UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA

# Topics in Image Processing and Applications to Structural Biology and Object Segmentation

A DISSERTATION SUBMITTED TO THE FACULTAD DE INGENIERÍA OF THE UNIVERSIDAD DE LA REPÚBLICA BY

## Federico Lecumberry Ruvertoni

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR EN INGENIERÍA ELÉCTRICA.

### ADVISOR/DIRECTOR DE TESIS
Dr. Guillermo Sapiro . . . . . . . . . . . . . . . . . . . University of minnesota, USA
Dr. Álvaro Pardo . . . . . . . . . Universidad Católitca del Uruguay, Uruguay
Dr. Alberto Bartesaghi . . . . . . . . . . . . . National Institutes of Health, USA

### COMMITTEE
Dr. Gadiel Seroussi . . . . . . . . . . . . . . . . . . . . . Universidad de la República
Dr. Pablo Musé . . . . . . . . . . . . . . . . . . . . . . . . . Universidad de la República
Dr. Andrés Almansa* . . . . . . . . . . . . . CNRS & Télécom ParisTech, France
Dr. Sriram Subramaniam* . . . . . . . . . National Institutes of Health, USA

* External reviewer

### DIRECTOR ACADÉMICO
Dr. Gregory Randall . . . . . . . . . . . . . . . . . . . . . Universidad de la República

Montevideo
March 6, 2012

**UNIVERSIDAD DE LA REPÚBLICA**

ACTA DE DEFENSA

TESIS DE DOCTORADO

**Fecha:** Martes 6 de Marzo de 2012.-

**Lugar:** Montevideo, Facultad de Ingeniería – Universidad de la República

**Plan de Estudio:** Doctorado en Ingeniería (Ingeniería Eléctrica).-

**Aspirante:** Federico Lecumberry Ruvertoni

**Documento de Identidad:** 2769975-7

**Director/es de Tesis:** Dr. Guillermo Sapiro

**Co-Directores de Tesis:** Dr. Álvaro Pardo

Dr. Alberto Bartesaghi

**Director Académico:** Dr. Gregory Randall

**Tribunal:** Dr. Andrés Almansa (Telecom Paris Tech e INCO, Fac. de Ingeniería, UdelaR)

Dr. Pablo Musé (IIE, Fac. de Ingeniería, UdelaR);
Dr. Gadiel Seroussi (IIE, INCO, Fac. De Ingeniería, UdelaR)

**Revisores:** Dr. Sriram Subramaniam (National Institutes of Health, USA)

Dr. Andrés Almansa (Telecom Paris Tech e INCO, Fac. de Ingeniería, UdelaR)

Los miembros del Tribunal hacen constar que en el día de la fecha el **Sr. Ing. Federico Lecumberry Ruvertoni** ha sido **APROBADO** en la defensa de su **Tesis de Doctorado** titulada: **"**Topics in Image Processing and Applications to Structural Biology and Object Segmentation**".**

Los reportes producidos por los dos revisores, que son muy favorables, se adjuntan.

La resolución del Tribunal se fundamenta en los puntos detallados a continuación:

Para un doctorado en ingeniería eléctrica es una fortaleza mostrar, como es el caso, que se pueden abordar dos problemas de naturaleza bien diferente, como ser resolver un gran desafío

llevando la tecnología a los límites fundamentales para obtener resultados concretos (observar ciertas estructuras moleculares en este caso) y por otro lado abordar un problema de un nivel de generalidad mucho mayor en donde las posibles aplicaciones son diversas pero no se aborda de forma específica una de ellas.

Los méritos principales de este trabajo radican en la amplitud de los temas abordados y de las herramientas utilizadas. Los resultados son excelentes. En el caso de la primera parte representa un aporte excepcional en la frontera del conocimiento en el campo de la cryomicroscopía electrónica. A juicio del Dr. Subramanian, como se señala en el reporte adjunto, los métodos desarrollados por Lecumberry tienen el potencial de revolucionar la manera en la que se trabaja en el análisis estructural de complejos macromoleculares.

La tesis está muy bien escrita, la presentación fue excelente, se aprecia la capacidad del candidato para comunicar ideas científicas y analizar profundamente los resultados. El candidato respondió con solvencia a las preguntas planteadas por los miembros del tribunal, demostrando gran madurez y conocimiento de los temas estudiados.

Para que conste,

**Dr. Andrés Almansa**

**Dr. Pablo Musé**

**Dr. Gadiel Seroussi**

ii

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892
Laboratory of Cell Biology
Center for Cancer Research
National Cancer Institute
Bldg. 50 Rm. 4306

Tel:    (301) 594-2062
Fax :  (301)480-3834
e-mail: ss1@nih.gov

February 29, 2012.
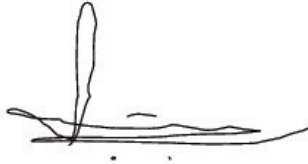
To: Members of the Thesis Approval Committee.

Dear Members of the Committee:

It is with great enthusiasm that I write in support of the thesis prepared by Federico Lecumberry as part of the requirements to obtain a Ph.D. degree at your University.

I have had the pleasure of working with Federico over the last several years through his association with my colleague Dr. Alberto Bartesaghi and collaborator Prof. Guillermo Sapiro. Although my laboratory is composed predominantly of more senior research personnel, a number of students have also recently completed their Ph. D. thesis work in the laboratory, so I have a good sense of the scholarship and output expected from a Ph.D. candidate working on interdisciplinary projects at the interface of physics, engineering and biology. Between 2007 and 2011, five graduate students who worked in the lab have obtained doctoral degrees and moved on successfully to positions in academia and industry; there are currently three graduate students in the lab pursuing their doctoral degrees under the auspices of the NIH-Oxford-Cambridge program.

Federico's thesis represents an outstanding piece of research that is at the cutting edge of current developments in 3D electron microscopy and image processing. It represents a synthesis of sophisticated computational methods combined with meticulous experimental work using the most advanced instrumentation currently available for cryo-electron microscopy. The written thesis demonstrates Federico's mastery of the complex nature of the project and well as the scientific nature of the challenge involved in reconstructing 3D structures using cryo-electron tomography and single particle cryo-electron microscopy. The methods presented in this work that Federico developed under the supervision of Dr. Alberto Bartesaghi have the potential to revolutionize the way we carry out structural analysis of complex macromolecular assemblies, and will be a key tool that that we and others in the field will be likely to use in the coming decade. I can say without hesitation that the quality of the thesis is easily comparable to those coming from the best graduate students at any top tier University in the United States. I give it my whole-hearted and unqualified recommendation.

Sincerely,

Sriram Subramaniam, Ph. D.
Senior Investigator and Chief, Biophysics Section
Laboratory of Cell Biology
Bldg. 50, Rm. 4306
National Cancer Institute, NIH
Bethesda, MD 20892.

# Informe sobre la tesis de doctorado presentada por Federico Lecumberry :
## *Topics in Image Processing and Applications to Structural Biology and Object Segmentation*

Informe Realizado por Andrés Almansa [***]

21 de febrero de 2012

La tesis de doctorado presentada por Federico Lecumberry aborda dos problemas de procesamiento de imagenes que estructuran la tesis en dos partes: (I) La reconstruccion tomografica microscopica en biologia estructural, y (II) la segmentacion de objetos en imagenes y secuencias de imagenes.

## Parte I: Structural Biology

Esta parte de la tesis aborda un problema de reconstrucción de imágenes biológicas en el cual quien suscribe posee escaso conocimiento. Esta parte del informe sera por lo tanto abordada desde el punto de vista de un investigador en procesamiento de imágenes que descubre las aplicaciones en biología estructural.

El trabajo presentado pretende llenar un vacío en las técnicas de reconstrucción tridimensional de estructuras moleculares a una escala intermedia entre los finos detalles alcanzables mediante rayos X y resonancia magnética, y la escala mas gruesa alcanzable mediante las técnicas actuales de microscopia electrónica.

La estrategia consiste en aumentar al máximo la capacidad de resolución de ésta última, mediante una técnica híbrida entre dos tecnologías existentes: (a) la *reconstrucción tomográfica* (ET) por un lado la cual posee la ventaja preservar las características de la estructura in vivo, pero se encuentra limitada en resolución a unos 20Å, y (b) *single particle* (SP) que permite alcanzar resoluciones de hasta 10 o 5Å.

Dicha técnica híbrida consiste en mantener la geometría de adquisición de ET, pero con métodos de registrado y de estimación de la función de transferencia similares a

---

[*]Investigador - CNRS LTCI - Telecom ParisTech, 46 rue Barrault, 75013 Paris, France
[**]Profesor Adjunto - InCo - Facultad de Ingenieria, UdelaR, Uruguay

los que se utilizan en SP. El desafío consiste en que las técnicas habituales de registrado utilizadas en SP no funcionan en estas condiciones debido los niveles de ruido significativamente más altos. Y esto constituye una restricción fundamental debido a que la dosis máxima admitida de radiación debe repartirse entre un número mayor de imágenes.

El dilema se resuelve mediante una la introducción de restricciones que permiten reducir drásticamente en el número de parámetros de registrado a resolver. Esto se logra mediante una manera astuciosa y conveniente de imponer localmente una condición de rigidez débil. De esta manera se mejora el condicionamiento del sistema y se admiten niveles de ruido mas elevados en cada una de las proyecciones tomográficas. Una segunda modificación respecto a los métodos conocidos, consiste en la utilización de solo una parte de las imágenes disponibles durante la fase de registrado, mientras que el conjunto completo de imágenes es utilizado durante la fase de reconstrucción. El método de selección de imágenes, a su vez utiliza criterios biológicos.

La pertinencia de ambas modificaciones es ilustrada mediante experimentos en datos simulados y en datos reales, la comparación de resultados frente a un ground-truth, el análisis cualitativo de resultados, así como diversas medidas objetivas de resolución que indican de manera unánime que las técnicas propuestas permiten aumentar en un orden de magnitud la resolución de ET.

Aunque los resultados son impresionantes y seguramente muy prometedores en el área de biología estructural, una buena parte del trabajo que permitió obtener estos resultados queda relativamente poco desarrollada. En efecto, la presentación algorítmica se concentra en una descripción a alto nivel de los métodos de ET, SP, y del método híbrido, y algunos detalles matemáticos y de implementación son desarrollados en los apéndices A-D. Pero la descripción del método de optimización en sí es bastante superficial. Si bien se sugiere que el problema global es altamente no convexo, no se explica en detalle la estrategia para encontrar un buen mínimo local cercano al optimo global. Tampoco se explican las propiedades que hacen que la función objetivo elegida sea preferible frente a otras.

Un desarrollo mas amplio de estos temas podría revelar vínculos (con aportes en ambos sentidos) entre los temas estudiados y problemas similares en tratamiento de imágenes, optimización robusta y registrado semi-rígido como son la superresolución, la identificación biométrica, la sismología, entre otras.

## Parte II: Object Segmentation

La segunda parte de la tesis se concentra en el problema de la segmentación de objetos en una imagen o secuencia de imágenes.

Siguiendo una gran efervescencia de trabajos en la segunda mitad de la década del 2000-2010 (que incluyen trabajos de Cremers, Faugeras, Vese, entre otros), Lecumberry adopta un modelo estadístico de alto orden para la forma a segmentar, lo cual resulta

indispensable para que el algoritmo de segmentación sea robusto frente a deformaciones y oclusiones.

Como generalización de los trabajos anteriores, Lecumberry extiende la técnica al caso de formas dinámicas suficientemente complejas como para no poder ser modeladas estadísticamente por un simple análisis en componentes principales (PCA). Lecumberry propone entonces la utilización de técnicas de reducción de dimensionalidad (diffusion maps), seguidas de un particionamiento del espacio reducido de formas en partes lineales, cada una de las cuales puede ser modelada individualmente por una gaussiana.

Esto conduce entonces a la formulación del problema en términos de M modelos estadísticos de forma, entre los cuales hay que elegir el que mejor representa la imagen a segmentar. El problema se transforma entonces en un problema acoplado de segmentación y clasificación de formas. Problema que Lecumberry resuelve mediante un método variacional que reúne elementos de level-sets, modelado e inferencia estadística, optimización, invariancia a transformaciones rígidas.

La solución adoptada es sumamente pragmática dada la complejidad del problema, y los resultados presentados son impresionantes, tanto por su variedad, como por la robustez de los métodos presentados frente a diversos tipos de degradación de la imagen.

No obstante queda claro que el trabajo sugiere un rico panorama de extensiones diversas tanto desde el punto de vista aplicativo, como desde el punto de vista teórico:

- Cual es el espacio de formas subyacente a esta modelización numérica, y qué propiedades posee que puedan ser explotadas para mejorar la solución?

- Qué garantías de convergencia a un optimo local o global ofrece el método numérico utilizado para la minimización de la energía? Algunos subproblemas son convexos, otros no. Como pueden utilizarse mejor estas propiedades para dar garantías de convergencia o para acelerar el algoritmo? Esta pista fue parcialmente explorada (posteriormente o en paralelo al trabajo de Lecumberry) por Cremers & Schmidt (2008).

- Como puede extenderse el método para su aplicación en casos de la vida real donde: (i) existen múltiples objetos con formas diversas y oclusiones mutuas, (ii) los shape-priors son desconocidos y solo pueden aprenderse a partir de la propia secuencia a segmentar?

- Como extender el modelo de formas dinámicas, a la segmentación de secuencias de imágenes, teniendo en cuenta la regularidad de la dimension temporal? Por el momento el çonjunto de formas"que componen un modelo o un conjunto de modelos es explotado sin ningún tipo de orden.

Las preguntas anteriores muestran que el tema abordado con éxito por Lecumberry es sumamente vasto y amerita un estudio mas profundo, con varias pistas de investigación que son sugeridas por el propio trabajo de Lecumberry.
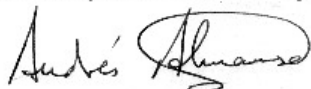
<div align="center">3</div>

## Valoración global

Para resumir, la presente tesis (que solo representa una parte del trabajo por Federico Lecumberry durante su doctorado) muestra su inquietud ingenieril y científica por diversas aplicaciones del tratamiento de imagen, asi como su capacidad y gusto por el dialogo multidisciplinario y para servirse de desarrollos recientes en áreas tan diversas como la geometria y estadistica de altas dimensiones, la optimizacion, el analisis de Fourier. También muestra su capacidad para hacer interactuar estas diversas areas de manera original para proponer soluciones innovadoras.

El manuscrito muestra igualmente un trabajo riguroso de validación de resultados, con criterios pertinentes a diversas areas de aplicacion como el analisis de video y la biologia estructural.

Por ultimo queda claro que las cuestiones abordadas por Lecumberry en este trabajo llegan a un grado de profundidad que lo situan en una buena posición para abrirse nuevas vias de investigacion en diversas areas de aplicacion.

Por todas estas razones, quien suscribe considera que la tesis de Federico Lecumberry constituye un excelente trabajo que merece ser defendido frente al jurado.

Andrés Almansa

4

viii

"Its five-year mission: to explore strange new worlds, to seek out new life and new civilizations, to boldly go where no man has gone before."

James T. Kirk

"Nuestras horas son minutos cuando esperamos saber y siglos cuando sabemos lo que se puede aprender."

Proverbios y Cantares
Antonio Machado

"Adiskide onekin, orduak labur."

Basque proverb.

# I would like to thank...

*A Julia y nuestra hija Maite.*

# Abstract

This thesis is divided in two parts, addressing quite different topics and fields such as Cryo-Electron Microscopy and Object Segmentation.

**Cryo-Electron Microscopy**   The first part addresses the problem of 3D reconstruction of macromolecular assemblies from their 2D random projection images obtained following a tomographic acquisition in a Transmission Electron Microscope. The challenge of this problem comes from the extremely noisy and low contrasted projection images obtained with this procedure. The problem of *reconstruction from random projections* or *multiple view geometry* is well known in Computer Vision, however the usual techniques based in identifying correspondent point or lines in several views are discarded, due to the noisy characteristics of the images (see for example figure 2.1b in page 7). Thus, usually thousands of *aligned* projections are averaged in order to increase the SNR and compute the reconstruction.

The main reason for the low SNR is the low electron doses allowed to irradiate the macromolecules of interest, otherwise severe radiation damage accumulates, deforming the macromolecules and affecting the projections. Embedding the specimens in vitreous ice and preserving it at cryogenic temperatures (below -150° Celsius) helps reduce the accumulated radiation damage. In this scenario, Cryo-Electron Microscopy has proved to be a powerful technique to obtain 3D reconstructions with a wide range of resolutions. Two of the most popular techniques in Cryo-Electron Microscopy are Single Particle Analysis and Cryo-Electron Tomography, and are described in this work. Their main differences for the goal of this work are that the former is capable to routinely obtain reconstruction $\sim$8Å or better of *in vitro* specimens, and the later allows to analyze *in vivo* specimens but hardly recovers reconstructions better than 20Å.

This work proposes a new framework combining the Single Particle Analysis and Cryo-Electron Tomography approaches, that combines the best of both worlds, adding the high resolution feature in the maps obtained with the Cryo-Electron Tomography data collection procedure without sacrificing its desired features. This is achieved through a new refinement algorithm and an innovative adaptation of the Single Particle Analysis reconstruction procedure to this

kind of data.

The framework is validated by a set of experiments, first, synthesizing the data collection procedure, allowing to have access to the complete groundtruth (parameters and macromolecule structure) for the comparison. The second set of experiments use data acquired in the Transmission Electron Microscope and only the macromolecule structure is known. The reconstructed density map for an homogeneous macromolecule shows details of about 10Å of resolutions.

This work was performed in collaboration with Alberto Bartesaghi and Sriram Subramaniam from the Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, USA.

This manuscript was prepared before the publication of this work, further information can be found in my web page: `http://iie.fing.edu.uy/~fefo/`.

**Object segmentation.** The second part addresses one of the most interesting and fundamental tasks in Computer Vision and Image Processing such as object segmentation. In this work we use the idea of Shape Models (SMs). SMs, capturing the common features of a set of training shapes, represent a new incoming object based on its projection onto the corresponding model. Given a set of learned SMs representing different objects classes, and an image with a new shape, this work introduces a joint classification-segmentation framework with a twofold goal. First, to automatically select the SM that best represents the object, and second, to accurately segment the image taking into account both the image information and the features and variations learned from the on-line selected model. A new energy functional is introduced that simultaneously accomplishes both goals. Position and transformation invariance is included as part of the modeling as well.

The model selection is performed based on a shape similarity measure, on-line determining which model to use at each iteration of the steepest descent minimization, allowing for model switching and adaptation to the data. High-order SMs are used in order to deal with very similar object classes and natural variability within them.

The presentation of the framework is complemented with examples for the difficult task of simultaneously classifying and segmenting closely related shapes, such as stages of human activities, in images with severe occlusions.

This work was presented in the IEEE International Conference on Image Processing, ICIP 2009 [1], and appears in the IEEE Transactions on Image Processing [2]. The contents of the second part is almost the same that appears in the later reference.

[1]  Federico Lecumberry, Álvaro Pardo, and Guillermo Sapiro.  Multiple shape models for simultaneous object classification and segmentation. In IEEE International Conference on Image Processing, ICIP 2009. Cairo, Egypt, pages 3001–3004. IEEE, 2009.

[2]  Federico Lecumberry, Álvaro Pardo, and Guillermo Sapiro.  Simultaneous object classification and segmentation with high-order multiple shape models. *IEEE Transactions on Image Processing*, 19(4), apr 2010.

# Preface

I started working on my PhD in 2007 with Álvaro Pardo and Gregory Randall. We studied different techniques for the analysis of high dimensional data, applying them to different datasets. In July 2007 Guillermo Sapiro gave a short course on this subject in Montevideo, and in October I visited his Research Group in Minnesota where we continued this line of work with the goal of applying these techniques to building models of the variation of shapes.

