by 4 sensors, where each sensor captures one spectral band. To illustrate this, consider the spectral bands of the Pléiades sensors [Bar+14] as shown in Fig. 2.1. Let i_B be a pixel belonging to the blue band acquired by the equivalent “blue” sensor of the Pléiades satellite, corresponding to an object with physical properties P , and $R(P, \lambda)$ be a function indicating the reflectivity of P at the wavelength λ . Then, the value of $T_B(P)$ corresponding to i_B can be obtained as

$$T_B(P) = \int_{-\infty}^{\infty} F_B(\lambda) R(P, \lambda) d\lambda \quad (2.12)$$

where $F_B(P)$ is the amplitude of the spectral filter corresponding to the blue color. In the same way we can obtain $T_R(P)$, $T_G(P)$ and $T_{IR}(P)$, which are independent projections of P into a 4-dimensional space.

2.1.2 Synthetic aperture radar images

When a radar image is acquired, the resulting pixel intensity is commonly assumed to have an exponential distribution with rate parameter $T_{SAR}(P)$. The main cause of noise in SAR images is due to the coherence of the signal used to illuminate the scene. This signal is emitted from the radar, bounces on the objects in the ground, and returns to the radar. The coherence of the signal leads to constructive and destructive interferences

in the image. To model this behavior, the resulting pixel intensity is usually modeled as a random variable with an exponential distribution as follows

$$i_{\text{SAR}}|P \sim \exp[T_{\text{SAR}}(P)]. \quad (2.13)$$

where $\exp(\lambda)$ represents an exponential distribution with scale parameter λ . To mitigate the effect of this noise, several images of the scene are captured and averaged to obtain the final SAR image [OQ04; Fjo+88], so that

$$i_{\text{SAR}}|P = \frac{1}{L} \sum_{l=1}^L i_{\text{SAR},l} \quad (2.14)$$

$$i_{\text{SAR},l} \sim \exp[T_{\text{SAR}}(P)] \quad (2.15)$$

where i_{SAR} is the result of averaging L images denoted $i_{\text{SAR},l}$ with $1 \leq l \leq L$. These kind of images are the so-called “multi-look” SAR images, where L is the number of looks. If the different variables $i_{\text{SAR},l}$ are assumed statistically independent, i_{SAR} can be described as

$$i_{\text{SAR}}|P \sim \Gamma(L, T_{\text{SAR}}(P)L^{-1}) \quad (2.16)$$

where $\Gamma(k, \theta)$ represents the gamma distribution with shape and scale parameters k and θ respectively. If the different looks $i_{\text{SAR},l}$ are not independent, non integer values for L are usually considered.

The result in (2.16) can be expressed following the structure of (2.2) as

$$i_{\text{SAR}} = T_{\text{SAR}}(P) \times \eta_{\text{SAR}} \quad (2.17)$$

$$\eta_{\text{SAR}} \sim \Gamma(L, L^{-1}) \quad (2.18)$$

where η_{SAR} is called the multiplicative speckle noise.

2.2 Joint behavior of the images

Section 2.1 introduced a model describing the statistical behavior of a pixel intensity. However, in order to build a change detector, we are interested in modeling the joint statistical behavior of the pixel intensities arising from the different sensors. To achieve that, this section studies the joint statistical properties of the pixel intensities across the different modalities, i.e., the statistical properties of $\mathbf{i} = [i_1, \dots, i_D]^T$, a vector containing the intensities obtained by each sensor S_d , for any $d = 1, \dots, D$, for a particular

pixel. Particularly, given an analysis window W , we are interested in the joint intensity probability density function (pdf) $p(\mathbf{i}|W)$ in the absence of change, which is studied in Section 2.2.2. In order to derive this distribution, Section 2.2.1 studies the characteristics of $p(\mathbf{i}|P)$ on unchanged areas.

2.2.1 Uniform regions

Assume that we have acquired D images using D different sensors, and that no change has occurred between these D acquisitions. Moreover, consider a uniform region, i.e., a region where the physical properties P of the objects being observed do not change (e.g., a corn field with P_{corn} , or an area covered by water with P_{water}). Under this assumption, P is not a random variable; it takes a particular value, which depends on the chosen uniform area. Studying the statistical distribution of these areas can be achieved by studying $p(\mathbf{i}|P)$.

Since our interest focuses on the intensity distribution conditionally on P , we can consider $T_d(P)$ for any $d = 1, \dots, D$ as deterministic (in the sense that it is not a random variable). Recalling that $i_d = f_d(T_d(P), \eta_d)$ from (2.2), the joint distribution of \mathbf{i} can be written

$$p(\mathbf{i}|P) = p(i_1, \dots, i_D|P) \quad (2.19)$$

$$= p[f_1(T_1(P), \eta_1), \dots, f_D(T_D(P), \eta_D)|P] \quad (2.20)$$

Since the sensors capturing the images are not related, we propose to assume that the random variables η_1, \dots, η_D associated with the sensor noises are statistically independent. Consequently, $p(\mathbf{i}|P)$ can be written as

$$p(\mathbf{i}|P) = \prod_{d=1}^D p[f_d(T_d(P), \eta_d)|P] \quad (2.21)$$

$$= \prod_{d=1}^D p(i_d|P). \quad (2.22)$$

Following this model, the family distribution $\mathcal{F}_{[S_1, \dots, S_D]}(P)$ for $p(\mathbf{i}|P)$ is

$$\mathcal{F}_{[S_1, \dots, S_D]}(P) = \prod_{d=1}^D \mathcal{F}_{S_d}(P) \quad (2.23)$$

where each $\mathcal{F}_{S_d}(P)$ can be obtained in a similar way as explained in Section 2.1 for the optical and SAR images. For conciseness, $\mathcal{F}_{[S_1, \dots, S_D]}(P)$ will be denoted simply as $\mathcal{F}_D(P)$.

2.2.2 Non uniform regions

A classical way of handling the change detection problem is to analyze the image using a sliding window and to define a change indicator for each window position [Ing02]. In this case, we are particularly interested in the statistical properties of the pixel intensities within a sliding window W , which we will denote $p(\mathbf{i}|W)$.

To obtain this pdf, we assume that the region of interest (covered by the sliding window) is composed of a finite number K of uniform areas with different physical properties P_1, \dots, P_K . Because of this, we propose to describe the physical properties of the region of interest by a discrete random variable with distribution

$$p(P|W) = \sum_{k=1}^K w_k \delta(P - P_k) \quad (2.24)$$

where w_k is the weight or probability of P_k which represents the relative area of W covered by P_k . Using (2.22) and the total probability theorem, the joint distribution of the pixel intensity can be expressed as

$$p(\mathbf{i}|W) = \int_P p(\mathbf{i}|P)p(P|W)dP \quad (2.25)$$

$$= \int_P p(\mathbf{i}|P) \left(\sum_{k=1}^K w_k \delta(P - P_k) \right) dP \quad (2.26)$$

$$= \sum_{k=1}^K w_k \int_P p(\mathbf{i}|P) \delta(P - P_k) dP \quad (2.27)$$

$$= \sum_{k=1}^K w_k p(\mathbf{i}|P_k) \quad (2.28)$$

which leads to the equation

$$p(\mathbf{i}|W) = \sum_{k=1}^K w_k \prod_{d=1}^D p(i_d|P_k). \quad (2.29)$$

In this case, the family of distributions associated with the pixel intensities within an analysis window W , denoted as $\mathcal{F}_W(w_1, P_1, \dots, w_K, P_K)$, is a mixture of distributions belonging to the family $\mathcal{F}_D(P)$. Moreover, according to (2.22), each component of this mixture is the product of densities associated with independent random variables.

Note that previous models proposed in the literature are mainly obtained by trying to match the statistical properties of the marginal distributions. On the contrary, the proposed model is derived by proposing a physical model for the pixel intensity acquisition under realistic assumptions concerning the sensors.

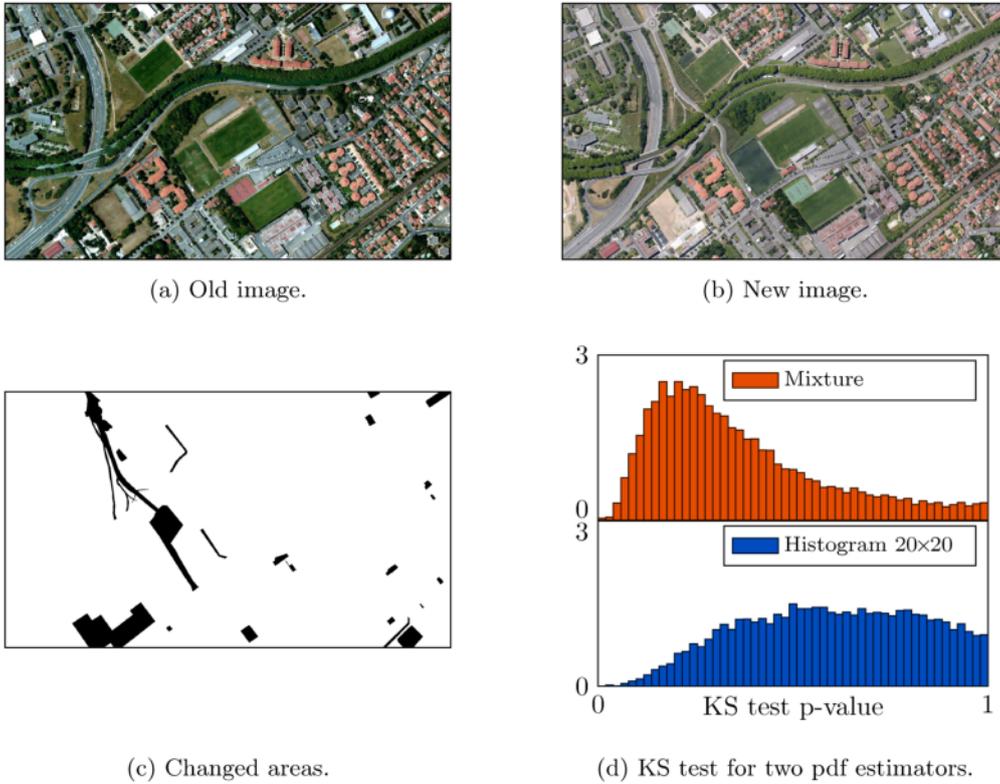


Figure 2.2: (a) Optical image from the south of Toulouse before the construction of a new road, (b) after its construction, and (c) the mask indicating the changed areas. (d) Shows the p-value distribution for the proposed mixture model and a histogram description.

2.2.3 Model validation

To assess the model introduced in Sections 2.2.1 and 2.2.2, we propose to estimate the joint distribution of the pixel intensities in different analysis windows. The dataset consists of two real registered optical images shown in Figs. 2.2(a) and 2.2(b). Only the windows outside the changed areas shown in Fig. 2.2(c) were considered. The joint pdf of this dataset has been estimated in two different ways. First, a so-called empirical pdf is obtained using a bivariate histogram with 20×20 classes. Second, a so-called estimated pdf is obtained from (2.23) replacing the model parameters by their estimates obtained using a modified version of the expectation maximization algorithm (please refer to Section 3.3 for more details on the algorithm). To quantify the validity of our model we used a two dimensional generalization of the Kolmogorov-Smirnov (KS) goodness of fit test [Pea83]. This test measures whether a sample population has been produced by a

given distribution or not, by assuming that the sample based cumulative density function (cdf) can be expressed as the population cdf with an additive random walk constrained to start and end in zero. An hypothesis test is designed, where the test statistic is the maximum difference between the sample based histogram and the population cdf (i.e., the maximum absolute value taken by the random walk). In the bivariate generalization of the test, the fact that three non linearly dependent cdf can be obtained (depending on the integration direction of the random variable) is taken into account by keeping the biggest of the three possible differences. The hypothesis H_0 indicating that the sample population was produced by the same distribution is accepted whenever the p-value is less than the significance level α . The histogram for the p-value obtained with the proposed model and with the histogram description or the joint distribution are compared in Fig. 2.2(d), where it can be observed that the proposed model produces better descriptions of the data than the histogram. The algorithm used to estimate the parameters of the proposed model will be presented in Section 3.3. Based on this we can assert that the proposed method obtains a better fit to the data than a histogram description. Moreover, the histogram description requires $20 \times 20 = 400$ scalar numbers to represent the data, while in this case the proposed method requires 5 parameters per component (two means, two variances, and a weight). Since for this example the algorithm in Section 3.3 was limited to produce a maximum of 10 component, this estimated model is always represented by 50 or less scalar parameters.

2.3 Dependency based measures

Based on the results obtained in Section 2.2, this section analyzes the behavior of the correlation coefficient and the mutual information as change detectors by analyzing some relevant examples. Moreover, it is also shown that these examples are in good agreement with the proposed model.

2.3.1 Unchanged optical images

Figs. 2.3(a) and 2.3(b) display examples of two optical images acquired by the same sensor associated with an unchanged area, where three kinds of objects can be clearly seen: a red roof, grass, and parts of trees. According to the proposed model, the joint distribution of these images should be a mixture of three Gaussian components. This mixture can be observed in Fig. 2.3(c), where the joint pixel intensity distribution, i.e., $p(i_{\text{old}}, i_{\text{new}})$ was

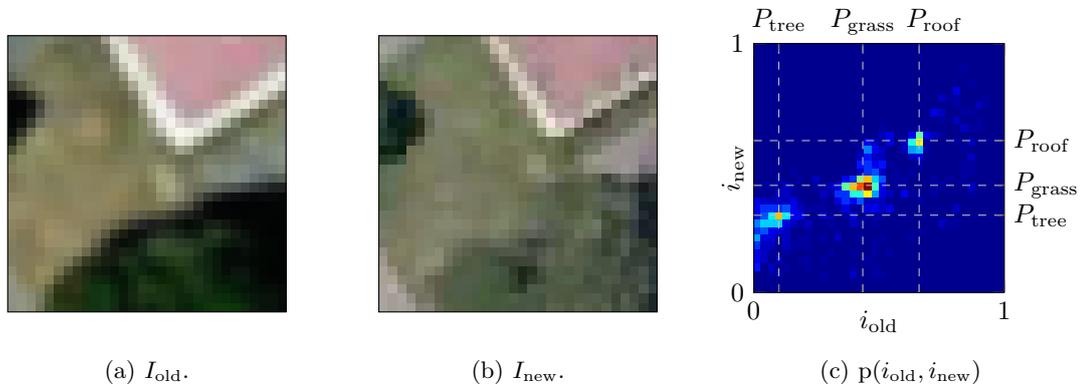


Figure 2.3: Images of an unchanged area in south of Toulouse and the corresponding joint distribution estimated from a histogram.

estimated by using a bivariate histogram. In order to be able to display this distribution in a two dimensional space we only considered the luminance of each optical image. The histogram was then constructed by considering each pair of corresponding pixels in both images as a realization of a random variable. The space was divided in a grid of classes, and the number of realizations corresponding to each class was counted. The class division used to compute the histogram was obtained by cross validation. Red color represents a high count, while dark blue color means a count close to zero. As expected from the proposed model, three main clusters are observed in the joint distribution. With dashed lines the physical property corresponding to each cluster has been highlighted, where (by recalling (2.3) and (2.4)) the parameters of the given cluster can be obtained through $[T_{old}(P), T_{new}(P)]$. In this case, the centroid of the three clusters in Fig. 2.3(c) are located very close to a straight line defined by $\mu_{S_1} = \lambda \mu_{S_2} + \beta$. Since these images have been acquired by the same kind of sensor, the observed relationship is the one associated with images with differences in brightness and contrast [Woj98], where the parameters λ and β account for contrast and brightness respectively.

The estimated correlation coefficient for these images is close to 0.77, while the estimated mutual information is 0.84 (which was obtained by replacing the joint intensity distribution by its histogram). By themselves these values do not provide any useful information. However, we can remark that the correlation coefficient is closer to 1 than to 0, and that the mutual information is different from 0. Nevertheless, these results should be compared to the results obtained in changed areas. A hypothesis test allowing one to show whether these results are significant, different or not in presence and absence

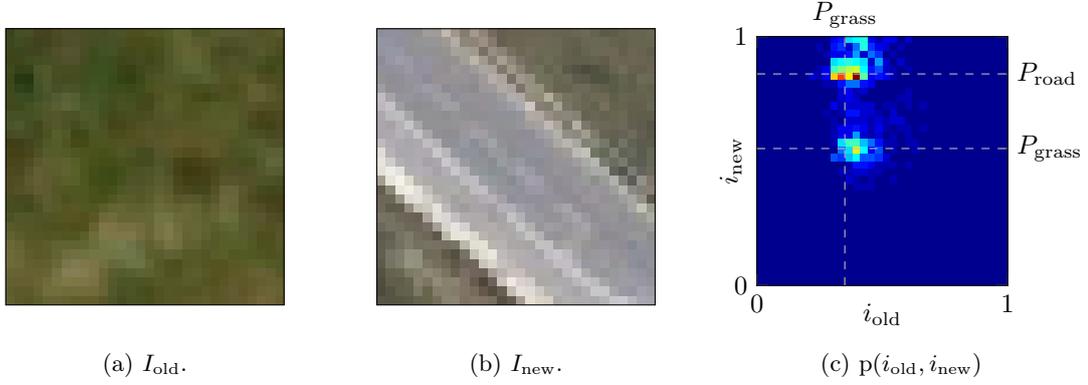


Figure 2.4: Images before and after the construction of a road in south of Toulouse and the corresponding joint distribution estimated from a histogram.

of change could also be performed.

2.3.2 Changed optical images

Figs. 2.4(a) and 2.4(b) show a pair of optical images captured from the same sensor and corresponding to a changed area. The first image is quite uniform and is mainly composed of grass, while a new road takes most of the central portion of the second image. In order to apply the proposed model for this image, we can consider that both images are composed by the intersection of all the objects in the image, i.e., in the second image we have an upper and a lower grass triangle and a central stripe with the road, whereas in the first image we have an upper and lower grass triangle and a central stripe which is also made of grass. Following this method, if there are K_{old} different objects in I_{old} and K_{new} in I_{new} we can conclude that the number of clusters obtained in the joint distribution will belong to the interval $[\max(K_{old}, K_{new}), K_{old} \times K_{new}]$.

In this case, the mixture distribution is expected to have two components, where the parameters for one of them can be obtained from $[T_{old}(P_{grass}), T_{new}(P_{grass})]$, and for the other they can be obtained from $[T_{old}(P_{grass}), T_{new}(P_{road})]$. Since both components share the same parameters in the dimension of i_{old} , both components are expected to be vertically aligned. This result can be clearly seen in Fig. 2.4(c), which was obtained by the same procedure as Fig. 2.3(c). In this situation (i.e., when both components are vertically or horizontally aligned), the correlation coefficient as well as the mutual information are expected to yield estimates close to zero.

For these images, the estimated correlation coefficient is 0.011 while the estimated

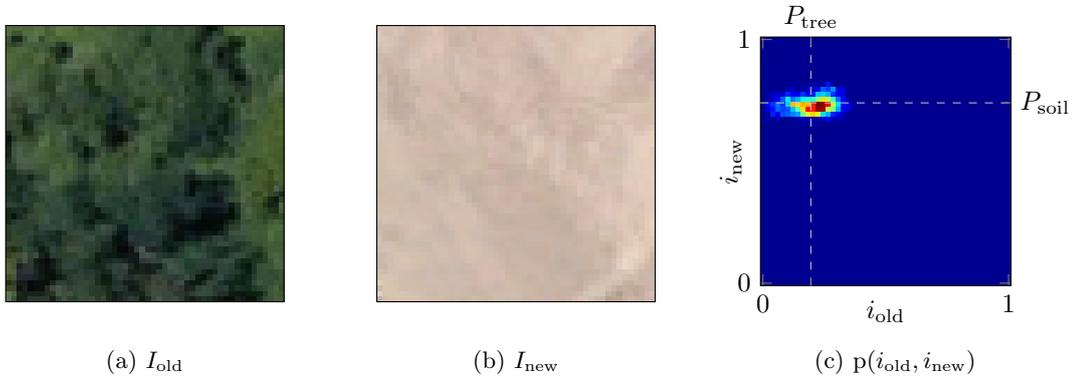


Figure 2.5: Old image with a group of trees, new image preparing the ground for a new construction, and their corresponding joint distribution estimated from a histogram.

mutual information is 0.14 (computed using the joint histogram of Fig. 2.4(c) as the joint distribution estimation). In both cases, this value is closer to zero than the result obtained for unchanged areas. Finally, one could quickly conclude that in circumstances such as those presented in these images the mutual information and the correlation coefficient are good similarity measures to detect changes. However, as it will be shown in what follows, the mutual information and the correlation coefficient are not always good similarity measures for change detection.

2.3.3 Uniform regions

Previously we presented two situations where the correlation coefficient and the mutual information produced significantly different estimates for changed and unchanged areas. However, this is not always the case. Based on the proposed statistical model we can construct cases where both similarity measures always detect a change regardless of the presence or absence of this change. One of these situations can be constructed by considering two different images whose mixture distribution only contains one component, which can be easily achieved by taking two images with only one object each. We would like to recall here that we are assuming that the intensity distribution of each component of the mixture is constructed as the distribution of two independent random variables. In such scenario, we expect both, the correlation coefficient and the mutual information, to detect such region as changed.

Consider two images with a unique object such as those displayed in Figs. 2.5(a) and 2.5(b) where the first image shows a forest area and the second image shows the soil

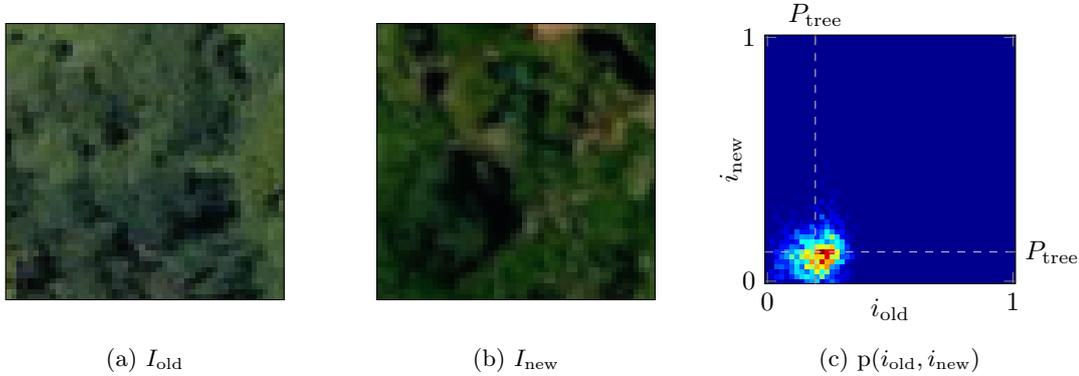


Figure 2.6: (a) and (b) Optical images for an unchanged homogeneous area with different brightnesses and contrasts, and (c) the corresponding joint distribution (estimated using a 2D histogram).

ground of a construction site. These images are clearly different and a similarity measure computed between these two images should clearly detect a change between them. The joint histogram, which is shown in Fig. 2.5(c) (computed using the same procedure as previously described for Fig. 2.3(c)), shows that their distribution consists of only one component whose parameters can be obtained from $[T_{old}(P_{tree}), T_{new}(P_{soil})]$. It can be seen that their joint distribution is elliptical, due to both images having different variances. However they are still independent. In this case the estimated correlation coefficient is 0.019 while the mutual information is 0.062, which is in agreement with the results obtained for the changed images of Fig. 2.4. As expected, a change between these two images is correctly detected.

However, we can also consider the case where the only object present in both images is the same, as depicted in Figs. 2.6(a) and 2.6(b), where both images show the same forest area captured by the same sensor at different dates. These images present no change and should be considered as unchanged. Their joint histogram, which is shown in Fig. 2.6(c) (computed using the same procedure previously described for Fig. 2.3(c)), clearly shows that their distribution consists of only one component whose parameters can be obtained from $[T_{old}(P_{tree}), T_{new}(P_{tree})]$. In this case the computed correlation coefficient is 0.051 while the mutual information is 0.048, which are very similar to those obtained for the images of Fig. 2.5, i.e., corresponds to a change between the two images. In these cases the mutual information, the correlation coefficient, or any dependency measure are clearly bad similarity measures, since they will detect a change

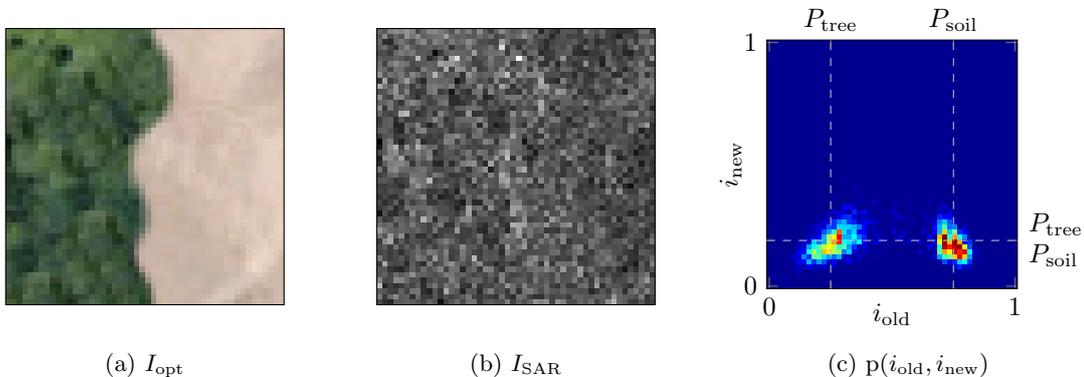


Figure 2.7: (a) Optical image with two different objects, (b) the unchanged image corresponding to a sensor with $T_S(P) = T_{Opt}(P)[1 - T_{Opt}(P)]$ and (c) the corresponding joint distribution (estimated using a 2D histogram).

between Figs. 2.6(a) and 2.6(a) whereas these two images are clearly not affected by any change.

2.3.4 Non linearly related sensors

Another situation where measures based on the dependency between two random variables fail to detect changes can be constructed by choosing a pair of sensors which are related by a function which is not bijective. In this case, we can consider a scene consisting of two objects with physical properties P_1 and P_2 such that $T_S(P_1) = T_S(P_2)$ for at least one of the sensors.

Such a situation is displayed in Figs. 2.7(a) and 2.7(b). Fig. 2.7(a) represents an optical image with half of it is covered by trees, whereas the other half is made of soil. Fig. 2.7(b) represents a SAR image with $T_{SAR}(P) = T_{Opt}(P)[1 - T_{Opt}(P)]$ affected by multiplicative speckle noise. In this case, the two images are not affected by any change and the visual differences in Figs. 2.7(a) and 2.7(b) are only due to differences between $T_{Opt}(P)$ and $T_{SAR}(P)$. Note that the way T_{SAR} has been constructed yields $T_{SAR}(P_{tree}) = T_{SAR}(P_{soil})$. This can be observed in Fig. 2.7(b), where no difference can be seen between the two halves of the image. The joint histogram of the two Figs. 2.7(a) and 2.7(b), which is shown in Fig. 2.6(c) (computed using the same procedure previously described for Fig. 2.3(c)), clearly shows that their distribution consists of two components, where the parameters of each component can be obtained from $[T_{Opt}(P_{tree}), T_{SAR}(P_{tree})]$, and the parameters of the other can be obtained from $[T_{Opt}(P_{soil}), T_{SAR}(P_{soil})]$, which happens

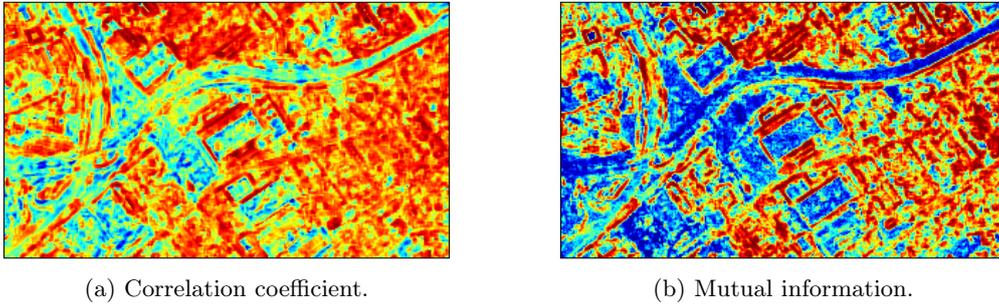


Figure 2.8: Similarity of the images in Figs. 2.2(a) and 2.2(b) as measured by (a) the correlation coefficient and (b) the mutual information.

to be the same as $[T_{\text{opt}}(P_{\text{soil}}), T_{\text{SAR}}(P_{\text{tree}})]$. This situation is very similar to the one observed in Fig. 2.4(c), and thus we expect that the correlation coefficient and the mutual information will produce a measure indicating the presence of a change. The estimated correlation coefficient is close to -0.058 for this example, while the mutual information yields a similarity of 0.13 , which are in accordance with those in Fig. 2.4. In this situation, as in Fig. 2.6, the correlation coefficient and the mutual information are not appropriate to detect the absence of change between the two images.

2.3.5 Real images

Previously we analyzed the behavior of the correlation coefficient and of the mutual information (which can be extended to any measure based on statistical dependency) in different situations and detected several situations where these measures are not suited to detect changes. This phenomenon can also be observed in real images, as those shown in Fig. 2.8. These results are based on the images in Fig. 2.2. Figs. 2.8(a) and 2.8(b) shows the estimated mutual information and correlation coefficient for all the moving windows within the image, where we would expect to see in blue (low similarity) a pattern similar to the change highlighted in Fig. 2.2(c). However, the most prominent changes detected by both methods are found on the football fields in the center of the image, the channel across the image, and the green area at the side of the channel. This behavior is easily explained by the previous analysis, since the football fields consist in uniform areas bigger than the analysis window, so that these windows show the behavior described in Section 2.3.3. This problem has been partially addressed by [IM09] by considering variable window sizes, trying to avoid sizes of the moving windows that

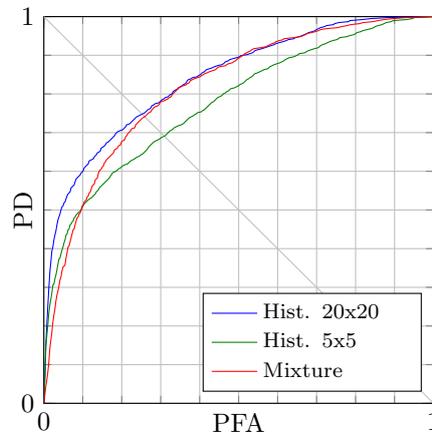


Figure 2.9: Even if the proposed model results in an improved model of the joint pixel intensities, its use does not reflect in any measurable improvement in the change detection performance compared to a histogram description when the mutual information is used as a similarity measure, as shown by these ROC curves.

result in them containing only uniform regions.

Another factor to consider is that whenever we compute a similarity measure, we discard some information considered as irrelevant and we keep some information considered as relevant. These irrelevant and relevant quantities are summarized into a single quantity, namely the similarity measure. When considering the mutual information or the correlation coefficient, we arbitrarily decide that the relevant information is contained solely in the dependency of the two random variables. This property yields a limit on the performance that can be obtained using this similarity measure. Even when the joint distribution estimation is improved, as shown in Section 2.2.3, the resulting change detection performance is not necessarily improved when using the mutual information as a similarity measure. To prove this, we applied a mutual information based change detector on the images presented in Fig. 2.2. In order to compute the mutual information, we estimated the joint distribution in 3 ways: using a histogram description with 5×5 classes, using 20×20 classes, and using the proposed method and the parameter estimation algorithm from Section 3.3. We used the receiver operating characteristic (ROC) curves depicted in Fig. 2.9 as a performance evaluation, where it can be observed that the performance obtained when using the mutual information to detect changes is not necessarily improved when a better statistical model is used for the images, motivating the definition of a new similarity measure for change detection.

2.4 Conclusions

This chapter introduced a statistical model to describe the joint distribution of any finite number of images independently of the kind of sensors used for their acquisition. This model is based on the assumption that the image acquired by different sensors are corrupted by independent noises. This implies that the observed pixel intensities are independent conditional to an unobserved variable P that represents the physical properties of the object being imaged. Extending this pixel intensity model to an analysis window results in a mixture distribution model.

The proposed model was used to analyze the performance of some classical change detection algorithms. Not only significant flaws were detected on dependency based similarity measures, but we have also shown that improving the statistical model used to describe the joint pixel intensity distribution does not necessarily improve the detector performance compared to simpler methods, such as those using a histogram. This allows us to conclude that improving the statistical description of an analysis window is worthless if it is only used to obtain a statistical dependency based measure. As a consequence a new similarity measure is required if we want to improve the detection performance.

CHAPTER 3

A new manifold based similarity measure

Contents

3.1	Manifold definition	38
3.1.1	Unchanged areas	39
3.1.2	Changed areas	40
3.2	A new similarity measure	41
3.2.1	Distance to the manifold	42
3.2.2	Estimation of the distance to the manifold	45
3.3	Parameter estimation	47
3.4	Performance on synthetic images	49
3.5	Conclusions	56

As stated in Chapter 2, similarity measures based on the statistical dependency between the pixel intensities are not necessarily best suited to detect changes between images acquired by heterogeneous sensors. To fill this gap, the present chapter proposes a new similarity measure based on the statistical model introduced in Chapter 2. More precisely, we assume that some properties of the mixture model parameters hold on unchanged areas only. This finding leads to the definition of a new a similarity measure based on the parameters of this statistical model.

Section 3.1 introduces the assumption that in the absence of change the parameters of the mixture model clusters defined in (2.29), should belong to a given manifold. This manifold depends upon the physical properties P of the objects in the scene and upon the sensor types. In Section 3.2, this assumption is used to build a similarity measure based on the distance between the observed parameters and the manifold. A simple approach based on the expectation maximization (EM) algorithm is proposed in Section 3.3 to estimate the parameters of the mixture model, which is required to obtain the similarity measure. Section 3.4 evaluates the performance of the method on synthetic images.

Finally conclusions are drawn in Section 3.5.

3.1 Manifold definition

Consider the vector $\mathbf{v}(P_1, \dots, P_D) = [T_1(P_1), \dots, T_D(P_D)]^T$ where P_d represents the physical properties of the object observed by the sensor S_d . The vector $\mathbf{v}(P_1, \dots, P_D)$ defines a parametric function of $[P_1, \dots, P_D]^T$ from \mathbb{P}^D (where \mathbb{P} represents the space where P lives) in the space $T_1(\mathbb{P}) \times \dots \times T_D(\mathbb{P})$, i.e., the space containing all the possible parameters of the joint distribution defined by

$$p(\mathbf{i}) = \prod_{d=1}^D p(i_d | P_d). \quad (3.1)$$

Now consider the particular case where all the sensors observe the same object. In other words, the object has not changed from one acquisition to the other. We can express this as $P_1 = P_2 = \dots = P_D = P$ and the parameter vector denoted as $\mathbf{v}(P)$ is a parametric function from \mathbb{P} to the space $T_1(\mathbb{P}) \times \dots \times T_D(\mathbb{P})$. Even though $\mathbf{v}(P_1, \dots, P_D)$ and $\mathbf{v}(P)$ have the same image space, the domain space of $\mathbf{v}(P)$ is reduced. Our assumption is that this space reduction leads the vector $\mathbf{v}(P) = [T_1(P), \dots, T_D(P)]^T$ to live within a manifold denoted as $\mathcal{M} \subset T_1(\mathbb{P}) \times \dots \times T_D(\mathbb{P})$. In what follows we will refer to areas corresponding to $P_1 = P_2 = \dots = P_D = P$ as unchanged areas. Another assumption is that in changed areas there is no reason for the vector $\mathbf{v}(P_1, \dots, P_D)$ to belong to \mathcal{M} . This difference will allow the manifold \mathcal{M} to be used as a reference for change detection.

The manifold \mathcal{M} characterizes the relationships between the involved sensors. It should be noted that, from (2.2) and (2.3), the equation (2.23) in unchanged areas can be expressed using $\mathbf{v}(P)$ as

$$\mathcal{F}(P) = \mathcal{F}[\mathbf{v}(P)] \quad (3.2)$$

i.e., the distribution family describing a component of the mixture (2.29) can be completely parametrized by $\mathbf{v}(P)$.

For instance, consider two sensors S_1 and S_2 differing only by their brightness and contrast such that $T_2(P) = \lambda_1 T_1(P) + \lambda_2$ [Woj98], the vector $\mathbf{v}(P)$

$$\begin{aligned} \mathbf{v}(P) &= [T_1(P), T_2(P)]^T \\ &= [T_1(P), \lambda_1 T_1(P) + \lambda_2]^T \end{aligned}$$

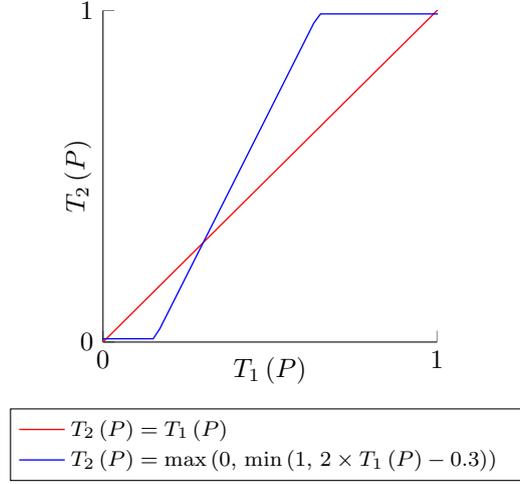


Figure 3.1: Manifolds \mathcal{M} obtained with simple sensor combinations (Red curve: two identical sensors, Blue curve: sensors with different brightness and contrast with a saturation).

defines a manifold \mathcal{M} in the space $T_1(P) \times T_2(P)$ whose representation is a straight line. The case where $\lambda_1 = 1$ and $\lambda_2 = 0$ is depicted by the red line in Fig. 3.1. Note that situations such as sensor saturations would lead to more complex expressions for $\mathbf{v}(P)$ leading to non linear manifold, as depicted by the blue curve in Fig. 3.1. However, in general the relationship between $T_i(P)$ and $T_j(P)$ is unknown for $j \neq i$. Moreover, the generic (i.e., for any sensor S) transformations $T_S(P)$ are generally intractable. Section 3.2 proposes a method to build a similarity measure based on the manifold \mathcal{M} defined by $\mathbf{v}(P)$.

3.1.1 Unchanged areas

As previously discussed, since an object is the same for all images associated with an unchanged area, the physical properties P for a given pixel do not change for all the images. Such object will produce a component in the mixture distribution (2.29) following the result in (3.2). Consequently, the parameter vector describing this component belongs to the manifold \mathcal{M} , defined under the “no change” hypothesis.

Figs. 3.2(a) and 3.2(b) displays two synthetic images that were generated from a ground truth noiseless image by adding independent additive Gaussian noises with the same variance for both images. These images represent two optical images acquired with the same sensor and captured with the same settings (i.e., brightness and contrast). If we consider the different objects contained in these images we can retrieve the ground truth

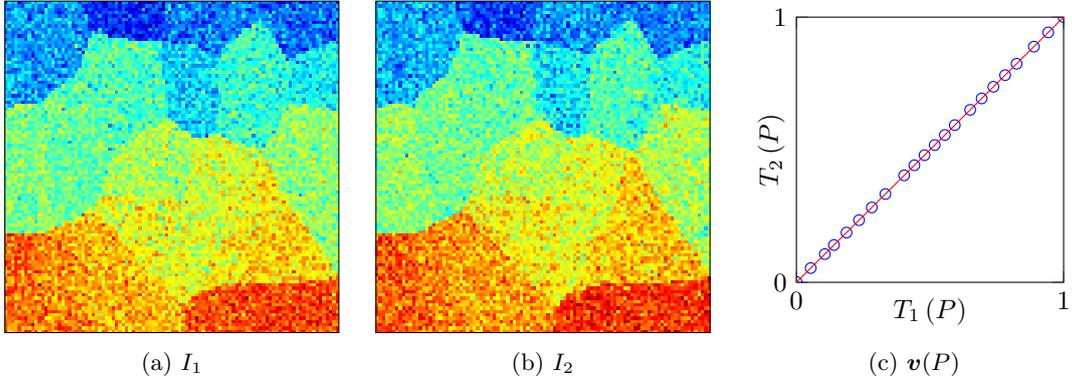


Figure 3.2: (a) and (b) are two (different) synthetic images obtained for the same sensor on a unchanged area. The $\mathbf{v}(P)$ of every each region is shown in (c).

values $\mathbf{v}(P)$ that generated them. Fig. 3.2(c) shows the location of the vectors $\mathbf{v}(P)$ in the space $T_1(P) \times T_2(P)$. Since both images were generated by adding noise to the same ground truth image, the manifold relating them is the straight line $T_1(P) = T_2(P)$, as depicted by the red line in Fig. 3.2(c). As expected, the values of $\mathbf{v}(P)$ follow exactly the relationship described by the red manifold \mathcal{M} in Fig. 3.1.

Fig. 3.3 illustrates a slightly more complex situation. Fig. 3.3(a) was generated in the same way as in Fig. 3.2(a). However, in order to generate Fig. 3.3(b) the ground truth noiseless image was transformed using the brightness and contrast transformation $\min(1, \max(0, 2x - 0.3))$. The blue areas of Fig. 3.3(b) are those with low pixel intensity and saturated towards zero, and the red areas are those with high pixel intensity and saturated towards one. These images correspond to two images captured by the same sensor, but having different brightness and contrast. Fig. 3.3(c) shows the location of the vectors $\mathbf{v}(P)$ in the space $T_1(P) \times T_2(P)$. Because of the way the images have been generated, the manifold relating $T_1(P)$ and $T_2(P)$ is the blue curve of Fig. 3.1, also depicted by the red line in Fig. 3.3(c). As expected, the values of $\mathbf{v}(P)$ follow exactly the relationship described by manifold.

3.1.2 Changed areas

Now let observe what happens when a change occurs in the observed scene between the two acquisitions. On changed areas, the object present on each image is not the same, and thus the physical properties observed by the sensor S_D are not the same. A pixel in this situation cannot belong to a mixture component with distribution $\mathcal{F}[\mathbf{v}(P)]$.

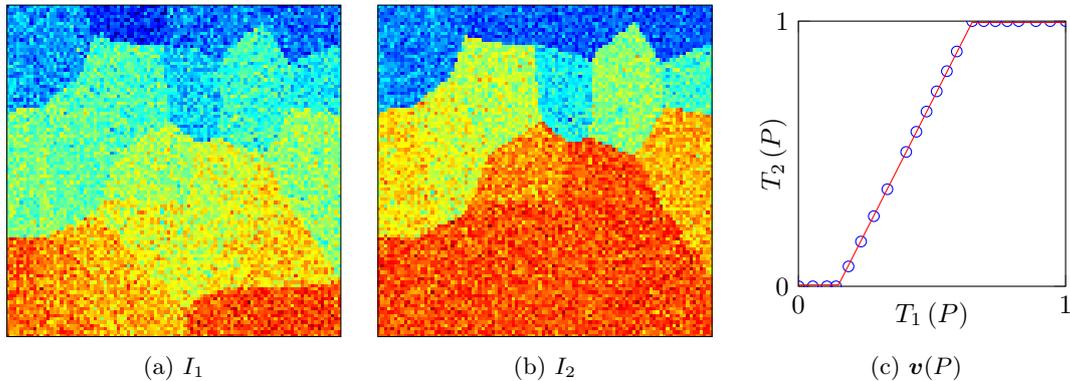


Figure 3.3: (a) and (b) are two synthetic images obtained for the same sensor with different brightness and contrast settings on an unchanged area. The $\mathbf{v}(P)$ of every each region is shown in (c).

Such pixel belongs to a component of a mixture distribution with parameters given by $\mathbf{v}(P_1, \dots, P_D)$.

This can be easily observed in Figs. 3.4 and 3.5. In this case, the second image in Figs. 3.2 and 3.3 has been rotated by 90° in order to introduce changes in the scene. Figs. 3.2(c) and 3.3(c) show the location of the corresponding vectors $\mathbf{v}(P_1, P_2)$ that produce each component of the mixture distribution. As observed, the vectors $\mathbf{v}(P_1, P_2)$ are no longer constrained to belong to the manifold \mathcal{M} highlighted with a red line in each case. We would like to highlight that some vectors $\mathbf{v}(P)$ do belong to the manifold. However, these vectors correspond to the areas present in the middle of the image which are not affected by the applied rotation.

3.2 A new similarity measure

Sections 3.1.1 and 3.1.2, suggest that the distance between $\mathbf{v}(P_1, \dots, P_D)$ and the manifold \mathcal{M} can be exploited for change detection. Several parameter vectors \mathbf{v} can be estimated on each analysis window, one for each object within the window. When an object within the analysis window does not change in all images, the estimation $\hat{\mathbf{v}}$ of \mathbf{v} for the component representing that area should belong to the manifold \mathcal{M} . When considering a changed areas, $\hat{\mathbf{v}}$ is not subject to the same constraint, which can be expressed

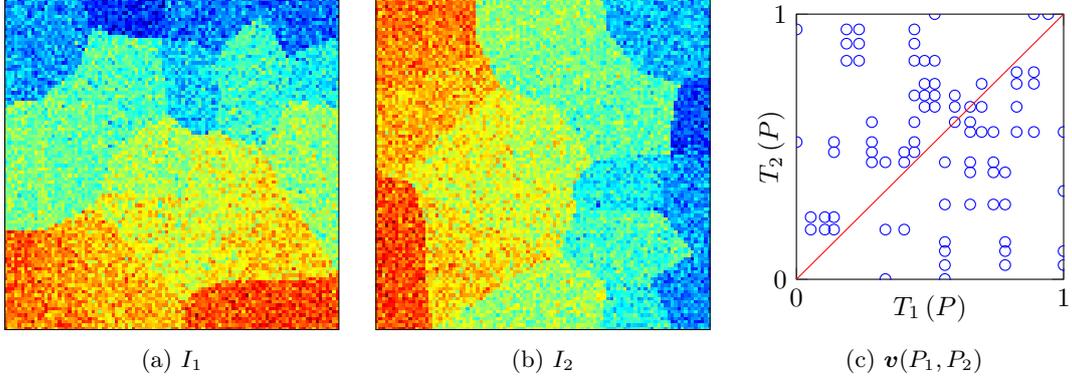


Figure 3.4: (a) and (b) are two synthetic images obtained for the same sensor on a changed area. The $\mathbf{v}(P_1, P_2)$ of every each region is shown in (c).

as

$$\begin{cases} \hat{\mathbf{v}} = \hat{\mathbf{v}}(P) & \text{if absence of change} \\ \hat{\mathbf{v}} = \hat{\mathbf{v}}(P_1, \dots, P_D) & \text{if presence of change} \end{cases} \quad (3.3)$$

where $[P_1, \dots, P_D]^T \neq [P, \dots, P]^T$. In other words, if the estimated vector $\hat{\mathbf{v}}$ belongs to \mathcal{M} , then the two scenes are the same; otherwise, they are different.

Since the transformations T_{S_d} for $d = 1, \dots, D$ are *a priori* unknown, it is not possible to obtain an analytical description of the manifold \mathcal{M} . To overcome this situation, it is proposed to estimate the manifold \mathcal{M} from a training dataset composed of unchanged images, considering P as a hidden variable.

3.2.1 Distance to the manifold

The parameter vectors associated with different components of the mixture distribution can be estimated from the pixel intensities observed within an analysis window W to obtain several estimated vectors $\hat{\mathbf{v}}$. Since these vectors $\hat{\mathbf{v}}$ are estimates, they will be subject to an estimation error, i.e., a variance, and this will not exactly lie on \mathcal{M} , but in a small neighborhood around it. Consequently, testing whether $\hat{\mathbf{v}}$ belongs exactly to \mathcal{M} is not a good strategy, since doing so would result in a high rate of false positives, i.e., detecting changes for unchanged areas. This effect can be observed in Fig. 3.6. On the left hand side of Fig. 3.6(a) the vectors $\mathbf{v}(P)$ obtained from the ground truth images used to generate Figs. 3.2(a) and 3.2(b) are plotted in the space $T_1(P) \times T_2(P)$. The right hand side of Fig. 3.6(a) displays the vectors $\hat{\mathbf{v}}(P)$ obtained after estimating $\mathbf{v}(P)$

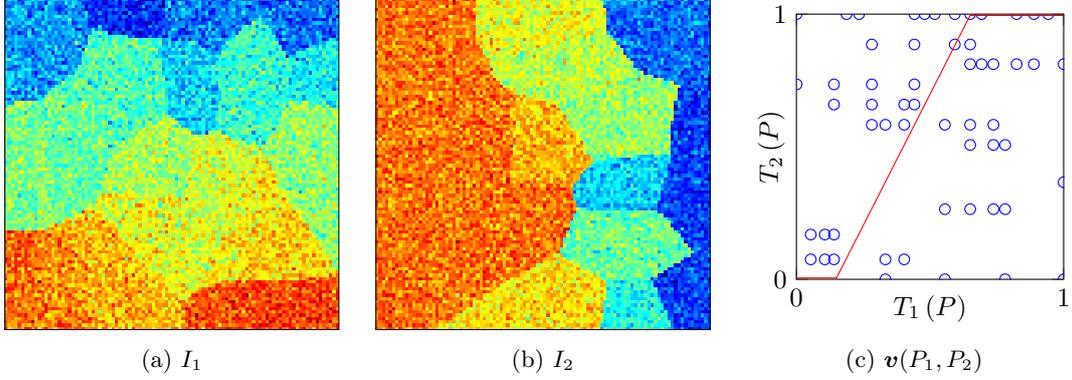


Figure 3.5: (a) and (b) are two synthetic images obtained for the same sensor with different brightness and contrast settings on a changed area. The $v(P_1, P_2)$ of every each region is shown in (c).

from the data associated with an unchanged area. For this estimation we considered the ground truth information about the location of the objects within the image in order to reduce the complexity of this example and focus just on discussed matter, however it should be noted that it is unknown in real case scenarios (which is considered later in Section 3.3). More precisely, the values of $\hat{v}(P)$ were computed using a maximum likelihood estimator. It can be clearly seen in Fig. 3.6(a) that the vectors \hat{v} (obtained from unchanged areas) do not lie exactly on the manifold. The same remark can be made for Fig. 3.6(b), where the same process was applied to the images in Figs. 3.3(a) and 3.3(b).

The distance between a vector \hat{v} and the manifold \mathcal{M} will be denoted as $d_{\mathcal{M}}(\hat{v})$ in what follows. We propose to investigate a hypothesis test to decide whether there is a change or not by thresholding this distance measure as follows

$$d_{\mathcal{M}}(\hat{v}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau \quad (3.4)$$

where H_0 and H_1 are the two hypotheses

H_0 : Absence of change

H_1 : Presence of change

and where τ is related to the probability of false alarm (PFA) (i.e., the probability of detecting an unchanged area as changed) and the probability of detection (PD) (i.e., the probability of correctly identifying a change) of the detector. The change detection

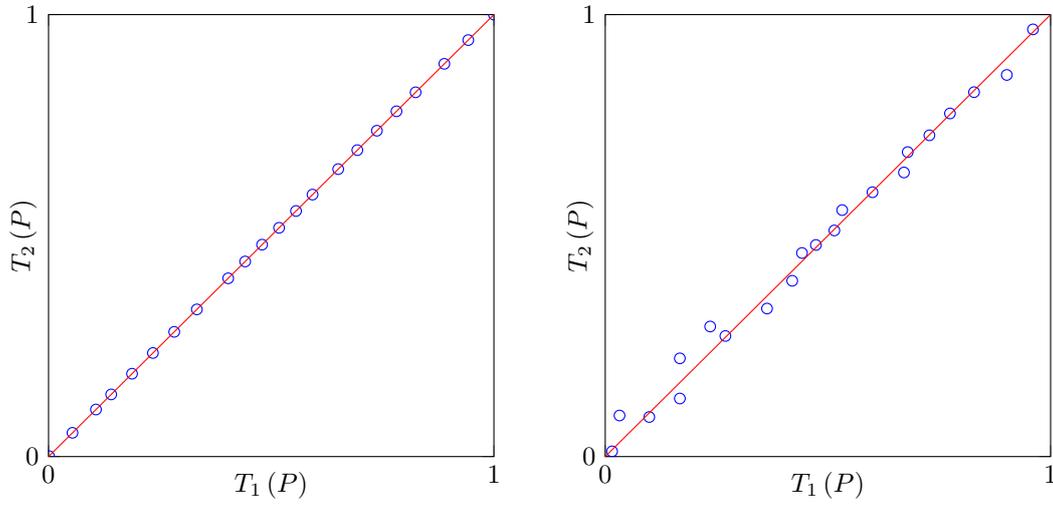
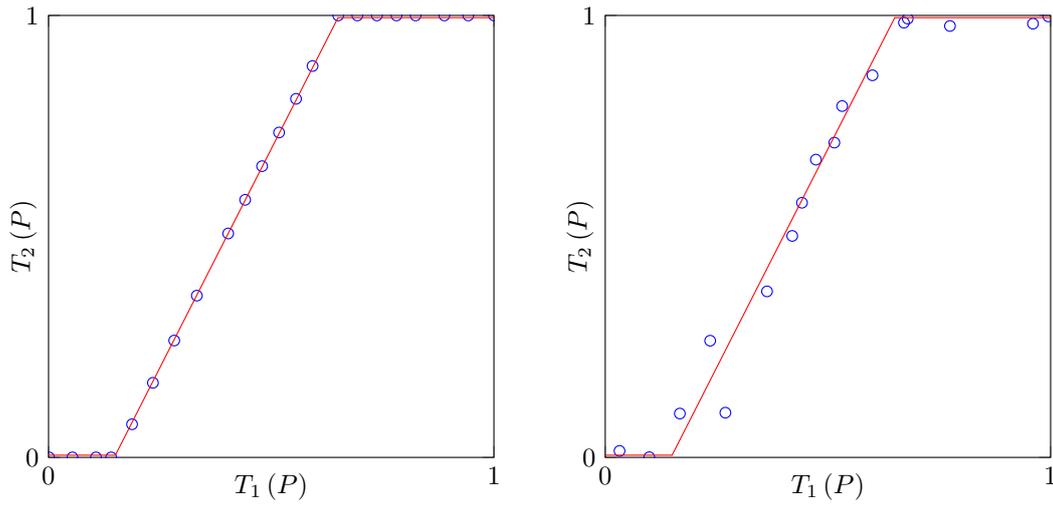
(a) Comparison between $\mathbf{v}(P)$ and $\hat{\mathbf{v}}$ for the images in Fig. 3.2.(b) Comparison between $\mathbf{v}(P)$ and $\hat{\mathbf{v}}$ for the images in Fig. 3.4.