On February 1st, 2008, the coldest day of that winter in Minnesota, my wife Julia and I arrived to Minneapolis directly from sunny Montevideo for my two years internship.

During 2008 the work on object segmentation was performed. It can be found in the second part of this manuscript, where all the previous work on high dimensionality analysis was condensed into one section, which was published in Transactions on Image Processing and was presented in the 2009 International Conference on Image Processing in Cairo, Egypt. I also took some courses in optimization and statistical learning. During the first month of 2009, with Guillermo and Ignacio Ramírez, we worked on a learning methods to enhance properties of the dictionaries for sparse modeling. It was published in the Third International Workshop on Computational Advances in Multi-Sensor Adaptive Processing as an invited paper. This work is not included in this manuscript.

In June 2009, Guillermo offered me some options on how to continue my work, one of which was working (*again*) with Alberto Bartesaghi, continuing Guillermo's collaboration with Sriram Subramaniam's research group at the Laboratory of Cell Biology, NCI/CCR, National Institutes of Health, USA. It took me one femtosecond to make my decision, but only the following day did I answer Guillermo. Ever since I've made several visits to Sriram's Lab where I met him and his team of collaborators. It was the first time I was in contact with this kind of high level research (my previous experience being academic and industry research), and the human quality and the high stakes of the subject, made this a very enriching experience at many levels. Getting into a completely new, exciting and challenging world such as Structural Biology and in particular Cryo-Electron Microscopy, and also propose some contributions to the area, along with Sriram, Guillermo and Alberto, took me about two

years where every day was worth it, and the reward is much more than this manuscript.

FLR

# Contents

Contents

# Part I

# Constrained Single Particle: Merging Cryo-Electron Tomography and Single Particle

# Chapter 1

# Introduction

Structure determination of macromolecular complexes is key for understanding the mechanisms and functioning of cell's "molecular machinery" which can in turn lead to the development of rational vaccine design strategies to combat diseases, for example. Various different techniques exist that can reach these goals, for example, X-ray Crystallography, Nuclear Magnetic Resonance (NMR) and Electron or Optical Microscopy, among others. Which technique should be used to obtain high resolution images for the biologic structure of interest depends strongly in its size or molecular weight. In other words, each technique is best suited for a certain range of atomic masses, usually measured in Daltons[1] (Da). X-ray and NMR are suited to the smaller masses up to $10^5$ Da, structures like aminoacids or secondary structures (as $\alpha$-helix or $\beta$-sheets), and for human scaled structures higher than $10^{26}$ Da. Microscopy techniques are limited by the wavelength of the beamed particle used, the lower the wavelength the higher the resolution. Light in optical microscopy has a wavelength of 600 nanometers and the range of masses where it is applicable starts at $10^{12}$ Da. Electron microscopy, uses electrons instead of light, with a wavelength of 6 picometers and its range of masses expands from $10^8$ to $10^{22}$ Da. The mentioned techniques cover almost all the range of masses of biological imaging except for a relative small gap between $10^6$ to $10^{10}$ Da. This range is extremely important in the field of cell biology because it is the mass range of cellular entities, sub-cellular organelles and organisms such as viruses and bacterias, whose sizes and complex structures are too big to be analyzed with the preceding techniques. Thus, achieving a molecular level resolution insight view into this area is key to understand and interpret their functions and interactions at cellular level. At this point, Cryo-Electron Microscopy (Cryo-EM) stands as an effective framework to fill this gap [Sub05].

However, the data acquisition technique is not the only factor that might

---

[1]One Dalton is approximately equal to the proton or neutron's mass, it is defined as one twelfth of the rest mass of an unbound neutral atom of carbon-12 in its nuclear and electronic ground state, $1\text{Da} = 1.660538921 \times 10^{-27}\text{kg}$

help to achieve this exciting goal and new algorithms and procedures must be developed accordingly. In this sense, a wide range of problems such as denoising, reconstruction, visualization, etc., need to be addressed implying developments covering several areas of applied mathematics and engineering such as Signal and Image Processing, Quantitative Analysis, Pattern Recognition, Multivariate Statistics, among others.

In Cryo-EM, the determination of high-resolution three dimensional (3D) structure of biological macromolecules requires the application of robust and effective algorithms for averaging the information obtained from the acquired two dimensional (2D) random projections. The resolution of the reconstructed maps is mainly affected by the acquisition procedure and the protocol of distribution of the dose of electrons while imaging the target. This acquisition procedure also determines the characteristics of the different methods, imposing a tradeoff between the achieved resolutions and the advantages of each method. The present work proposes an innovative framework combining two of the most successful Cryo-EM techniques (Single Particle and Cryo-Electron Tomography) that allow to increase the resolution of the 3D reconstructions without sacrificing their highly rated features. Also a new protocol of dose fractionating is needed to better exploit the advantages of the proposed framework. The framework is named *Constrained Single Particle* (CSP).

The remainder of this part is organized as follows. Chapter 2 briefly describes the Cryo-Electron Microscopy characteristics including further descriptions of the Single Particle and Cryo-Electron Tomography techniques. Chapter 3 describes the proposed framework. Chapter 4 presents validation experiments and their discussion. Chapter 5 concludes this part. Appendix A describes the Euler angles convention used in this work, Appendix B introduces a metric for measuring the aligment error in the reconstructions proposed in this work. Appendix C describes the image formation process in Electron Microscopy and the CTF correction. Finally, appendix D introduces the usual methods for measuring resolution.

# Chapter 2

# Cryo-Electron Microscopy

Cryo-Electron Microscopy is the technique where a frozen-hydrated specimen is exposed in a Transmission Electron Microscope (TEM) to obtain an image of its projection. The sample in the specimen grid is embedded in a very thin layer of vitreous ice preserving its unembedded (native) characteristic features, allowing to obtain projection images of fully hydrated macromolecules; generally thousand of projections from different views are collected from identical copies of the macromolecule or particle of interest. Using these set of projection images a three dimensional (3D) reconstruction of the specimen can be obtained, once provided the 3D orientation of every projection. The whole procedure can be thought as the reconstruction from multiple random projections, a well known problem in Image Processing and Computer Vision [Fau93, HZ04]. However, as it will be shown, in EM the projections are extremely noisy and low contrasted images, and the usual procedures (such as searching for correspondences) fail.

The mathematical foundations of the projection and reconstruction procedures are based on Radon Transform theory and on the *Fourier Central Slice Theorem*, that states that the 2D Fourier Transform of the projection of a 3D object is equal to a central 2D section (slice) of the 3D Fourier Transform of the object. Following this theorem, all the 2D transforms of the projection images can be aligned to a common reference in the 3D Fourier reciprocal space, and then inverted to obtain the 3D structure of the specimen in the real space.

The two most widely used Cryo-EM techniques are Single Particle Analysis (SP) and Cryo-Electron Tomography (Cryo-ET). Both techniques share the same basic principles, but otherwise present major differences and applications. The most notorious is that SP is applied to study *in vitro* homogeneouses specimens, where only one projection from each particle in the sample is taken with relatively high dose, achieving reconstructions up to a resolution between 15 to 4Å [Gri07]. On the other hand, Cryo-ET is applied to study *in vivo* specimens, and several projections of each particle are taken tilting the

sample grid forcing to fractionate the allowable dose between the tilts[1], and therefore lower Signal to Noise Ratio (SNR). The resolution achieved by the standard ET reconstruction are between 100 to 40Å, and can be increased up to 20Å with further processing as will be described later (see averaging of sub-volumes [LBB$^+$08, BSL$^+$08] in page 16). The rest of this section describes the common Cryo-EM's characteristics and the next sections concentrate on their particularities.

The image formation process in a TEM is based on the interaction between the electrons and the object. The planar wave in the electron beam illuminates the object, introducing a local phase shift by an amount given by the integral of the potential distribution in the wave propagation direction. Therefore, the densities and thickness of the material in the traversed sample affects the phase shift of the wave and then, the contrast in the image. Appendix C briefly presents the image formation process in a TEM. Finally, the image is recorded by detecting the electrons that pass through the sample to a system of electromagnetic lenses which focus and enlarge the image with hundred of thousand times magnifications. This image is usually called a *micrograph*.

The interaction between the electrons and the object originates elastic or inelastic scattering of the electrons, involving or not energy transference to the object, leading to variations in the obtained resolution. This energy transfer accumulates, breaking apart molecular bonds, damaging and deforming the object (an effect known as radiation damage) producing lower resolution images with undesired background and noise effects. The dose of electrons also affects the amount of damage, in particular the damage to the secondary structure occurs with doses greater than $1e^-/\text{Å}^2$ [Fra06a]. These effects are temperature-dependent and lowering the temperatures to cryogenic levels (below -150° Celcius) helps to reduce the radiation damage.

Thus, in order to preserve the specimen from radiation damage in Electron Microscopy, it is embedded in a plastic, glucose or vitreous ice, similar to the fossilization of animals in amber. Embedding the specimen in vitreous ice in Cryo-EM has some advantages, since the scattering densities of water and proteins are sufficiently different to produce contrasted images even at low resolution, compared with the embedding in glucose. However, the Cryo-EM's sample preparation procedure to obtain the sample grid has many manual steps that makes it an "art". One major key point is the cooling step, where the specimen experiments a cooling gradient close to $10^5$ degrees per second, which avoids the formation of ice crystals that modify the finer details in the structure. This step is performed using a freeze-plunger that holds the specimen grid and plunges it in the cryogen (liquid ethane), which is then transferred to liquid nitrogen, and finally mounted in the specimen holder.

---

[1]The usual doses in SP are about 30 $e^-/\text{Å}^2$, while the doses in Cryo-ET are about 1 or 2 $e^-/\text{Å}^2$

(a)                       (b)
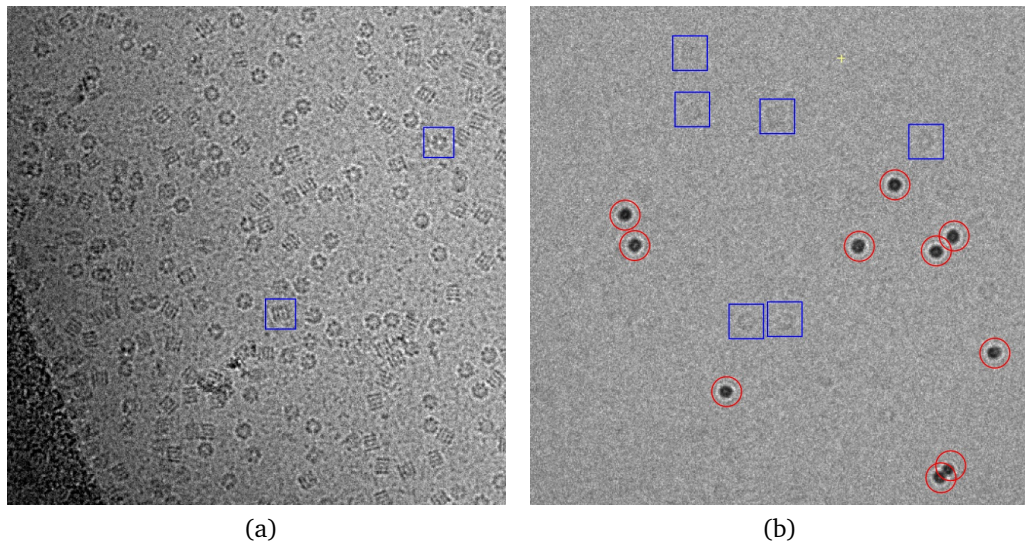
Figure 2.1: Section of an image (a micrograph) obtained in (a) Single Particle and (b) Cryo-Electron Tomography. Projections of several copies of the same complex in random orientations and positions are marked by blue squares. In (b) also gold particles about 10nm wide are marked by the red circles, acting as fiducial markers to help in the alignment of the tilt series. Note the difference in contrast and SNR between both images due, mainly, by the doses used in each technique (see text).

This ultra fast cooling procedure is possible due to the high heat capacity given by the small size of the specimen. Another advantage of the embedding at cryotemperatures is the fact that the biological structure of the processed specimen is similar to its native (unembedded) structure, i.e., the specimen does not collapse and the contrast in the images is related to the biological structure of the specimen, rather than to an external substance as in the negative stain procedures, where an "opaque" substance (usually a heavy metal salt) is added to the preparation and surrounds the molecules in the specimen increasing their contrast. Moreover, the specimen in the grid never comes in contact with other surface or substances, therefore, the observed shape is the true shape of the hydrated specimen in the solution and has no distortion due to an attaching process. However, the procedure does not guarantee keeping the specimen unaltered, since a shrinkage is produced in the normal direction to the sample plane, leading to an extra reduction in resolution in this direction.

Despite Cryo-EM's artisan specimen preparation procedure for data collection, compared to X-ray's, it is much simpler and faster, avoiding the necessity of creating good crystals in the process. This also impacts in the time scale of the target specimen: since the whole cooling process takes seconds or less, short-lived functional aspects of the particles or macromolecules can be analyzed. This advantage is compensated by the high number of parameters

needed to completely determine the geometry to compute the reconstruction.

As a direct consequence of the low electron doses, the Signal to Noise Ratio (SNR) of the micrograph is very low, which added to the already low contrast produced by the specimens usually makes it practically impossible to see any specimen by a non experienced user, *see* figure 2.1. Therefore, the procedure to obtain a high resolution reconstruction is to average a large number of particles, since the SNR for the average of $N$ *aligned* samples contaminated by uncorrelated noise, increases with $N$. Usually $N$ is over ten thousand or one hundred thousand, depending on the technique. Clearly, the averaging should be done with an aligned version of the particle, otherwise, a rough, low resolution reconstruction of the particle is obtained. However given the low images' SNR their alignment is a challenging task.

As for any real imaging system, the TEM is characterized by a Point Spread Function (PSF) or its Fourier Transform, the Contrast Transfer Function (CTF), as is preferred in EM, governing how contrast in the image is formed by mapping the amplitudes of the input. The Contrast Transfer Theory states that [Fra06b], under the weak-phase approximation, the relationship between the object and the bright field image can be described by the linear transfer theory, this is, the Fourier transform of the observed image is the product of the CTF and the Fourier transform of the object being imaged. The CTF depends on many factors, among them, defocus, spherical aberration of the lenses and the size of the source, Appendix C briefly describes the image formation process in the TEM and the mathematical definition of the CTF. The estimation of the CTF for a given image and its correction is a crucial step in order to reach high resolution reconstructions (helical[2] or atomic), since the information at frequencies higher than the first CTF's zero are fundamental, and their frequency dependent amplitude and phase produce an intolerable contrast distortion. CTF correction is a well studied subject and a fundamental procedure in the SP framework. However, its use in Cryo-ET is less well established mainly for the following reasons: first, the estimation of the defocus is more complex in Cryo-ET than in SP given its acquisition procedure and the variation of the defocus over the image due to the tilted geometry, second, the lower amount of signal present in the micrographs due to the lower doses is not enough to accurately estimate the CTF, and third, there are more critical issues (like the low SNR) limiting the resolution obtained by Cryo-ET, therefore, full correction of the CTF does not guarantee its improvement.

Along with the development of these techniques various software packages have been developed since the beginnings of the 1980, most of them open source or freeware software. Among them MRC [CHS96], IMAGIC [vHHO$^+$96], SPIDER [FRP$^+$96], EMAN [LBC99], FREALIGN [Gri07] have been defining file

---

[2]Resolution where secondary structures, such as alpha helices and beta sheets, can be distinguishable.
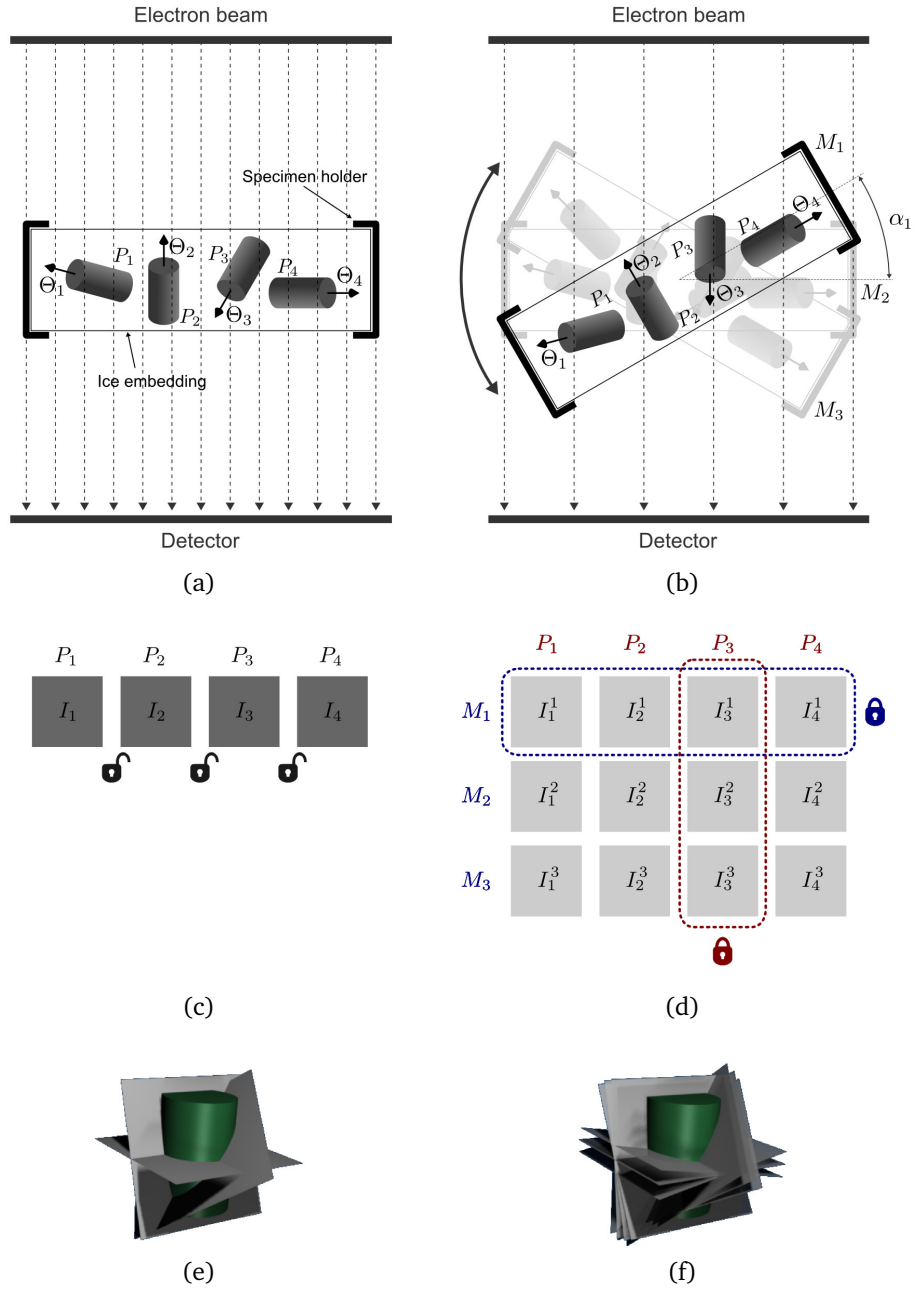
Figure 2.2: Comparison between Single Particle (left) and Tomography (right) data acquisition geometric configuration. (a) Single Particle acquisition geometry: $N_P{=}4$ particles $P_k$ are randomly and independently distributed with orientation $\Theta_k$. (b) Tomography acquisition geometry: Same as in Single Particle but the specimen holder is tilted an angle $\alpha_i$, obtaining $N_M{=}3$ micrographs $M_i$, fractionating the total dose between them. (c) In Single Particle, only one particle projection image $I_k$ by particle with relative high dose is taken. (d) In Tomography, $N_M$ $N_P$ particle projection images $I_k^i$ are taken with lower SNR. Constraints between the projections of a particle and between projections in the same micrograph arise from the tilting geometry (see text). (e) and (f) Relative alignment of the particle projection images with respect to the reconstruction.

formats, results visualization, standardizing and automatizing many aspects of
the SP and ET reconstruction procedure.

## 2.1   Single Particle Analysis

Single Particle Analysis is the Cryo-EM framework applied to study an *in vitro*
specimen of an individual non-crystallized macromolecular assembly.   This
characteristic gives the technique's name, a single kind of particle is present in
the specimen and is the object of interest. The in vitro characteristic is assured
by a purification procedure, where the particle of interest is isolated and only
copies of it are present in the specimen grid without any contaminants.

Usually, only one micrograph is obtained from each specimen imaged area,
containing a large number of identical copies of the particle of interest which
are randomly distributed.  SP assumes that the single particles have even dis-
tribution of orientations on the specimen grid, allowing to compute a 3D aver-
aged reconstruction structure of the studied particle. The homogeneity of the
target particle and its projections is an essential requirement to guarantee the
success of SP reconstruction, otherwise their heterogeneity will be averaged
degrading the reconstruction.  Once the micrograph is obtained, the particle
projections are, usually, manually selected and cropped from the micrograph
(a process called boxing), obtaining a stack of raw images of particle projec-
tions. Therefore, the challenging problem for obtaining the reconstruction of
the particle is to accurately compute the alignment of their projections despite
the low SNR and low contrast present in the images.

Figure 2.2a shows a schematic representation of the SP geometrical acqui-
sition configuration.  Four particles $P_k$ ($k = 1, \ldots, 4$) are distributed with ran-
dom orientations $\Theta_k$ in the ice embedding supported by the specimen holder.
After acquisition and boxing, four particle projection images $I_k$ are obtained
(Figure 2.2c), with no other connection between them than being projections
of different instances of the same entity.  Figure 2.2e shows how the parti-
cle projections are assembled in the 3D structure of the studied particle given
their determined alignment.

Particles are randomly distributed within the ice, with no constraints in
their orientations and positions, leaving six degrees of freedom per particle
projection, three *rotational parameters* Euler angles $\Theta$ and three *translational
parameters*, in the 3D coordinate reference frame of the ice volume. The pro-
jection maps the translational parameters of the particle from this 3D coor-
dinate reference system to the 2D coordinate reference system of the micro-
graph.  Thus, in order to obtain the alignment for each particle projection,
five parameters should be computed $(\vartheta, s)$, where the rotational Euler angles
$\vartheta = (\phi, \theta, \psi)$ are mapped from $\Theta$ by the projection and $s = (s_x, s_y)$ are the
"in-plane" translational parameters in the 2D coordinate reference system of

each boxed image. The first two angles from $\vartheta$ define the 3D "out-of-plane" orientation of each particle projection $(\phi, \theta)$, while the third one $\psi$ encodes its "in-plane" rotation; Appendix A explains the Euler Angles convention used in this work. Therefore, if $N$ particle projection images are collected, a grand total of $5N$ parameters should be computed in order to obtain the reconstruction.

The determination of the alignment parameters for each particle projection is performed following a typical implementation of an iterative Expectation-Maximization algorithm, alternating two steps of *refinement* and *reconstruction*. In the first step the alignment parameters $(\vartheta, s)$ are *refined* for each particle projection. In the second step a 3D structure is *reconstructed* using the projection images and their alignment parameters just computed. This 3D structure is used as a reference map for the refinement step in the next iteration, increasing its resolution through the refinement-reconstruction iterations.

The EM techniques can be grouped in two different classes based in how the initial reference map is obtained: the *reference-free*, where it is computed directly from the data without external information, and second, the *reference-based*, where a model building method is used in order to compute it. SP belongs to the second class. Among the reference-free techniques, two of them are the most used in SP as model builders [JSB08]: the *common lines* methods [vHGM+00, Fra06b, Fra09] and *Random Conical Tilt* (RCT) [Rad88]. The common-line techniques exploit the common line shared by the Fourier Transforms of 2D projections of a 3D object, measuring the angle between those lines, their relative orientations can be determined and used to compute the 3D reconstruction. In RCT each particle is irradiated twice, first a tilted image is obtained with the specimen holder tilted, about 45°, and then an untilted image, the zero tilt micrograph; these two images are a *tilt pair* related by the tilted geometry. RCT assumes that the particle has a preferred orientation in the embedding ice, viewed for the top in the untilted micrograph, for example. Then, aligning the untilted images, the alignment of the tilted images can be determined according to their relative orientation given by the untilted images and the geometry of each tilt pair (see Figure 2.3a).

The refinement of the alignment parameters has been addressed by many different algorithms, the most common approach used in EM is [vHGM+00, vH87] the "projection matching", which considers several re-projections images of the reference model in different orientations, and compares them to the particle projection image. Two kind of methodologies to sample the possible orientations to be tested are used [JSB08]. The first method is based in an uniform sampling of the five $(\vartheta, s)$ dimensional manifold of parameters. Usually, the sample period is reduced along with the iterations, in order to improve the alignment accuracy. The second group of refinement methods

compute the next set of parameters to be tested based in the local behavior of the objective function, a kind of gradient descent method, where the gradient could be explicitly computed or not. From the Image Processing point of view this is the minimization of some measure of dissimilarity or "distance" between the image and the re-projection of the reference model. Many possible options could be considered, but the noisy and low contrasted characteristics of the acquired images turns most of them unhelpful. The most common dissimilarity function found in the EM literature computes dissimilarity properties based in the spectral characteristics of the images. The (Linear) Cross-Correlation Coefficient (CC) is one of the most used tools for the comparison of two images; other popular cost function is the Phase Residual (PR). A more robust version to noise bias of the Cross-Correlation, proposed by Stewart and Grigorieff in [SG04], is based in weighting down the signals components with low SNR, leading to the following expression

$$R = \sum_i \left| CC_{f,i}^3 \right| \quad \text{with} \quad CC_{f,i} = \frac{CC_i^2}{CC_i + \omega_f}. \tag{2.1}$$

where $CC_i$ is the average CC in a region, usually non-overlapped spherical shells covering the Fourier space, $CC_{f,i}$ is a filtered version of $CC_i$ and $\omega_f$ is a constant to reduce the dependence of the alignments on small correlation coefficients. Any of these dissimilarity functions, or new proposals, can be used in the SP framework but they must provide the discrimination and robustness needed for the particular characteristics of the EM's images.

In the refinement step the particle projection images can be treated separately, refining their parameters independently from the others, or they can be classified in clusters given their view and refine the parameters for each class. The first method refines successively the parameters of one image at a time, independently from the parameters of the other images in the stack. On one hand, this approach allows a massive parallelization in the refinements of the particles given their decoupling; however, their SNR might not be enough to accurately refine their parameters. The second method looks for increasing the SNR of the views by averaging several images facing the same orientation. In order to do so, clustering the set of images ensures that images from the same orientation are classified together, then, an average is obtained for each class and their relative orientations are computed. Another class of refinement algorithms is based in the maximization of a log-likelihood function [Sig98, SMVC09, SC09]. These algorithms allow to model in a natural way the noisy characteristics of the input data and the hidden unknown parameters in the problem, calculating probability-weighted assignments for all possible orientations. Thus, instead of assign the parameters for each projection individually or for a class of projections, the refined structure contains each image at all possible orientations, weighted by the probability of the

structure given any particular set of alignment parameters. Usually these algorithms show more robustness to the reference-bias and have superior convergence to the correct structure than a correlation-based algorithm [Sig98], on the other hand the computation of the refined structure is computationally more expensive.

Once the set of alignment parameters is determined the reconstruction is ready to be computed. It is a common practice in SP to select a subset of projection images to be used in the reconstruction step. This subset includes only the "best" projections, based on the same dissimilarity function used in the refinement step to determine the quality of the projection. This allows to discard damaged or improperly aligned projections adding heterogeneity in the 3D averaged structure. As an extreme case Borgnia et al. [BSZM04] show that using only the best 139 (or even 9!) projections the reconstruction is superior to the reconstructed with a higher number of projection, supporting the quality vs. quantity approach. The techniques to compute the 3D reconstruction given the selected particle projection images and their alignment parameters can be grouped in three categories: (a) weighted back-projection, (b) interpolation in Fourier reciprocal domain, and (c) iterative algebraic methods. The weighted back-projection algorithms are based in the *back-projection* operation, considered as the inverse of the projection, going from the 2D projection to a 3D volume where all the voxels in the projection direction have the same value equal to the value of the correspondent pixel in the 2D projection. When several projections in different arbitrary directions are considered, they are linearly added obtaining the 3D reconstruction of the particle. However, this raw reconstruction is like a blurred version of the original object, since the density of sampling points decreases with increasing spatial frequency, resulting in an overemphasis of the low frequencies in comparison with the higher spatial frequencies. Therefore, to solve this problem a weighting function (a filter) is applied to the 2D projections before reconstruction, specially designed for arbitrary projection geometries. The Fourier reconstruction method is based on the Fourier Central Slice Theorem and the main drawback of these methods is the arbitrary position of the projections that do not coincide with the regular 3D Cartesian sample grid in the Fourier domain. This leads to complicated interpolation schemes to estimate the set of coefficients in the sampling grid that best fits the given measurements, and then obtain the 3D structure applying the inverse Fourier transformation. The iterative algebraic methods are based in the formation of the discrete projection image, considering the object densities and the projection direction. Thus, every pixel in the projection image is modeled as a weighted sum of the density values of the object; the weights are computed taking into account the contribution of each voxel given the projection angle and the interpolation rule. The weighted back-projection and Fourier interpolation methods are computationally more efficient than the it-
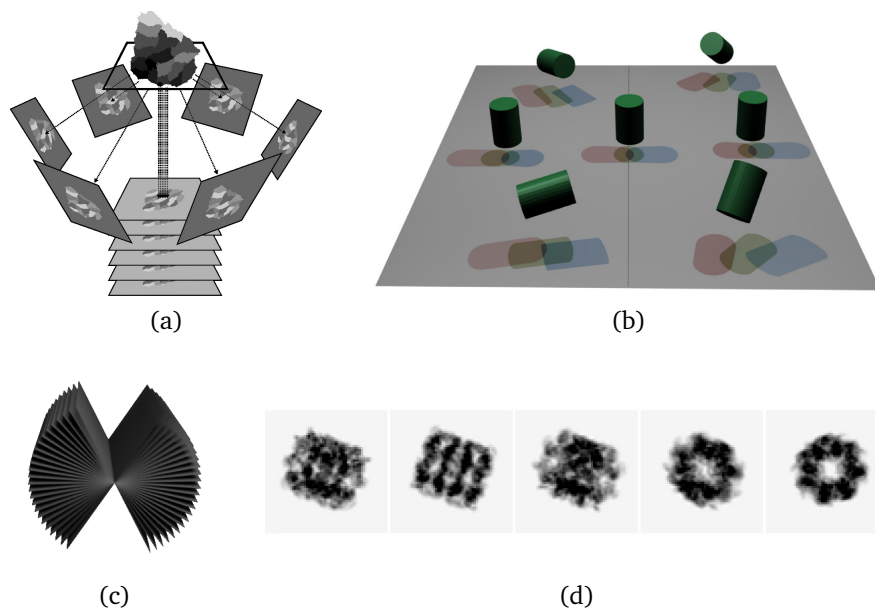
Figure 2.3: (a) Random Conical Tilt acquisition geometry: all the untilted projections from one view are aligned together, determining the orientation of the tilted projections used for the reconstruction. (b) Simulation of the projections obtained in a tomographic framework. To simplify the image, the electron beam has been tilted with a fixed plane, projections to different micrographs are identified by different colors. (c) Missing wedge in Fourier reciprocal space for the tomographic acquisition, each plane corresponds to the Fourier transform of one micrograph. (d) Projections in five different tilt angles of a simulated GroEL particle (see figure 4.7).

erative techniques, however the later ones can obtain superior reconstructions according to the sensitivity to angular gaps and smoothness; finally the quality of Fourier interpolation methods critically depends on the interpolation method used [Fra06b]. Many reviews cover the mathematical details of this topic, in particular [Fra06b] (and references therein) analyze its application to Electron Microscopy.

The multivariate statistical techniques and signal processing tools which are present in the refinement and reconstruction procedures in SP cover a wide spectrum of techniques, many surveys and books [vHGM$^+$00, Fra09, Fra06a, Fra06b] are dedicated to go through the variations and successful EM applications.

## 2.2   Cryo-Electron Tomography

Cryo-Electron Tomography (Cryo-ET) is the Cryo-EM framework applied to the tomographic data collection procedure, see figures 2.2b and 2.3b. The main advantage of Cryo-ET over SP is that the former allows to study speci-

men maintaining their *in vivo* structural integrity characteristics. This allows to apply Cryo-ET to analyze the native state of specific sections of macromolecular assemblies, like cells, sub-cell's organelles, isolated macromolecular complexes or spikes on top of viruses' membranes [JSB08, LBB$^+$08, Fra06a]. This avoids the necessity of purification or synthesis of the sample to isolate the target entity, because the projections into the micrograph will include contributions from the whole macromolecule, invalidating the hypothesis of SP. Another advantage of Cryo-ET is that the tilted geometry allows to have many views of the specimen avoiding ambiguities in the determination of the handedness, which can be used to build the initial reference, transforming Cryo-ET in a reference-free method and no reference bias is present in the reconstruction.

ET's data collection geometry is shown in figures 2.2b and 2.3b. The electron beam direction is fixed as in SP, but the specimen holder (specimen grid) is tilted a *tilt angle α* around a single *tilt axis*. Every time the specimen grid is tilted, one micrograph $M_i$ is captured, originating a set of micrographs sharing a common line (the tilt axis), this set of images is know as a *tilt series T*. As the tilt angle increases, the thickness of the ice cross section increases as the inverse of $\cos \alpha$, doubling it at $\alpha = 60°$. Therefore, commonly, the set of tilt angles covers the range ±60° with steps about 2°, leaving a gap in viewing angles producing a *missing wedge* of information in Fourier domain requiring specialized processing for its compensation (see figure 2.3c). The radiation damage and its consequence in the amount of allowed radiation, dramatically affects the acquisition of the images, forcing to fractionate the total amount of allowable dose between the micrographs, leaving each micrograph with an even lower SNR than in SP. These lower SNR in Cryo-ET's images compared to SP's images, imply that the number of images to be collected and averaged should be much higher than in SP, thus, usually the number of micrographs to be collected in Cryo-ET is higher than in SP. Another caveat that should be addressed is that, besides the radiation damage producing slightly variations in the particle between micrographs, also different sections of the specimen preparation might move relative to the support grid, even if the fiducials markers (see below) attached are well aligned, therefore, some misalignment between the particle's projections may occur, reducing the high-frequency information in the reconstructed volumes leading to a further reduction in the resolution of the final reconstruction.

The reconstruction of the 3D tomographic volume given the set of tilted micrographs follows the same ideas of Radon Transform and the Fourier Central Slice Theorem as in SP. In order to compute the reconstruction, the complete alignment parameters of the set of micrographs should be obtained, determining their common-line (the tilt axis) given by a point in the 2D plane and an orientation angle (the tilt axis angle $\beta$) and the tilt angle $\alpha$, totalizing four pa-

rameters per micrograph. Since the SNR of the micrographs is extremely low, signal features are barely present in an individual micrograph, thus, to introduce some external features, high-density fiducial markers (colloidal gold particles) are added in the specimen preparation step to determine the refinement of micrograph's parameters. The reconstruction process implies rotations and translations of the micrographs to align them and build the 3D volume; all these computations involve interpolation of the data reducing their high frequency components, being another source of reduction in the achievable resolution limits. Figure 2.1b shows a section of a micrograph from a tilt series with the usual SNR in a tomographic acquisition, gold particles (about 10 nanometers wide) are marked by the red circles while the blue squares ($35 \times 35$ pixels approximately) mark a few particles to be analyzed; note the noisy and low contrasted acquired image.

Cryo-ET has been successfully used to study 3D structures of heterogeneous one-of-a-kind specimens, such as whole cells where the SP and other techniques fail [Fra06a]. However, some complex structures such as cells and viruses contain, within their heterogeneity, multiple copies of specialized macromolecules, such as sub-cell's organelles or viruses' spikes, that can be individualized in the 3D reconstructed tomogram and their sub-volumes (also called sub-tomograms) cropped. Following the same idea as in SP with 2D images, a 3D averaging of these sub-volumes is the key to obtain higher-resolution maps of this macromolecules that cannot be purified to undergo the in vitro SP analysis. But, as pointed out before, the success of the averaging relies in the classification and alignment of, in this case, the sub-volumes. The alignment of 3D volumes has been extensively studied in Image Processing and Computer Vision, where it is referred to as image or volume registration [HZ04], however, the missing wedge is a too strong feature biasing the alignment of the sub-tomograms. Therefore, it should be explicitly taken into account in the alignment procedure, otherwise the resolutions will be compromised. Stoffler et al. [SFF+03] overcome this problem by "coarse visual determination" of the axial direction in each sub-volume, this is used to obtain an initial average which is refined several times adding different features (as weighting functions, symmetry, etc.) every time. Bartesaghi et al. [BSL+08] propose a framework for the alignment, classification, and averaging of tomographic data, explicitly modeling the missing wedge as an occluding mask in the Fourier reciprocal domain not affecting the alignment and classification process, achieving a resolution of 20Å.

The sub-volume alignment requires the determination of six parameters per particle, the same six degrees of freedom per particle as in the SP, three Euler angles defining the orientation $\Theta = (\phi, \theta, \psi)$ and three translational parameters $\mathcal{X} = (X, Y, Z)$. To formalize the notation, if the tilt series has $N_P$ particles of interest $P_k$ ($k = 1, \ldots, N_P$) and $N_M$ micrographs $M_i$ ($i = 1, \ldots, N_M$) are

taken, and all the particles have projections in all the micrographs, the whole reconstruction process implies the computation of a grand total of $4N_{\text{M}} + 6N_{\text{P}}$ parameters. These parameters are $(\alpha_i, \beta_i, x_i, y_i)$ for each micrograph $M_i$ and $\{\Theta_k, \mathbf{X}_k\} = (\phi_k, \theta_k, \psi_k, X_k, Y_k, Z_k)$ for each particle $P_k$.

Despite the listed drawbacks, ET and Cryo-ET have led to a significant progress in applications ranging from material science to bio-materials and life science (structural, cellular and tissue biology). Advances in vitreous sectioning of a variety of specimens (such as bacteria, yeast and skin cells), description of the structure and conformational changes in the bacterial chemotaxis, structural analysis of the architectures of a wide range of intact viruses without staining, among others are successful examples of its application [Sub05, BS09]. A special emphasis should be done in the structural analysis of symmetric and non-symmetric viruses with immediate medical relevance in designing strategies to combat viral diseases [SBL$^+$07, LBC$^+$08, BSL$^+$08, LBB$^+$08]. Combined with image averaging methods (specially in the case of highly symmetric viruses) have shown to be a powerful method providing complementary structural information to be used with atomic structures of individual components determined by X-ray.