Figure 3.6: Comparison between the (left) ground truth vectors $\mathbf{v}(P)$ or $\mathbf{v}(P_1, P_2)$ with the (right) vectors $\hat{\mathbf{v}}$ obtained from noisy data.

problem now reduces to estimating a distance from $\hat{\mathbf{v}}$ to the manifold \mathcal{M} . As shown in Section 3.2.2, this estimation can be directly performed without estimating the manifold \mathcal{M} .

3.2.2 Estimation of the distance to the manifold

The manifold \mathcal{M} representing objects belonging to unchanged areas is *a priori* unknown, and could be estimated by using training data associated with unchanged areas. However, we propose to directly estimate the distance to \mathcal{M} rather than its exact location in the space.

First we would like to note that $\mathbf{v}(P)$ is not a random variable, but a deterministic unknown value. However, we can think of a deterministic value x_0 as a random variable X with a pdf such that

$$p_X(x) \propto \mathbb{1}_{x_0}(x) \quad (3.5)$$

where $\mathbb{1}_{x_0}(\cdot)$ is the indicator function located in x_0 . This way we can express $\mathbf{v}(P)$ as a random variable. Moreover, we know that the indicator function used to represent this random variable will indicate a value belonging to the manifold \mathcal{M} . If we now consider $\mathbf{v}(P)$ to be unknown, it can be seen as a random variable taking values on \mathcal{M} , such that

$$p(\mathbf{v}(P)) \propto \mathbb{1}_{\mathcal{M}}[\mathbf{v}(P)] \quad (3.6)$$

Let consider $p(\hat{\mathbf{v}}|\text{no change})$, i.e., the pdf of $\hat{\mathbf{v}}$, in the absence of change. Since each $\hat{\mathbf{v}}$ is estimated using a maximum likelihood estimator from several pixels, the asymptotic distribution of this estimator can be used (when enough samples are used to compute the estimate) to approximate $p(\hat{\mathbf{v}}|H_0)$. As a consequence

$$\hat{\mathbf{v}} - \mathbf{v}(P) \sim \mathcal{N}(0, \sigma_N^2) \quad \text{under } H_0 \text{ and for } N \text{ large enough.} \quad (3.7)$$

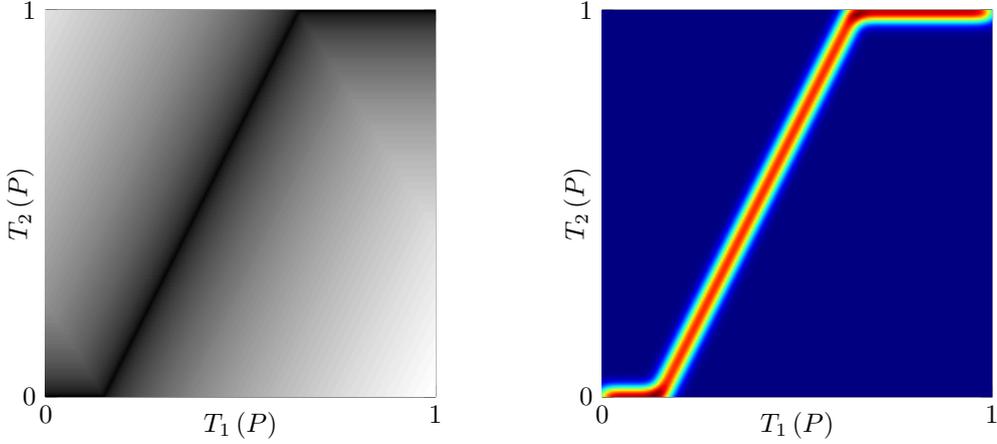
where σ_N is a decreasing function of N . In what follows we will assume that N is always large enough so that the Gaussian approximation in (3.7) holds.

Since the training data arises from vectors $\hat{\mathbf{v}}$ associated with different values of P (and hence, of $\mathbf{v}(P)$), the distribution of $\hat{\mathbf{v}}$ can be expressed as the convolution between \mathcal{M} and a zero mean Gaussian distribution. Indeed,

$$p(\hat{\mathbf{v}}) = \int_{\mathbf{v}(P)} p[\hat{\mathbf{v}} - \mathbf{v}(P)|\mathbf{v}(P)] p[\mathbf{v}(P)] d\mathbf{v}(P) \quad (3.8)$$

$$p(\hat{\mathbf{v}}) \propto \int_x p_{\mathcal{N}(0, \sigma^2)}(\hat{\mathbf{v}} - x) \mathbb{1}_{\mathcal{M}}(x) dx \quad (3.9)$$

where we considered $x = \mathbf{v}(P)$, and replaced $p[\hat{\mathbf{v}} - \mathbf{v}(P)|\mathbf{v}(P)]$ and $p[\mathbf{v}(P)]$ by their pdf, a zero mean Gaussian pdf and a density proportional to the manifold indicator function. The resulting density $p(\hat{\mathbf{v}})$ can be seen as a low-pass version of the manifold indicator



(a) Euclidean distance from the space $T_{S_1} \times T_{S_2}$ to \mathcal{M} for the sensors described by the blue curve in Fig. 3.1. (b) Density $p(\hat{\mathbf{v}})$ for the sensors described by the blue manifold in Fig. 3.1.

Figure 3.7: Relationship between Fig. 3.7(a) the distance to the manifold \mathcal{M} and Fig. 3.7(b) the density $p(\hat{\mathbf{v}})$. Black (blue on the right) color represent a value near zero while white (red on the right) represents a high value.

function, which means it will take higher values near \mathcal{M} and low values far from \mathcal{M} , behaving like an inverse distance measure, or a “closeness” measure.

This relationship is illustrated in Fig. 3.7 for the manifold \mathcal{M} representing two similar sensors with different brightness and contrast. Precisely, Fig. 3.7(a) shows the euclidean distance from any point in the space $T_1(P) \times T_2(P)$ to \mathcal{M} , with black representing a distance of zero, and white representing a high distance. Fig. 3.7(b) shows the convolution of \mathcal{M} with a zero mean normal distribution and represents the theoretical pdf of $p(\hat{\mathbf{v}})$, where blue color represents a probability of zero and red color represents a high probability.

Based on this observation, we propose to use $p(\hat{\mathbf{v}})$, or more precisely its inverse $\frac{1}{p(\hat{\mathbf{v}})}$, a similarity measure for change detection. Note that $\frac{1}{p(\hat{\mathbf{v}})}$ assigns low values to points near \mathcal{M} and high values to points far from \mathcal{M} as a pseudo distance between $\hat{\mathbf{v}}$ and \mathcal{M}

$$d_{\mathcal{M}}(\hat{\mathbf{v}}) = \frac{1}{p(\hat{\mathbf{v}})} \quad (3.10)$$

It should be noted that since we are considering (3.7) to be valid, the density $p(\hat{\mathbf{v}})$ is never zero, so that its inverse is well defined. We consider important to note that $\frac{1}{p(\hat{\mathbf{v}})}$ does not define a distance in the strict mathematical sense.

As discussed in the beginning of Section 3.2, this pseudo distance can thus be used as a similarity measure for change detection. The density $p(\hat{\boldsymbol{v}})$ can be estimated from a set of vectors $\hat{\boldsymbol{v}}$ obtained from unchanged area used as training data. The estimation of $p(\hat{\boldsymbol{v}})$ can be conducted using well-known density estimation techniques such as Parzen windows [KC02]. In such case, we can redefine the similarity measure as

$$d_{\mathcal{M}}(\hat{\boldsymbol{v}}) = \frac{1}{\hat{p}(\hat{\boldsymbol{v}})}. \quad (3.11)$$

The proposed similarity measure has several interesting advantages when compared to the Euclidean distance. The first one is that less resources are required to estimate $p(\hat{\boldsymbol{v}})$ since it is not required to estimate the manifold, but just $p_{\mathcal{M}}(\boldsymbol{v})$. The second and most interesting reason is that contrary to the Euclidean distance, $p(\hat{\boldsymbol{v}})$ is the probability of finding an unchanged parameter in that area of the space. Taking into account (3.7), this allows us to define a threshold as a function of a given probability of false alarm. However, it should be noted that this threshold would be dependent on the number of samples used to obtain the vectors $\boldsymbol{v}(P)$.

3.3 Parameter estimation

In order to compute the similarity measure proposed in Section 3.2.2, it is first required to estimate the vectors $\hat{\boldsymbol{v}}$. The algorithm introduced in [FJ02], which is based on the classical EM algorithm is used for this estimation. The choice for this algorithm is due to its capability to determine the number of components in the mixture model (i.e., objects in the scene), within a predefined range, using a heuristic approach based on an information measure. The algorithm, described in Algo. 1, begins with an upper bound of the number of components, and gradually removes those that do not describe enough samples. In Algo. 1, *InfLogLikelihood* is a modified *LogLikelihood* based on information theory criteria, and $\dim(\boldsymbol{v}')$ is the dimension or number of elements in the vector \boldsymbol{v}' .

As previously mentioned, each component of the mixture distribution describes one object. However, even if the component parameters are well estimated, it is not always clear which pixels belong to which component, specifically when two or more components share similar parameters. This effect can be observed in Fig. 3.8, where Fig. 3.8(d) shows a significant noise even though the cluster parameters are well detected. This problem becomes more important when the number of objects in the image increases, and thus, estimating a pixel-wise distance measure is not recommended. Instead, as it is the case

Algorithm 1: Expectation Maximization based algorithms from [FJ02].

Input: $I = \{i_1, \dots, i_N\}$, K_{\max} , K_{\min} , \max_{iter} , τ

Output: $V' = \{v'_1, \dots, v'_K\}$, $w = \{w_1, \dots, w_K\}$

```

1  $K \leftarrow K_{\max}$ ;
2 while  $K \geq K_{\min}$  do
3    $n_{\text{iter}} \leftarrow 0$ ;
4   while  $\text{LogLikelihood} - \text{LogLikelihood}_{\text{old}} > \tau$  and  $n_{\text{iter}} < \max_{\text{iter}}$  do
5      $\text{LogLikelihood}_{\text{old}} \leftarrow \text{LogLikelihood}$ ;
6      $\{V', w, \text{LogLikelihood}\} \leftarrow \text{EM-Step}(I, V', w)$ ;
7      $k_{w \min} \leftarrow \arg \min_k w_k$ ;
8     if  $w_{k_{w \min}} < \frac{\dim(v'_k)}{2N}$  then
9        $\left[ \text{remove component } k_{w \min}; K \leftarrow K - 1; \right.$ 
10       $n_{\text{iter}} \leftarrow n_{\text{iter}} + 1;$ 
11       $k_{w \min} \leftarrow \arg \min_k w_k$ ;
12       $\text{remove component } k_{w \min}; K \leftarrow K - 1;$ 
13       $\text{InfLogLikelihood} \leftarrow$ 
            $\text{LogLikelihood} - \frac{\dim(v')}{2} \sum_1^K \log(w_k) - \frac{\dim(v')+1}{2} K \log(N)$ ;
14      if  $\text{InfLogLikelihood} > \text{InfLogLikelihood}_{\text{best}}$  then
15         $V'_{\text{best}} \leftarrow V'$ ;
16         $w_{\text{best}} \leftarrow w$ ;
17  $V' \leftarrow V'_{\text{best}}$ ;
18  $w \leftarrow w_{\text{best}}$ ;

```

with other classic distance measures (correlation coefficient and mutual information for example), it is proposed to compute a unique window-based similarity measure that indicates the similarity between two windows. To achieve that, it is proposed to obtain a window similarity as the average similarity of each component, i.e.,

$$d_W = \sum_{k=1}^K w_k d_{\mathcal{M}}(\hat{v}_k) \quad (3.12)$$

where d_W is the window similarity, $d_{\mathcal{M}}(\hat{v}_k)$ is the distance to the manifold \mathcal{M} of the k -th component of the mixture distribution, as defined in (3.11), and w_k is the component weight from (2.29). Other ways of addressing the estimation of the mixture distribution

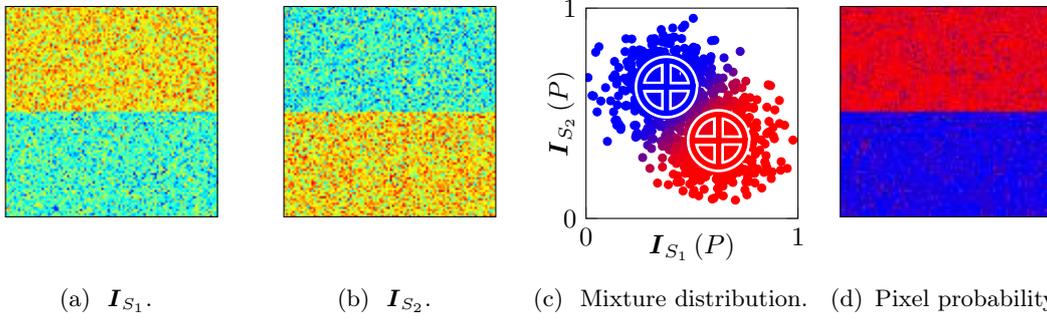


Figure 3.8: Figures (a) and (b) show two images captured by sensors S_1 and S_2 respectively. Figure (c) shows the mixture distribution of their pixel intensity, and their estimated parameters represented by a circle with a cross. Figure (d) presents the resulting probability of each pixel belonging to each component.

that overcome the main limitations of the method just presented will be proposed in Chapter 4.

3.4 Performance on synthetic images

The validity of the proposed similarity measure is assessed by evaluating its performance through the receiver operating characteristic (ROC) curves obtained when detecting changes on a dataset consisting of synthetic images.

The change detection results are compared with those obtained with different classical methods, namely, mean pixel difference, mean pixel ratio, correlation coefficient and mutual information. The first two reference change detection methods were provided by the ORFEO Toolbox [OTB14]. The change detection results are compared with those obtained with classical methods.

The images shown in Figs. 3.9(a) and 3.9(b) were created by generating a synthetic scene P composed of triangular patches representing the different objects. The objects generated on the bottom half of Fig. 3.9(a) are the same as those of Fig. 3.9(b), while the upper half has changed. This difference is shown in Fig. 3.9(c), where the areas that changed between the two images are indicated in black. The first synthetic scene was corrupted by additive Gaussian noise with $\text{SNR} = 30\text{dB}$ to form the synthetic optical image Fig. 3.9(e). To generate the SAR image, a known transformation was first applied to the scene P to simulate the relationship between optical and SAR sensors. More

precisely, the following transformation

$$\mathbf{v}_P = [T_{\text{opt}}(P), T_{\text{SAR}}(P)] \quad (3.13)$$

with $T_{\text{SAR}}(P) = T_{\text{opt}}(P)[1 - T_{\text{opt}}(P)]$ (as depicted in Fig. 3.9(d)) was used for this experiment. Note that the relationship between $T_{\text{opt}}(P)$ and $T_{\text{SAR}}(P)$ is nonlinear. The resulting image was corrupted by multiplicative gamma noise with shape parameter $L = 5$ to obtain the image in Fig. 3.9(f).

The images displayed in Fig. 3.10 compare the similarity measures obtained with the proposed method, the correlation coefficient, the mutual information, the mean pixel difference and the mean pixel ratio. These results were obtained using window sizes optimized by cross validation to produce the best performance for each method. For the correlation coefficient and the mutual information, a window size of 50×50 pixels was found to yield better results, whereas a window size of 21×21 pixels was used for the mean pixel difference and the mean pixel ratio. Note that the difference in the window sizes is due to the inefficiency of the correlation coefficient and the mutual information for small homogeneous windows (as described in Section 2.3), thus requiring bigger (and thus more likely heterogeneous) windows. The mutual information was computed by integrating numerically the joint distribution derived in Chapter 2. A window size of 20×20 pixels was selected for the proposed method. It is important to note that an undersized window means a small number of pixels to estimate the mixture distribution. Obviously, this results in a high variance of the model parameter estimates. On the other hand, an oversized window results in too many components, reducing the identifiability of the different objects.

Fig. 3.11(a) displays the ROC curves for the different methods. In order to compare the performance of the different methods, we propose to choose a detection threshold corresponding to $\text{PFA} = 1 - \text{PD} = \text{PND}$, located in the diagonal line displayed in Fig. 3.11(a). Table 3.1 shows the values of PFA obtained with the different methods, confirming the good performance of the proposed method.

We evaluated also the performance of the proposed strategy for different values of the signal to noise ratio (SNR) associated with the optical image. The ROC curves of the proposed method obtained for different SNRs are shown in Fig. 3.11(b), where it can be observed that the change detection performance is not affected for $\text{SNR} \geq 10\text{dB}$. The performance drop obtained for lower SNR is due to errors in the estimation of the mixture parameters, which are difficult to estimate in the presence of significant noise.

Table 3.1: Performance of the different methods for detecting changes on the images of Fig. 3.11(a).

Method	PFA = PDN
Proposed method	5.52%
Correlation Coefficient	27.68%
Mutual Information	26.35%
Mean Pixel Difference	38.47%
Mean Pixel Ratio	30.60%

Finally, to demonstrate the importance of accurately estimating the number of component of the mixture distribution, we intentionally overestimated K , considering $\hat{K} = K_0 + l$, for $l = 1, \dots, 4$, and K_0 the true number of components. Fig. 3.11 displays the impact of this overestimation on the change detector performance. In this case, the proposed change detector was applied to the images displayed in Figs. 3.9(e) and 3.9(f). The distribution of the estimated distance d_W for changed and unchanged windows when the number of components is forced to be equal to the real number of objects is displayed in Fig. 3.12(a). The effect of overestimating K by 1 for all windows is shown in Fig. 3.12(b). It can be observed that the distributions of the changed and unchanged windows are closer when the number of components is overestimated. Since both distribution are closer, any chosen decision threshold for the hypothesis test will result in a higher error rate. This can be observed in Fig. 3.12(c), where the performance decreases when K is overestimated for the same dataset of Figs. 3.9(e) and 3.9(f). Underestimating the number of components of the mixture distribution leads to the missclassification of objects, assigning different objects to a unique component. This underestimation has an impact on the estimation of the component parameters, since $\hat{\nu}$ is then obtained from different objects with different values of P , invalidating the proposed analysis. The method performance clearly benefits from an accurate estimation of K . This statement leads us to propose an alternative method in Chapter 4.

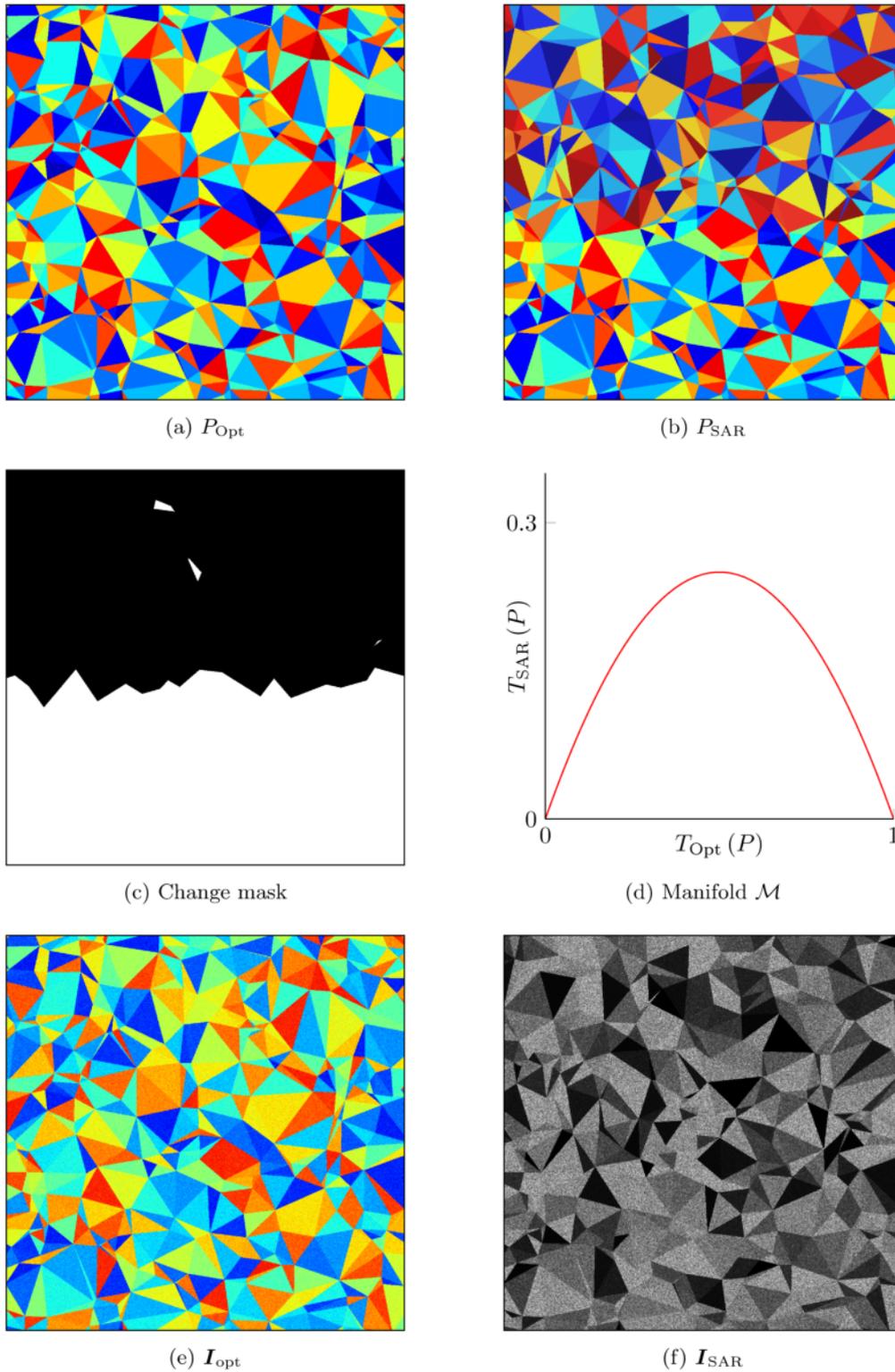


Figure 3.9: Example of synthetic images with changed and unchanged areas, where $T_{\text{opt}}(P) = P$ and $T_{\text{SAR}}(P) = P(1 - P)$.

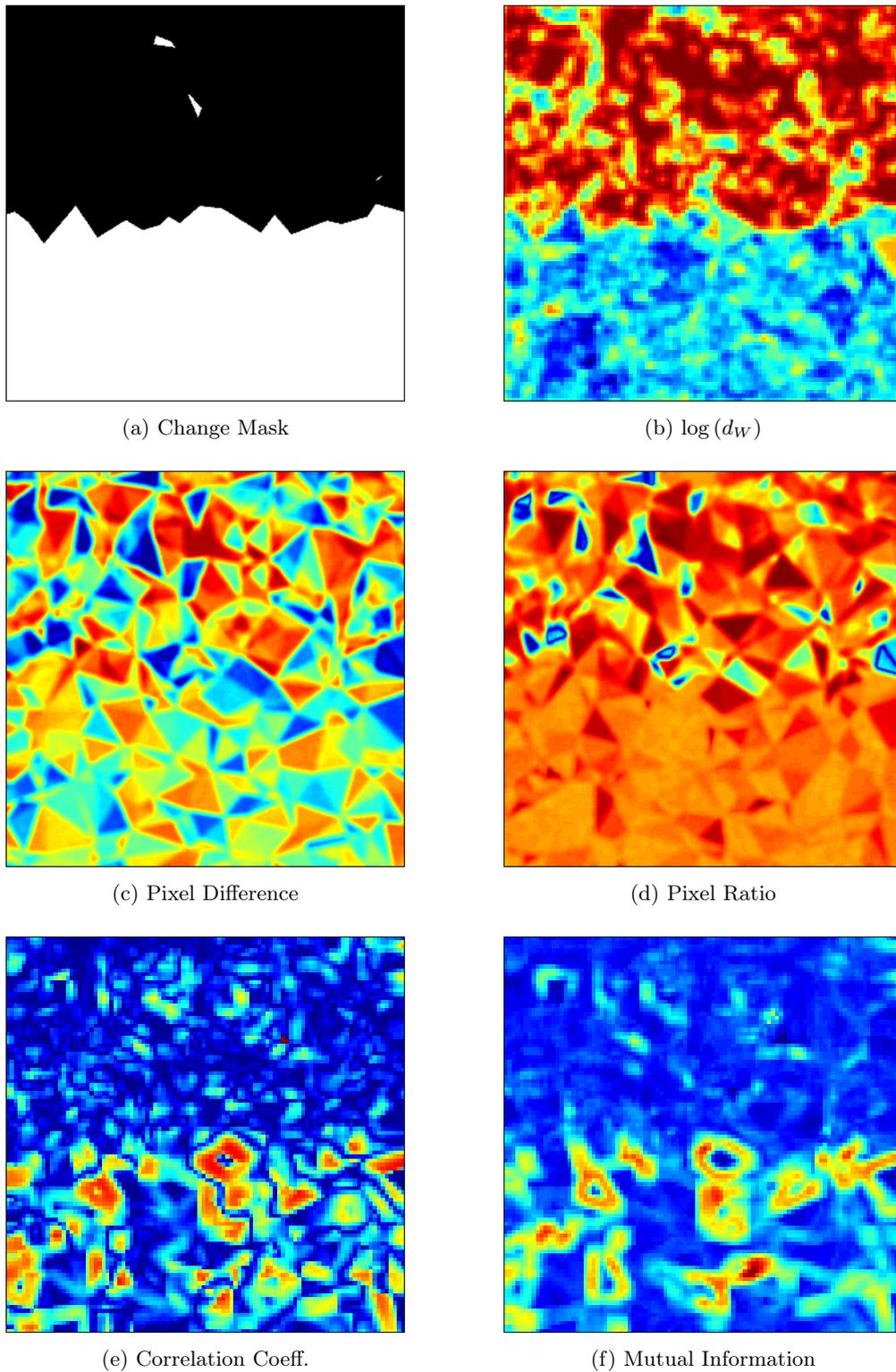


Figure 3.10: Estimated change maps for the images of Fig. 3.9. Red areas indicate high similarity, while blue areas indicate low similarity.

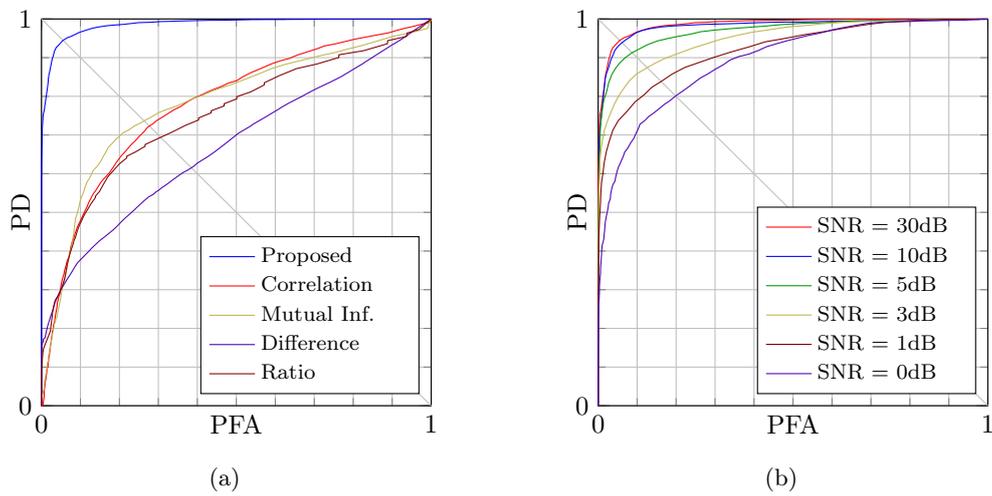
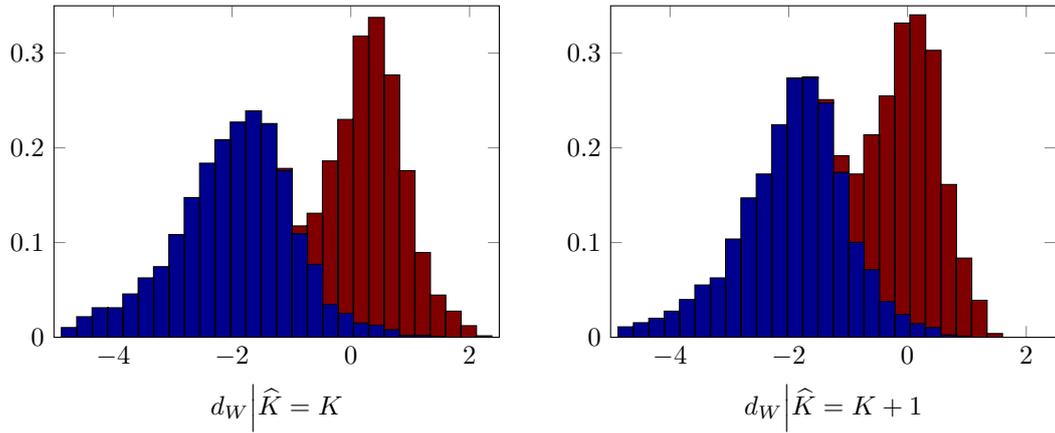
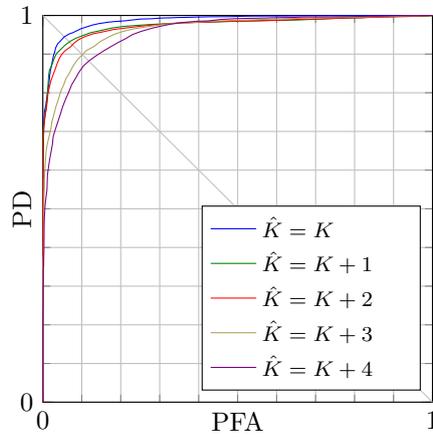


Figure 3.11: ROC curves for synthetic images (a) for different methods, (b) for the proposed method with different SNRs.



(a) Distribution of the estimated manifold distance d_W for changed (red) and unchanged (blue) areas when correctly estimating K .

(b) Distribution of the estimated manifold distance d_W for changed (red) and unchanged (blue) areas when overestimating estimating K by 1.



(c) Performance drop on the detector by overestimating K by different amounts.

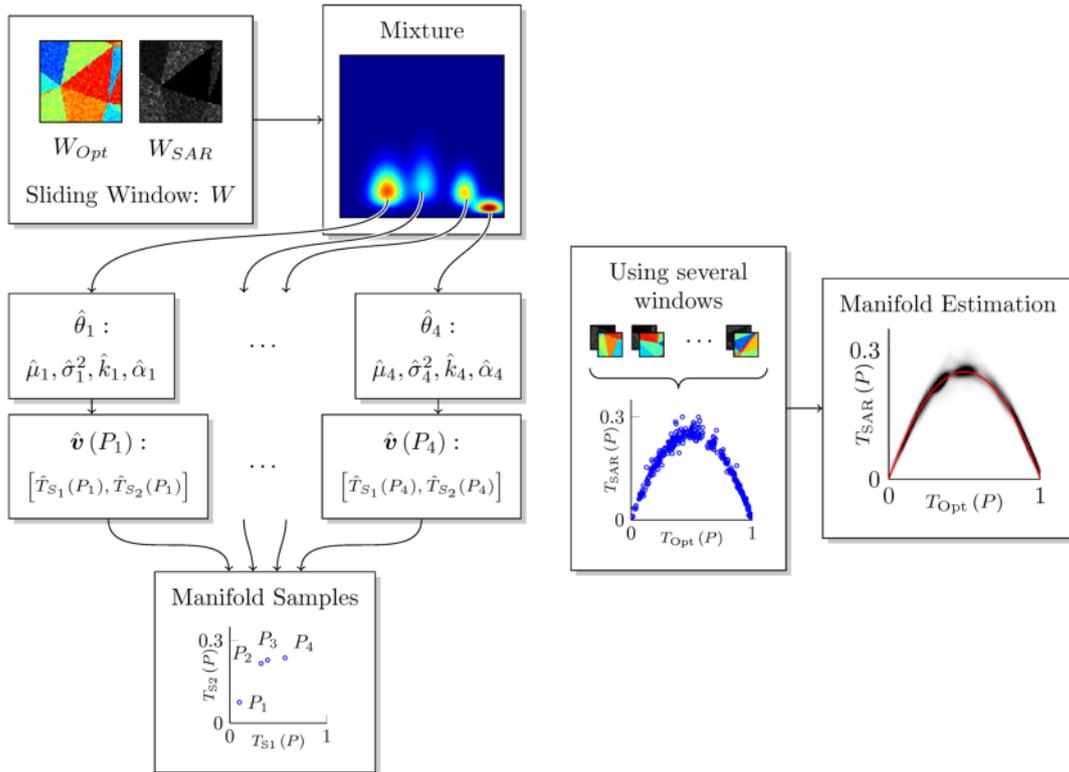
Figure 3.12: Impact of overestimating the number of components K in the mixture distribution.

3.5 Conclusions

A similarity measure suitable to detect changes between images acquired by homogeneous and heterogeneous sensors can be defined based on the parameters of the mixture model studied in Chapter 2. Section 3.1 introduces the manifold \mathcal{M} within the space generated by these parameters. This manifold describes the joint behavior of the sensors in the absence of change. In other words, in the absence of change, the component parameters are constrained to belong to \mathcal{M} , while this constraint is no longer satisfied in the presence of changes. Based on this statement, Section 3.2 proposes to use the distance to \mathcal{M} as a similarity measure. However, the manifold \mathcal{M} , is *a priori* unknown. A quantity that behaves as a distance to the manifold is proposed in Section 3.2.2, where a window-based strategy for its estimation is developed.

A summary of this change detection strategy is shown in Fig. 3.13, explaining how to go from the training unchanged image to the parameter space, and how the distance to the manifold is estimated from several parameters obtained throughout the training image. The performance of this approach was evaluated in Section 3.4 on synthetic images. The proposed similarity measure performs much better than classical measures when applied to the detection of changes between heterogeneous datasets.

The estimation of the vectors $\hat{\mathbf{v}}$ is of crucial importance for the proposed approach, since it is involved in both, the training process and the change detection. It is important that each component represents only one object, and does not arise as the estimation of several different objects at the same time. It is also important for each component to be estimated from as many pixels as possible, since this reduces the variance of the estimator $\hat{\mathbf{v}}$. The results shown in Section 3.4 show that the proposed change detection method outperforms the other methods studied in this theses. However, these results depend on an EM based algorithm to estimate the mixture parameters. This algorithm has two main drawbacks. First, it requires to know a range for the number of objects within the analysis window. An overestimation on this parameter results in more components estimated from less pixels each. An underestimation results in different objects represented by the same component. Second, it does not take into account the spatial correlation present on the images which provides valuable information to improve the parameter estimation. Chapter 4 studies a way of addressing these two issues.



(a) Process performed on each analysis window W of the training area to estimate $\mathbf{v}(P)$ from the image. Each component of the estimated mixture distribution produces a different vector $\mathbf{v}(P)$, corresponding to the value of P associated to the object that the component represents.

(b) The process (a) is repeated on all W in the training area to obtain several vectors $\mathbf{v}(P)$ for different values of P . These estimations are then used to learn the underlying manifold (red curve).

Figure 3.13: Process to estimate the manifold \mathcal{M} from training data. A pair of unchanged optical and SAR images are used as an example.

CHAPTER 4

A Bayesian non parametric model for change detection

Contents

4.1 Bayesian modeling	60
4.1.1 Finite mixture distributions	60
4.1.2 Non parametric mixture distribution	62
4.1.3 Parameter estimation	66
4.1.4 Concentration parameter	68
4.2 Spatial correlation	73
4.2.1 Markov random field	75
4.2.2 Markov random field as a prior in a DPMM	76
4.3 Performance on synthetic images	79
4.4 Conclusion	82

A statistical model was introduced in Section 2.2.2 to describe the joint distribution of pixel intensities associated with a set of remote sensing images captured by homogeneous and heterogeneous sensors. As previously mentioned, the estimation of the mixture parameters is of crucial importance for the change detection algorithm. However, the EM based algorithm used for this (estimation presented in Chapter 3) has some limitations. First, the number of components of the mixture distribution has to be contained within a given range that has to be fixed *a priori*. This limits the practical window size, since a bigger window usually requires increasing the number of components of the mixture, while the effects of under or overestimating this quantity has been discussed in Section 3.4. Second, the spatial correlation has not been considered into the estimation. The importance of considering it is that it allow us to deal with problems like the one presented in Fig. 3.8. Moreover, it would improve the general identifiability of the mixture components, which in turn would allow bigger window sizes to be used.

This chapter studies a different approach to estimate the distribution (2.29) based on a Bayesian non parametric approach.

Section 4.1.1 begins by presenting a simple Bayesian approach for the model described in Section 2.2.2. The first limitation of the EM based algorithm motivates the need for a BNP framework. Section 4.1.2 introduces the BNP model by extending the model in Section 2.2.2 to account for a possible infinite number of dimensions. Section 4.1.3 develops a collapsed Gibbs sampler to sample the resulting posterior distribution. The generated samples are then used to estimate the parameters of the proposed model. The use of a prior distribution to estimate the concentration parameter α that arises in Section 4.1.2 is investigated in Section 4.1.4, and an algorithm to sample and estimate this parameter is also derived.

To account for the spatial correlation between adjacent pixels of the image, Section 4.2 proposes to include a Markov random field (MRF) prior into the previous BNP model. Section 4.3 assess the performance of a change detector based on this approach and compares the results with those obtained in Section 3.4. Finally, Section 4.4 presents some conclusions regarding the topic covered in this chapter.

4.1 Bayesian modeling

4.1.1 Finite mixture distributions

In Section 2.2.2 we introduced a statistical model describing the pixel intensity vector $\mathbf{i} = [i_1, \dots, i_D]$, where i_d is the intensity produced by the d -th sensor S_d within a moving window W . This model is given by the following mixture model (refer to Section 2.2.2 for its derivation)

$$p(\mathbf{i}|W) = \sum_{k=1}^K w_k \prod_{d=1}^D p(i_d|P_k) \quad (4.1)$$

where P_k are the physical properties of the k -th object contained in the scene, and w_k is the proportion of W covered by such object.

A Bayesian description of (4.1) can be obtained by including priors on the mixture

parameters, namely, on the physical properties P , or equivalently $\mathbf{v}(P)$, and the component weights. The mixture distribution of Section 2.2.2 can be represented as follows

$$\mathbf{i}_n | \mathbf{v}_n \sim \mathcal{F}(\mathbf{v}_n) \quad (4.2)$$

$$p(\mathbf{v}_n | \mathbf{V}') = \sum_{k=1}^K w_k \delta(\mathbf{v}_n - \mathbf{v}'_k) \quad (4.3)$$

where $\mathbf{i}_n = [i_{n,S_1}, \dots, i_{n,S_D}]^T$ (for $1 \leq n \leq N$) is the pixel intensity vector of the n -th pixel, \mathbf{v}_n is the parameter vector associated with the object containing the n -th pixel, $\mathbf{V}' = [\mathbf{v}'_1, \dots, \mathbf{v}'_K]^T$ where \mathbf{v}'_k is the parameter vector associated with the k -th cluster or object and $\mathcal{F}(\mathbf{v}_n)$ is a distribution on the family \mathcal{F} identified by the parameter vector \mathbf{v}_n (which was previously introduced at the end of Section 2.2.2). Introducing a cluster label for each pixel in the observation window $\mathbf{z} = [z_1, \dots, z_N]^T$, one obtains an equivalent model given by

$$\mathbf{i}_n | z_n \sim \mathcal{F}(\mathbf{v}'_{z_n}) \quad (4.4)$$

$$z_n \sim \text{Cat}_K(\mathbf{w}) \quad (4.5)$$

where $\mathbf{w} = [w_1, \dots, w_K]^T$ and $\text{Cat}_K(\cdot)$ represents the K -th dimensional categorical distribution. It is important to note that both representations are equivalent to the model defined in (4.1).

By defining a prior for \mathbf{v}'_k and \mathbf{w} , we can build a Bayesian framework [Har03] to estimate the mixture parameters. A common approach is to consider the following prior information

$$\mathbf{v}'_k \sim \mathcal{V}_0 \quad (4.6)$$

$$\mathbf{w} \sim \text{Dir}_K(\alpha) \quad (4.7)$$

where \mathcal{V}_0 is a prior for the parameter vector \mathbf{v}'_k that depends on the application (usually taken as a conjugate prior), and $\text{Dir}_K(\alpha)$ denotes the classical conjugate prior of categorical distributions: the symmetric Dirichlet distribution of dimension K and concentration parameter α . Applying these priors to a mixture model results in the so-called Dirichlet mixture model [BZV04]. However, this model requires to know the parameter K *a priori*, which can be a problem. This problem was heuristically solved in Chapter 3 by testing different values of K within a predefined range. The BNP framework investigated in this chapter allows this limitation to be removed, by making K a random variable taking its value in the set of natural numbers \mathbb{N} .

Algorithm 2: Generation of samples from a realization of a Dirichlet process.

Input: \mathcal{V}_0, α

Output: v_1, v_2, v_3, \dots

```

1 for  $n \geq 1$  do
2    $u \sim \text{Uniform}(1, \alpha + n)$ ;
3   if  $u < n$  then
4      $v_n \leftarrow v_{[u]}$ ;
5   else
6      $v_n \sim \mathcal{V}_0$ ;

```

Considering K as a random variable introduces some difficulties, specifically the fact that the model exhibits a variable complexity. A common way of dealing with this problem is to consider a non parametric approach (where non parametric means that the number of parameters is not finite). In this case we consider a mixture of an infinite number of components, where only a finite set of them has a non-zero weight. Note that the effective number of non-zero components is finite, but the model descriptions is always non-parametric. However, considering $K \rightarrow \infty$ yields an infinite dimensional parameter vector \mathbf{w} and a matrix \mathbf{V}' with an infinite number of columns. Estimating all the parameters of this mixture model is clearly intractable.

To overcome this problem the model presented in the next section considers these parameters as intermediate and does not require their estimation. The model leads to a finite number of vectors \mathbf{v}_n through the estimation of a finite number of parameters, namely the concentration parameter α and the distribution \mathcal{V}_0 .

4.1.2 Non parametric mixture distribution

Define as $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$ the matrix containing the N random vectors \mathbf{v}_n , $n = 1, \dots, N$, where \mathbf{v}_n is associated with the n -th pixel \mathbf{i}_n of the observation window. In order to understand how our Bayesian non parametric method works, consider first Algo. 2 that generates vectors \mathbf{v}_n by using an iterative algorithm. Since some of these vectors are possibly repeated, we finally obtain $K \leq N$ different vectors that are denoted as \mathbf{v}'_k , $k = 1, \dots, K$ associated with each object of the observation window. We can compute the joint distribution of the N vectors $\mathbf{v}_1, \dots, \mathbf{v}_N$ by using the chain rule as explained

in [GB12]

$$\begin{aligned} p(\mathbf{v}_1, \dots, \mathbf{v}_N) &= \prod_{m=1}^N p(\mathbf{v}_m | \mathbf{v}_1, \dots, \mathbf{v}_{m-1}) \\ &= \frac{\prod_{k=1}^K (N_k - 1)! \alpha p_{\mathcal{V}_0}(\mathbf{v}'_k)}{\Gamma(\alpha + N - 1)} \end{aligned} \quad (4.8)$$

where $p_{\mathcal{V}_0}$ is the pdf associated with the distribution \mathcal{V}_0 , N_k is the number of vectors \mathbf{v}_n taking the value \mathbf{v}'_k . This distribution can be factorized into two terms: the first one is related to the particular values of \mathbf{v}'_k and the second one is related to the data partitioning. The resulting conditional distribution of \mathbf{V} given \mathbf{V}' (where $\mathbf{V}' = [\mathbf{v}'_1, \dots, \mathbf{v}'_K]$) can be written

$$p(\mathbf{V} | \mathbf{V}') = \frac{\prod_{k=1}^K \alpha (N_k - 1)!}{\Gamma(\alpha + N - 1)}. \quad (4.9)$$

It can be observed that the distribution of $\mathbf{V} | \mathbf{V}'$ only depends on the cardinal of each partition set, and not on the order the vectors have been drawn. Thus, any random vector can be thought as if it was the last drawn vector, meaning that these random vectors are exchangeable. Using the “de Finetti’s theorem” [OT10], one can show that the vectors \mathbf{v}_n are conditionally independent given a latent distribution \mathcal{V} . In this case, \mathcal{V} is defined by the pdf $p_{\mathcal{V}}$ defined as

$$p_{\mathcal{V}}(\mathbf{v}_n) = \sum_{k=0}^{\infty} w_k \delta(\mathbf{v}_n - \mathbf{v}'_k) \quad (4.10)$$

with

$$\mathbf{v}'_k \sim \mathcal{V}_0 \quad (4.11)$$

$$w_k = w'_k \prod_{j=1}^{k-1} (1 - w'_j) \quad (4.12)$$

$$w'_k \sim \text{Beta}(1, \alpha) \quad (4.13)$$

where (4.12) and (4.13) are known as a stick breaking process (SBP) and denoted as $\text{SBP}(\alpha)$ [GB12; IJ01], which can be thought as a generalization of $\text{Dir}_K(\alpha)$ for $K \rightarrow \infty$. The parameter α controls the concentration of the variables w_k . A small value of α provides a few values for the discrete distribution \mathcal{V} containing most of the probability, while a high value of α provides a reduced sparsity in the distribution \mathcal{V} , leading to a uniform distribution. The vectors \mathbf{v}_n defined by (4.10) tend to be grouped into clusters.

As it will be discussed in Section 4.1.4, the expected number K of different clusters given that the windows contains N pixels, is

$$E(K|\alpha, N) = \alpha[\psi(\alpha + N) - \psi(\alpha)] \quad (4.14)$$

where $\psi(\cdot)$ denotes the digamma function. Note that $\lim_{\alpha \rightarrow \infty} E(K|\alpha, N) = N$, since N samples can generate at most N clusters.

A Dirichlet process $DP(\mathcal{V}_0, \alpha)$ is a stochastic process whose realizations are probability distributions \mathcal{V} as defined by the pdf (4.10), so that the vectors \mathbf{v}_n in Algo. 2 can be described as

$$\mathbf{v}_n \sim \mathcal{V} \quad (4.15)$$

$$\mathcal{V} \sim DP(\mathcal{V}_0, \alpha). \quad (4.16)$$

Note that the Dirichlet process is often chosen as the conjugate prior for infinite discrete distributions. The relevance of Algo. 2 is that it provides a method to generate samples \mathbf{v}_n from a distribution \mathcal{V} with an infinite number of parameters only from its Bayesian priors without requiring the direct computation of its parameters. The finite mixture model in Section 2.2.2 can thus be extended through a BNP framework into a Dirichlet process mixture model (DPMM)

$$\mathbf{i}_n | \mathbf{v}_n \sim \mathcal{F}(\mathbf{v}_n) \quad (4.17)$$

$$\mathbf{v}_n \sim \mathcal{V} \quad (4.18)$$

$$\mathcal{V} \sim DP(\mathcal{V}_0, \alpha) \quad (4.19)$$

where \mathcal{V}_0 is the base distribution and α is the concentration parameter.

Introducing the cluster labels for each pixel of the observation window $\mathbf{z} = [z_1, \dots, z_N]$, we obtain the equivalent model

$$\mathbf{i}_n | z_n \sim \mathcal{F}(\mathbf{v}'_{z_n}) \quad (4.20)$$

$$z_n \sim \text{Cat}_\infty(\mathbf{w}) \quad (4.21)$$

$$\mathbf{w} \sim \text{SBP}(\alpha) \quad (4.22)$$

$$\mathbf{v}'_k \sim \mathcal{V}_0 \quad (4.23)$$

where $\text{Cat}_\infty(\mathbf{w})$ is an infinite dimensional generalization of the categorical distribution such that $p(z_n = k) = w_k$ for any $k \in \mathbb{N}^*$. This parametrization is equivalent to the parametrization in (4.17) to (4.19), but makes explicit that the pixels \mathbf{i}_n , for

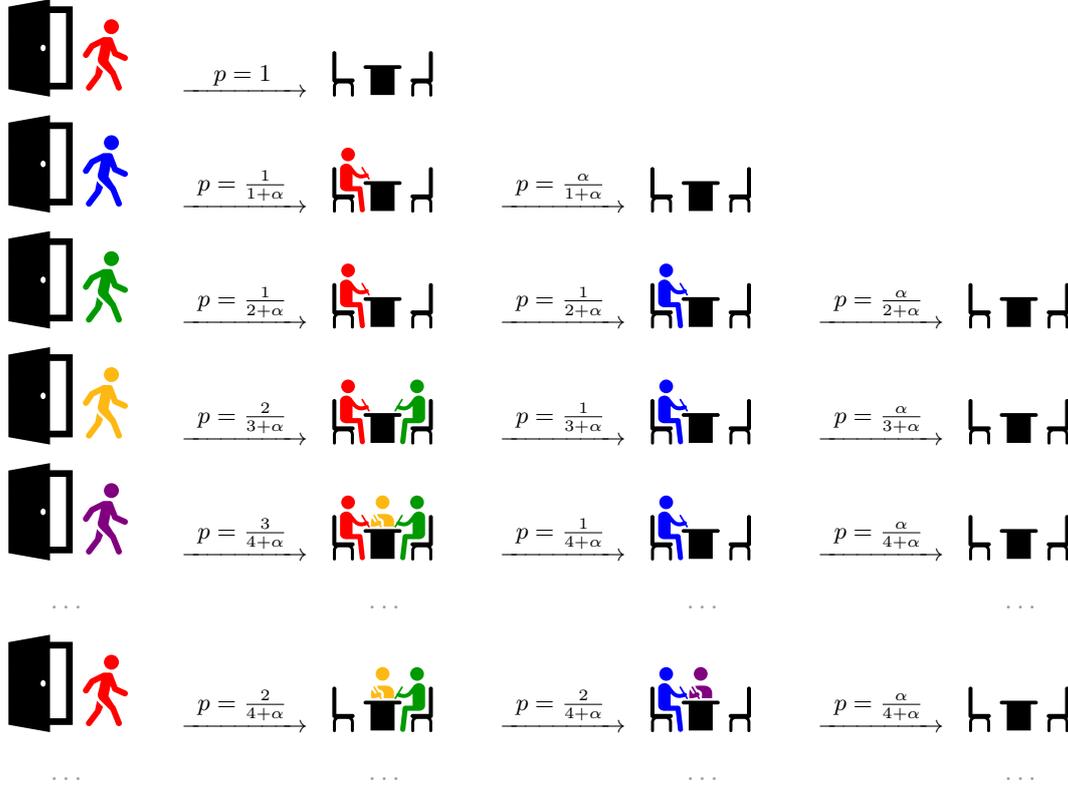


Figure 4.1: Graphical representation of a Chinese restaurant process.

$n = 1, \dots, N$, are partitioned into different clusters. Moreover, (4.21) and (4.22) define the so-called Chinese restaurant process $\text{CRP}(\alpha)$ [GB12]. Thus, the model can be reduced to

$$\mathbf{i}_n | z_n \sim \mathcal{F}(\mathbf{v}'_{z_n}) \tag{4.24}$$

$$\mathbf{z} \sim \text{CRP}(\alpha) \tag{4.25}$$

$$\mathbf{v}'_k \sim \mathcal{V}_0. \tag{4.26}$$

The advantage of this last parametrization is that it allows the parameters \mathbf{v}'_k to be integrated out, and that it only requires the estimation of \mathbf{z} , which is a vector of discrete random variables.

The name of this process arises from the idea of a Chinese restaurant, where the same table is shared by different customers, as displayed in Fig. 4.1. In this imaginary restaurant, there is an infinite number of tables. However, customers are prone to share tables. Moreover, the probability of a new customer to occupy a table is proportional to the number of people in that table. However, the probability of a new customer sitting

in a new table always exist, and is proportional to the concentration parameter α , with $\alpha \in \mathbb{R}^+$. A way of interpreting the parameter α can be expressed as: “the probability of a new customer to sit in a new table is the same as that of a table with α customers”. In our example, illustrated in Fig. 4.1, the red customer is the first arriving. Since the restaurant is empty, he will sit in an unoccupied table with probability 1. When the blue customer arrives, he can choose to share the table with the red customer, with probability $\frac{1}{1+\alpha}$, or to sit in a new table, with probability $\frac{\alpha}{1+\alpha}$. In general, a new customer will choose to sit in the k -th table with probability $\frac{N_k}{N+\alpha}$, where N_k is the number of people in that table, or in a new table with probability $\frac{\alpha}{N+\alpha}$. If we are interested in sampling from the probability distribution of a restaurant with N customers, once the restaurant is full we remove one customer and choose one new random table for him as if he was a new customer.