# Chapter 3

# Merging Tomography and Single Particle

The main advantage of Cryo-ET is the potential to individualize sub-volumes with homogeneous particles in heterogeneous macromolecular assemblies, allowing to align and average them in order to increase their resolution. This unique characteristic comes along with the advantage of being a reference-free method, the capacity of analyze *in vivo* specimens and the disambiguation in the handedness determination being given by the tilted geometry. Furthermore, the geometrical data collection configuration (see figure 2.2b) imposes a set of constraints that are strongly used in the tomographic reconstruction. On the other hand, SP cannot deal with this heterogeneous conformations and no constraints are imposed in its optimization procedure. However the reconstruction is performed with the CTF corrected original raw images, where the tolerated doses (therefore the SNR) are higher, allowing to routinely achieve resolutions around 10 to 12 Å, and in some cases achieve 4 Å or better [LBC+08]. Meanwhile, the usual resolutions for Cryo-ET with sub-volume averaging are about 20 Å [BS09, and references therein].

## 3.1   The goal of this work

The analysis of the advantages of Cryo-ET over SP shows that its benefits are due to the tilted geometry of the acquisition procedure. On the other hand, besides the higher doses that are used, the benefits of SP in high-resolution maps are due to the fact that the reconstruction is computed with the original particle projection images, along with the standardized CTF estimation and correction.

   This work proposes a new framework combining these two approaches, trying to get the best of both worlds, adding the high resolution feature in the maps obtained with the Cryo-ET data collection procedure, through a new

refinement algorithm and an innovative adaptation of the SP reconstruction procedure to this kind of data. Their main characteristics rely on (a) the acquisition of a tilt series of micrographs (b) a refinement procedure of the parameters of each projection image imposing the geometrical tomographic constraints, and (c) a reconstruction procedure based on the original CTF corrected projections instead of sub-tomograms. The requirements of fractionation of the allowable dose leads to a new protocol of distribution of dose giving flexibility in the combination of what data is used in the refinement and/or the reconstruction.

Section 3.2 introduces the reconstruction step, while Section 3.3 introduces the refinement in the alignment procedure.

## 3.2 Reconstruction of tomographic data with a Single Particle framework

The tomographic reconstruction recovers the full 3D position and orientation parameters for each particle in the embedding ice, allowing to crop the sub-volumes and compute their averages. This procedure, aligning the tilt axes for all the micrographs and then building the 3D volume, generates a mapping function between the coordinates of each particle projection and the particle reconstruction, that will be explained in the next section. See [WTN$^+$97] for a similar derivation of the coordinates relation.

Considering only the horizontal (0-tilt) micrograph, as in the case of SP, the mapping function relates the coordinates of the SP framework and the ET 3D reconstruction. Based in this mapping function it is possible to revert the path, relating the coordinates from the ET to the SP space, allowing to assign the 3D orientation and 2D position in the micrograph for each particle projection. In other words, the five parameters for the SP reconstruction for the projections in the 0-tilt micrograph can be computed from the ET's set of six parameters. In that way, an *à la* SP reconstruction based in the original raw images with their parameters assigned mapping the ET reconstruction can be computed. Moreover, not only the projections in the 0-tilt micrograph can be used. Also the particle projections in the tilted micrographs can be related by incorporating the transformation given by the tilt axis and the tilt angle to the inverse mapping function, as well. Using all the micrographs allows to use all the particle projections in the tilt series, considerably increasing the number of images, and thus the number of particle's views, used in the reconstruction.

Among the advantages of this new method is that the reconstruction is done with the original raw data in the images as in SP. This is a main advantage over ET, since all the signal frequency components are present in the data used in the reconstruction. Second, although the CTF estimation and

correction can be performed in SP as well as in ET, the former is much more stable and accurate, obtaining better compensations and avoiding the zero inversion problem (see Appendix C). In the ET case, assuming and horizontal tilt axis, the CTF (therefore the defocus), varies with the distance to the tilt axis; Fernández et al. analyze [FLC06] the defocus variation in the tilted micrograph, proposing a technique for estimation and correction of the CTF of tilted micrographs.

The acquisition methodology in ET implies a protocol in the fractionating and distribution of the dose among the micrographs, taking into account the radiation damage and its effects in the deformation of the particle. Following the direction of SP the highest possible dose should be concentrated in the first (or firsts) micrographs, obtaining images with higher SNR with less damage. On the other hand, as the tilt angle increases, the amount of dose should be increased in order to obtain similar contrast in the images since the thickness of the ice increases. One possible solution to overcome this compromise in the distribution of the dose, is to use only the firsts, less damaged, micrographs in the reconstruction, avoiding to incorporate particle projections that have accumulated more radiation damage. Thus, a higher fraction of the dose can be irradiated in these selected micrographs, in comparison to the dose used in classical ET. The rest of the available dose is distributed among the higher tilted micrographs following the classical ET protocol. With this approach the number of particle projections used in the reconstruction is reduced, becoming a set of less images but with higher SNR and with less accumulation of radiation damage. The next section shows how to use other projections than this set, in order to improve the accuracy in the determination of the alignment parameters. Note that, albeit the higher tilted micrograph will have less SNR and accumulated damage, they are still useful in the initial ET reconstruction providing more views of the particle and reducing the missing wedge, finally helping to obtain an initial accurate estimation of the orientation and translation parameters.

The main tradeoff of this approach lies in the much lower SNR present in each particle projection image, due to the dose fractionation, as was just discussed. Besides this, there is a major change in the acquisition methodology with respect to SP that could become a possible caveat in the reconstruction procedure. This is, usually the target particle in their native state is not isolated, for example, an organelle in a cell or a spike on top of a virus' membrane, therefore, when it is tilted and projected into the micrograph, other parts of the whole macromolecule occlude the particle and its projection image gets contaminated with contributions from other physiologically relevant background. Beniac et al. [BDA+07] applied the SP framework directly to a whole virus and processed the cropped image of their spikes over the membrane obtaining a 18Å 3D reconstructed map. The studied virus has a spherical

shape with the spikes randomly distributed over the membrane in its normal direction, therefore the projection of the spikes are not isolated, as is assumed by SP, but contaminated with the projection of the whole virus. Despite this, even the top view of the spikes can be recovered without appreciable degradation.

Beyond this SP-like reconstruction, also the refinement of the alignment parameters for each particle projection can be performed following a SP approach. Despite being the origin of all the differences between SP and Cryo-ET, the tilted geometry of the acquisition procedure provides important features that are incorporated in a new refinement process that helps to overcome the lower SNR in the images. This is considered and described in the next section, which boldly goes where no Cryo-EM refinement algorithm has gone before.

## 3.3   Constrained Single Particle

The data collection process introduces "structure" in the stack of particle projection images that is not present in the standard SP, allowing to regroup the images in sub-stacks given by the particle or the micrograph where they belong. For each one of these sub-stacks a geometric constraint relates their particle projection images, transforming the procedure in a Constrained Single Particle (CSP) refinement. This idea is schematically shown in figure 2.2d, where, assuming all the particles have projections in all the micrographs, the stack of $N = N_M N_P$ images of particle projections is ordered in $N_P$ columns and $N_M$ rows. All the particle projections images $I_k^i$ belong to this matrix, where the $k$-th column is the sub-stack corresponding to the $N_M$ particle projections of the $k$-th particle $\mathscr{S}(P_k) = \{I_k^i : i = 1,\ldots,N_M\}$ and the $i$-th row is the sub-stack corresponding to the $N_P$ particle projections into the $i$-th micrograph $\mathscr{S}(M_i) = \{I_k^i : k = 1,\ldots,N_P\}$.

For each sub-stack the geometric constraints over the particles projections can be summarized as:

**Constraint on the micrographs**  All the particle projections images into the $i$-th micrograph $(I_k^i \in \mathscr{S}(M_i))$ belong to the two dimensional plane given by the micrograph, thus they should all share the same orientation.

**Constraint on the particles**  All the particle projection images for the $k$-th particle $(I_k^i \in \mathscr{S}(P_k))$ are projected into the set of micrographs, thus they should be tilted following their corresponding tilt angles. In other words, the 3D angle between two particle projection images is the same angle between the tilt angles of the corresponding micrographs.

The constraints imply that all the projection images of one particle are restricted to move "together" if their orientation changes with the refinement
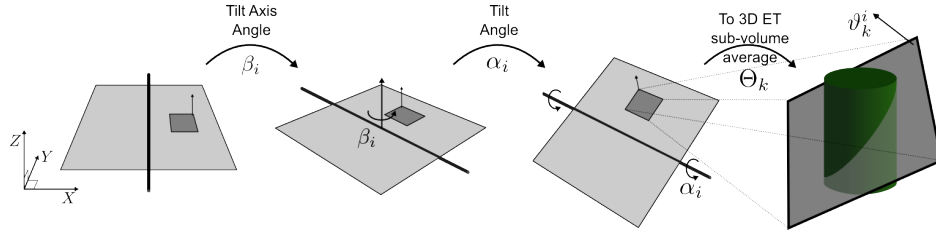
Figure 3.1: The concatenation of the shown rotations relates the orientation of the particle projection in the registered micrograph with the coordinates in the particle reconstruction, allowing to compute a mapping function between the $\vartheta_k^i$ and $\Theta_k$, $\alpha_i$ and $\beta_i$.

process. Similarly, the projection images in each micrograph should move together facing all of them the same direction preserving coplanarity. Moreover, the orientation $\vartheta_i^k$ for the particle projection image $I_k^i$ of a particle $P_k$ into a micrograph $M_i$ is completely determined by the orientation of the particle, the tilt angle and tilt axis angle of the micrograph, see [WTN$^+$97], i.e.

$$\vartheta_i^k = f(\Theta_k, \alpha_i, \beta_i), \tag{3.1}$$

where $f(\cdot)$ is a mapping function relating the orientations of one particle and micrograph with the orientation of the corresponding particle projection. The orientation of the particle projection $\vartheta_i^k$ needed to compute the 3D reconstruction, corresponds to the concatenation of the rotations going from the registered micrographs to the 3D reconstruction. These rotations are shown in figure 3.1 and lead to $R = R(\Theta_k)R_Y(\alpha_i)R_Z(\beta_i)$. Representing the orientation $\vartheta_i^k$ as a rotation matrix given the Euler angles convention (ZYZ in this work, see Appendix A) yields the same matrix $R$, therefore $R_{ZYZ}(\vartheta_i^k) = R(\Theta_k)R_Y(\alpha_i)R_Z(\beta_i)$, allowing to determine $\vartheta_i^k$ as a function of $\Theta_k$, $\alpha_i$ and $\beta_i$ as is given by equation (3.1).

The refinement of an unconstrained SP (USP) procedure following the projection matching criteria, looks for minimize some dissimilarity function $v(\cdot)$ measuring the discordance between the projection image $I$ and the projection of the reference map in a certain direction given by $\vartheta_k^i$. Therefore, when all the projection images are treated independently, the effective cost function to be minimized is the (weighted) average of the dissimilarity function, this is,

$$J(\{I_k^i\}, \{\vartheta_k^i\}, \{s_k^i\}) = \frac{1}{W} \sum_{\substack{i=1,\dots,N_M \\ k=1,\dots,N_P}} w_{I_k^i} v\left(I_k^i, \vartheta_k^i, s_k^i\right) \tag{3.2}$$

where a weighting coefficient $w_I$ is added to generalize, and $W = \sum_I w_I$. Therefore, when the set of images acquired in the tomographic framework, represented by the matrix of images in Figure 2.2d, is processed by USP the sum in the previous equation is computed term by term in any order, for example, following the index in the stack of images. The same result is obtained,

if the sum is computed first row by row (or column by column) and then these intermediate results are added. These harmless change in the way of computing Equation (3.2) implies a major change in the objects being optimized, since, again following the representation given in Figure 2.2d, summing row by row (or column by column) corresponds to accumulate the cost function for the corresponding micrograph (or particle). Hence, the definition of the dissimilarity function can be extended for micrographs and particles as the sum of the cost function of their corresponding particle projections, i.e., the sum in the corresponding row or column

$$\bar{v}_M\left(\alpha_i, \beta_i\right) = \sum_{I_k^i \in \mathscr{S}(M_i)} w_{I_k^i} v\left(I_k^i, \vartheta_k^i, s_k^i\right) = \sum_{k=1,\dots,N_P} w_{I_k^i} v\left(I_k^i, f(\Theta_k, \alpha_i, \beta_i), s_k^i\right)$$

(3.3)

and

$$\bar{v}_P\left(\Theta_k\right) = \sum_{I_k^i \in \mathscr{S}(P_k)} w_{I_k^i} v\left(I_k^i, \vartheta_k^i, s_k^i\right) = \sum_{i=1,\dots,N_M} w_{I_k^i} v\left(I_k^i, f(\Theta_k, \alpha_i, \beta_i), s_k^i\right), \quad (3.4)$$

were Equation (3.1) was used in the last equality on both equations. Using this definitions Equation (3.2) can be written as

$$J(\{I_k^i\}) \;=\; J(\{P_k\}, \{M_i\}) \;=\; \frac{1}{W} \sum_{i=1}^{N_M} \bar{v}_M\left(\alpha_i, \beta_i\right) \qquad (3.5)$$

$$= \; \frac{1}{W} \sum_{k=1}^{N_P} \bar{v}_P\left(\Theta_k\right) \qquad (3.6)$$

Rewriting Equation (3.2) as Equations (3.5) and (3.6) modifies the rotational parameters being optimized. Instead of determining $\vartheta_k^i$ for all the particle projection images, only the $\Theta_k$ for each particle and the $(\alpha_i, \beta_i)$ for each micrograph must be determined. This implies a change in the objects being "rotated" in the optimization, moving from the particle projection images in Equation (3.2) to the micrograph in Equation (3.5) and particle in Equation (3.6). This will be strongly used in the optimization procedure decoupling the refinement of particles and micrograph. Note that not only the number of objects is reduced, a dramatic reduction in the number of parameters to determine is obtained, since, instead of computing the orientation parameters $\vartheta_k^i$ for all the particle projections, a total of $3N_PN_M$ parameters, only the orientation of the particles $\Theta_k$ and the orientation of the micrographs $(\alpha_i, \beta_i)$ should be computed, a total $3N_P + 2N_M$ parameters. For the usual number of particles $N_P$ and micrographs $N_M$ used in Cryo-ET, this gives a tremendous reduction in the number of parameters to compute since $3N_PN_M \gg 3N_P + 2N_M$.[1] Finally, the two

---

[1] Considering $N_M = 60$ micrographs and $N_P = 3,000$ particles, the total number of rotational parameters to be determined is $3N_MN_P = 540,000$ for USP and $2N_M + 3N_P = 9,120$ for CSP. Both methods needs to compute the $2N_MN_P = 360,000$ translational parameters.

dimensional translational parameters $s_k^i$ are determined when the dissimilarity function $v(\cdot)$ is computed in Equations (3.3) and (3.4).

In this way, the constraints through the mapping function (Equation (3.1)) perform a dimensionality reduction of the search space, meaning that not all the dimensions in the original SP space are free to move once multiple tilts are included, and the dependency between projections is bonded by the orientations of the particles and the micrographs. Furthermore, imposing these geometric constraints might approximate the dimensionality of the space to the intrinsic dimensionality of the data. From the point of view of the optimization procedure, reducing the dimensionality of the search space is an advantage and alleviates the problems related to work in a high dimensional space known as the curse of dimensionality.

The weights $w_I$ could reflect different aspects of the reconstruction problem, in particular the radiation damage of the specimen when exposed to the electron beam or the variation in the amount of material due to the tilt angle, moreover, they can be used to explicitly leave out projections, for example thresholding by the values of cost function $v(\cdot)$. This weighting procedure is also applied in the reconstruction step, where usually only a percentage of the best images are used.

Being a constrained SP refinement, CSP uses a reference map which is refined in each iteration, needing as in SP an initial reference, however, as a main difference, and advantage compared to SP, CSP can use the tilt series to build an initial low-resolution reference map in the same way as in ET. Furthermore, a set of alignment parameters for both, the micrographs and the particles is obtained from this reconstruction and are ready to be used as an initial guess for the optimization. Moreover, this initial set of parameters reduces the refinement to a local search in the low dimensionality space for the rotational parameters while the translational parameters are refined as in SP.

Finally, considering the refinement of the micrographs and particles parameters separately, the number of particle projection images in a sub-stack for a particle ($\mathscr{S}(P_k)$) and in a sub-stack for a micrograph ($\mathscr{S}(M_i)$) is very different, since the number of particles $N_P$ in a tilt series is higher than the number of projections per particle. This might imply a different behavior in the refinement of particles and micrographs due to the number of terms and the relative variation of the cost function for a certain particle projection alignments.

### 3.3.1 Optimization procedure

Given a tilt series with $N_P$ particles projected into $N_M$ micrographs, the goal for CSP is to determine the alignment for these $N_P + N_M$ objects, thus, mapping the orientation parameters using equation (3.1) permits to determine
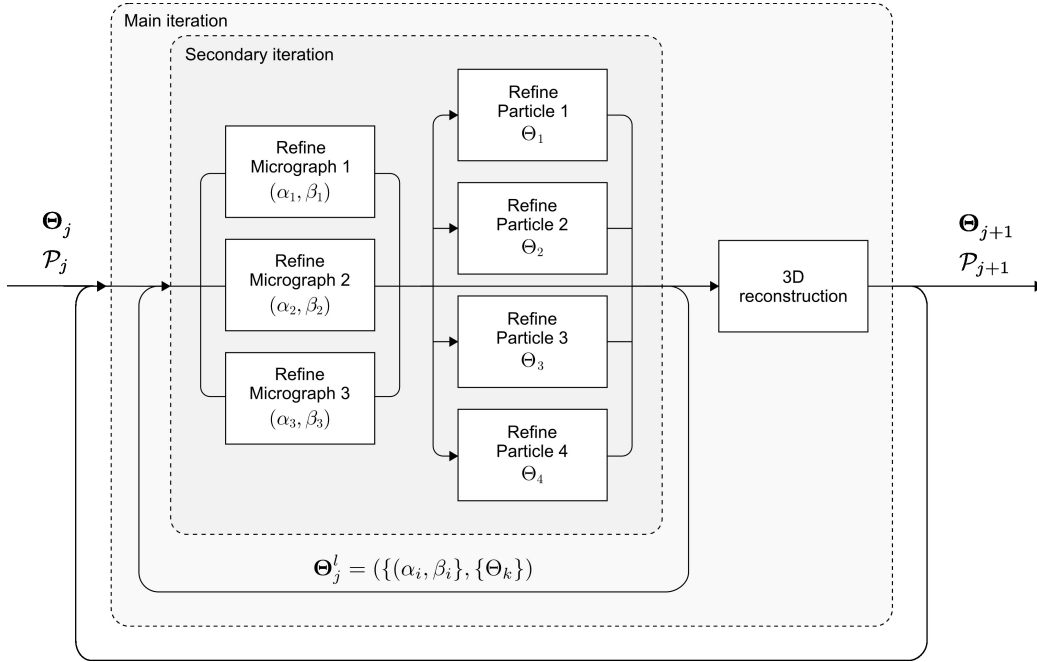
Figure 3.2: CSP main and secondary iterations (see text).

the orientations for the $N_{\mathrm{P}}N_{\mathrm{M}}$ particle projections images in order to compute the reconstruction; also the translational parameters for the projection images should be determined. Therefore, the parameters to be determined by CSP are

$$\Theta = \left( \{\Theta_k\}, \{\alpha_i, \beta_i\}, \{s_k^i\} \right), \tag{3.7}$$

where $\Theta_k \in \mathbb{R}^3$ are the Euler angles of the $k$-th particle, $\alpha_i \in \mathbb{R}$, $\beta_i \in \mathbb{R}$ are the tilt angle and the tilt axis angle of the $i$-th micrograph respectively, and $s_k^i \in \mathbb{R}^2$ is the translation of each particle projection image; here $i = 1, \ldots, N_{\mathrm{M}}$ and $k = 1, \ldots, N_{\mathrm{P}}$. For simplicity, the indexes in the notation of the particle projection images are omitted from now on, i.e., $I$ means $I_k^i$.

The cost function expressed as Equations (3.5) and (3.6) decomposed Equation (3.2) based in the imposition of the constraints. This kind of decomposition is suitable for a block coordinate descent method [Ber99], where the blocks are given by the coordinates of each optimized object (particles and micrographs); the procedure optimizes each block of parameters separately, holding all the others fixed. Few cycles of this procedure are performed, as a secondary iteration level, until the parameters stabilize. Then, the optimized parameters are used to update the reconstructed map, used as a reference for the next main iteration. Figure 3.2 shows graphically the complete optimization procedure.

In the optimization of every block of coordinates, several orientations for each object are tested, corresponding to different points in the space of lower

dimensionality; each point corresponds to a particular configuration of orientations for every particle and micrograph, and then, for every particle projection, as a set of related points in the higher dimensionality space. Hence, in order to evaluate the cost function for every object $\bar{v}_{M,P}(\cdot)$, using equations (3.3) and (3.4), the cost function $v(\cdot)$ is computed for the corresponding particle projection images fixing their orientations through the mapping function.

To further formalize the basic step, let $\Theta_{j,l}$ be the set of parameters at the $j$-th main iteration and the $l$-th secondary iteration; at main iteration $j$, the secondary iteration starts with $\Theta_{j,0} = \Theta_j$, and at the end, after $r$ iterations $\Theta_{j+1} = \Theta_{j,r}$. For example, when the coordinate corresponding to the tilt angle $\alpha_i$ for $M_i$ is refined, the constraint on the micrograph is applied on its particle projections images $I \in \mathscr{S}(M_i)$ and the following optimization problem is solved

$$\hat{\alpha}_i = \arg\min_{\alpha_i} \bar{v}_{M_i}(\alpha_i, \beta_i) = \arg\min_{\alpha_i} \sum_{I \in \mathscr{S}(M_i)} w_I v\left(I, f(\Theta_k, \alpha_i, \beta_i), s_k^i\right). \quad (3.8)$$

For every candidate point $\tilde{\alpha}_i$ in the previous equation the $s_k^i$ parameters are determined in order to correspond with the proposed $\tilde{\alpha}_i$. The same procedure is done in the refinement of $\beta_i$ for each micrograph $M_i$ with $I \in \mathscr{S}(M_i)$ and for each Euler angle in $\Theta_p$ for every particle $P_k$ with $I \in \mathscr{S}(P_k)$, $\forall i$. This procedure should be cycled a few times since, for example, $\alpha_i$ in $\Theta_{j,1}$ was computed to be optimal with the $\Theta_k$ parameters at $\Theta_{j,0}$ and should not be optimal with their new values in $\Theta_{j,1}$. This is the reason for the secondary loop, which runs until no significant improvement is made $\Theta_{j,r} \approx \Theta_{j,r-1}$.

Once the secondary iteration refined the parameters for all the micrographs and particles, with the orientations $\vartheta_k^i$ computed using Equation (3.1) the new density map $\mathscr{P}_{j+1}$ is computed.

# Chapter 4

# Validation experiments

In order to validate the framework a set of experiments with synthesized and collected data were performed. The goal of using synthesized data is to have full control of the groundtruth alignment parameters as well as the atomic structure of the particle, therefore several measures, introduced in the next section, can be obtained to evaluate the framework performance. In the case of real collected data, the only groundtruth is the atomic structure of the particle.

Several copies of an homogeneous macromolecule complex are synthesized or acquired following the tilted geometry and dose fractionating protocol of the Cryo-ET framework. The obtained data can be processed by three different approaches. First, following the sub-volume averaging procedure described in Section 2.2. Second, by the proposed framework CSP, using the density map obtained by the previous method as an initial reference and imposing the constraints between the particle projection images. And, third, it can be processed following the approach of SP, without imposing the constraints, refining each particle projection independently. The later procedure was called Unconstrained SP (USP) in this work, in opposition to the Constrained SP (CSP) proposed.

The cost function used in this experiment is the Phase Residual (PR) defined in the FREALIGN [Gri07] software package,

$$v\left(I_k^i, \vartheta_k^i, s_k^i\right) = \mathrm{PR}(I, \vartheta_k^i, s_k^i),$$

corresponding to use the dissimilarity function proposed by Stewart and Grigorieff (Equation (2.1)). Also FREALIGN was used as the USP implementation for the comparison.

## 4.1  Validation with phantoms

These experiments were performed using the crystal structure of the RNA Polymerase II-TFIIB complex [LBW+10], filtering its electron density map to about
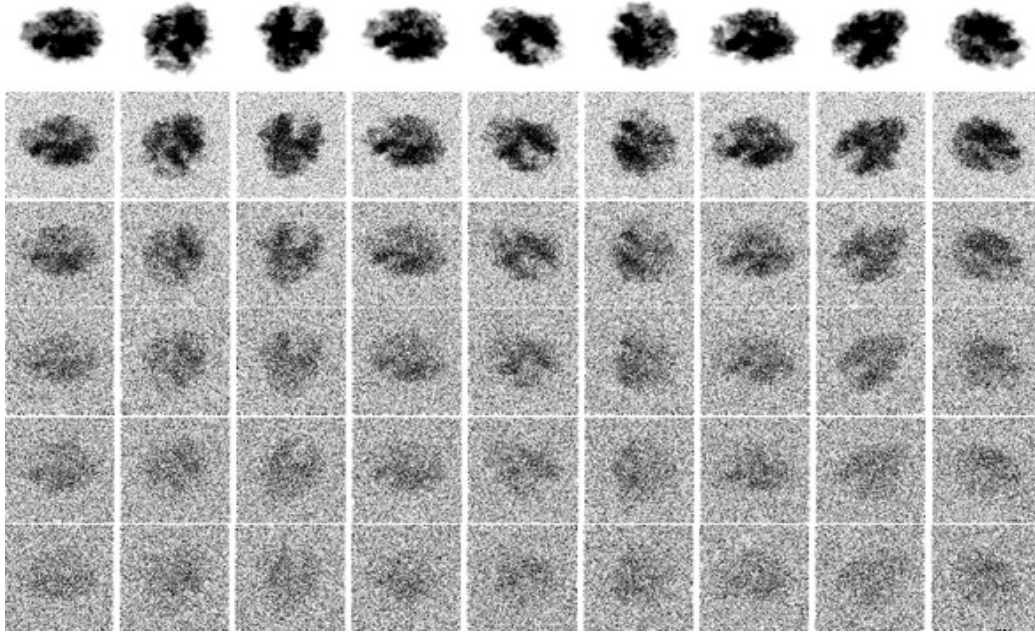
Figure 4.1: Particle projection images of a map generated by filtering the crystal structure of RNA Polymerase II-TFIIB complex (PDB ID: 3K7A) with decreasing SNR (top to bottom). Top row are the noiseless projection images generated from the synthesized particle.

3Å (see 3K7A in the PDB database [LBW$^{+}$11]), and the ET's acquisition procedure was simulated. This macromolecular assembly was selected, among other reasons, because it does not present any kind of symmetry, avoiding multiple "valid" solutions for the particle projection's orientation, allowing the comparison of the orientation angles determined by both methods. Furthermore, since CSP performs a local optimization using the parameters computed by the tomographic reconstruction as an initial point, USP should be restricted to a local search too. Using a non-symmetric particle guarantees that even doing a wide search no other coherent view will be selected, besides the errors produced by the noise.

A set of four tilt series was created, each one with five particles randomly oriented and sixty one micrographs, leaving a stack of noiseless particle projection images (see first row in figure 4.1). In these conditions, the groundtruth of the particle density map and all the parameters of the particles and micrographs are known. The configuration leaves a set of $N_\mathrm{P} = 20$ particles with $N_\mathrm{M} = 61 \times 4$ micrographs, giving a total of $N = N_\mathrm{M}N_\mathrm{P} = 1220$ particle projection images. From these micrographs and using the groundtruth parameters, a low resolution map was reconstructed by cropping and averaging the subvolumes of the corresponding particles from the reconstructed tomograms of each tilt series; this map was used as the initial reference for the CSP and USP refinement-reconstruction procedures. Two different sources of perturbation were tested. First, uncorrelated, zero mean Gaussian noise with increasing

variances was added to the two dimensional projections. Figure 4.1 shows nine columns with different 2D projection images. Top row shows the original noiseless projection images obtained; the other rows show the noisy images corrupted by additive Gaussian noise (increasing its variance within the column). Depending on the SNR levels in the particle projections this number of images could be low compared with the usual number needed in SP, but with the aim of comparing the performance of both methods is adequate. However, for the reconstruction of the maps, having few projections could represent a disadvantage, specially with the noisier images. In order to overcome this, the reconstructed maps are obtained with the noiseless images. This is a way to wean the result from the number of images, since taking their average does not lead to improved reconstructions which can only be obtained with a better alignment.

The second source of error is a perturbation in the initial condition in the optimization, i.e., an error in the parameters of the particles and micrographs, and therefore in the parameters of the particle projections images. The goal of this perturbation is to test the convergence of the algorithm and the tolerance to errors in the estimation of the initial values from the ET procedure.

A total of 121 ($11 \times 11$) different experiments are performed varying the amount of perturbation added by each source of error. The range of variance for both, the added Gaussian noise and the perturbation of the orientation angles varies from 0 to 100 in steps of 10, reaching for example a SNR of $^1/_{100}$. The results obtained are consistent throughout the experiments, for the discussion the one with the highest perturbations in the initial conditions and the lowest SNR is going to be used; an example of the noise corrupted particle projection images that are used are displayed in the last row of Figure 4.1.

In these "controlled" experiments the groundtruth parameters are known and a numerical analysis of their error can be performed. In particular, the relative orientation between all the particle projection images in the 3D coordinate system are known. When a global 3D rotation $R_G$ affects the stack of images, the reconstructed map will be affected by the same rotation $R_G$, but the relative orientation between the images in the stack will be preserved. A quantitative measure of the alignment error is proposed in Appendix B using the rotation $R^k$ between the determined and groundtruth orientations for every particle projection $I_k$. This measure is based on the fact that, if there is no error, then $R^k = R_G \ \forall k$, and also the mean rotation will be $\bar{R} = R_G$. The numerical value, called "variance of the alignment error" (VAE), is the Frobenius norm of the difference between the rotation $R^k$ and the mean rotation $\bar{R}$ matrices. See Appendix B for further explanation of its computation.

Another indicator of the performance of the framework and quality of the reconstruction is its resolution. Appendix D introduces and compares the most frequently used resolution estimators in EM, among them, the Fourier Shell
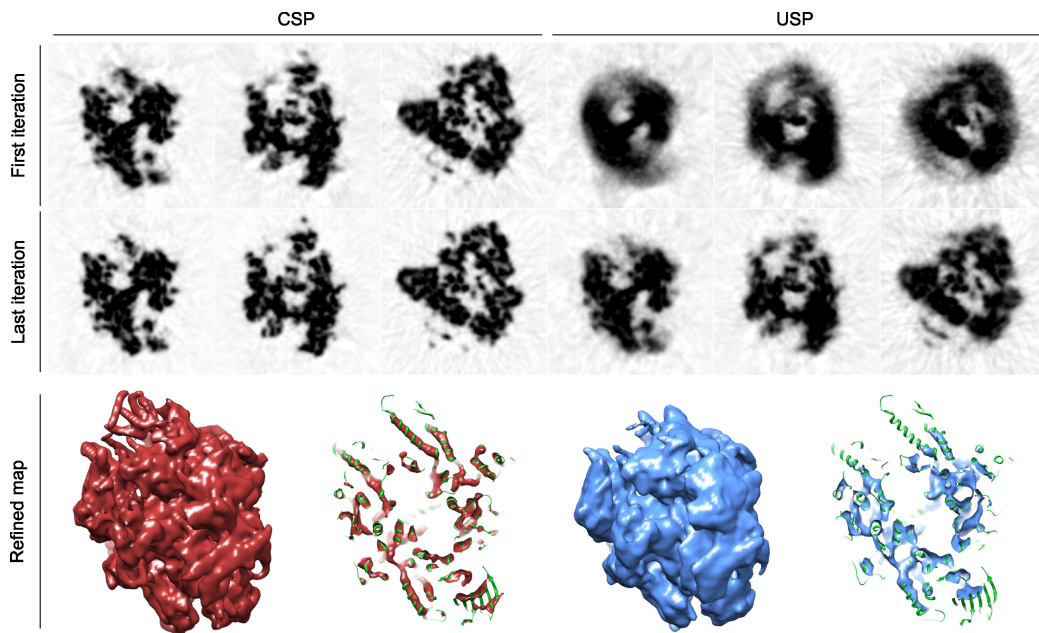
Figure 4.2: Map refinement for 3K7A phantom data, comparison of CSP vs. USP refinement for SNR = 0.01. Top row shows the axial, coronal and sagital cuts of the refined map after one iteration, by CSP (left) and USP (right). Middle row shows same cuts for the last iteration. Last row shows one view of the final refined map and the docking of the X-ray coordinates for CSP (left, in red) and USP (right, in blue).

Correlation (FSC) is the most widely used. The FSC curve varies with the frequency, measuring the coherence between two 3D maps computing an averaged Cross-Correlation (CC) coefficient in a set of non-overlaped thin shells in the Fourier reciprocal domain (see Section D.1.2). As a CC the values of FSC lie between 0 and 1, the FSC resolution estimator corresponds to the highest frequency where FSC=0.5, known as the $f_{0.5}$ cut-off frequency. The standard FSC computation is performed by splitting the set of images in two independent sets, computing two different reconstructed maps and computing the FSC between these two *half-set* reconstructions. Besides this FSC estimation, in this synthesized experiments it is possible to compare the reconstructed map with the one used to compute the projections. Thus, in addition to the traditional FSC, the FSC between the reconstructed map and the groundtruth can be computed in order to analyze the coherence between both maps. Moreover, since the reconstruction is computed from the whole set of images instead of half of it, the estimation should be more accurate due to the higher number of views used in reconstruction.

The mean Phase Residual for the whole stack is also monitored in the experiments.

Figure 4.2 shows the refined maps obtained by CSP and USP. Also the axial, coronal and sagital cuts of the models are shown for the first and last

iteration for both methods. More precisely, the three images at the left side of the top row are the model cuts of the map after one main iteration of the CSP refinement, similarly the three images at the right are the model cuts for the map after one iteration of the USP refinement. The middle row shows the same cuts for the last iteration. The bottom row shows one view of the reconstructed density map and the docking of the groundtruth X-ray coordinates for CSP (left) and USP (right).

Figure 4.3 plots the comparative results between CSP and USP for one experiment. Figure 4.3a shows five different FSC curves, the green dash-dotted curve is the initial FSC, i.e., is the FSC for the ET sub-volume reconstruction, starting point for CSP and USP. The blue dashed curves correspond to the USP refinement, the thin line is the curve at the end of the first iteration, the thick one is the curve for the last iteration. The same applies to the red solid curves corresponding to the CSP refinement. In the CSP case the thin curve is at the end of the first main iteration. Figure 4.3b shows the FSC curves computed with respect to the groundtruth model for each method in the same iterations. The abscissa's axis is the Fourier frequency, which correspond to the inverse of the map's resolution in Angstroms.

The FSC curves consistently show a considerable improvement in resolution obtained by CSP compared to the resolution obtained by USP. Considering the $f_{0.5}$ cut-off criteria the initial resolution is about 17.7Å, which is improved by USP to 16.1Å after one iteration, reaching 9.3Å by the end of the iterations. On the other hand, the resolution estimation for CSP are 6.7Å after one main iteration and 5.7Å by the end of the iterations. Beyond the numerical values, the shape of the curves support the improvement achieved by both CSP and USP. USP shows a substantial improvement in the resolution and in the correlation of lower frequencies, but falling off rapidly around 12Å. On the other hand, CSP shows an extraordinary boost in the first iteration, then it mildly improves with subsequent the iterations. Moreover, after the first main iteration with CSP, the remarkable improvement strongly supports the imposition of the geometric constraints, being noticeable that the final resolution achieved by USP is much lower than the resolution achieved by CSP with only one main iteration. Also shown in these plots is that the improvement in USP is given by the iterations, slowly aligning each image to the reference; however, iterating with CSP does not achieve the same amount of improvement. This can also be seen in the maps and their cuts in Figure 4.2: an unrecognizable map after the first iteration and a rough approximation in the final iteration for USP, meanwhile the map after CSP's first main iteration has many more details than USP's reconstruction, which is improved with the iterations, emphasizing sharp details. This is noted too, in the 3D views where the $\alpha$-helices can be easily recognized and a better docking of the X-ray coordinates is obtained. A well known issue in classification is that obtaining an improvement
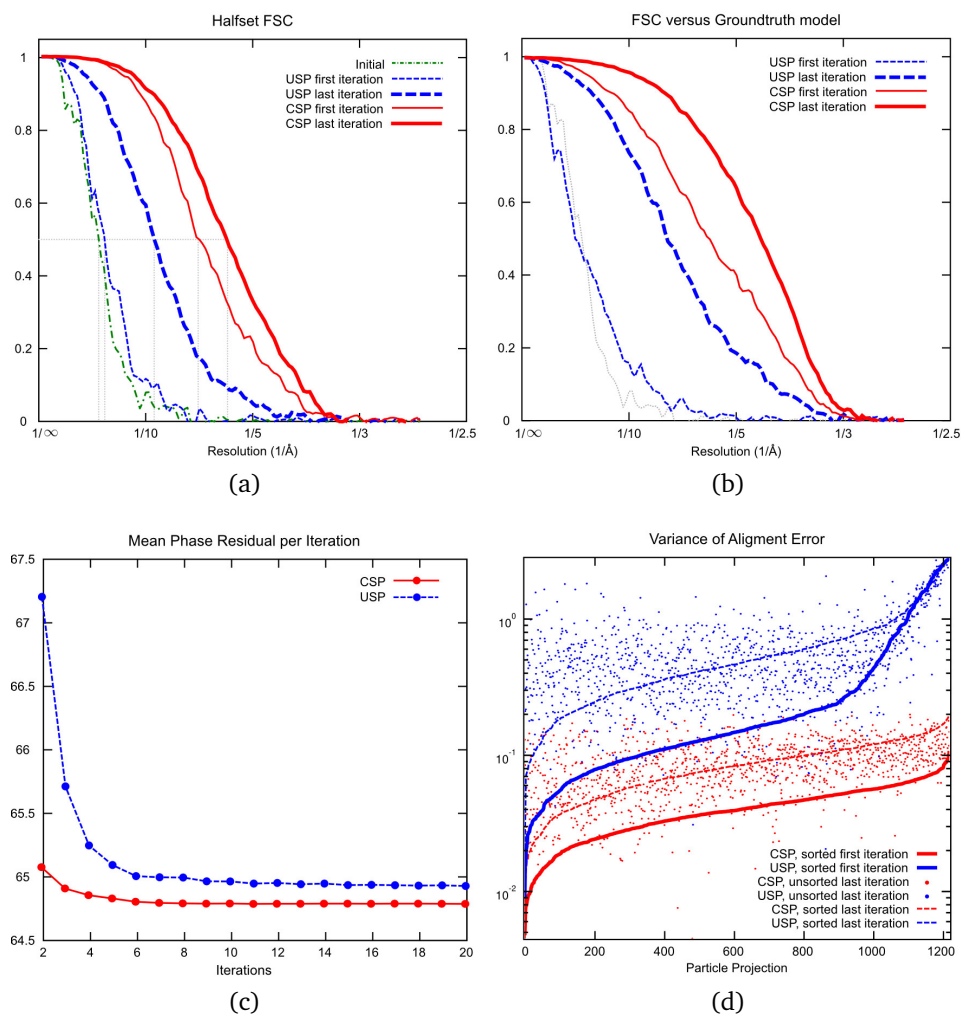
Figure 4.3: Comparison of CSP vs. USP refinement for SNR = 0.01. (a) (Half-set) FSC plots after one iteration and for the final refinement. Also the (Half-set) FSC for the initial map is plotted. (b) FSC between whole set reconstruction and groundtruth map, after one iteration and for the final refinement. (c) Evolution of the global mean Phase Residual with the iterations. (d) Variance of the Alignment Error (VAE) per particle projection; continuous and dashed curves are sorted by its value, dots are the VAE on the first iteration ordered by their index in the last iteration (continuous curve).

from 70% to 80% is much simpler than obtaining an improvement from 95% to 98%. A similar issue might be happening here: the higher the resolution of the starting map, tougher is to increase the resolution based on the same data. This could be the case in CSP, since, after one main iteration the resolution boosts up and then the resolution increments have small steps, but the refined map is much better than the USP map.