4.1.3 Parameter estimation

To estimate the parameters \mathbf{v}_n of a DPMM, we suggest to use a Markov chain Monte Carlo (MCMC) algorithm based on a collapsed Gibbs sampler [CG92]. To implement this algorithm we choose the CRP parametrization of the DP. As previously stated, the advantage of this parametrization is that it allows the parameters \mathbf{v}'_k to be integrated out leading to a partially collapsed Gibbs sampler [VP08]. More precisely, to estimate the latent variables \mathbf{z} , we can sample from $p(\mathbf{z}|\mathbf{I}, \mathcal{V}_0)$, where $\mathbf{I} = [\mathbf{i}_1, \dots, \mathbf{i}_N]$. The Gibbs sampler is an iterative algorithm that samples sequentially the conditional probabilities of each variable with respect to (w.r.t.) the other variables. For the proposed problem, samples from $p(z_n|\mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0)$ are generated, where $\mathbf{z}_{\setminus n} = [z_1, \dots, z_{n-1}, z_{n+1}, \dots, z_N]$. This conditional probability can be obtained as follows

$$p(z_n|\mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0) \propto p(\mathbf{I}|\mathbf{z}, \mathcal{V}_0)p(z_n|\mathbf{z}_{\setminus n}) \quad (4.27)$$

where \propto means “proportional to”, $p(z_n|\mathbf{z}_{\setminus n})$ can be obtained using $p(\mathbf{z}) = p(\mathbf{V}|\mathbf{V}')$ and (4.9). More precisely, the following result can be obtained

$$\begin{aligned} p(z_n|\mathbf{z}_{\setminus n}) &= p(\mathbf{z}) \times p(\mathbf{z}_{\setminus n})^{-1} \\ &= \frac{\prod_{k=1}^{K^*} \alpha (N_k - 1)!}{\Gamma(\alpha + N - 1)} \times \frac{\Gamma(\alpha + N - 2)}{\prod_{k=1}^{K^*} \alpha (N_k - \mathbb{1}_k(z_n) - 1)!} \end{aligned} \quad (4.28)$$

$$= \begin{cases} \frac{\alpha}{\alpha + N - 2} & \text{if } z_n = 0 \\ \frac{N'_{z_n}}{\alpha + N - 2} & \text{if } 1 \leq z_n \leq K^* \end{cases} \quad (4.29)$$

where $\mathbb{1}_k(z_n)$ is the indicator function, taking the value 1 when $z_n = k$ and 0 otherwise, K^* is the number of different values in \mathbf{z} in the current moment, N'_{z_n} is the number of pixels in the cluster indicated by z_n (excluding z_n), and $z_n = 0$ when a new cluster is created. The probability $p(\mathbf{I}|\mathbf{z}, \mathcal{V}_0)$ is obtained by marginalizing out \mathbf{V}' from the likelihood $p(\mathbf{I}|\mathbf{z}, \mathbf{V}')$ as follows

$$\begin{aligned} p(\mathbf{I}|\mathbf{z}, \mathcal{V}_0) &= \int p(\mathbf{I}|\mathbf{z}, \mathbf{V}')p(\mathbf{V}'|\mathcal{V}_0)d\mathbf{V}' \\ &= \prod_{k=1}^K \int p(\mathbf{I}_{\{k\}}|\mathbf{v}'_k)p(\mathbf{v}'_k|\mathcal{V}_0)d\mathbf{v}'_k \end{aligned} \quad (4.30)$$

$$= \prod_{k=1}^K p(\mathbf{I}_{\{k\}}|\mathcal{V}_0) \quad (4.31)$$

where $\mathbf{I}_{\{k\}} = \{\mathbf{i}_n : z_n = k\}$. Note that \mathbf{V}' is not required to estimate the vector of latent variables \mathbf{z} . We can now compute $p(z_n|\mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0)$ from (4.27) as follows

$$p(z_n|\mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0) \propto p(z_n|\mathbf{z}_{\setminus n}) \times \prod_{k=1}^K p(\mathbf{I}_{\{k\}}|\mathcal{V}_0) \quad (4.32)$$

so we can factorize out all the terms that do not depend on z_n . To do this we should remember that z_n has an influence on $\mathbf{I}_{\{k\}}$ by defining the k to which \mathbf{i}_n will be assigned, resulting in

$$p(z_n|\mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0) \propto p(z_n|\mathbf{z}_{\setminus n}) \frac{p(\mathbf{I}_{\{z_n\}}|\mathcal{V}_0)}{p(\mathbf{I}_{\{z_n\}\setminus n}|\mathcal{V}_0)} \quad (4.33)$$

where $\mathbf{I}_{\{z_n\}\setminus n} = \{\mathbf{i}_m : z_m = z_n, m \neq n\}$, i.e., we only consider the $p(\mathbf{I}_{\{k\}})$ to which \mathbf{i}_n belongs, and divide by the influence of all the other \mathbf{i}_m in the same cluster (which from the point of view of z_n is no more than a proportionality constant). By replacing $p(z_n|\mathbf{z}_{\setminus n})$ with (4.29), we obtain

$$p(z_n|\mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0) \propto \begin{cases} \alpha p(\mathbf{i}_n|\mathcal{V}_0) & \text{if } z_n = 0 \\ N'_{z_n} \frac{p(\mathbf{I}_{\{z_n\}}|\mathcal{V}_0)}{p(\mathbf{I}_{\{z_n\}\setminus n}|\mathcal{V}_0)} & \text{if } 1 \leq z_n \leq K. \end{cases} \quad (4.34)$$

Moreover, when $\mathbf{I}_{\{k\}}|\mathcal{V}_0$ belongs to a distribution family that can be described using a sufficient statistic $\mathbf{T}(\mathbf{I}_{\{k\}})$ such that $p_{\mathbf{I}}(\mathbf{I}_{\{k\}}|\mathcal{V}_0) = p_{\mathbf{T}}(\mathbf{T}(\mathbf{I}_{\{k\}})|\mathcal{V}_0)$ (e.g., for distributions belonging to an exponential family), we have $\mathbf{T}(\mathbf{I}_{\{k\}}) = \mathbf{T}(\mathbf{I}_{\{k\}\setminus n}) + \mathbf{T}(\mathbf{i}_n)$. This means that (4.34) can be easily computed just by keeping track of $\mathbf{T}(\mathbf{I}_{\{k\}})$ for each cluster, and avoiding recomputing the likelihood at each iteration, leading to significant speed improvements.

Algorithm 3: A collapsed Gibbs sampler to sample from the partition distribution.

Input: $\mathbf{I} = \{\mathbf{i}_1, \dots, \mathbf{i}_N\}$, \mathcal{V}_0 , α , j_{\max} , j_{\min}

Output: $\mathbf{z} = \{z_1, \dots, z_N\}$

- 1 $z_n^{(0)} \leftarrow 0$, $\forall 1 \leq n \leq N$;
- 2 $K \leftarrow 0$;
- 3 **for** $j : 1 \leq j \leq j_{\max}$ **do**
- 4 $z_n^{(j)} \leftarrow z_n^{(j-1)}$, $\forall 1 \leq n \leq N$;
- 5 **for** $n : 1 \leq n \leq N$, *in random order* **do**
- 6 /* Remove the n-th pixel from its current class */
- 7 **if** $z_n^{(j)} \neq 0$ **then**
- 8 $\mathbf{T}_{z_n^{(j)}} \leftarrow \mathbf{T}_{z_n^{(j)}} - \mathbf{T}(\mathbf{i}_n)$; $N_{z_n^{(j)}} \leftarrow N_{z_n^{(j)}} - 1$;
- 9 /* Sample a new class for the n-th pixel */
- 10 $p_0 \leftarrow \alpha p_{\mathbf{T}}(\mathbf{T}(\mathbf{i}_n) | \mathcal{V}_0)$;
- 11 $p_k \leftarrow N_k \frac{p_{\mathbf{T}}(\mathbf{T}_k + \mathbf{T}(\mathbf{i}_n) | \mathcal{V}_0)}{p_{\mathbf{T}}(\mathbf{T}_k | \mathcal{V}_0)}$, $\forall 1 \leq k \leq K$;
- 12 $z_n^{(j)} \sim \text{Cat}(p_0, p_1, \dots, p_K)$;
- 13 /* Place the n-th pixel in its new class */
- 14 **if** $z_n^{(j)} = 0$ **then**
- 15 $K \leftarrow K + 1$; $z_n^{(j)} \leftarrow K$;
- 16 $\mathbf{T}_K \leftarrow \mathbf{T}(\mathbf{i}_n)$; $N_K \leftarrow 1$;
- 17 **else**
- 18 $\mathbf{T}_{z_n^{(j)}} \leftarrow \mathbf{T}_{z_n^{(j)}} + \mathbf{T}(\mathbf{i}_n)$; $N_{z_n^{(j)}} \leftarrow N_{z_n^{(j)}} + 1$;
- 19 $z_n \leftarrow \text{mode}(z_n^{(j_{\max})}, \dots, z_n^{(j_{\min})})$, $\forall 1 \leq n \leq N$;

Algo. 3 shows the implementation of the described approach. The parameters j_{\max} and j_{\min} are related to the MCMC nature of the Gibbs sampler, where j_{\max} is the maximum number of samples to generate and j_{\min} defines the end of the so called burn-in period, while \mathbf{T}_k is the sufficient statistic vector for the k -th cluster.

4.1.4 Concentration parameter

The selection of the concentration parameter α in the model presented in Section 4.1.2 has a direct influence on the number of objects detected within an image. The probability

$p(K|\alpha, N)$ has been studied in [EW95; Ant74] leading to

$$p(K|\alpha, N) = S_N(K)\alpha^K \frac{\Gamma(\alpha)}{\Gamma(N + \alpha)} \quad (4.35)$$

where $S_N(K)$ is the unsigned Stirling number. We recall that the unsigned Stirling numbers follow the recursive relationship

$$S_N(K) = (N - 1)S_{N-1}(K) + S_{N-1}(K - 1) \quad (4.36)$$

with the boundary conditions $S_0(0) = 1$ and $S_N(K) = 0$ for $K > N$. Using (4.35), the expected value of K given by (4.14) can be derived. This implies that the choice of α provides information about the resulting number of objects. However, the number of objects within an analysis window depends on several factors, including the image resolution, the window size, and the particular scene being imaged (e.g., a rural area, an urban area, etc.). These conditions can be *a priori* unknown or might exhibit strong variations within the image. As we did with K , we can use a Bayesian approach to estimate α jointly with the other parameters. In this case, we need to assign a prior distribution $p(\alpha)$ to the concentration parameter α . A gamma prior was proposed in [EW95]. However, this prior has the disadvantage that it still assigns higher probability to some values of K than others.

To demonstrate this effect, we ran the algorithm with different parameters for this gamma prior on a synthetic image generated for $\alpha_0 = 1$. After each iteration of Algo. 3 the value of α is updated according to its conditional distribution given the number of non zero components produced by the CRP. Since α is a random variable, we ran 100 iterations for each set of parameters to be able to obtain the sample mean and variance. We choose the parameters such that the distribution has a fixed shape ($k = 2$ in the gamma distribution with shape parameter k), and different mean values. Fig. 4.2 shows the evolution of α at each iteration for $E[\alpha]$ ranging from 10^{-3} to 10^3 . It can be observed that $\alpha \rightarrow \alpha_0$ only when $E[\alpha] \geq \alpha_0$. This is due to the gamma distribution having exponentially decreasing tails, which implies that the probability of $\alpha > E[\alpha]$ reduces very drastically. It can be argued that this is not a problem, since we can always propose a prior with a “high enough $E[\alpha]$ ”. However, it should be noted that the computational time required by the Gibbs sampler is (at best) linearly proportional to the current number of components in the mixture distribution. This is illustrated by the fact that the curves in Fig. 4.2 for $E[\alpha] \leq 1$ were computed in a few seconds, while the first iterations of the curves for $E[\alpha] > 10$ lasted around 1 hour.

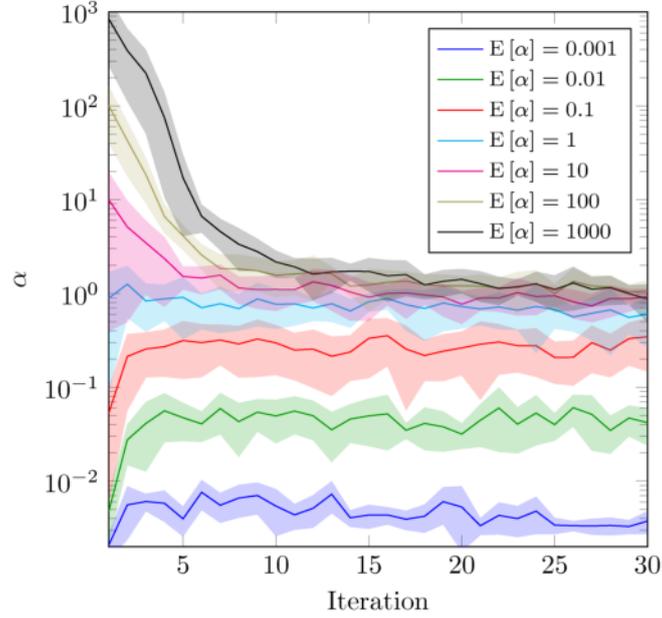


Figure 4.2: Evolution of the parameter α with a gamma prior for $\alpha_0 = 1$. The lines indicate the values of α for different iterations and the shaded area corresponds to 1 standard deviation respect to the mean.

To avoid problems arising from an informative prior we propose a non-informative Jeffreys prior for α . The use of a Jeffreys prior is very common in Bayesian inference [Gel09], specially as priors for variances, where it takes the form $p(\sigma) \propto \sigma^{-1}$, which is an improper prior (i.e., its integral diverges) but yields proper posterior distributions. In the more general case, the Jeffreys prior [Jef46] is defined as

$$p(X) \propto \sqrt{E_Y \left[\left(\frac{d}{dX} \log p(Y|X) \right)^2 \right]} \quad (4.37)$$

where the prior for the variable X provides no information about the parameter Y . In our case, for $p(\alpha|N)$ we have

$$p(\alpha|N) \propto \sqrt{E_K \left[\left(\frac{d}{d\alpha} \log p(K|\alpha, N) \right)^2 \right]}. \quad (4.38)$$

To compute this expression, we start by deriving $\frac{d}{d\alpha} \log p(K|\alpha, N)$ using (4.35)

$$\frac{d}{d\alpha} \log p(K|\alpha, N) = \frac{K}{\alpha} - \left(\Delta \psi_N^{(0)}(\alpha) + \alpha^{-1} \right) \quad (4.39)$$

where $\Delta\psi_N^{(i)}(\alpha) = \psi^{(i)}(N + \alpha) - \psi^{(i)}(1 + \alpha)$, the function $\psi(\cdot)$ is the digamma function, and $\psi^{(i)}(\cdot)$ is its i -th derivative. Denoting $m = \Delta\psi_N^{(0)}(\alpha) + \alpha^{-1}$, we obtain

$$E_K \left[\left(\frac{K}{\alpha} - m \right)^2 \right] = \frac{\Gamma(\alpha)}{\Gamma(N + \alpha)} \sum_{K=1}^N S_N(K) \alpha^K \left[\frac{K}{\alpha} - m \right]^2. \quad (4.40)$$

Defining the number \mathcal{S}_N^i for $i = 0, 1, 2, \dots$ as the following series

$$\mathcal{S}_N^i = \sum_{K=1}^N S_N(K) \alpha^K K^i \quad (4.41)$$

the squared binomial in (4.40) can be expanded to obtain

$$E_K \left[\left(\frac{K}{\alpha} - m \right)^2 \right] = \frac{1}{\alpha^2} \mathcal{S}_N^2 - 2 \frac{m}{\alpha} \mathcal{S}_N^1 + m^2 \mathcal{S}_N^0. \quad (4.42)$$

By using (4.36), we obtain a closed form expression for \mathcal{S}_N^0

$$\begin{aligned} \mathcal{S}_N^0 &= \sum_{K=1}^N \alpha^K S_N(K) \\ &= (N - 1) \mathcal{S}_{N-1}^0 + \alpha \mathcal{S}_{N-1}^0 \\ &= (N + \alpha - 1) \mathcal{S}_{N-1}^0. \end{aligned} \quad (4.43)$$

Solving the recursive relationship leads to

$$\mathcal{S}_N^0 = \frac{\Gamma(N + \alpha)}{\Gamma(\alpha)}. \quad (4.44)$$

Differentiating (4.41) and (4.44) w.r.t. α , we obtain

$$\mathcal{S}_N^1 = \mathcal{S}_N^0 \alpha m \quad (4.45)$$

$$\mathcal{S}_N^2 = \mathcal{S}_N^0 (\alpha m + \alpha^2 m^2 + \alpha^2 m') \quad (4.46)$$

where $m' = \frac{dm}{d\alpha} = \Delta\psi_N^{(1)}(\alpha) - \alpha^{-2}$. This leads to

$$E_K \left[\left(\frac{K}{\alpha} - m \right)^2 \right] = \frac{\Delta\psi_N^{(0)}(\alpha)}{\alpha} + \Delta\psi_N^{(1)}(\alpha) \quad (4.47)$$

yielding the following Jeffrey's prior

$$p(\alpha|N) = \mu_N \sqrt{\frac{\Delta\psi_N^{(0)}(\alpha)}{\alpha} + \Delta\psi_N^{(1)}(\alpha)} \quad (4.48)$$

where μ_N is a normalization constant greater than $(N-1)^{-1}\pi^{-1}$, i.e., $p(\alpha|N)$ is a proper prior. It is important to note that for $\alpha < N$ this distribution behaves as $\alpha^{-\frac{1}{2}}$, while for $\alpha > N$ it behaves as $\alpha^{-\frac{3}{2}}$, which in both cases yields a much heavier tail than the gamma prior.

To include this prior for α in the Gibbs sampler, we need to compute its conditional distribution. The conditional distribution of α depends exclusively on the partitioning \mathbf{z} , particularly, on the number of pixels to be partitioned, and the number of clusters in the partition

$$\begin{aligned} p(\alpha|K, N) &\propto p(K|\alpha, N) p(\alpha|N) \\ &\propto S_N(K) \alpha^K \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)} p(\alpha|N) \end{aligned} \quad (4.49)$$

$$\propto \alpha^K B(\alpha, N) p(\alpha|N) \quad (4.50)$$

where $B(\alpha, N) = \frac{\Gamma(\alpha)\Gamma(N)}{\Gamma(N+\alpha)}$ is the beta function. As described in [EW95], we can simplify this expression by introducing a new random variable. Since the beta function can also be expressed as $B(\alpha, N) = \int_0^1 t^{\alpha-1}(1-t)^{N-1} dt$, we obtain

$$p(\alpha|K, N) \propto \alpha^K p(\alpha|N) \int_0^1 t^{\alpha-1}(1-t)^{N-1} dt \quad (4.51)$$

where t is a variable living in the interval $[0, 1]$. Note that we can interpret (4.51) as a marginal distribution of

$$p(\alpha, t|K, N) \propto \alpha^K t^{\alpha-1}(1-t)^{N-1} p(\alpha|N). \quad (4.52)$$

In the MCMC scheme, we can sequentially sample t and α from

$$p(t|\alpha, K, N) \propto t^{\alpha-1}(1-t)^{N-1} \quad (4.53)$$

$$p(\alpha|t, K, N) \propto \alpha^K t^{\alpha-1} p(\alpha|N) \quad (4.54)$$

where (4.53) is a beta distribution, and (4.54) is defined as

$$p(\alpha|t, K, N) \propto \alpha^K t^{\alpha-1} \sqrt{\frac{\Delta\psi_N^{(0)}(\alpha)}{\alpha} + \Delta\psi_N^{(1)}(\alpha)}. \quad (4.55)$$

It can be easily shown that the marginal distribution $p(\alpha|t, K, N)$ behaves as a power law distribution for $\alpha \rightarrow 0$, and as a power law distribution with an exponential cutoff for $\alpha \rightarrow \infty$, i.e.,

$$p(t|\alpha, K, N) = L_\infty(\alpha|t, K, N) \alpha^{K-\frac{3}{2}} e^{\alpha \log t} \quad (4.56)$$

$$p(\alpha|t, K, N) = L_0(\alpha|t, K, N) \alpha^{K-\frac{1}{2}} \quad (4.57)$$

Algorithm 4: Sampling the posterior distribution of α .

Input: α, K, N

Output: α

- 1 $t \sim \text{B}(\alpha, N); M \leftarrow M(N, K, t);$
 - 2 $\alpha \sim \Gamma\left(K + \frac{1}{2}, -\frac{1}{\log t}\right);$
 - 3 **while** $u \sim \text{U}(0, 1), u \text{p}_{\Gamma}(\alpha) < M \text{p}(\alpha|N, K, t)$ **do**
 - 4 $\alpha \sim \Gamma\left(K + \frac{1}{2}, -\frac{1}{\log t}\right);$
-

where $L_{\infty}(\alpha|N)$ and $L_0(\alpha|N)$ are slowly varying functions in ∞ and 0 respectively. For the positive scale factor d , we obtain

$$\lim_{\alpha \rightarrow \infty} \frac{L_{\infty}(\alpha|N)}{L_{\infty}(d\alpha|N)} = 1 \quad (4.58)$$

$$\lim_{\alpha \rightarrow 0} \frac{L_0(d\alpha|N)}{L_0(\alpha|N)} = 1. \quad (4.59)$$

This behavior is difficult to replicate with most well known distributions, specially due to the mixed power law and exponential cutoff for $\alpha \rightarrow \infty$. However, this distribution can be upperbounded by removing the power law behavior for $\alpha \rightarrow \infty$. Indeed

$$\text{p}(\alpha|t, K, N) \leq L(\alpha|t, K, N) \alpha^{K-\frac{1}{2}} e^{\alpha \log t} \quad (4.60)$$

where $L(\alpha|t, K, N)$ is a bounded slowly varying function of α in ∞ and 0 and $\alpha^{K-\frac{1}{2}} e^{\alpha \log t}$ is a gamma distribution with shape parameter $K + \frac{1}{2}$ and scale parameter $-\frac{1}{\log t}$. A rejection sampling approach [RC05] can be easily implemented as shown in Algo. 4 and is included at the end of each iteration in Algo. 3.

The influence of this prior is shown in Fig. 4.3 where 100 simulations were run on the same image as Fig. 4.2 to obtain the sample mean and variance of α when using the proposed Jeffreys prior. It can be observed that regardless of the initial conditions ($\alpha \ll 1$ or $\alpha \gg 1$) the algorithm converges to the same value close to α_0 .

4.2 Spatial correlation

In Section 3.5 two major problems were identified for the estimation of \mathbf{v} . The first one, linked to the estimation of the number of objects in the scene, was addressed in Section 4.1.1. The second one, related to the spatial correlation of the images, is studied in this section. Consider the example illustrated in Fig. 4.4 where the same pair of

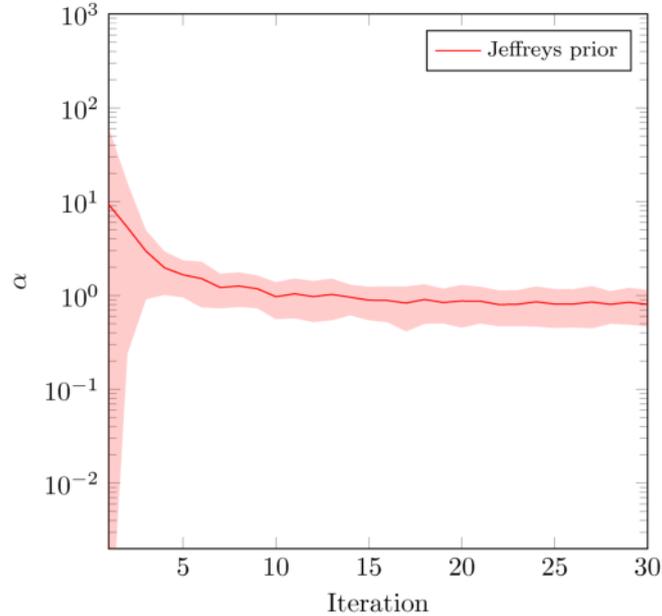


Figure 4.3: Evolution of the parameter α with a Jeffreys prior for $\alpha_0 = 1$. The red line indicates the values of α for different iterations and the red shaded area corresponds to 1 standard deviation respect to the mean.

images is presented twice, but the pixels in Fig. 4.4(b) have been randomly shuffled in both images at the same time. Applying the proposed change detection strategy in both datasets provides the same result, since the joint distribution of the pixel intensities remains unchanged. However, it is clear that Fig. 4.4(a) corresponds to a situation where the images have not been affected by any change or by a few minor changes, whereas it is not possible to say the same thing about Fig. 4.4(b) (it is even very difficult to identify which pixel belongs to each object in the image). In conclusion, it is clear that there is much information in the images from Fig. 4.4(a) that do not lie solely in its pixel intensities, but also on their pixel spatial arrangement.

A classic tool to model spatial correlation is the Markov random fields (MRF) [BKR11]. We can account for this correlation by introducing an MRF prior for the label parameter $\mathbf{z} = [z_1, \dots, z_N]$ of the DPMM presented in Section 4.1.2. At this point, it is interesting to mention the work of [OB08; CT10] that defined an MRF for the cluster distribution parameters \mathbf{v}'_k . However, since we want to prevent sampling from \mathbf{v}'_k by using a collapsed Gibbs sampler, we have to apply the MRF to the labels \mathbf{z} . Section 4.2.1 reviews some MRF concepts needed to build an appropriate prior for the

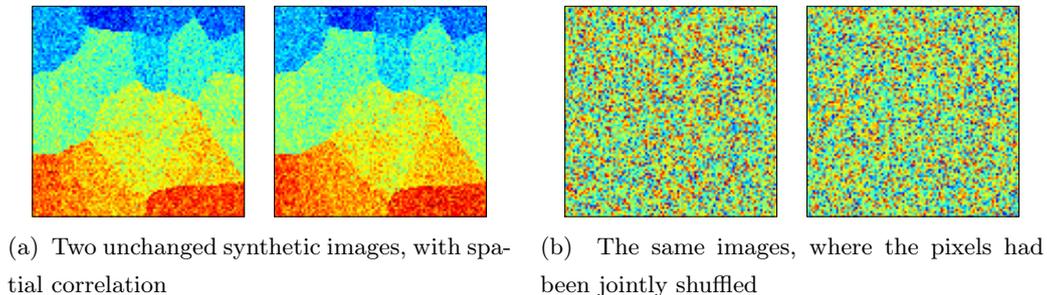


Figure 4.4: Two pairs of unchanged synthetic images. In (a) the images present spatial correlation. In (b) the spatial correlation had been broken by randomly shuffling the pixels jointly in both images.

latent variables. Section 4.2.2 explains how to couple this prior with the DPMM model introduced in Section 4.1.2 and presents the resulting algorithm.

4.2.1 Markov random field

As mentioned, before a classical tool to capture spatial correlation between adjacent pixels of an image is the MRF, which allows a joint distribution to be defined using a neighborhood graph. Let $\mathbf{z} = \{z_1, \dots, z_N\}$ be a group of random variables, and G be a simple weighted graph [Har94] (i.e., an undirected graph with no loop, no more than one edge between any two vertices, and with a weight associated with each edge). The vertices of G represent the different random variables z_n , while the weighted edges represent some sort of affinity between the connected random variables. The random vector \mathbf{z} is an MRF if the distribution of one variable z_n of this vector conditionally to the other variables $\mathbf{z}_{\setminus n}$ is only dependent on the variables belonging to its neighborhood, i.e.,

$$p(z_n | \mathbf{z}_{\setminus n}) = p(z_n | \mathbf{z}_{\text{ne}(n)}) \quad (4.61)$$

where $\mathbf{z}_{\text{ne}(n)}$ is the group of random variables that belong to the neighborhood of z_n , i.e., that are connected by an edge to z_n .

However, constructing a joint distribution such that its conditional distribution verifies (4.61) is not trivial. In particular, defining the conditional distributions independently can result in an improper joint distribution. The Hammersley-Clifford theorem [Gri73] gives a necessary and sufficient condition ensuring the existence of the joint distribution. This condition states that $p(\mathbf{z})$ should factorize over the cliques \mathcal{C} of G (we

recall that a clique of an undirected graph is a subset of vertices where any two vertices are connected by an edge). This is equivalent to defining $p(\mathbf{z}) = \exp[H(\mathbf{z})]$, and requiring that

$$H(\mathbf{z}) = \sum_{C \in \mathcal{C}} H_C(\mathbf{z}_C) \quad (4.62)$$

where $H(\cdot)$ is the so-called graph cost function, $H_C(\cdot)$ is a local cost function for the clique C , and $\mathbf{z}_C = \{z_n : n \in C\}$. Defining $H_n(\mathbf{z})$ as all the terms in $H(\mathbf{z})$ involving z_n , i.e.,

$$\begin{aligned} H_n(\mathbf{z}) &= \sum_{C \in \mathcal{C}, n \in C} H_C(\mathbf{z}_C) \\ &= h_{\text{ne}(n)}(\mathbf{z}_{\text{ne}(n)}) + h_n(z_n) \end{aligned} \quad (4.63)$$

where $h_n(z_n)$ is a cost function for the 1-vertex clique associated with the n -th vertex and $h_{\text{ne}(n)}(\mathbf{z}_{\text{ne}(n)})$ is a cost function associated with its neighborhood. The resulting conditional distribution $p(z_n | \mathbf{z}_{\setminus n})$ can be written as

$$p(z_n | \mathbf{z}_{\setminus n}) \propto \exp[H_n(\mathbf{z})]. \quad (4.64)$$

4.2.2 Markov random field as a prior in a DPMM

The random variables \mathbf{z} define a partition of the data. They are discrete random variables holding a categorical value associated with the pixel clusters (i.e., they represent an identification or label, and their actual numeric value is irrelevant). Since they take categorical values, the sole interaction between z_n and z_m should consist of evaluating whether they have the same value or not. This is obtained by a cost function that follows a Potts model [Pot52]

$$H(z_n | \mathbf{z}_{\setminus n}) = H_n(z_n) + \sum_{m \in \text{ne}(n)} \omega_{nm} \mathbb{1}_{z_n}(z_m) \quad (4.65)$$

$$= H_n(z_n) + \sum_{\substack{m \in \text{ne}(n) \\ z_n = z_m}} \omega_{nm} \quad (4.66)$$

where $\mathbb{1}_{z_n}(\cdot)$ is the indicator function, and ω_{nm} is the weight of the edge connecting vertices n and m . Note that an arbitrary cost function $H_n(z_n)$ can be chosen.

It is necessary, however, to relate the prior given by the MRF with the prior obtained through the CRP. This can be achieved by simply considering the cost function

$$H_n(z_n) = \log p(z_n | \mathbf{i}_n, \mathbf{V}') \quad (4.67)$$

where \mathbf{V}' is the infinite dimensional parameter vector containing the vectors \mathbf{v}' that describe the components of the DPMM. It is important to note that z_n is independent of any other z_m conditionally to $\mathbf{i}_n, \mathbf{V}'$, since using (4.20) to (4.23) we can write

$$p(z_n | \mathbf{i}_n, \mathbf{V}') \propto p(\mathbf{i}_n | z_n, \mathbf{V}') \times p(z_n) \quad (4.68)$$

$$\propto p_{\mathcal{F}}(\mathbf{i}_n | \mathbf{v}'_{z_n}) \times w_{z_n}. \quad (4.69)$$

As such, $H_n(z_n)$ does not depend on any z_m for $m \neq n$, so that this cost function defines a valid MRF, as required by the Hammersley-Clifford theorem. The MRF defined by this cost function is given by

$$p(z_n | \mathbf{z}_{\setminus n}, \mathbf{i}_n, \mathbf{V}') \propto p(z_n | \mathbf{i}_n, \mathbf{V}') \prod_{\substack{m \in \text{ne}(n) \\ z_n = z_m}} e^{\omega_{nm}}. \quad (4.70)$$

The next step is to integrate out \mathbf{V}' from (4.70) and replacing by (4.34) to obtain the following result

$$\begin{aligned} & \int p(z_n | \mathbf{i}_n, \mathbf{V}') p(\mathbf{V}' | \mathbf{z}_{\setminus n}, \mathbf{I}_{\setminus n}, \mathcal{V}_0) d\mathbf{V}' \\ &= p(z_n | \mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0) \end{aligned} \quad (4.71)$$

$$\propto \begin{cases} \alpha p(\mathbf{i}_n | \mathcal{V}_0) & \text{if } z_n = 0 \\ N'_{z_n} \frac{p(\mathbf{I}_{\{z_n\}} | \mathcal{V}_0)}{p(\mathbf{I}_{\{z_n\} \setminus n} | \mathcal{V}_0)} & \text{if } 1 \leq z_n \leq K. \end{cases} \quad (4.72)$$

As a consequence, integrating out \mathbf{V} from (4.70) leads to

$$p(z_n | \mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0) = \int p(z_n | \mathbf{z}_{\setminus n}, \mathbf{i}_n, \mathbf{V}) p(\mathbf{V} | \mathbf{z}_{\setminus n}, \mathbf{I}_{\setminus n}, \mathcal{V}_0) d\mathbf{V} \quad (4.73)$$

$$= \int p(z_n | \mathbf{i}_n, \mathbf{V}) \prod_{\substack{m \in \text{ne}(n) \\ z_n = z_m}} e^{\omega_{nm}} p(\mathbf{V} | \mathbf{z}_{\setminus n}, \mathbf{I}_{\setminus n}, \mathcal{V}_0) d\mathbf{V} \quad (4.74)$$

$$\propto \begin{cases} \alpha p(\mathbf{i}_n | \mathcal{V}_0) & \text{if } z_n = 0 \\ N'_{z_n} \frac{p(\mathbf{I}_{\{z_n\}} | \mathcal{V}_0)}{p(\mathbf{I}_{\{z_n\} \setminus n} | \mathcal{V}_0)} \prod_{\substack{m \in \text{ne}(n) \\ z_n = z_m}} e^{\omega_{nm}} & \text{if } 1 \leq z_n \leq K \end{cases} \quad (4.75)$$

which is the conditional probability of the DPM-MRF model that has to be included in the partially collapsed Gibbs sampler. It should be noted that by integrating out \mathbf{V}' the density $p(z_n | \mathbf{z}_{\setminus n}, \mathbf{I}, \mathcal{V}_0)$ is no longer independent from any other z_m . However, the MRF is not defined conditional to \mathcal{V}_0 but conditional to \mathbf{V}' , so that the conditions for the Hammersley-Clifford theorem hold. We would like to remark that the only difference between (4.75) and (4.34) is that the DPM-MRF conditional distribution requires to

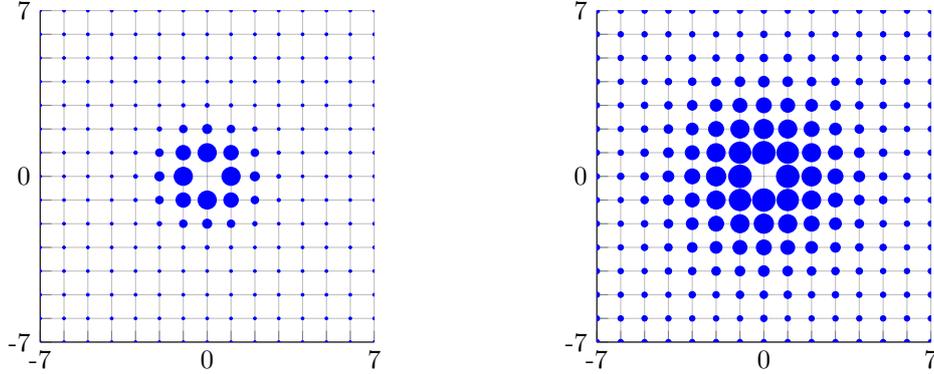


Figure 4.5: Example of two neighborhood weights, for $\sigma = 1$ (left) and $\sigma = 2$ (right). The size of the circles represents the weight of that pixel with respect to the pixel in $(0, 0)$.

define a set of weights relating a pixel with its neighbors and to keep track of which pixels in the neighborhood belong to the same class. An expected result of such similarity is that setting all the weights to $\omega_{nm} = 0$ (i.e., by removing spatial correlation), the suggested model reduces to the DPMM model described in Section 4.1.2.

In order to guarantee the MRF homogeneity, the weight relating a pixel to its neighborhood is defined by an isotropic function of the spatial distance between the pixels m and n (denoted as d_{mn}) such that

$$\lim_{d_{mn} \rightarrow \infty} \omega(d_{mn}) = 0 \quad (4.76)$$

which means “the more distant two pixels, the less correlated their classes”. This function is generally chosen as a constant value on the Moore neighborhood [G+03; OB08] (pixels with an L_∞ distance of 1) and zero anywhere else. However, we propose to use a Gaussian pdf as a weighting function, which verifies (4.76) and has a maximum for $d_{mn} = 0$, which is expressed as

$$\omega(d_{mn}|\lambda, \sigma) = \lambda \exp\left(-\frac{d_{mn}^2}{\sigma^2}\right) \quad (4.77)$$

where λ controls the influence of the MRF in the model, and σ controls the neighborhood size. We have chosen this weighting function since it promotes smoothness and since it provides a simple parametrization allowing to modify the radius of the neighborhood. For computational simplicity, we have chosen to set $\omega(d_{mn}) = 0$ for $d_{mn} > 5\sigma$.

Algo. 5 summarizes the resulting parameter estimation algorithm for the DPM-MRF based on a collapsed Gibbs sampler. Note that each cluster keeps track of the pixels it

contains through the sufficient statistics \mathbf{T}_k and that each pixel \mathbf{i}_n keeps track of the influence of the surrounding pixels through $w(d_{mn})$. Note also that the DPMM parameter α is estimated from its posterior distribution at the end of the algorithm.

4.3 Performance on synthetic images

Section 3.4 presented the performance of the proposed change detector using an EM based algorithm to obtain the vectors $\hat{\mathbf{v}}$. As it is mentioned there, and shown in Fig. 3.8, deciding what object (or mixture component) each pixel belongs to based only on the pixel intensity is very error prone, specially when the different components of the mixture distribution are close one to the other. Due to the limitations observed in Fig. 3.8, we were obliged to define a window based distance measure d_W , which in turn, limits the resolution of the change detector.

Section 4.2 introduces an MRF prior into the model, so that the cluster assignment of each pixel does not depend solely on its intensities, but also on its spatial location. Fig. 4.6 shows the pixel assignment obtained with EM and the one obtained with BNP-MRF. It is clear that the introduction of the spatial correlation significantly improves the pixel assignment. This is explained by the fact that even if a pixel intensity is ambiguous and thus, difficult to assign, the neighboring pixels influence the decision. This can be clearly seen in Fig. 4.6(e), where red pixels can be found in areas closer to the blue component of the mixture, and vice-versa. These pixels would have been missassigned by the EM based estimator proposed in Section 3.3. Due to this, the change detection can now be performed pixel-wise instead of window-wise, so that the distance $d_{\mathcal{M}}(\hat{\mathbf{v}})$ can be used as defined in (3.11).

Moreover, we previously discussed a problem regarding the window size, which is also greatly improved by the BNP-MRF method. The analysis window size should not be too small, otherwise the number of samples used to estimate $\hat{\mathbf{v}}$ is reduced, resulting in a high variance of $\hat{\mathbf{v}}$. However, an increased window size presented two major disadvantages. First, it reduced the change detector resolution, which is not a problem anymore since now we perform a pixel-wise change detection. Second, it increased the number of objects in the mixture, which has two consequences

- The number of objects could easily be outside of the predefined range.
- Having several components increases the probability of finding two components that are similar, which are prone to be merged by the EM estimator. Moreover,

Algorithm 5: A collapsed Gibbs sampler implementing a DPM-MRF parameter estimator with unknown α .

Input: $\mathbf{I} = \{\mathbf{i}_1, \dots, \mathbf{i}_N\}$, \mathcal{V}_0 , λ , σ , j_{\max} , j_{\min}

Output: $\mathbf{z} = \{z_1, \dots, z_N\}$

- 1 $z_n^{(0)} \leftarrow 0$, $\forall 1 \leq n \leq N$;
- 2 $K \leftarrow 0$;
- 3 $\alpha \leftarrow 1$;
- 4 $\omega_{nm} \leftarrow \lambda \exp\left(-\frac{d_{mn}^2}{\sigma^2}\right)$, $\forall 1 \leq n \leq N$, $m \in \text{ne}(n)$;
- 5 **for** $j : 1 \leq j \leq j_{\max}$ **do**
- 6 $z_n^{(j)} \leftarrow z_n^{(j-1)}$, $\forall 1 \leq n \leq N$;
- 7 **for** $n : 1 \leq n \leq N$, *in random order* **do**
- 8 /* Remove the n-th pixel from its current class */
- 9 **if** $z_n^{(j)} \neq 0$ **then**
- 10 $\mathbf{T}_{z_n^{(j)}} \leftarrow \mathbf{T}_{z_n^{(j)}} - \mathbf{T}(\mathbf{i}_n)$; $N_{z_n^{(j)}} \leftarrow N_{z_n^{(j)}} - 1$;
- 11 $h_{m, z_n^{(j)}} \leftarrow h_{m, z_n^{(j)}} - \omega_{mn}$, $\forall m \in \text{ne}(n)$;
- 12 /* Sample a new class for the n-th pixel */
- 13 $p_0 \leftarrow \alpha \text{p}_{\mathbf{T}}(\mathbf{T}(\mathbf{i}_n) | \mathcal{V}_0)$;
- 14 $p_k \leftarrow N_k \frac{\text{p}_{\mathbf{T}}(\mathbf{T}_k + \mathbf{T}(\mathbf{i}_n) | \mathcal{V}_0)}{\text{p}_{\mathbf{T}}(\mathbf{T}_k | \mathcal{V}_0)} e^{h_{n,k}}$, $\forall 1 \leq k \leq K$;
- 15 $z_n^{(j)} \sim \text{Cat}(p_0, p_1, \dots, p_K)$;
- 16 /* Place the n-th pixel in its new class */
- 17 **if** $z_n^{(j)} = 0$ **then**
- 18 $K \leftarrow K + 1$; $z_n^{(j)} \leftarrow K$;
- 19 $\mathbf{T}_K \leftarrow \mathbf{T}(\mathbf{i}_n)$; $N_K \leftarrow 1$;
- 20 $h_{m,K} \leftarrow \omega_{mn}$, $\forall m \in \text{ne}(n)$;
- 21 $h_{m,K} \leftarrow 0$, $\forall m \notin \text{ne}(n)$;
- 22 **else**
- 23 $\mathbf{T}_{z_n^{(j)}} \leftarrow \mathbf{T}_{z_n^{(j)}} + \mathbf{T}(\mathbf{i}_n)$; $N_{z_n^{(j)}} \leftarrow N_{z_n^{(j)}} + 1$;
- 24 $h_{m, z_n^{(j)}} \leftarrow h_{m, z_n^{(j)}} + \omega_{mn}$, $\forall m \in \text{ne}(n)$;
- 25 /* Estimate the new α */
- 26 $t \sim \text{B}(\alpha, N)$; $M \leftarrow M(N, K, t)$;
- 27 $\alpha \sim \Gamma\left(K + \frac{1}{2}, -\frac{1}{\log t}\right)$;
- 28 **while** $u \sim \text{U}(0, 1)$, $u \text{p}_{\Gamma}(\alpha) < M \text{p}(\alpha | N, K, t)$ **do**
- 29 $\alpha \sim \Gamma\left(K + \frac{1}{2}, -\frac{1}{\log t}\right)$;
- 30 $z_n \leftarrow \text{mode}\left(z_n^{(j_{\max})}, \dots, z_n^{(j_{\min})}\right)$, $\forall 1 \leq n \leq N$;

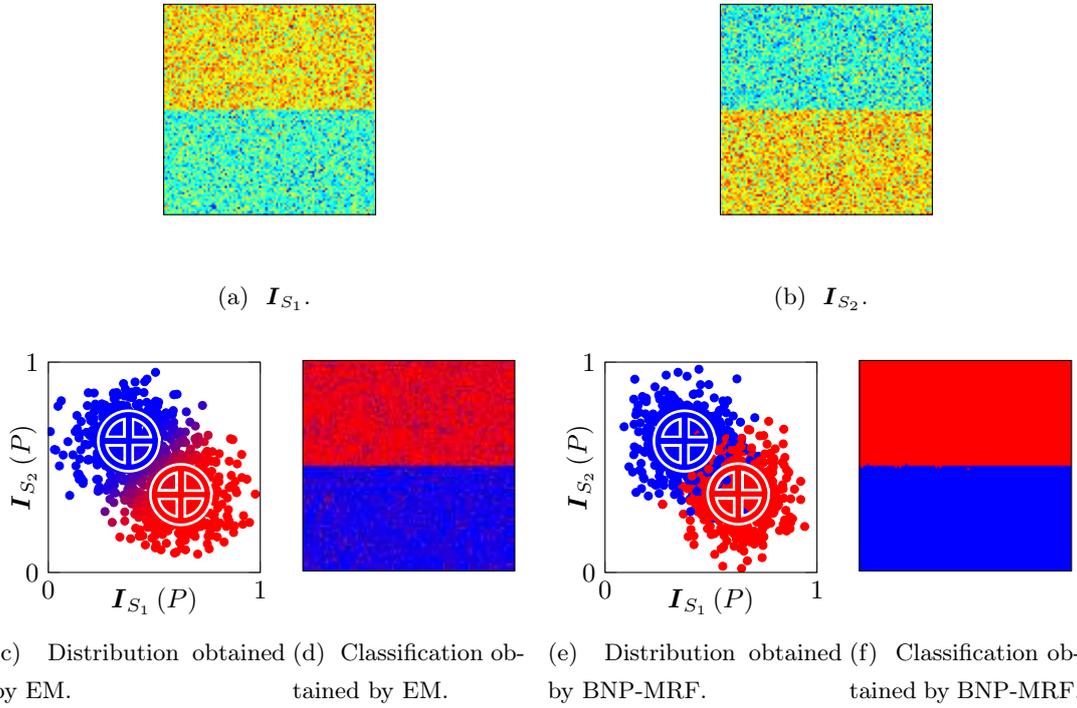


Figure 4.6: Figures (a) and (b) show the same images as in Fig. 3.8 captured by sensors S_1 and S_2 respectively. Figure (c) and (e) show the mixture distribution of their pixel intensity, and their estimated parameters represented by a circle with a cross for the EM and the BNP-MRF estimations respectively. Figure (d) and (f) presents the resulting classification of each pixel into each mixture component for the EM and the BNP-MRF estimations respectively.

event if these components are not merged, it leads to the problem depicted in Fig. 3.8.

The problems mentioned above do not affect the BNP-MRF approach, since the number of objects is unbounded thanks to the BNP, and similar components do not risk being merged if they are not spatially close thanks to the MRF.

Considering this, there is no problem to work with bigger windows when using the BNP-MRF algorithm other than memory or computational time issues (i.e., bigger windows implies tracking more pixel interactions and mixture components at the same time). Window sizes between 100×100 and 200×200 were found to present a good trade off, providing a reasonable computational time (which deserves attention when dealing with

MCMC methods), as well as being able to contain most objects present in the analyzed images. Moreover, with such a window size, we can process several windows in parallel, allowing one to increase the analysis speed.

Fig. 4.7 presents the simulation results obtained for a dataset consisting of two heterogeneous synthetic images. Fig. 4.7(b) is a synthetic image representing the intensity channel of a SAR image, while Fig. 4.7(a) is a synthetic optical image. Fig. 4.7(e) displays the ground truth showing in black the areas of the image that have been affected by changes. To obtain $d_{\mathcal{M}}(\hat{\mathbf{v}}_n)$, the labels z_n were computed using windows of size 100×100 , which was chosen to optimize the processing time. The neighborhood graph defined by $\omega(d_{mn}|\lambda, \sigma)$ was obtained with $\lambda = 30$ and $\sigma = 6$, with $\omega(d_{mn}) = 0$ for $d_{mn} \geq 30$. This choice of σ was done taking into account the average size of the triangles features in the images. The windows were processed using an overlap of 30 pixels in order to ensure the continuity of the spatial correlation in z_n . The values $\hat{\mathbf{v}}_n$ were obtained by computing the maximum likelihood estimator of \mathbf{v}_n using $\mathbf{I}_{\{z_n\}}$, while the manifold \mathcal{M} was estimated using 1% of the pixels randomly selected from unchanged areas.

Fig. 4.7(c) shows $\log(d_{\mathcal{M}}(\hat{\mathbf{v}}))$, the computed distance to the manifold \mathcal{M} , while Fig. 4.7(d) shows the corresponding distance $\log d_W$ obtained with the change detector based on EM, where a red color corresponds to a large value of $d_{\mathcal{M}}(\hat{\mathbf{v}})$ and d_W respectively. It can be observed that Fig. 4.7(c) provides a more accurate change detection compared to Fig. 4.7(d) (which is highlighted in the circular zoom area) due to BNP-MRF producing a pixel-wise change detection, while EM produces a window-wise change detection. Fig. 4.7(f) compares the ROC [PBF54] curves obtained with our method as well as other classical methods. The proposed BNP-MRF model provides better results than the EM based model for this example. If we consider the working situation defined by equal probability of false alarm (PFA) and probability of non detection ($1 - \text{PD}$), the EM method provides an error rate of 5.52%, while the BNP-MRF method yields an error rate of 4.18%, which represents a reduction of 24%.

4.4 Conclusion

As discussed in Section 3.5, the estimation of \mathbf{v} is of crucial importance to the performance of the proposed change detection method. In Section 3.3, we proposed to use an EM based algorithm to compute the mixture parameters. However, two main limitations were highlighted, namely

Table 4.1: Change detector performance for different methods for the images in Fig. 3.11(a).

Method	PFA = PDN
BNP-MRF	4.18%
EM	5.52%
Correlation Coefficient	27.68%
Mutual Information	26.35%

- The number of components was heuristically obtained from a predefined range.
- The spatial correlation was not considered, disregarding valuable information that can help improve the estimation of the mixture parameters of our model.

In this chapter we improved this estimation by using several modifications into the process. First, a non parametric approach was introduced to estimate the number of mixture components belonging to each estimation window. Second, a non informative prior was derived for the concentration parameter of the resulting mixture model. Including this prior in the change detection model made the resulting parameter estimation approach suitable for scenarios including different window sizes and having a highly varying number of objects within an analysis window. Finally, a Markov random field was coupled with the Bayesian non parametric mixture model to account for the spatial correlation between adjacent pixels of remote sensing images. Notably, we introduced a new weight function for the MRF, that produced smooth results and allowed us to parametrize the neighborhood size of the MRF according to the image resolution and the size of the features to be observed. We also derived a new Jeffreys prior to the concentration parameter α , which improved its estimation compared to previous priors proposed in the literature. Finally, a collapsed Gibbs sampler MCMC algorithm was proposed to estimate the model parameters removing the need to sample the parameters of our mixture model, and improving the general convergence of the resulting algorithm.

All these modifications increased the robustness of the parameter estimation algorithm, allowing bigger sizes for the analysis windows to be considered and thus improving the parameter estimation accuracy. Moreover, the introduction of a Potts model led to

a classification pixel map allowing a pixel-level change detection strategy. The change detection rule resulting from the proposed Bayesian non-parametric model showed improved detection performance when compared with other strategies.