Figure 4.3b plots the FSC curves between the map reconstructed with the whole stack of images and the groundtruth map used to take the projection

images. These curves show a better correlation between the maps, probably due to an improvement in the reconstructed maps since they are obtained with twice the number of images. Numerically, the $f_{0.5}$ cut-off criteria estimates resolutions at 19.9Å, 7.4Å, 5.7Å and 4.5Å respectively. Noteworthy, the estimated resolution and the FSC curve for USP's first iteration are worse than the ones for the half-set FSC estimation (the curve for the initial map in Figure 4.3a is also plotted as a reference). This denotes a smaller correlation in lower frequencies between the groundtruth and the reconstructed maps, or a kind of overfitting between both half-sets when the alignment is not accurate, a classical drawback of the FSC resolution estimation. The FSC curves between CSP and the groundtruth maps in Figure 4.3b exhibit the same behavior as in Figure 4.3a: the first iteration after imposing the constraints gives the main boost in the refinement. However, the resolution enhancement within CSP's iterations is more noticeable in this graph, specially the shape of the curve for the last iteration having a larger "flat" zone in lower frequencies and a slower fall-off, denoting its highest correlation with the groundtruth map.

Figure 4.3c shows the evolution with the iterations of the mean Phase Residual (PR) for all the projections in the data set, the dashed blue curve for USP and in solid red for CSP. The PR is the cost function being optimized by both methods, individually per image in USP and a weighted average in CSP. Both plots show a steady reduction of the mean PR due to the optimization, getting stable around iteration twelve. This behavior is consistent with the FSC curves through the iterations, although they are not numerically comparable; for USP there is an important reduction within the firsts iterations while for CSP the main reduction come in the first iteration imposing the constraints, then it slightly improves. In the experiment shown, the mean PR in CSP is less than the PR in USP, which could imply a better matching between the projection image and the reference model. However, this is the case with the noisiest images where the individual alignment of the projections is tougher than aligning an array of correlated projections, reflecting the advantage of CSP. When the initial perturbation is not too high, generally the mean PR for USP is less than the mean PR for CSP, having a better matching in terms of PR but without satisfying the constraints, which is not reflected in a better alignment.

Figure 4.3d plots the variance with the particle projection index in the alignment error introduced in Section B.2, sorted by its increasing values. The figure shows the same data representation for USP, in blue, and CSP, in red. The solid thick line represents the variance (VAE) for each particle projection of the alignment error at the last iteration sorted increasing its value; the VAE for the first iteration is plotted, too, as a dot for each particle projection respectively. Also the sorted version of the VAE for the first iteration is plotted as a dashed line; this corresponds to the previous dots plotted in an increasing
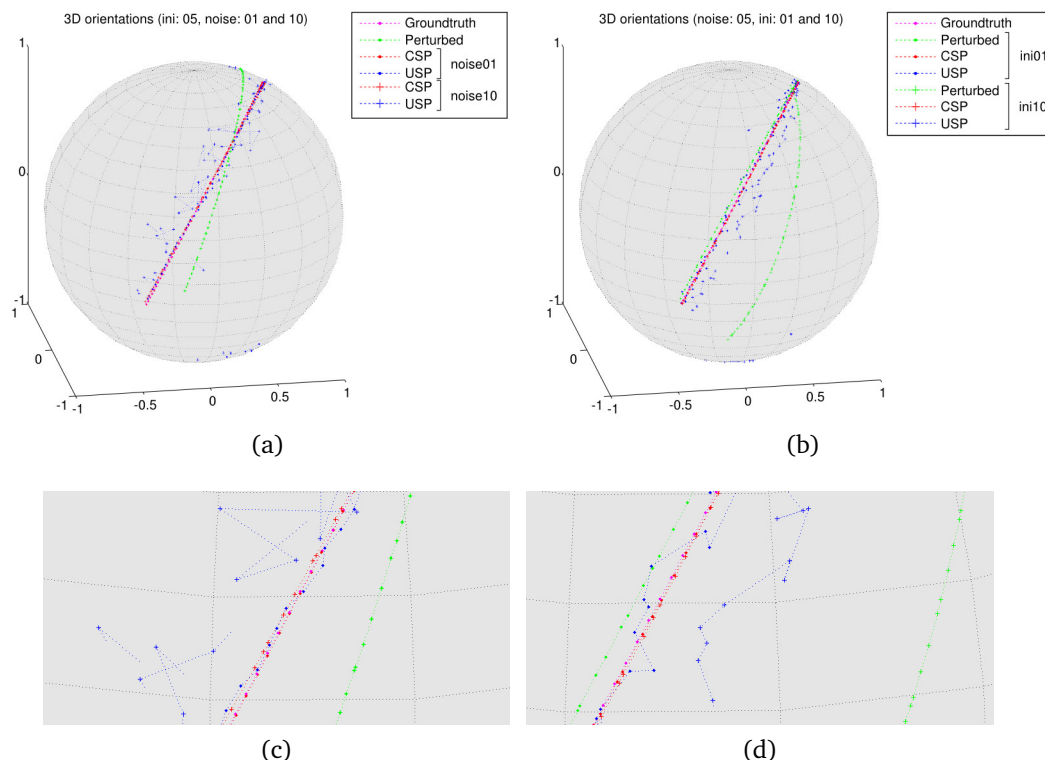
Figure 4.4: Particle projections plotted as points in the unit sphere with coordinates given by their "out-of-plane" angles. Groundtruth points are plotted in magenta following the tilt angles of the micrographs. The perturbed orientations are plotted in green. The refined orientations are plotted in red for CSP and blue for USP. (a) Fixing the amount of perturbation of the initial orientations in an intermediate step (ini05), two different levels of noise are shown (noise01 and noise10), see legend and text. (b) Fixing the amount of variance of added noise in an intermediate step (noise05), two different levels of perturbation are shown (ini01 and ini10), see legend and text. (c) Close view of the sphere in (a) (d). Close view of the sphere in (b).

order. It can be seen that the large majority of the projections reduces its alignment error variance, since most of the points are above the thick curve for each method. However, CSP has clearly a smaller variance for each projection than USP, meaning that the dispersion of the alignment error with respect to the groundtruth orientation is higher for USP, and the constraint of moving all the projections of each particle altogether helps reducing this spreading. Also this is stressed in USP's curve, where there is a subset of projections, the ones with higher VAE, that cannot be improved, while this effect does not happen in CSP.

The fulfillment of the constraints can be viewed in Figure 4.4, where each particle projection is represented as a point in the unit sphere given its "out-of-plane" angles. The initial groundtruth orientations are plotted in magenta while the perturbed ones are plotted in green; note that the perturbations
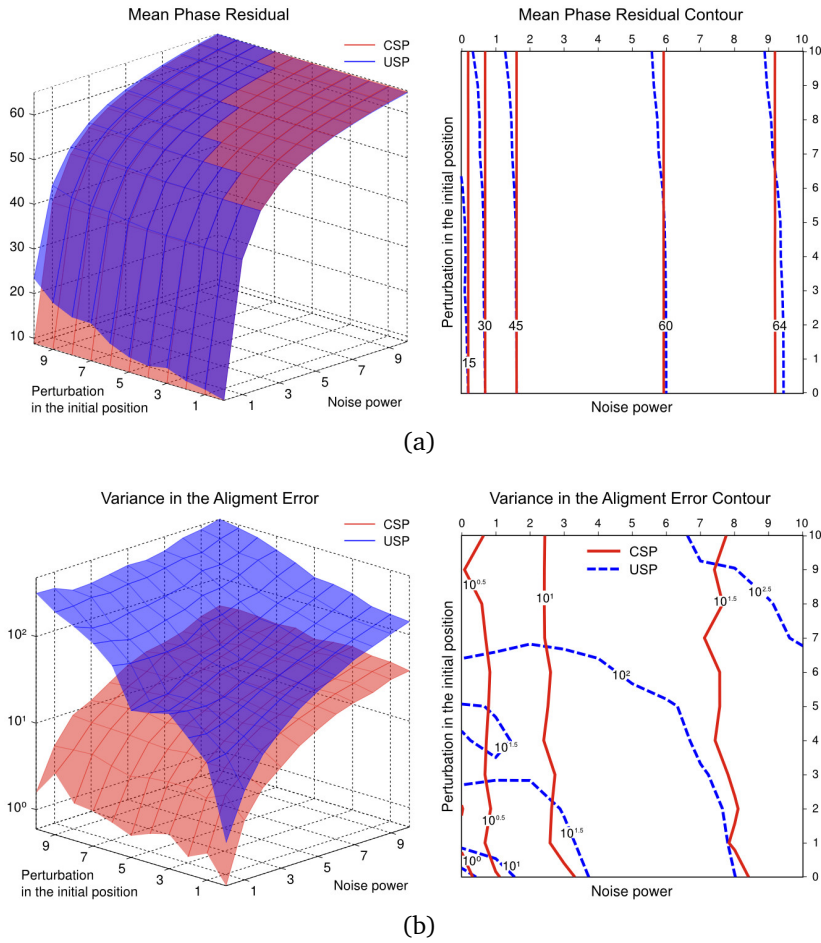
(a)



(b)
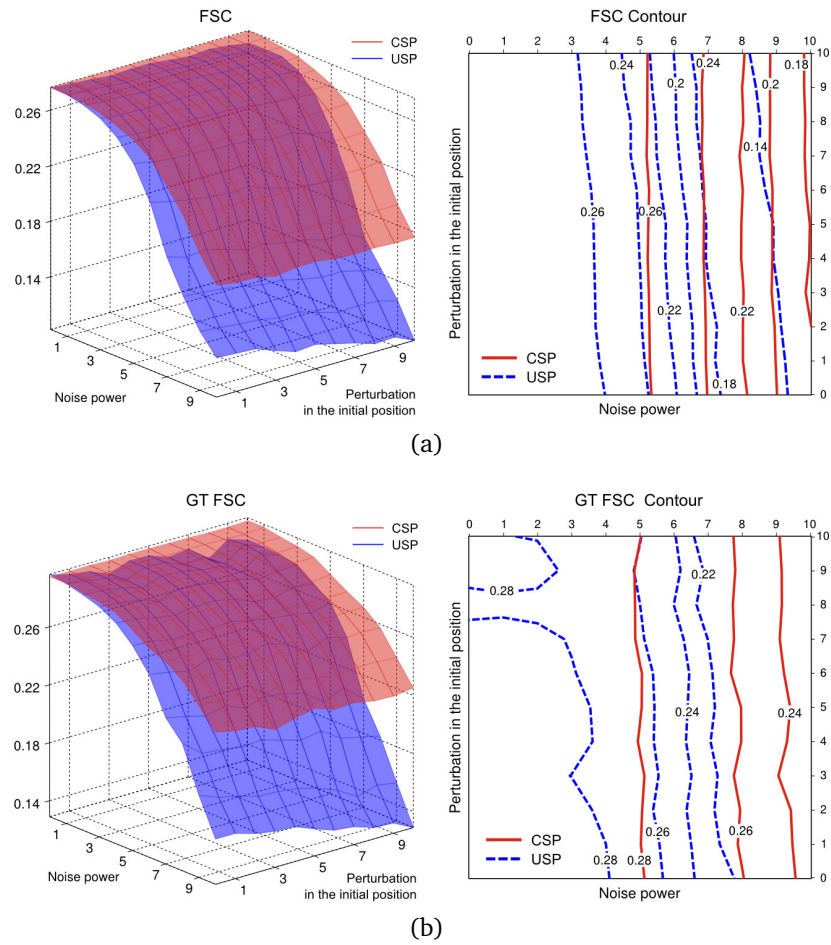
Figure 4.5: Comparative analysis of CSP vs. USP refinement as a function of SNR and perturbation of initial condition on 3K7A phantom. (a) Map resolution measured as $f_{0.5}$ (Halfset) FSC cutoff. (b) Map resolution measured as $f_{0.5}$ FSC cutoff (whole set vs. groundtruth map).

preserve the constraints since the perturbed objects are the particles and the micrographs. The orientations recovered by CSP are plotted in red and in blue for USP. Two different experiments are shown, first, for a given perturbation of the initial orientations, set to an intermediate amount, the results with the lowest and highest variance of added noise are shown (plotted with points and crosses, respectively in Figures 4.4a and 4.4c , also they are connected with a dotted line following the order of the micrographs). The second experiment shown, fixes the variance of added noise in an intermediate step and plots the results for the lowest and highest perturbation of the initial orientations, following the same color and markers convention, see Figures 4.4b and 4.4d. In all the experiments CSP fulfills the constraints, showing great robustness to both, the perturbation in the initial orientations and the noise. On the other hand, the orientations recovered by USP do not always fulfill the constraints, sometimes they change the order given by the micrographs and present an

(a)



(b)

Figure 4.6: Comparative analysis of CSP vs. USP refinement as a function of SNR and perturbation of initial condition on 3K7A phantom. (a) Sum of mean Phase Residual. (b) Sum of the Variance in the Alignment Error compared to ground truth (log-scaled).

erratic behavior; this is expected since the refinement of each projection is performed independently. The only experiment where USP leads to a result close to the groundtruth is for the lowest amount of added noise; again, this is expected, since the projection image has almost not degradation.

Figures 4.5 and 4.6 show the same plot as in Figure 4.3 for all the experiments. In all the cases the mesh and its contour are plotted in a grid varying the standard deviation of the perturbation from the groundtruth orientation and the standard deviation of the Gaussian noise added to the particle projection images. Figures 4.5a and 4.5b plot the $f_{0.5}$ cut-off frequency for the half-set FSC and FSC against the groundtruth respectively, Figure 4.6a plots the mean PR in the last iteration and Figure 4.6b plots the sum of the VAE for all the projection images (the area under the thick lines in Figure 4.3d). One of the main characteristics of these plots, in particular the contour plots, is the robustness of CSP to the perturbations in the initial position denoted by the
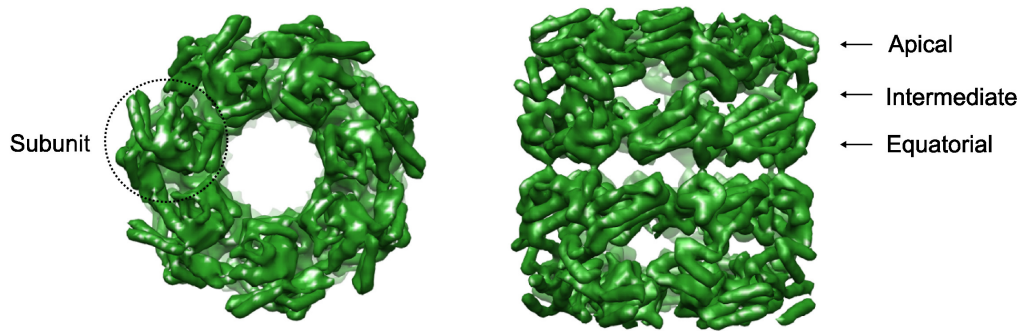
Figure 4.7: Top (left) and side (right) views of the GroEL complex at a resolution of 4Å. One of the fourteen subunits is marked in the top view. The side view marks the apical, intermediate and equatorial levels of one of the two rings. The views were generated with an alpha transparency proportional to the depth to improve its visualization.

almost vertical level lines. Concerning the FSC, CSP always reports better resolutions than USP, and their gap increases if the FSC against the groundtruth is considered. The PR shows the same behavior previously commented, with both methods reaching similar results. Finally, the sum of the VAE show that USP has a considerable error, increasing with the perturbation and the noise, which is a quantification of the dispersion observed in Figure 4.4; surprisingly, CSP almost preserves its robustness with the variation of the initial perturbation, specially in the noisiest experiments.

## 4.2 Validation with real data

The experiments with data acquired following the Cryo-ET procedure reveal the complexity of the problem due to the extremely low SNR. The selected particle was the unliganded GroEL macromolecular assembly, widely used as a well-know structure, whose atomic map is available (Id 3e76 in the PDB database [KLP11]). The atomic map was filtered in order to obtain a high-resolution groundtruth map to be used in the comparisons. From the geometric point of view, GroEL is a tetradecamer composed of two back-to-back, seven-subunit rings [Hor11]. Its top and side views are shown in Figure 4.7. The GroEL macromolecule is a chaperonin protein, i.e., a protein that is required for the proper folding and unfolding of many other proteins, translating the polypeptide chains into theirs characteristic three-dimensional active structures. In the case of GroEL, it requires the co-chaperonin protein complex GroES, acting as a lid.

In this experiment the real orientations of the particles and micrographs are not known; therefore the VAE measure can not be computed, and only

the refined maps can be used for comparison. Despite the advantage of being a well-know structure, GroEL as a highly (C7) symmetric complex, adds two drawbacks in the comparison. First, multiple "valid" solutions for the particle projection's orientation parameters might exist, avoiding the direct comparison of the orientation angles. Second, the reconstruction procedure is improved by repeating the image in the symmetric orientation increasing the number of averaged images, and then the SNR, which can hide the improvements given by the refinement methods. However, both, USP and CSP are subject to this issue, and if a better alignment is obtained by one of the methods, the multiplication of the number of images will act as an amplification factor in both cases.

Two different datasets were acquired, varying the dose fractionating protocol. In the first one, GroEL_01, the total dose (70 $e^-/\text{Å}^2$) was equally divided throughout sixty one micrographs. In the second dataset, GroEL_02, an increased total dose (105 $e^-/\text{Å}^2$) the dose was divided throughout forty five micrographs, leading to a higher dose per micrograph.

## 4.2.1 Experiments with the GroEL_01 dataset

The GroEL_01 dataset has a total of 25 tilt series, about 100 particles in each one and 60 micrographs, totaling about 150,000 particle projection images. This data was used with the standard ET reconstruction procedure to reconstruct the corresponding tomograms, generating an initial map with the sub-volume averaging procedure previously described; also a set of initial orientations for each particle and micrograph, and therefore, for each particle projection via the mapping function (Equation (3.1)) was obtained. Figure 4.8a shows a subset of the particle projection images used in the experiment; each particle projection is located approximately on the center of the image.

The refinement and reconstruction procedure was as follows. First, a low resolution map was reconstructed following the sub-volume averaging procedure using the cropped particles from the tomographic reconstruction, this map is shown in Figure 4.8b. Using this map as an initial reference, only the translational parameters of each image were refined, preserving the orientations given by ET, and reconstructing using the original raw images following the procedure described in Section 3.2. This procedure can be seen as a partial CSP refinement (and will be noted as CSP0), were the orientations parameters are not refined, given by the ET orientations. The obtained map is shown in Figure 4.8c. Using this map as an initial reference map, the full USP and CSP refinement procedures were applied. The obtained maps are shown in Figures 4.8d and 4.8e, respectively. Visual inspection of these maps reveals the improvement obtained from the sub-volume averaged map, showing that the whole procedure in the ET reconstruction leaves aside high frequency in-

(a)



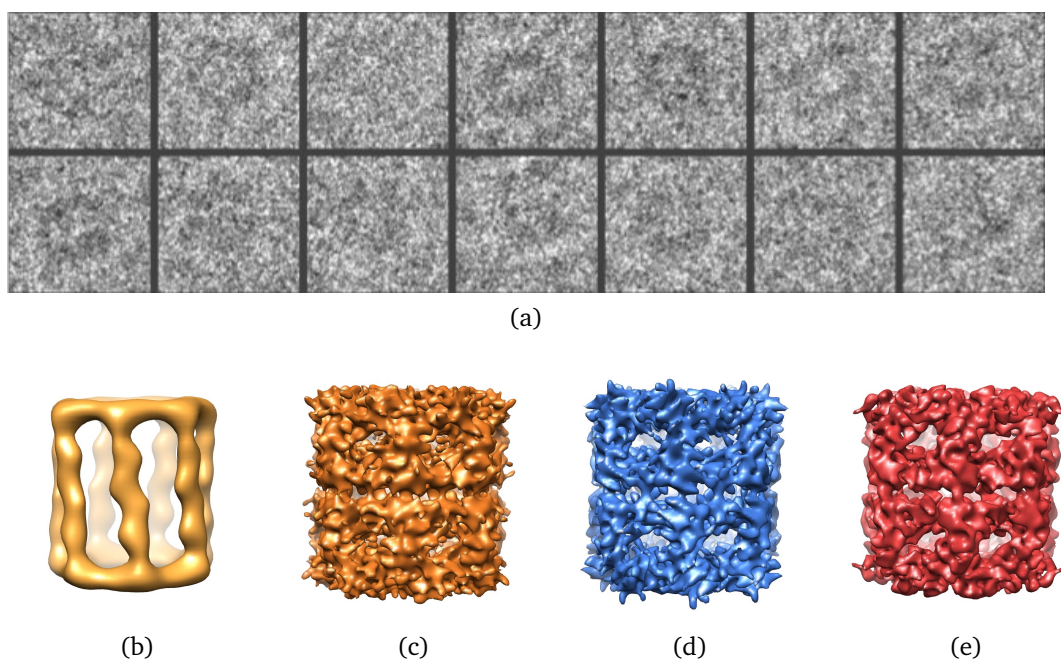(b)             (c)             (d)             (e)

Figure 4.8: (a) Particle projection images from the GroEL_01 dataset. Maps obtained with (b) tomographic sub-volume averaging, (c) partial CSP, (d) USP and (e) CSP (see text).

formation due to the averages and interpolations. This range of data is "recovered" when the original projection images are used in the reconstruction and only the translational parameters are refined. It should be noted, being an important issue in the analysis and interpretation of the results, that the maps reconstructed with the alignments determined by CSP0, USP and CSP where computed taking into account only the *best* particle projections, using their PR as a discriminative feature; only the best 25% were used. Actually, this could be viewed as an advantage to the USP reconstruction, since CSP looks for a compromise between all the projections of a particle, probably slightly degrading the alignment of its *best* projections in order to keep a "mutual" coherence. This could be controlled with the weighting scheme $w_I$ in the particle cost function definition given in Equation (3.4). This apparently drawback is CSP's strongest feature, since as will be shown in the next section with the GroEL_02 dataset, in extremely low SNR's scenarios, by including several tilted views and imposing the constraints it is possible to recover better alignments.

Figures 4.9, 4.10 and 4.11 present the results for the GroEL_01 dataset. Figure 4.9a plots the half-set FSC curves for the following maps: CSP0 in thin solid red, USP in thick dashed blue and CSP in thick solid red, the FSC curve for the map obtained after one iteration of the partial CSP refinement over the ET map is plotted in thin dashed green. Also the CTF profile, i.e. the profile along a single radial line of the estimated CTF for one tilt series, is plotted as a
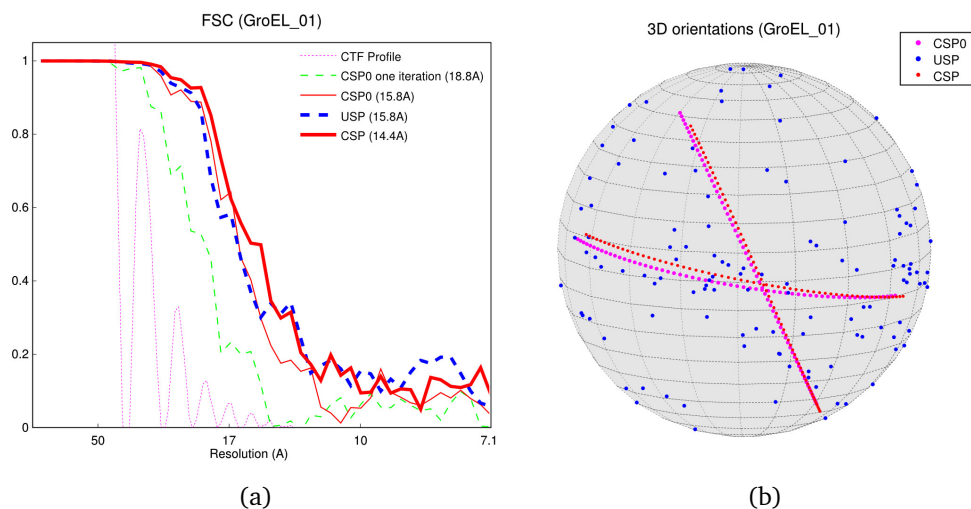
Figure 4.9: Results for the GroEL_01 dataset. (a) Half-set FSC curves. (b) Orientations of some particle projections, note the fulfillment of the constraints for the CSP case.

dashed magenta curve, see Appendix C for further description of the CTF. Note the frequency-dependent oscillations and the dampening envelope attenuating high-resolution information, also note that the first zero crossing is around 40Å, the usual resolution achieved by ET without CTF correction and sub-volume averaging; also note that beyond 13Å the CTF profile lacks significant components. The achieved resolution estimations are 15.8Å for both USP and the partial CSP refining only the translation parameters, and 14.4Å for the full CSP refinement. Figure 4.9b plots the fulfillment of the constraints, where each particle projection is represented as a point in the unit sphere given its orientation. Two different particles are shown for the partial CSP, USP and full CSP. Obviously, partial and full CSP (in magenta and red, respectively) fulfill the constraints while USP (in blue), starting from the partial CSP does not fulfill them. Figure 4.10a shows the axial, coronal and sagital cuts for (top to bottom) CSP, groundtruth and USP maps, respectively. Figure 4.10b shows the side view of (left to right) the CSP, groundtruth filtered at 10Å and USP maps with the X-ray coordinates docked (in purple). Figure 4.10c zooms at one $\alpha$-helix in the apical region of the top ring marked with a red square in Figure 4.10b. Figures 4.11a, 4.11b and 4.11c show a thin section of the maps in the apical, intermediate and equatorial region of the top region, respectively, and the X-ray coordinates docked.

Figure 4.10 *(following page)*: Results for the GroEL_01 dataset. (a) Left to right: axial, coronal and sagital cuts of the map by (top to bottom) CSP, groundtruth at 10Å and USP, respectively. (b) Side view of the (left to right) CSP, groundtruth at 10Å and USP with the X-ray coordinates docked. The areas marked by the red squares are zoomed in (c).
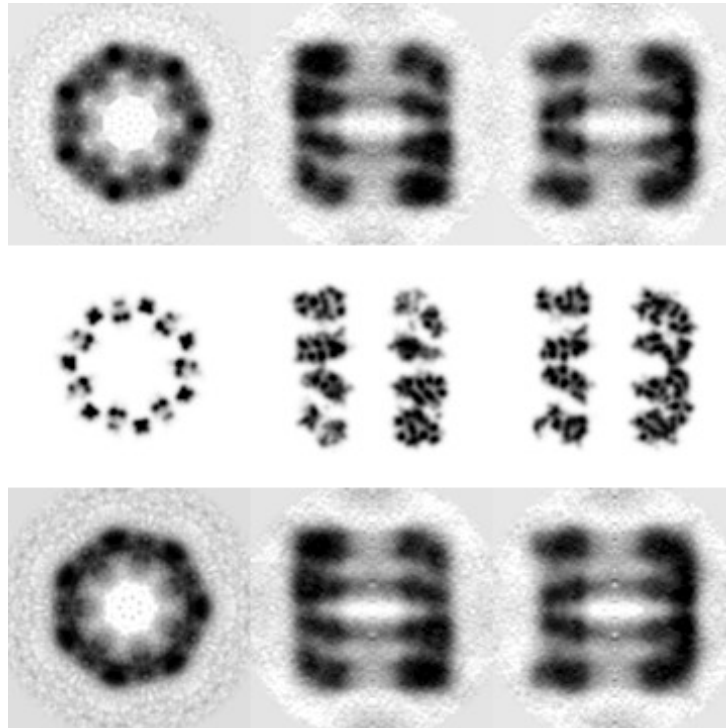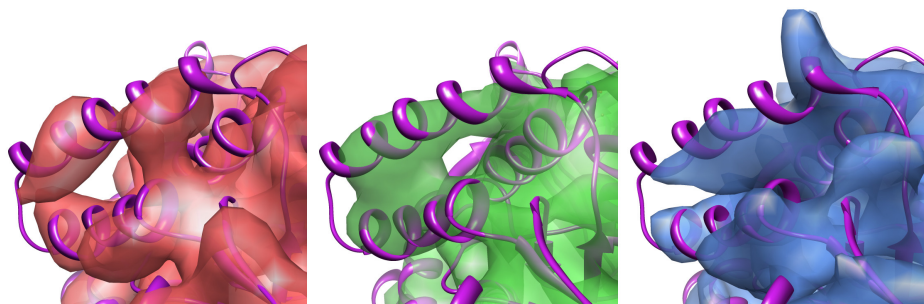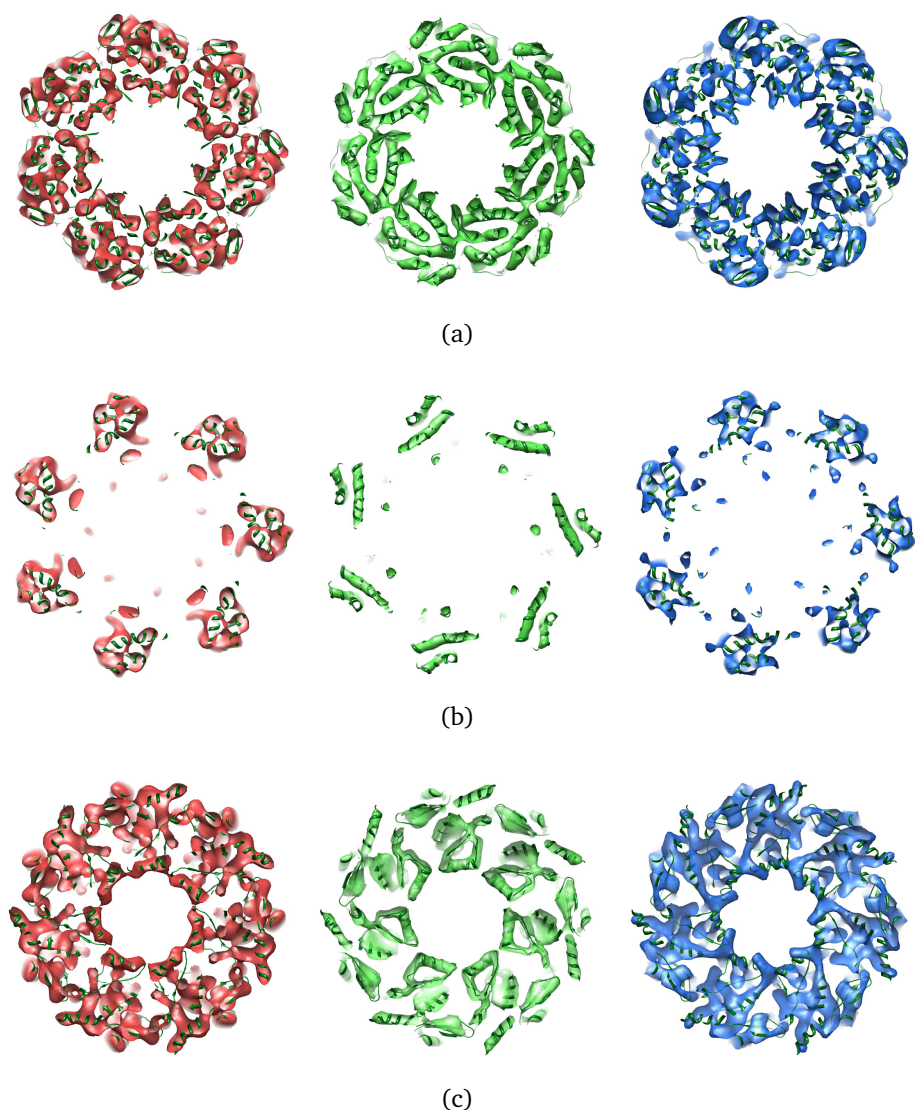
(a)
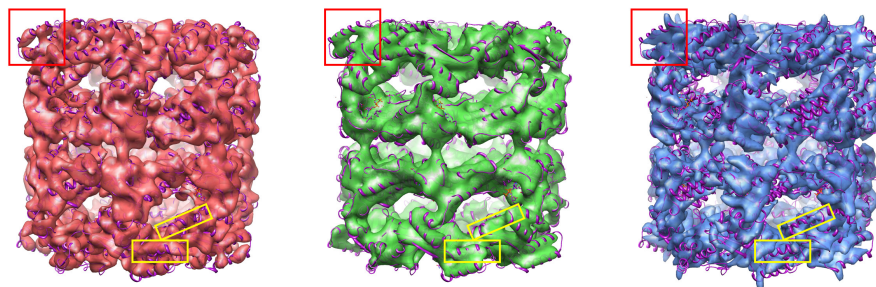


(b)



(c)

Figure 4.11: (a) Apical (b) intermediate and (c) equatorial slices from the top ring of the density maps obtained with the GroEL_01 dataset shown in figure 4.10 for (left to right) CSP, groundtruth at 10Å and USP. Also the docked X-ray coordinates are shown.

The main conclusions from these results is that CSP slightly but clearly, improves the results obtained by USP. The improvement is marginal in the FSC curves, although the resolution estimation gives more than one Angstrom of difference. Also in the model cuts, CSP shows a much cleaner map with less artifacts, and more well defined zones than USP. In the density maps with the docked coordinates, CSP shows better distribution of the density around several helices, clearly in Figure 4.10c, where USP also shows a strange "horn" in a direction outside the particle structure. Also in various regions marked in Figures 4.11a, 4.11b and 4.11c the CSP better matches the direction and posi-
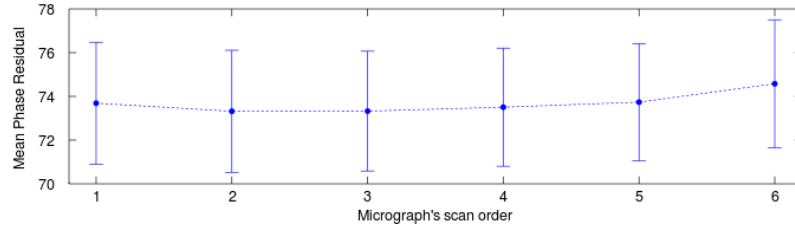
Figure 4.12: Mean PR for the first six micrographs in the GroEL_02 dataset.

tion of helices while USP usually misses them or wrongly determines another direction.

This experiment shows some clear improvements achieved by the proposed framework, nevertheless, is it possible to further exploit the advantage of the CSP framework? Several issues affect the results showed with the GroEL_01 dataset favoring the USP reconstruction, first, the initial reference and initial orientations are the ones computed by the partial CSP, approaching the solution to the local minima defined by CSP. Second, GroEL is a highly symmetric particle, which could be exploited by USP determining alignments for projections of the same particle in different symmetric subunits, while CSP explicitly avoids it. Third, only the best 25% of the projections were used in the reconstruction. In USP, this PR's thresholding discards the improperly aligned projections which do not affect the ones used in the reconstruction. Meanwhile, in CSP the rejected projections could degrade the alignment of the selected ones, since all the projections of one particle affect their alignment. This can be controlled by the weighting scheme $w_I$. And finally, is there enough signal present in the data in order to achieve higher resolutions? Several factors affect how much signal is present in the data, and in particular up to what frequency (resolution) there is *available* signal; some of these factors are: the dose that is used, the magnification of the images, the accumulated radiation, among others. The CTF profile for the GroEL_01 dataset plotted in Figure 4.9a indicates that it is expected to find signal up to 13Å, and then, CSP managed to extract almost all the *information* available in the data. Hence, there seems to be room for improvement of the CSP reconstruction; this will be addressed in the next section with the GroEL_02 dataset.

### 4.2.2 Experiments with the GroEL_02 dataset

The GroEL_02 dataset has a total of 144 tilt series, with a total of 11,105 particles. The increased total dose (105 $e^-/\text{Å}^2$) was equally distributed between the 45 micrographs used, leaving about 2.33 $e^-/\text{Å}^2$ for each micrograph. The number of micrographs should be reduced from the previous dataset due to the total radiation accumulated, note that the last projection in this dataset has accumulated 50% more radiation than in the previous dataset. Thus, some
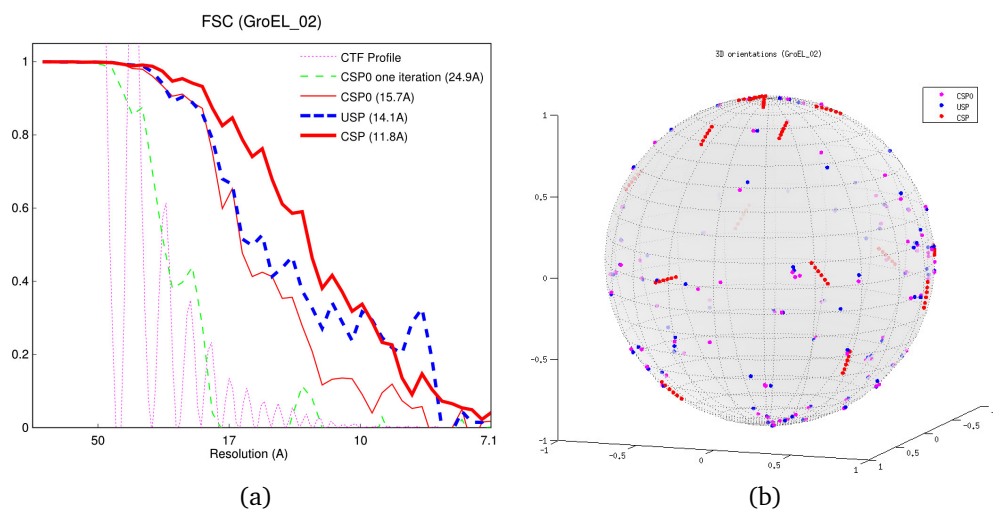
Figure 4.13: Results for the GroEL_02 dataset. (a) Half-set FSC curves. (b) Orientations of some particle projections, note the fulfillment of the constraints for the CSP case.

particle projection images have better SNR, but also causes a faster degradation of the particle, therefore, not all the projections could be useful and some of them may degrade the reconstruction. The SNR is higher than the SNR in the GroEL_01 dataset, but still much lower than the SP standards since the tilted geometry requires it.