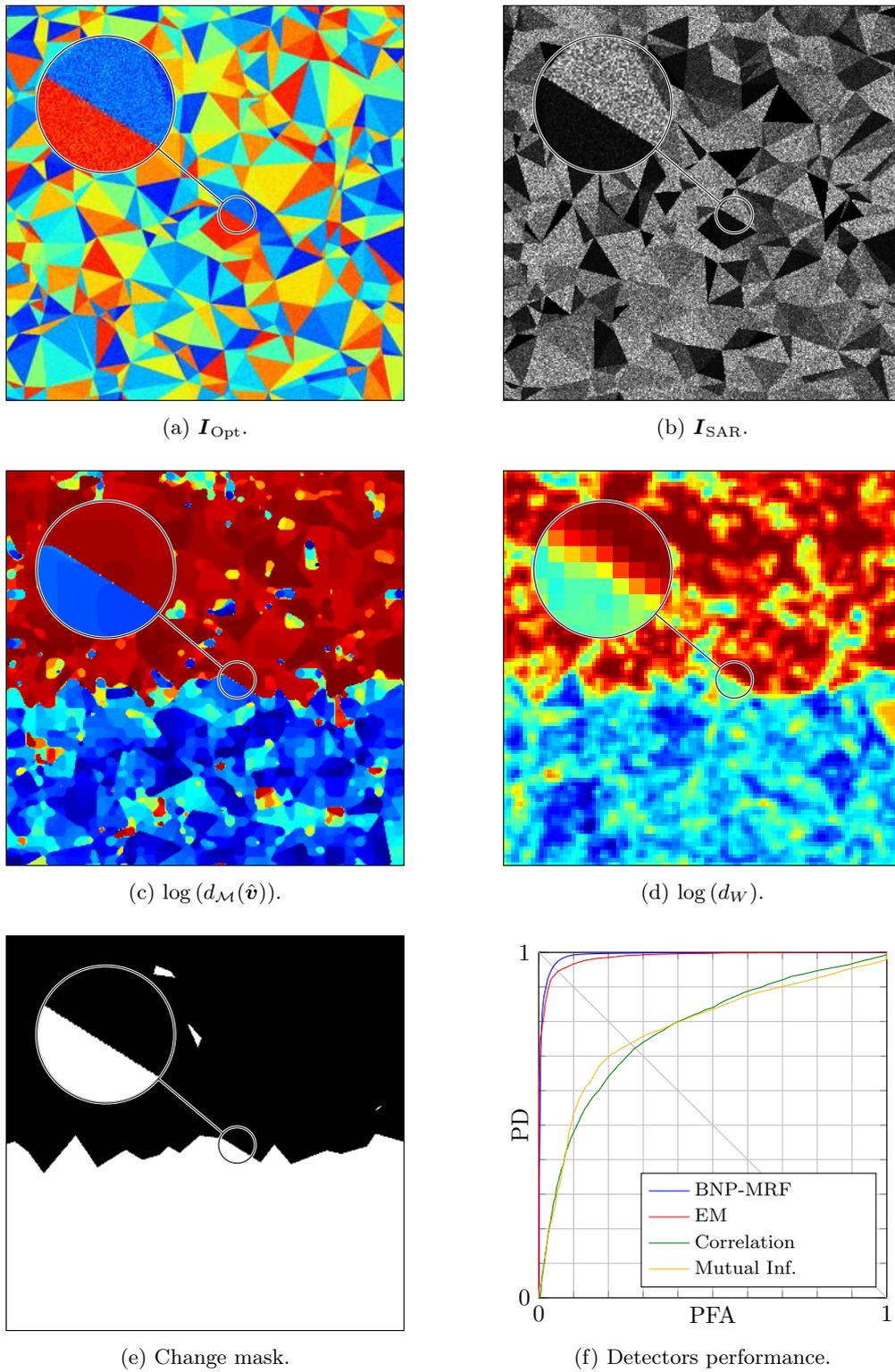


Figure 4.7: Results obtained by applying the proposed method into a dataset consisting of two synthetic heterogeneous optical and SAR images.

CHAPTER 5

Validation on real images

Contents

5.1	Homogeneous optical images	88
5.2	Heterogeneous optical images	92
5.3	Heterogeneous optical and SAR images	96
5.4	Conclusions	100

Sections 3.4 and 4.3 presented the performance of the proposed change detector on synthetic images. However, these images were generated based on the model presented in Section 2.2.2, which inspired the proposed change detector. Consequently, it is reasonable to think that the performance obtained on these datasets was favorable.

In order to validate the proposed method on more general conditions, this chapter evaluates the performance of the algorithm on different datasets consisting of real images. The change detection results obtained with the EM and BNP-MRF based estimators are compared with those obtained with different classical methods based on the mean pixel difference, mean pixel ratio, correlation coefficient and mutual information. The first two reference methods were provided by the ORFEO Toolbox [OTB14]. When available, the detector performance is also compared with the method of [MMS08] based on conditional copulas. Note that the method presented in [MMS08] is one of the most recent change detection methods that can be applied to both homogeneous and heterogeneous images.

Three datasets were chosen and are presented in Sections 5.1 to 5.3, that allows the detector to be tested in different conditions:

- Two homogeneous optical images, where the manifold \mathcal{M} is expected to be close to a straight line.
- Two heterogeneous optical images, where the manifold \mathcal{M} is expected to deviate from a straight line.
- Two heterogeneous optical and SAR images, where \mathcal{M} is completely unknown.

Finally, conclusions are highlighted in Section 5.4.

5.1 Homogeneous optical images

The first experiment considers a pair of images from an urban area in the south of Toulouse (France), acquired by Pleiades satellites within a time period of 16 months and characterized by some new constructions. Figs. 5.1(a) and 5.1(b) display multi-spectral images with 4 spectral bands (blue, green, red and infra-red) obtained by pan-sharpening the 2m resolution multispectral image and the corresponding 50cm resolution panchromatic image. Fig. 5.1(e) shows the ground truth provided by a photo-interpreter indicating in black the areas affected by changes.

To obtain $d_{\mathcal{M}}(\hat{\mathbf{v}})$ for the BNP-MRF approach, the labels z_n were computed on windows of size 200×200 , which was determined based on the system memory and processing power. The neighborhood graph defined by $\omega(d_{mn}|\lambda, \sigma)$ was obtained with $\lambda = 60$ and $\sigma = 1$, with $\omega(d_{mn}) = 0$ for $d_{mn} \geq 5$. The choice of a small value for σ is due to the fact that the scene corresponds to a urban area consisting mostly of small objects. The windows were processed using an overlap of 50 pixels in order to ensure the continuity of the spatial correlation in z_n . The values $\hat{\mathbf{v}}_n$ were obtained by computing the maximum likelihood estimator of \mathbf{v}_n using $\mathbf{I}_{\{z_n\}}$, while the manifold \mathcal{M} was estimated using 1% of the pixels randomly selected from unchanged areas. To obtain d_W for the EM approach, the mixture distribution was estimated using a window size of 10×10 pixels. The windows were processed using an overlap of 50% (5 pixels) in order to increase the output resolution of the detector. Note that the value of K was constrained to belong to the interval $[1, 10]$ for the EM based algorithm. The pixel difference and the pixel ratio measures were obtained as the L_2 norm of the measure obtained for each image channel.

Fig. 5.1(f) displays the manifold formed by the estimated vectors $\hat{\mathbf{v}}$. As expected, it lies within a straight line, although a second horizontal line can be seen at the bottom, whose presence is probably due to different illuminations casting different shadows on each image. This results in some objects being dark in one image, while bright in the other one. However, since these changes were not considered in the ground truth, they are included in the manifold.

Figs. 5.1(c) and 5.1(d) show the estimated distances $d_{\mathcal{M}}(\hat{\mathbf{v}})$ and d_W obtained with the proposed BNP-MRF and EM based estimators respectively. It can be observed that Fig. 5.1(c) provides a more accurate change detection compared to Fig. 5.1(d), which is highlighted in the circular zoom area. This improved performance is confirmed in the

Table 5.1: Performance of the different methods for detecting changes on the images of Fig. 5.1.

Method	PFA = PDN
BNP-MRF	9.62%
EM	16.51%
Correlation Coefficient	26.08%
Mutual Information	26.27%
Mean Pixel Difference	27.88%
Mean Pixel Ratio	29.67%

ROCs displayed in Fig. 5.2. Table 5.1 shows the error rates obtained with $PFA = 1 - PD$ for the different methods. In this condition, the EM based method reduces the error rate by 37% compared to the best performing classical method, while the BNP-MRF method reduces the error rate by 42% compared to the EM based method.

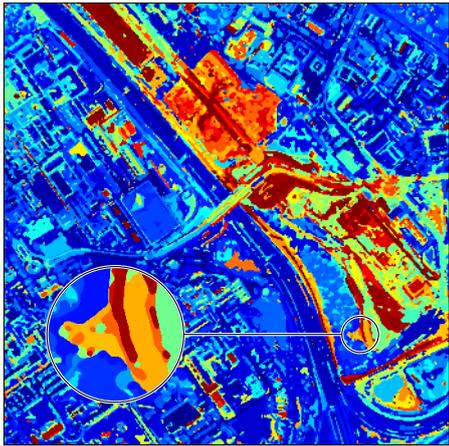
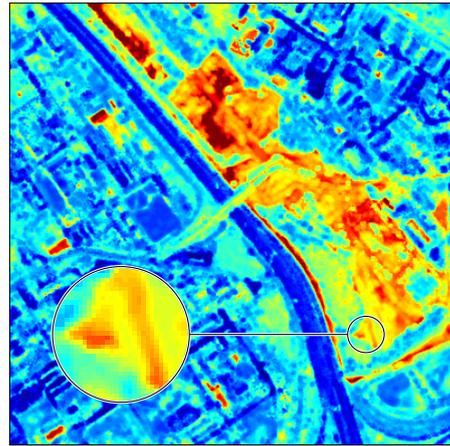
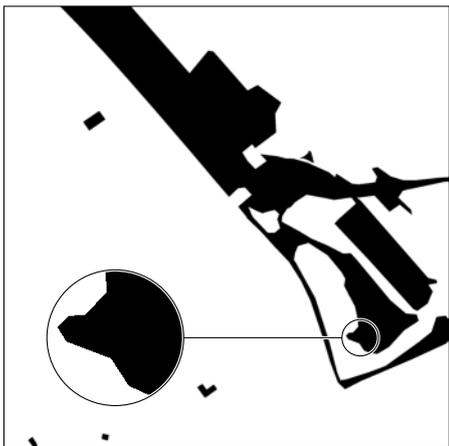
It is interesting to note that since the images are homogeneous, the pixel intensity of both images are linearly dependent. This remark explains why the correlation coefficient and the mutual information perform very similarly for this example.



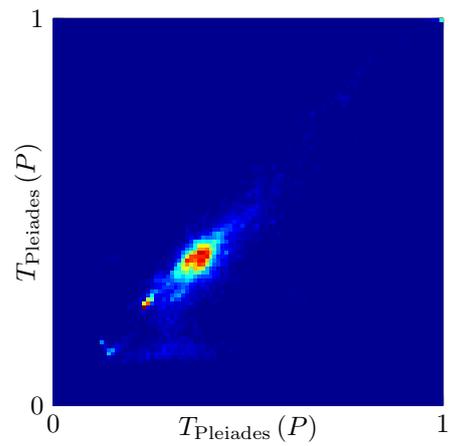
(a) Pleiades image – May 2012



(b) Pleiades image – Sept. 2013

(c) Distance $d_{\mathcal{M}}(\hat{v})$.(d) Distance d_W .

(e) Ground truth.



(f) Manifold estimation.

Figure 5.1: Results obtained by applying the proposed method into a dataset consisting of two real homogeneous optical images.

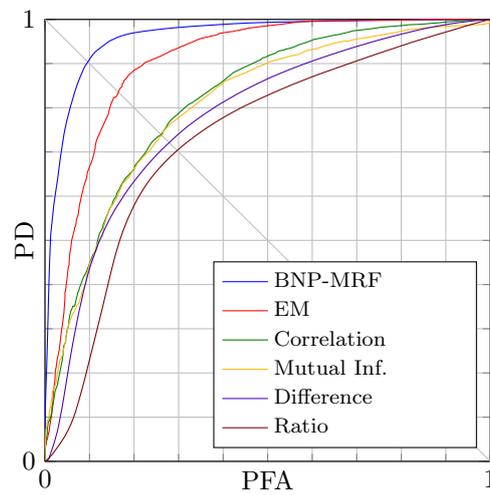


Figure 5.2: Change detection performance for the images of Fig. 5.1.

5.2 Heterogeneous optical images

The second simulation studies a pair of images from an urban area, in the same geographical position as the dataset used in Section 5.1 also characterized by some construction works made within a 14 month time interval. Fig. 5.3(a) is identical to Fig. 5.1(a), while Fig. 5.3(b) is a 3 channel image obtained from Google Earth which has been downsampled to match the 50cm resolution of the Pleiades image. Fig. 5.3(e) shows the ground truth provided by a photo-interpreter indicating in black the areas affected by changes. Note that the simulation scenario considered to obtain z_n , $d_{\mathcal{M}}(\hat{\mathbf{v}})$, d_W and $\hat{\mathbf{v}}_n$ was the same as the one used in Section 5.1. However, due to both images having a different number of channels (or spectral bands) the pixel difference and the pixel ratio measures were obtained as the L_2 norm of the measure obtained for the channels that both images have in common (i.e., red, green and blue).

Fig. 5.3(f) displays the manifold formed by the estimated vectors $\hat{\mathbf{v}}$. As expected, part of it lies within a straight line, but an important part of it deviates from the manifold, since the sensors present some differences. We believe that the main reason for this deviation is a big difference in the location of the green spectral filter employed by each sensor.

Figs. 5.3(c) and 5.3(d) show the estimated distances $d_{\mathcal{M}}(\hat{\mathbf{v}})$ and d_W obtained with the BNP-MRF and EM based models respectively. It can be observed that Fig. 5.3(c) provides an improved change detection compared to Fig. 5.3(d), which is highlighted in the circular zoom area. Moreover, some small changes that were not observed by the change mask were well detected by the BNP-MRF method. Fig. 5.4 shows the ROCs obtained with our method as well as other classical methods, illustrating the interest of the proposed model. If we consider the working situation defined by $\text{PFA} = 1 - \text{PD}$, the best performing classical method yields an error rate of 24.01%, while the EM based method yields an error rate of 15.6% and the BNP-MRF method leads to an error rate of 10.7%. This represents a reduction of 35% in the error rate by using the EM based method with respect to the classical methods, and a reduction of 31% by using the BNP-MRF approach with respect to the EM based method. Table 5.2 shows the error rates obtained for $\text{PFA} = 1 - \text{PD}$ using the different methods.

It should be noted that since these images are heterogeneous, the pixel intensities are not necessarily linearly dependent. This results in a significant performance reduction on the correlation coefficient. However, the performance of the manifold based approach as well as the mutual information methods, which are adapted to heterogeneous images,

Table 5.2: Performance of the different methods for detecting changes on the images of Fig. 5.3.

Method	PFA = PDN
BNP-MRF	10.71%
EM	15.58%
Correlation Coefficient	32.03%
Mutual Information	24.01%
Mean Pixel Difference	29.16%
Mean Pixel Ratio	33.87%

are very similar to what was previously obtained.

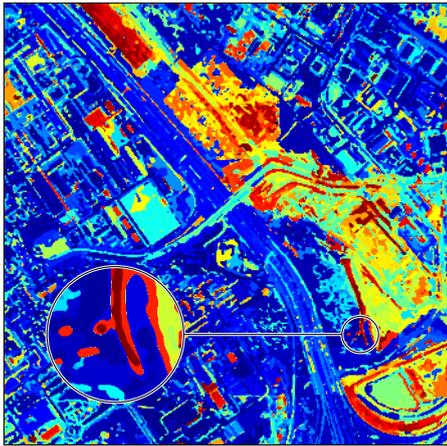
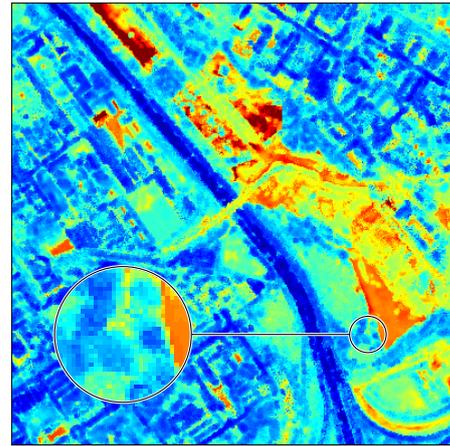
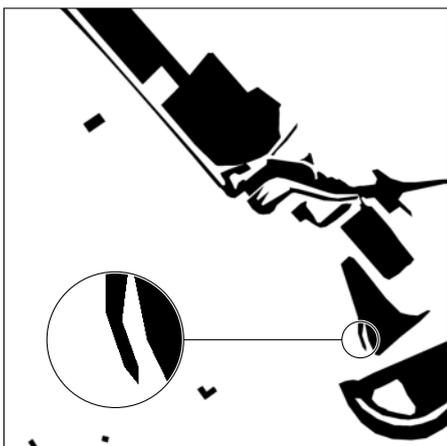
Since the BNP-MRF estimation is based on an MCMC algorithm, it is pertinent to mention the running time of these simulations. The images in Figs. 5.3(a) and 5.3(b) are 2000×2000 pixels, with 4 and 3 channels each. The running time of the EM based algorithm was 105s while the BNP-MRF method required a running time of 51m with 8 cores in parallel, which makes BNP-MERF 233 times slower than the EM method. However, since the processing can be parallelized this is not a huge drawback for its application in remote sensing.



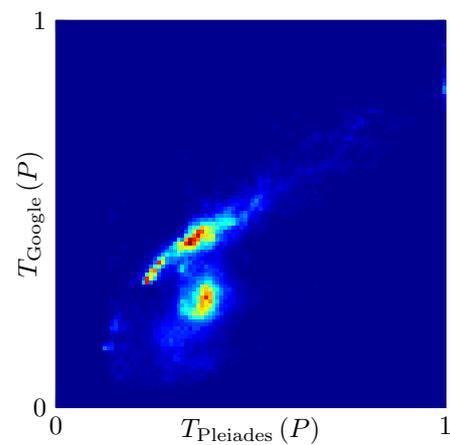
(a) Pleiades image – May 2012



(b) Google Earth – July 2013

(c) Distance $d_{\mathcal{M}}(\hat{v})$.(d) Distance d_W .

(e) Ground truth.



(f) Manifold estimation.

Figure 5.3: Results obtained by applying the proposed method into a dataset consisting of two real heterogeneous optical images.

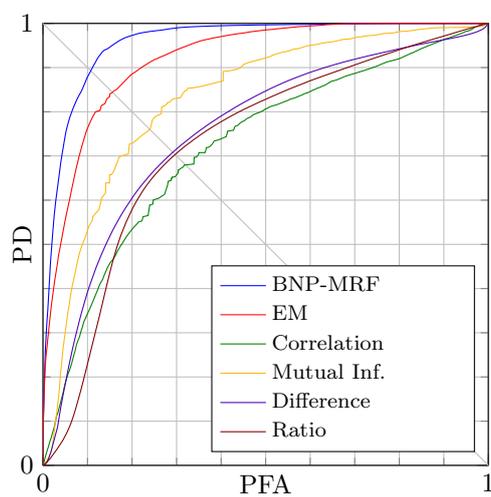


Figure 5.4: Change detection performance for the images of Fig. 5.3.

5.3 Heterogeneous optical and SAR images

The last experiment considers a pair of images from a mixture of urban and rural areas, near Gloucester before and during a flooding. Fig. 5.5(b) is the intensity channel of an image captured by a TerraSAR-X satellite with pixel resolution of 7.3m. Fig. 5.5(a) is a 3 channel image obtained from Google Earth which has been downsampled to match the pixel resolution of the TerraSAR-X image. Fig. 5.5(e) shows the ground truth provided by a photo-interpreter indicating in black the areas affected by the flooding.

To obtain $d_{\mathcal{M}}(\hat{\mathbf{v}})$ for the BNP-MRF approach, the labels z_n were computed on windows of size 200×200 , which was determined based on the system memory and processing power. The neighborhood graph defined by $\omega(d_{mn}|\lambda, \sigma)$ was obtained with $\lambda = 60$ and $\sigma = 6$, with $\omega(d_{mn}) = 0$ for $d_{mn} \geq 30$. The choice of σ was motivated by the fact that the scene corresponds mostly to a rural area with big homogeneous regions. The values $\hat{\mathbf{v}}_n$ were obtained by computing the maximum likelihood estimator of \mathbf{v}_n using $\mathbf{I}_{\{z_n\}}$, while the manifold \mathcal{M} was estimated using 1% of the pixels randomly selected from unchanged areas. To obtain d_W for the EM approach, the mixture distribution was estimated using a window size of 20×20 pixels. The windows were processed using an overlap of 50% (10 pixels) in order to increase the output resolution of the detector. Again, the presence of bigger objects motivated the selection of a bigger window size compared to that of Section 5.1. Note that the value of K was constrained to belong to the interval $[1, 10]$ for the EM based algorithm. The pixel difference and the pixel ratio measures were computed on the SAR intensity channel and the optical image luminance.

Fig. 5.5(f) displays the manifold obtained from the estimated \mathbf{v} . Since the speckle noise present in SAR tends to produce a higher variance in the resulting SAR image, it is expected to observe a higher variance in the estimated mixture parameters. Moreover, the result displayed in Fig. 5.5(f) is a 2-dimensional projection of a 4-dimensional space (since $T_{\text{opt}}(P)$ is actually a 3-dimensional vector) which is not as regular as in the previous cases.

Figs. 5.5(c) and 5.5(d) show the estimated distances $d_{\mathcal{M}}(\hat{\mathbf{v}})$ and d_W obtained with the proposed BNP-MRF and EM based models respectively. It can be observed that Fig. 5.5(c) provides a better change detection compared to Fig. 5.5(d), which is highlighted in the circular zoom area. Fig. 5.6 shows the ROCs obtained with our method and with alternative strategies including the method described in [MMS08]. This comparison is in favor of the proposed BNP-MRF model. If we consider the working situation where $\text{PFA} = 1 - \text{PD}$, the best performing classical method yields an error rate of 21.75%, while

Table 5.3: Performance of the different methods for detecting changes on the images of Fig. 5.5.

Method	PFA = PDN
BNP-MRF	8.19%
EM	14.58%
Conditional Copulas	23.96%
Correlation Coefficient	31.19%
Mutual Information	23.12%
Mean Pixel Difference	21.75%
Mean Pixel Ratio	18.61%

the EM based method yields an error rate of 14.6% and the BNP-MRF method leads to an error rate of 8.19%. This represents a reduction of 33% in the error rate by using the EM based method with respect to the classical methods, and a reduction of 44% by using the BNP-MRF approach with respect to the EM based method. which represents a reduction of 44%. Table 5.3 shows the error rates obtained for $PFA = 1 - PD$ using the different methods.

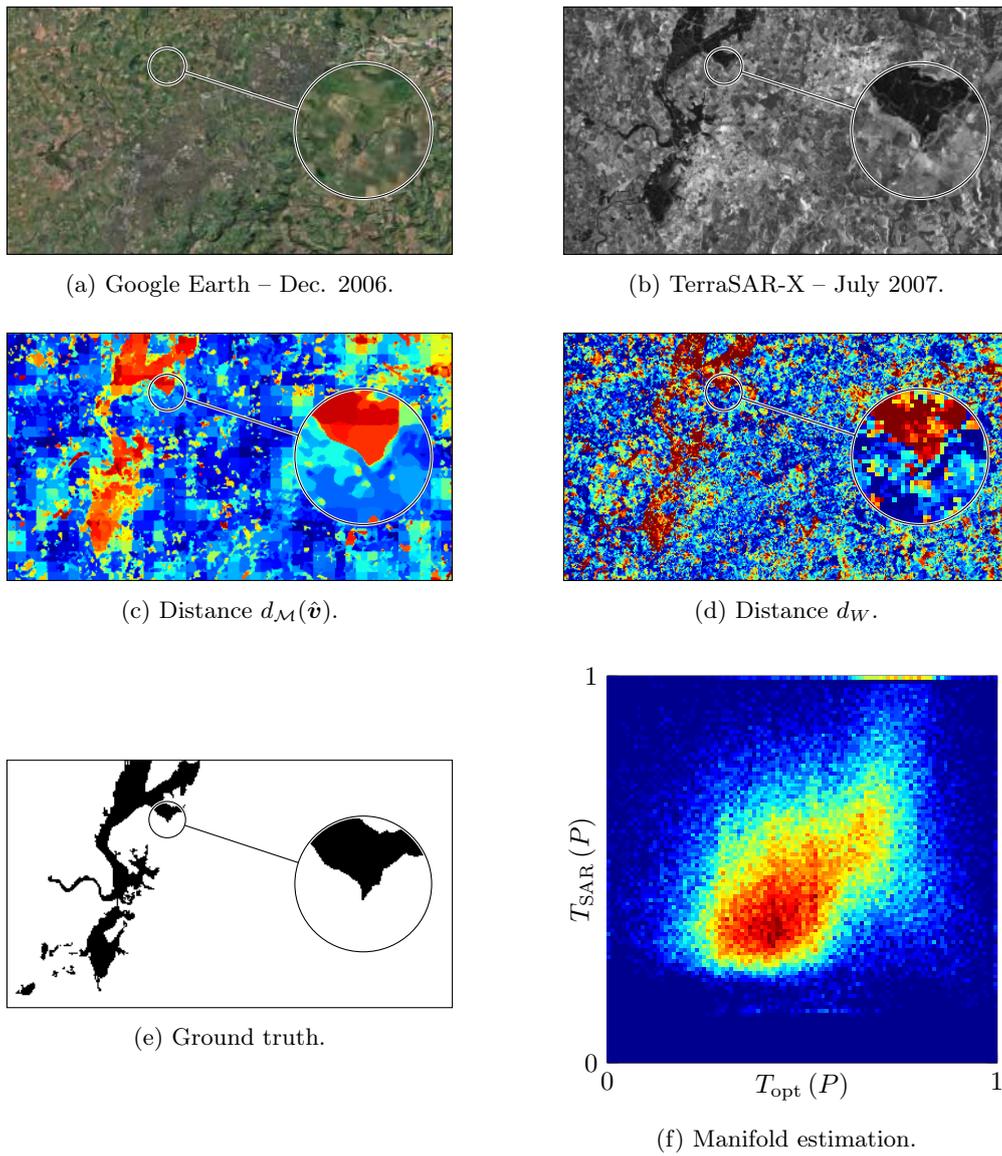


Figure 5.5: Results obtained by applying the proposed method into a dataset consisting of two real heterogeneous optical and SAR images.

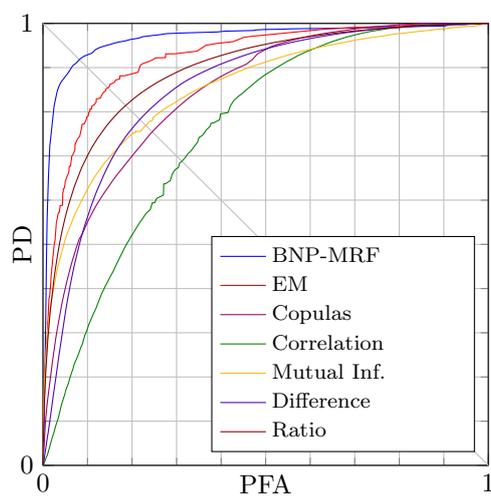


Figure 5.6: Change detection performance for the images of Fig. 5.5.

5.4 Conclusions

The results shown in Sections 3.4 and 4.3 were obtained for synthetic datasets generated based on the model presented in Section 2.2.2. These results were used to validate the theoretical viability of the proposed methods. However, they were not sufficient to assess the performance of the proposed method since the data was generated using the assumptions at the basis of the proposed change detector. In order to provide a more realistic performance overview, this chapter focused on testing the EM and BNP-MRF based approaches as well as other classic change detection algorithms on different real datasets. The performance was measured using the ROC curves of the different detectors, in favor of our algorithm.

The BNP-MRF method outperformed the EM based method in all cases, at the expense of a much higher computational time. The obtained improvement is not only due to the accuracy of the change detector, but also due to an improved resolution in the detection. This makes the BNP-MRF more suitable to detect small changes, as shown in Fig. 5.3(c), and also more robust to the selection of the window size. This was observed by the fact that the window size for the EM method had to be changed throughout the experiments, while the window size for the BNP-MRF method remained fixed. To finish, it is interesting to mention that the EM based estimator could still be useful in situations where the computational time has to be reduced.

CHAPTER 6

Conclusion and perspectives

Remote sensing images are becoming widely available nowadays, with many commercial and non-commercial services providing them. The sensor technology required to capture this kind of images is evolving fast. Not only classical sensors are improving in terms of resolution and noise level, but also new kinds of sensors are proving to be useful. The availability of different kinds of sensors allows us to combine their advantages for several image processing applications. One particular application that can take advantage of the complementarity of different sensors is the detection of changes between different remote sensing images. This application has received a significant interest in this PhD thesis, with a particular attention to the detection of changes between optical and SAR images.

This Phd manuscript is organized as follows. Chapter 1 introduces the different remote sensing images that have been considered in this work, with a specific attention to optical and SAR images. It also discusses the concept of heterogeneous and homogeneous sensors that are used to acquire these images. The motivations to detect changes between images produced by heterogeneous sensors, as well as the corresponding challenges are presented. Chapter 1 also briefly summarizes the different change detection methods proposed in the literature, their strengths and limitations. A specific attention is devoted to statistical change detection methods and to hypothesis tests, that have been used to decide the presence or absence of changes between different images. Several of these methods require the estimation of a joint pixel intensity distribution in order to compute an appropriate similarity measure.

Chapter 2 studies a statistical model to describe the joint distribution of several image intensities based on the assumption that the images are acquired by different sensors and are contaminated by independent noises. This model is first proposed for uniform areas of an image. It is then extended to the more general case of nonuniform areas, leading to a mixture model where each component represents a different object within the image. This mixture model is then used to analyze the performance of some classical change detection algorithms. Not only significant flaws are detected on dependency

based similarity measures, but we also show that improving the statistical model used to describe the joint pixel intensity distribution does not significantly improve the detection performance compared to simpler methods, such as those based on the histogram of the pixels located inside an analysis window. From this analysis we conclude that a new similarity measure is required in order to improve the change detection performance for remote sensing images, which will be the objective of the next chapters.

Chapter 3 begins by studying the behavior of the expectation maximization algorithm used to estimate the parameters of the mixture model introduced in Chapter 2 for change detection. We assume that in the absence of changes the parameters of this mixture model belong to a manifold, contrary to cases associated with the presence of changes. Based on this statement, we propose to use the distance to the manifold as a similarity measure for change detection between different images. Since this manifold is *a priori* unknown, we propose a strategy to estimate the distance between this manifold and a parameter vector associated with the model introduced in Chapter 2. More precisely, the mixture model parameters are estimated using a modified version of the expectation maximization algorithm, which allows the number of objects (or number of components in the mixture distribution) contained within an analysis window to be estimated. This strategy leads to a window based change detection, where changes are detected for any analysis window. The performance of the proposed change detector (referred to as EM detector) is assessed through its receiver operating characteristics computed from several datasets composed of synthetic images representing heterogeneous optical and SAR images, real homogeneous optical image, real heterogeneous optical images and heterogeneous optical and SAR images. In all studied cases the proposed method provides very competitive results when compared to other state-of-the-art change detection strategies. However, some limitations related to the parameter estimation method are identified, namely, the number of components in the mixture distribution has to be estimated heuristically from a predefined finite set, and the model does not consider the spatial correlation between adjacent pixels of the images.

Chapter 4 addresses the previously mentioned limitations about the parameter estimation method using a new Bayesian nonparametric model. The first limitation (unknown number of mixture components for any analysis window) is bypassed by considering the number of components as a random variable, which leads to a Bayesian non parametric mixture model. The parameters of this model can be estimated using a Markov chain Monte Carlo method. More precisely, we introduce a collapsed Gibbs

sampler which has improved convergence when compared with standard Markov chain Monte Carlo methods. This model depends in particular on a Jeffreys prior for the parameter of a Chinese restaurant process. The second limitation (taking advantage of the spatial correlations between adjacent pixels of the images) is addressed by including a Markov random field prior into the model. A non standard graph weighting function is defined for this Markov random field prior in order to introduce smoothness and control the influence of the spatial correlation. A fundamental consequence of this improved model is that change detection can be performed at a pixel level and no longer at a window level. This property leads to an increased output resolution of the change detector, and allows changes associated with smaller objects to be detected. The performance of the resulting change detector (referred to as BNP-MRF detector) is assessed through its receiver operating characteristics as for the EM detector. In all cases the change detector based on the proposed Bayesian non parametric algorithm clearly outperforms the algorithm studied in Chapter 2 and other state-of-the-art change detection methods.

Some perspectives for improving the change detection approach proposed in this thesis are finally presented. In Chapter 2, the parameters of the proposed mixture model are estimated with a maximum likelihood estimator, which is not robust to the presence of outliers (resulting for instance from dead pixels and specular reflections). The presence of these outliers might be mitigated by considering robust estimators such as M-estimators [Hub04] which would deserved to be studied in the context of the proposed change detector. Another approach that would deserve further analysis consists of considering the outliers as an additive error term subjected to sparsity constraints as in [Wri+09]. The proposed statistical model assumes that all the captured image noises are independent, which might not be realistic when several images are preprocessed together, as in the case of pansharpened optical images. In this case, considering a mixture model that allows certain correlation between groups of images could be considered. For instance, a 4-dimensional Gaussian distribution with full correlation matrix could be considered for pansharpened Pléiades images. Another research direction would be to extend the proposed model to features extracted from the image, i.e., the wavelet coefficients associated with a remote sensing image are known to be distributed according to a generalized Gaussian distribution [All12], independently of the sensor used to acquire the image. Using the wavelet coefficients would allow us to detect changes for images acquired from any sensor without having to build a statistical model for each particular sensor. The

proposed parameter estimation algorithm requires the a priori knowledge of the distribution of each image intensity, which may be a problem in some practical applications. In these cases, different approaches based on the empirical likelihood framework could be investigated. These approaches do not require to define a parametric family for the observed data [Owe01], i.e., we can think of estimating the mixture parameters related to the scene without having to define a distribution family for the mixture components. Finally, Chapter 4 studied a BNP-MRF model for change detection. The main drawback of this model is the computational cost associated with its estimation algorithm. However, the number of iterations of the MCMC algorithm might be reduced by providing an informative initialization instead of a random one. The mean shift algorithm [CM02], that has received a considerable attention in the literature, is an algorithm that would deserve to be embedded in the proposed estimation method.

Bibliography

- [Alb+07] V. Alberga et al. “Comparison of similarity measures of multi-sensor images for change detection applications”. In: *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*. Barcelona, Spain, July 2007, pp. 2358–2361.
- [Alb09] V. Alberga. “Similarity Measures of Remotely Sensed Multi-Sensor Images for Change Detection Applications”. In: *Remote Sensing* 1.3 (2009), pp. 122–143.
- [All12] M.S. Allili. “Wavelet Modeling Using Finite Mixtures of Generalized Gaussian Distributions: Application to Texture Discrimination and Retrieval”. In: *IEEE Trans. Image Process.* 21.4 (Apr. 2012), pp. 1452–1464.
- [Alp+07] L. Alparone et al. “Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest”. In: *IEEE Trans. Geosci. Remote Sens.* 45.10 (Oct. 2007), pp. 3012–3021.
- [Ant74] C. E. Antoniak. “Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems”. In: *Ann. Statist.* 2.6 (Nov. 1974), pp. 1152–1174.
- [Bar+14] C. Barbey et al. “Comparaison des images Pléiades et drone pour la viticulture de precision dans le cadre du programme Earthlab”. In: *R. Fr. Photogrammétrie Télédétection*. 208 (Oct. 2014), pp. 123–126.
- [BB07] F. Bovolo and L. Bruzzone. “A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain”. In: *IEEE Trans. Geosci. Remote Sens.* 45.1 (2007), pp. 218–236.
- [BBM05] Y. Bazi, L. Bruzzone, and F. Melgani. “An Unsupervised Approach Based on the Generalized Gaussian Model to Automatic Change Detection in Multitemporal SAR Images”. In: *IEEE Trans. Geosci. Remote Sens.* 43.4 (2005), pp. 874–887.

- [Ben+15a] Cs. Benedek et al. “Multilayer Markov Random Field Models for Change Detection in Optical Remote Sensing Images”. In: *ISPRS J. Photogrammetry and Remote Sensing* 107 (2015), pp. 22–37.
- [Ben+15b] Cs. Benedek et al. “Multilayer Markov Random Field Models for Change Detection in Optical Remote Sensing Images”. In: *ISPRS J. Photogrammetry and Remote Sensing* 107 (2015), pp. 22–37.
- [BF00a] L. Bruzzone and D. Fernández Prieto. “An adaptive parcel-based technique for unsupervised change detection”. In: *Int. J. Remote Sensing* 21.4 (2000), pp. 817–822.
- [BF00b] L. Bruzzone and D. Fernández Prieto. “Automatic Analysis of the Difference Image for Unsupervised Change Detection”. In: *IEEE Trans. Geosci. Remote Sens.* 38.3 (2000), pp. 1171–1182.
- [BF02] L. Bruzzone and D. Fernández Prieto. “An adaptive semi parametric and context-based approach to unsupervised Change Detection in multitemporal remote-sensing images”. In: *IEEE Trans. Image Process.* 11.4 (2002), pp. 452–466.
- [BF97] L. Bruzzone and D. Fernández Prieto. “An Iterative Technique for the Detection of Land-Cover Transitions in Multitemporal Remote-Sensing Images”. In: *IEEE Trans. Geosci. Remote Sens.* 35.4 (1997), pp. 858–867.
- [BKR11] A. Blake, P. Kohli, and C. Rother. *Markov Random Fields for Vision and Image Processing*. MIT Press, 2011.
- [BS09] Cs. Benedek and T. Szirányi. “Change Detection in Optical Aerial Images by a Multi-Layer Conditional Mixed Markov Model”. In: *IEEE Trans. Geosci. Remote Sens.* 47.10 (2009), pp. 3416–3430.
- [BS97] L. Bruzzone and S. B. Serpico. “Detection of Changes in Remotely-Sensed Images by the Selective Use of Multispectral Information”. In: *Int. J. Remote Sensing* 18.18 (Dec. 1997), pp. 3883–3888.
- [BZV04] N. Bouguila, D. Ziou, and J. Vaillancourt. “Unsupervised learning of a finite mixture model based on the Dirichlet distribution and its application”. In: *IEEE Trans. Image Process.* 13.11 (Nov. 2004), pp. 1533–1543.
- [Cas96] K. R. Castleman. *Digital image processing*. Prentice Hall, 1996. ISBN: 9780132114677.

- [CDB06] C. Carincotte, S. Derrode, and S. Bourennane. “Unsupervised change detection on SAR images using fuzzy hidden Markov chains”. In: *IEEE Trans. Geosci. Remote Sens.* 44.2 (2006), pp. 432–441.
- [Cel09a] T. Celik. “Multiscale change detection in multitemporal satellite images”. In: *IEEE Geosci. Remote Sens. Lett.* 6.4 (2009), pp. 820–824.
- [Cel09b] T. Celik. “Unsupervised change detection in satellite images using principal component analysis and k-means clustering”. In: *IEEE Geosci. Remote Sens. Lett.* 6.4 (2009), pp. 772–776.
- [CG92] G. Casella and E. I. George. “Explaining the Gibbs Sampler”. In: *The American Statistician* 46.3 (1992), pp. 167–174.
- [CGM95] W. G. Carrara, R. S. Goodman, and R. M. Majewski. *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*. Artech House signal processing library. Artech House, 1995. ISBN: 9780890067284.
- [Cha+07] F. Chatelain et al. “Bivariate Gamma Distributions for Image Registration and Change Detection”. In: *IEEE Trans. Image Process.* 16.7 (2007), pp. 1796–1806.
- [CK10] T. Celik and M. Kai-Kuang. “Unsupervised Change Detection for Satellite Images Using Dual-Tree Complex Wavelet Transform”. In: *IEEE Trans. Geosci. Remote Sens.* 48.3 (2010), pp. 1199–1210.
- [CM02] D. Comaniciu and P. Meer. “Mean shift: a robust approach toward feature space analysis”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 24.5 (May 2002), pp. 603–619.
- [CM91] J. C. Curlander and R. N. McDonough. *Synthetic Aperture Radar: Systems and Signal Processing*. Wiley Series in Remote Sensing and Image Processing. Wiley, 1991. ISBN: 9780471857709.
- [CT10] S. Chatzis and G. Tsechpenakis. “The Infinite Hidden Markov Random Field Model”. In: *IEEE Trans. Neural Netw.* (2010).
- [CT11] M. Chabert and J.-Y. Tournet. “Bivariate Pearson distributions for remote sensing images”. In: *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*. Vancouver, Canada, July 2011, pp. 4038–4041.

- [CVA03] H.-M. Chen, P. K. Varshney, and M.K. Arora. “Performance of mutual information similarity measure for registration of multitemporal remote sensing images”. In: *IEEE Trans. Geosci. Remote Sens.* 41.11 (Nov. 2003), pp. 2445–2454.
- [EW95] M. D. Escobar and M. West. “Bayesian Density Estimation and Inference Using Mixtures”. In: *J. Amer. Stat. Assoc.* 90.430 (1995), pp. 577–588.
- [FJ02] M. A. T. Figueiredo and A. K. Jain. “Unsupervised learning of finite mixture models”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 24.3 (Mar. 2002), pp. 381–396.
- [Fjo+88] R. Fjortoft et al. “An optimal multiedge detector for SAR image segmentation”. In: *IEEE Trans. Geosci. Remote Sens.* 36.3 (1988), pp. 793–802.
- [Fro+82] V. S. Frost et al. “A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 4.2 (Mar. 1982), pp. 157–166.
- [FS86] A. M. Fraser and H. L. Swinney. “Independent coordinates for strange attractors from mutual information”. In: *Phys. Rev. A* 33 (2 Feb. 1986), pp. 1134–1140.
- [Fun90] T. Fung. “An assessment of TM imagery for land-cover change detection”. In: *IEEE Trans. Geosci. Remote Sens.* 28.12 (1990), pp. 681–684.
- [G+03] L. Gray, A. New, et al. “A mathematician looks at wolfram’s new kind of science”. In: *Notices Amer. Math. Soc.* 2003.
- [GB12] S. J. Gershman and D. M. Blei. “A tutorial on Bayesian nonparametric models”. In: *J. Math. Psychol.* 56 (2012), pp. 1–12.
- [Gel09] A. Gelman. “Bayes, Jeffreys, prior distributions and the philosophy of statistics”. In: *Statistical Science* (2009), pp. 176–178.
- [Gho+07] S. Ghosh et al. “A context sensitive technique for unsupervised change detection based on Hopfield type neural networks”. In: *IEEE Trans. Geosci. Remote Sens.* 45.3 (2007), pp. 778–789.
- [Giu+13] L. Giustarini et al. “A Change Detection Approach to Flood Mapping in Urban Areas Using TerraSAR-X”. In: *IEEE Trans. Geosci. Remote Sens.* 51.4 (2013), pp. 2417–2430.

- [Gri73] G. R. Grimmett. “A Theorem about Random Fields”. In: *Bulletin of the London Mathematical Society* 5.1 (1973), pp. 81–84.
- [GSB13] A. Ghosh, B. N. Subudhi, and L. Bruzzone. “Integration of Gibbs Markov Random Field and Hopfield-Type Neural Networks for Unsupervised Change Detection in Remotely Sensed Multitemporal Images”. In: *IEEE Trans. Image Process.* 22.8 (2013), pp. 3087–3096.
- [Har03] H. L. Harney. *Bayesian Inference: Parameter Estimation and Decisions*. Advanced Texts in Physics. Springer, 2003.
- [Har94] F. Harary. *Graph Theory*. Addison-Wesley Series in Mathematics. Perseus Books, 1994.
- [HMC03] X. Huang, A.C. Madoc, and A.D. Cheetham. “Wavelet-based Bayesian estimator for Poisson noise removal from images”. In: *Int. Conf. Multim. Expo.* Vol. 1. July 2003, pp. 593–596.
- [Hub04] P. J. Huber. *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley, 2004.
- [Hus+13] M. Hussain et al. “Change detection from remotely sensed images: From pixel-based to object-based approaches”. In: *ISPRS J. Photogramm. Remote Sens.* 80 (2013), pp. 91–106.
- [IG04] J. Inglada and A. Giros. “On the possibility of automatic multisensor image registration”. In: *IEEE Trans. Geosci. Remote Sens.* 42.10 (Oct. 2004), pp. 2104–2120.
- [IJ01] H. Ishwaran and L. F. James. “Gibbs sampling methods for stick-breaking priors”. In: *J. Amer. Stat. Assoc.* 96.453 (2001).
- [IM09] J. Inglada and G. Mercier. “A New Statistical Similarity Measure for Change Detection in Multitemporal SAR Images and Its Extension to Multiscale Change Analysis”. In: *IEEE Trans. Geosci. Remote Sens.* 45.5 (May 2009), pp. 1432–1445.
- [Ing02] J. Inglada. “Similarity measures for multisensor remote sensing images”. In: *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*. Vol. 1. Toronto, Canada, 2002, 104–106 vol.1.
- [Jef46] H. Jeffreys. “An Invariant Form for the Prior Probability in Estimation Problems”. In: *Proc. R. Soc. A* 186.1007 (1946), pp. 453–461.

- [KC02] N. Kwak and C.-H. Choi. “Input feature selection by mutual information based on Parzen window”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 24.12 (Dec. 2002), pp. 1667–1671.
- [LFT12] C. Latry, S. Fourest, and C. Thiebaut. “Restoration technique for Pleiades-hr panchromatic images”. In: *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.* XXXIX-B1 (2012), pp. 555–560.
- [LJW08] L. Liu, Y. Jiang, and C. Wang. “Noise Analysis and Image Restoration for Optical Sparse Aperture Systems”. In: *Int. Workshop on Educ. Tech. and Training, and Geosci. and Remote Sensing*. Vol. 1. Shanghai, China, Dec. 2008, pp. 353–356.
- [Lon+12] N. Longbotham et al. “Multi-Modal Change Detection, Application to the Detection of Flooded Areas: Outcome of the 2009–2010 Data Fusion Contest”. In: *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* 5.1 (2012), pp. 331–342.
- [MAS11] G. Moser, E. Angiati, and S. B. Serpico. “Multiscale unsupervised change detection on optical images by Markov random fields and wavelets”. In: *IEEE Geosci. Remote Sensing Letters* 8.4 (July 2011), pp. 725–729.
- [MMS02] F. Melgani, G. Moser, and S. B. Serpico. “Unsupervised change detection methods for remote sensing images”. In: *Optical Engineering* 41.12 (Dec. 2002), pp. 3288–3297.
- [MMS08] G. Mercier, G. Moser, and S. B. Serpico. “Conditional Copulas for Change Detection in Heterogeneous Remote Sensing Images”. In: *IEEE Trans. Geosci. Remote Sens.* 46.5 (May 2008), pp. 1428–1441.
- [MS06] G. Moser and S. B. Serpico. “Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 44.10 (Oct. 2006), pp. 2972–2982.
- [MS09] G. Moser and S. B. Serpico. “Unsupervised change detection from multichannel SAR data by Markovian data fusion”. In: *IEEE Trans. Geosci. Remote Sens.* 47.7 (July 2009), pp. 2114–2128.
- [MSV07] G. Moser, S. B. Serpico, and G. Vernazza. “Unsupervised change detection from multichannel SAR images”. In: *IEEE Geosci. Remote Sensing Letters* 4.2 (Apr. 2007), pp. 278–282.

- [NC06] H. Nemmour and Y. Chibani. “Multiple support vector machines for land cover change detection: An application for mapping urban extensions”. In: *ISPRS J. Photogramm. Remote Sens.* 61.2 (2006), pp. 125–133.
- [OB08] P. Orbanz and J. M. Buhmann. “Nonparametric Bayesian Image Segmentation”. In: *Int. J. Comput. Vision* 77.1–3 (May 2008), pp. 25–45.
- [OQ04] C. Oliver and S. Quegan. *Understanding Synthetic Aperture Radar Images*. SciTech radar and defense series. SciTech Publ., 2004.
- [OT10] P. Orbanz and Y. W. Teh. “Bayesian nonparametric models”. In: *Encyclopedia of Machine Learning*. Springer, 2010, pp. 81–89.
- [OTB14] OTB Development Team. *The ORFEO Toolbox Software Guide*. Available at <http://orfeo-toolbox.org/>. 2014.
- [Owe01] A. B. Owen. *Empirical Likelihood*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, 2001.
- [Pac+07] F. Pacifici et al. “An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 45.9 (2007), pp. 2940–2952.
- [Paj06] G. Pajares. “A Hopfield Neural Network for Image Change Detection”. In: *IEEE Trans. Neural Netw.* 17.5 (Sept. 2006), pp. 1250–1264.
- [PBF54] W. W. Peterson, T.G. Birdsall, and W. Fox. “The theory of signal detectability”. In: *IRE Trans. Inf. Theory* 4.4 (Sept. 1954), pp. 171–212.
- [Pea83] J. A. Peacock. “Two-dimensional goodness-of-fit testing in astronomy”. In: *Monthly Notices of the Royal Astronomical Society* 202 (Feb. 1983), pp. 615–627.
- [PF10] F. Pacifici and F. Del Frate. “Automatic change detection in very high resolution images with pulse-coupled neural networks”. In: *IEEE Geosci. Remote Sensing Lett.* 7.1 (2010), pp. 58–62.
- [Pot52] R. B. Potts. “Some generalized order-disorder transformations”. In: *Math. Proc. Cambridge Ph. Soc.* 48 (01 Jan. 1952), pp. 106–109.
- [Pou+11] V. Poulain et al. “High Resolution Optical and SAR Image Fusion for Building Database Updating”. In: *IEEE Trans. Geosci. Remote Sens.* 49.8 (Aug. 2011), pp. 2900–2910.

- [Pra+13] C. Pratola et al. “Toward Fully Automatic Detection of Changes in Suburban Areas From VHR SAR Images by Combining Multiple Neural-Network Models”. In: *IEEE Trans. Geosci. Remote Sens.* 51.4 (2013), pp. 2055–2066.
- [PRC05] G. Pajares, J. J. Ruz, and J. M. de la Cruz. “Performance Analysis of Homomorphic Systems for Image Change Detection”. In: *Pattern Recognition and Image Analysis*. Vol. 3522. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005, pp. 563–570.
- [PV13] K. G. Pillai and R. R. Vatsavai. “Multi-sensor Remote Sensing Image Change Detection: An Evaluation of Similarity Measures”. In: *Proc. IEEE Int. Conf. Data Mining Workshops*. Dallas, USA, Dec. 2013, pp. 1053–1060.
- [QP07] Y. Qian and G. Ping. “Multispectral Remote Sensing Image Classification with Multiple Features”. In: *Proc. IEEE Int. Conf. Mach. Learn. Cybern.* Vol. 1. Hong Kong, China, Aug. 2007, pp. 360–365.
- [Qui+14] G. Quin et al. “MIMOSA: An Automatic Change Detection Method for SAR Time Series”. In: *IEEE Trans. Geosci. Remote Sens.* 52.9 (2014), pp. 5349–5363.
- [Rad+05] R. J. Radke et al. “Image change detection algorithms: a systematic survey”. In: *IEEE Trans. Image Process.* 14.3 (Mar. 2005), pp. 294–307.
- [RC05] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Texts in Statistics. Secaucus, NJ, USA: Springer, 2005.
- [RV93] E. J. M. Rignot and J. Van Zyl. “Change Detection Techniques for ERS-1 SAR Data”. In: *IEEE Trans. Geosci. Remote Sens.* 31.4 (1993), pp. 896–906.
- [Sch06] R. A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Elsevier Science, 2006.
- [Sen+13] S. Senturk et al. “Unsupervised classification of vineyard parcels using SPOT5 images by utilizing spectral and textural features”. In: *Proc. IEEE Int. Conf. Agro-Geoinf.* Fairfax, USA, Aug. 2013, pp. 61–65.
- [Sin89] A. Singh. “Digital change detection techniques using remotely-sensed data”. In: *Int. J. Remote Sensing* 10.6 (1989), pp. 989–1003.

- [SKZ14] P. Singh, Z. Kato, and J. Zerubia. “A Multilayer Markovian Model for Change Detection in Aerial Image Pairs with Large Time Differences”. In: *Proc. IEEE Int. Conf. Pattern Recognit.* Stockholm, Sweden, Aug. 2014, pp. 924–929.
- [SS14] M. Shadaydeh and T. Sziranyi. “An improved local similarity measure estimation for change detection in remote sensing images”. In: *Proc. IEEE Int. Conf. Aerospace Electronic Remote Sens. Technol.* Yogyakarta, Indonesia, Nov. 2014, pp. 234–238.
- [SSS12] C. D. Storie, J. Storie, and G. Salinas de Salmuni. “Urban boundary extraction using 2-component polarimetric SAR decomposition”. In: *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*. Munich, Germany, 2012, pp. 5741–5744.
- [Sta06] International Organization for Standardization. *ISO 12232:2006. Photography – Digital still cameras – Determination of exposure index, ISO speed ratings, standard output sensitivity, and recommended exposure index*. Geneva, Switzerland, 2006.
- [Sun+11] F. Sun et al. “Information extraction of fast-growing plantation based on remote sensing”. In: *Proc. IEEE Int. Conf. Geoinf.* Shanghai, China, June 2011, pp. 1–4.
- [Tis+04] C. Tison et al. “A new statistical model for Markovian classification of urban areas in high-resolution SAR images”. In: *IEEE Trans. Geosci. Remote Sens.* 42.10 (Oct. 2004), pp. 2046–2057.
- [TLB88] R. Touzi, A. Lopés, and P. Bousquet. “A statistical and geometrical edge detector for SAR images”. In: *IEEE Trans. Geosci. Remote Sens.* 26.6 (1988), pp. 764–773.
- [UY12] P. Uprety and F. Yamazaki. “Use of high-resolution SAR intensity images for damage detection from the 2010 Haiti earthquake”. In: *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*. Munich, Germany, 2012, pp. 6829–6832.
- [VFH93] J. D. Villasenor, D. R. Fatland, and L. D. Hinzman. “Change detection on Alaska’s north slope using repeat-pass ERS-1 SAR imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 31.1 (1993), pp. 227–235.

- [VP08] D. A. Van Dyk and T. Park. “Partially collapsed Gibbs samplers: Theory and methods”. In: *J. Amer. Stat. Soc.* 103.482 (2008), pp. 790–796.
- [Wei06] C. Weitkamp. *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*. Springer Series in Optical Sciences. Springer New York, 2006.
- [Woj98] L. Wojnar. *Image Analysis: Applications in Materials Engineering*. Taylor & Francis, 1998.
- [Wri+09] J. Wright et al. “Robust Face Recognition via Sparse Representation”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 31.2 (Feb. 2009), pp. 210–227.