The missing wedge in this case is too big in order to consider the subvolume averaging, therefore to obtain an initial set of parameters a not exhaustive random global optimization was performed keeping the orientations that minimize the PR; a standard procedure in the SP framework. The good results obtained (see later) indicate that even when there is no available initial condition, a random search will not degrade CSP performance. As in the previous experiment, a partial CSP was initially performed refining only the translational parameters of the projection images, from this map, the USP and full CSP refinement are performed.

**A new refinement and reconstruction scenario: CSP$_{N:1}$**

Increasing the dose per micrograph brings a faster degradation of the particles, and therefore, the high frequency signal is already lost in micrographs with lower tilt angle than in the previous dataset. Although it can be interpreted

Figure 4.14 *(following page)*: Results for the GroEL_02 dataset. (a) Left to right: axial, coronal and sagital cuts of the map by (top to bottom) CSP, groundtruth at 10Å and USP, respectively. (b) Side view of the (left to right) CSP, groundtruth at 10Å and USP with the X-ray coordinates docked. The marked regions are zoomed and viewed from top in (c).
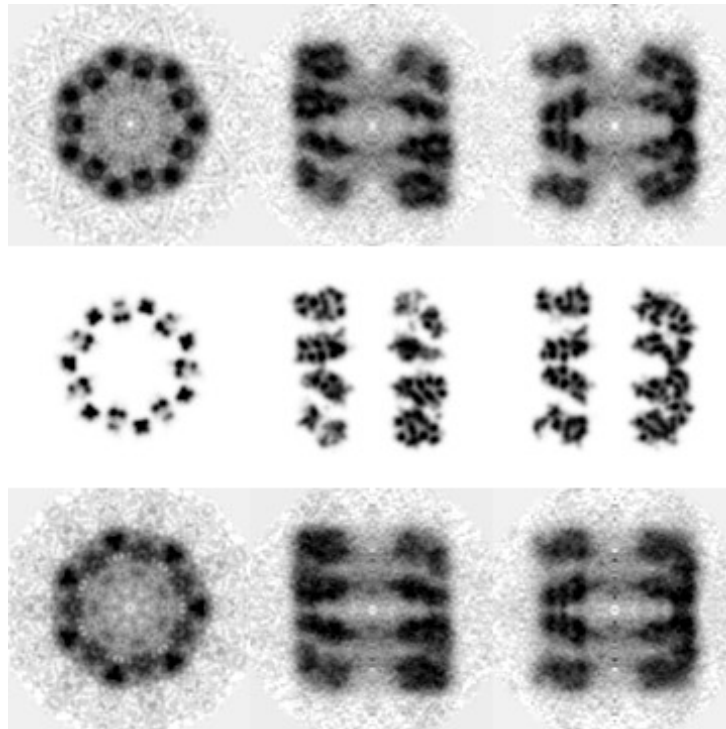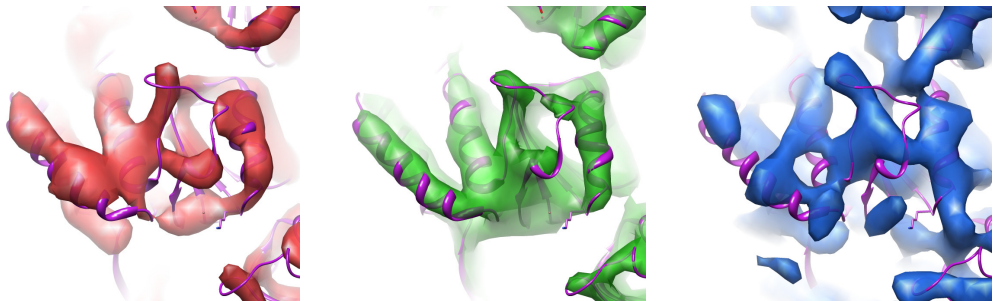
(a)
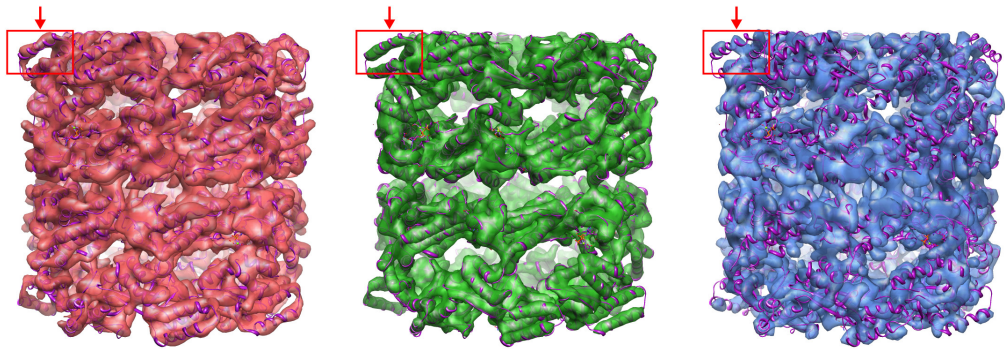


(b)



(c)

as an error, it is expected that the first micrograph has a higher mean PR than the second one, despite being less affected by the radiation damage. This can be explained by the fact that the first irradiation of the specimen burns out irregularities produced in the ice when it was created and destroys small residues in the ice affecting the projection images. After the second irradiation the PR of micrographs increases again. Figure 4.12 shows an error bar plot where the mean PR and its standard deviation for the first six micrographs is plotted. Therefore, if only one micrograph is going to be used, then the reasonable choice is to select the second one, where the projections with lower mean PR are located.

This proposes a new scenario for the refinement and reconstruction, reconstructing with the projections from one micrograph and using the extra projections to do the constrained refinement. In other words, strengthen the refinement of the parameters for the projections used in the reconstruction by doing the constrained refinement with more projections. When $N$ projections are used in the refinement and $M$ projections are used in the reconstruction, the CSP framework will be noted as $CSP_{N:M}$ and pronounced as "CSP $N$ to $M$". Following this notation, a $CSP_{6:1}$ experiment was done, using the first six micrographs for the parameters refinement, imposing the constraints between them and only the projection in the second micrograph was used for reconstruction. This scenario leaves a set of 11,105 particle projection images to be used in the reconstruction. On the other hand, USP refines and reconstruct only with the projections in the second micrograph. This is the corresponding comparison since the set of images which is used to build the maps to be compared remains the same, therefore, a better reconstruction implies a better alignment.

Figures 4.13, 4.14 and 4.15 present the results for the $CSP_{6:1}$ experiment on the GroEL_02 dataset. Figure 4.13a plots the half-set FSC curves for the following maps: CSP0 in thin solid red, USP in thick dashed blue and CSP in thick solid red. The FSC curve for the map obtained after one iteration of the partial CSP refinement over the ET map is plotted in thin dashed green, and the CTF profile is plotted as a dashed magenta curve. Note that, the higher dose used in this dataset takes the CTF profile up to 10Å, which should be reflected in the resolution of the obtained maps. Figure 4.13b plots the fulfillment of the constraints, where each particle projection is represented as a point in the unit sphere given its orientation. Figure 4.14a shows the axial, coronal and sagital cuts for (top to bottom) CSP, groundtruth and USP maps,

---

Figure 4.15 *(following page)*: (a) Apical (b) intermediate and (c) equatorial slices from the top ring of the density maps obtained with the GroEL_02 dataset shown in figure 4.14 for (left to right) CSP, groundtruth at 10Å and USP. The regions marked by a red square are zoomed below each slice (see text).
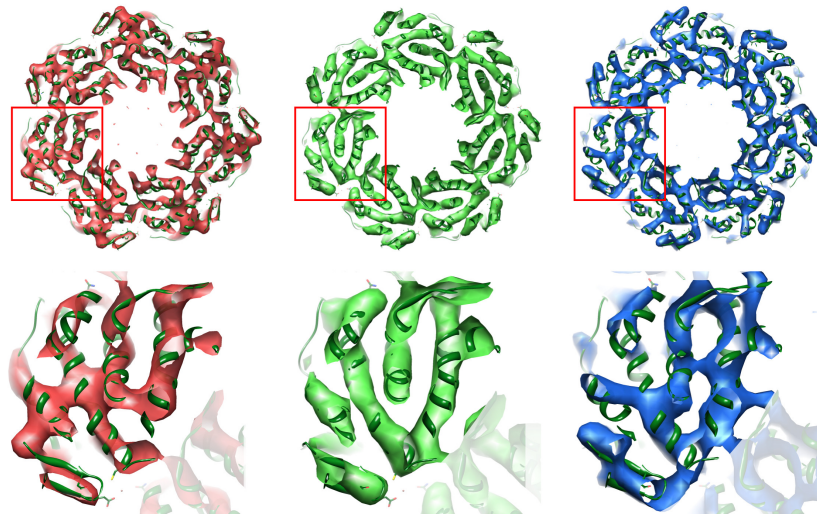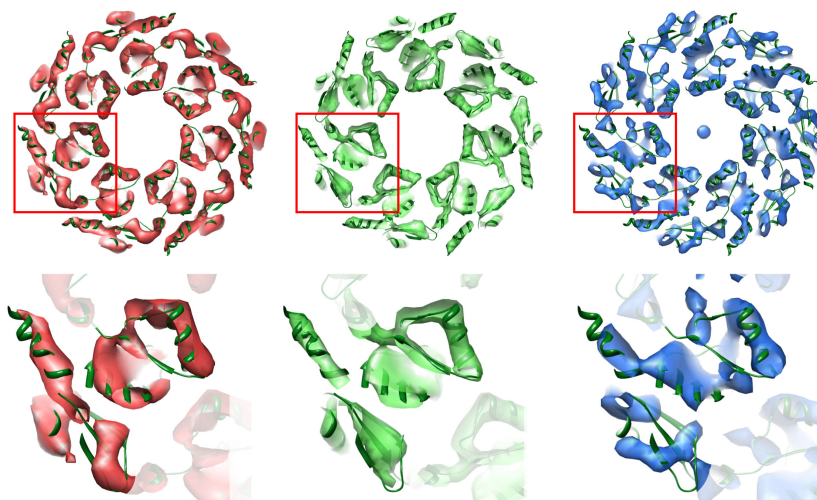
(a)
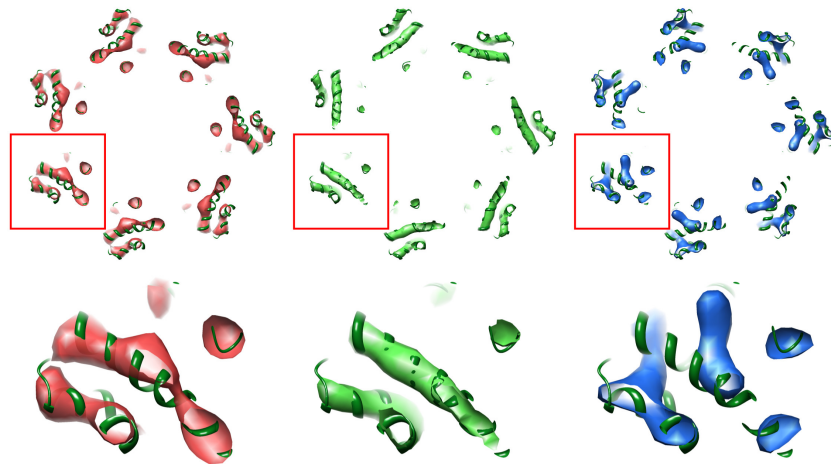


(b)



(c)

respectively. Figure 4.14c and 4.14b shows the top and side views of (left to right) the CSP, groundtruth filtered at 10Å and USP maps with the X-ray coordinates docked (in purple). Figures 4.15a, 4.15b and 4.15c show a thin section of the maps in the apical, intermediate and equatorial region of the top region, respectively, with the X-ray coordinates docked, also the region marked by red squares are shown immediately below each region.

The results of this experiment show that the map obtained with CSP achieves better resolution, has finer details and better defined regions than the map obtained with USP. From the FSC point of view, the curve for CSP shows more correlation at higher frequencies improving the results of the partial CSP in all the frequencies. Meanwhile the USP curve, shares with the partial CSP curve the same correlation for lower frequencies, with a slight increase for medium frequencies, in particular at the $f_{0.5}$ cut. The estimated resolution by the $f_{0.5}$ cut-offs are 11.8Å for CSP and 14.1Å for USP, more than 2Å difference in the range of 10 to 15 angstroms, and this is noticeable in the reconstructed maps. The model cuts in Figure 4.14a show much more coherence between the groundtruth and the CSP map than with the USP map. The *subunits* in the CSP map are well defined, and the features are much more clear, almost looking as a noisier version of the groundtruth model cuts. In the case of USP, there are more artifacts due to the alignment errors which in turn cause artifacts in the map because of the imposition of symmetry. Finally, the model views with the docked X-ray coordinates, support the much better correlation between the groundtruth and the CSP map, where some helical details are well defined and clearly denoted. Again, as in the GroEL_01 dataset those details are hardly guessed in the USP map. The zoomed regions shown in Figures 4.14 and 4.15 are clearly defined by CSP and the density maps follow the direction of the helices, on the other hand USP presents errors connecting closer different helices showing density in direction that does not correspond with the groundtruth.

Although the resolution estimation by the $f_{0.5}$ for the $CSP_{6:1}$ is far from 10Å, the CSP map looks quite similar to the groundtruth map, supporting that the *real* resolution is closer to 10Å. This is further supported by the fact that only 11,105 projections were used in the reconstruction, while other experiments use above 30,000 projections, and the FSC estimation accuracy increases with the number of images. Finally, the CSP achieved map is surprisingly good, being possible, again with the relatively few projection in the reconstruction, only if a good set of alignments is used, supporting the imposition of the geometrical constraints as in CSP.

# Chapter 5

# Concluding remarks

The possibility to study *in vivo* specimens, by identifying and reconstructing by sub-volume averaging macromolecules within them, the reference-free initialization and disambiguation in the handedness of the particles given by tilted geometry, are highly rated features of the Cryo-Electron Tomography framework. Nevertheless, they are sacrificed in the compromise for high-resolution reconstructions due to the experimental limit of about 20Å. The Single Particle framework, on the other side, does not have these features, but achieves much higher resolutions in its reconstructed maps, based on a highly tuned procedure of refinement and reconstruction using always the original particle projection images and the CTF correction. In this work, a new framework is proposed, for the refinement and reconstruction of data acquired with a Cryo-ET procedure that obtains high-resolution reconstructed maps. This framework, called CSP (standing for Constrained Single Particle) describes a new procedure for the refinement of the particles and micrographs' alignment parameters based on the Single Particle approach exploiting the geometric constraints of the tomographic acquisition. The imposition of these constraints, defines a mapping function relating the parameters of the particles and micrographs with the parameters of the particle projections, allowing to determine their alignment in a space with a much lower dimension, hence gaining robustness and accuracy.

The experiments performed on acquired data show that it is possible to achieve resolutions of about 10Å with the usual doses in ET and with a relatively reduced number of images, outperforming the reconstruction without imposing the constraints USP (standing for Unconstrained Single Particle). With synthesized and acquired data, CSP shows better response than USP in a scenario with low SNR and a set of extra tilted projections. In other words, CSP obtains a similar performance as USP but with lower doses in the data used for reconstruction, along with some (few) extra tilted projections used to do the constrained refinement. These extra projections allow to still recover the alignment of the particles even when the SNR is so low that USP fails to

align each projection individually. This reduction of the SNR requirements directly affects the quality of the reconstruction. If less dose is tolerable in order to obtain a reconstruction, then, less radiation damage occurs to the specimen preserving their high frequency details, thus, CSP will allow to analyze smaller particles, more susceptible to radiation damage. Also, the analysis of non-symmetric particles should benefit by CSP since the extra tilted projections used in the refinement would help to overcome the ambiguities in the orientations, where USP usually fails with low doses.

The proposed framework shows a great flexibility in the configuration of how many projections are used, both for the refinement of the parameters and the reconstruction. This is shown with a new configuration denoted as $CSP_{N:1}$ which cannot be reproduced by any current method (Single Particle, Random Conical Tilt or Tomography), being also the one with better comparative results. This flexibility must be accompanied with a coherent protocol of fractionation of the dose in the acquisition of the data in order to build a tilt series of micrographs with the dose needed for the task to be used.

A new generation of CMOS devices to imaging the projections in a TEM are being developed and evaluated [JMK+08, MML+10, MCM+11]. These devices known as Direct Detector Devices (DDD) avoid the need for light conversion as required by the CCD devices currently used, and promise the possibility of single electron counting, or the ability to compensate the specimen movements or variations in the dose fractionation [GMFH11]. Therefore, even lower doses could be used in the imaging process with less radiation damage to the specimen. CSP has shown to be a powerful method with lower SNR images and could be well suited to this new digitization technology.

Hence, CSP cannot only efficiently process data acquired as a tomographic tilt series with a single particle approach, but also offers new scenarios of data collection and processing, introducing itself as an advanced merge of Cryo-Electron Tomography and Single Particle.

# Appendix A

# Euler Angles Convention

Determining the orientation of a rigid object in 3D can be done through the definition of three angles, the Euler angles $(\phi, \theta, \psi)$ associated with three rotations transforming the 3D Cartesian coordinate system to the object coordinate system. Each rotation is defined by an angle around an axis, being in total twelve possible valid combinations of the three Cartesian axes $(X, Y, Z)$. The convention used in this work is ZYZ, the same convention used by FRE-ALIGN [SG04] and SPIDER [FRP$^+$96], representing the orientation of the object with the following matrix $R$

$$R = R_Z(\psi)\, R_Y(\theta)\, R_Z(\phi)$$

where

$$R_Z(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad R_Y(\alpha) = \begin{pmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{pmatrix}.$$

Figure A.1 shows the concatenation of these rotations and the transformation between the initial $(X, Y, Z)$ and final $(X''', Y''', Z''')$ coordinate reference sys-
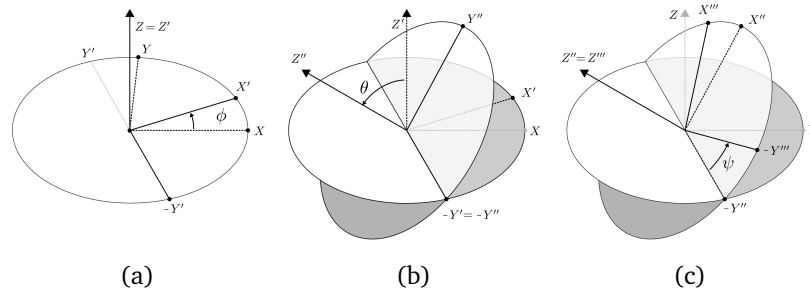


Figure A.1: Euler Angles convention used in this work. The object is (a) rotated by $\phi$ in the positive direction around $Z$, then (b) by $\theta$ in the negative direction around $Y'$, and finally (c) by $\psi$ in the positive direction around $Z''$.

tem. First, the object is rotated by $\phi$ in the positive direction around $Z$, then by $\theta$ in the negative direction around the new $Y$ ($Y'$), and finally by $\psi$ in the positive direction around the new $Z$ ($Z''$).

The angles $\{\phi, \theta\}$ completely determine the orientation of the object, and are called the *out-of-plane* rotation angles, in opposition to $\psi$, the *in-plane* rotation angle, which only defines a rotation around the orientation axis. For example, in Figure 4.4 each point in the unit sphere is completely determined by the out-of-plane angles $\{\phi, \theta\}$.

In the tomographic case, the geometry of the tilted micrographs is generated by setting $\psi = 0$, $\phi = \phi_i$ is the micrograph's tilt axis angle and $\theta = \theta_i$ is the micrograph's tilt angle. In the case of Random Conical Tilt (see Section 2.1) the geometry is defined by $\psi = 0$, $\theta = \theta_0$ (constant) and $\phi_i = 2\pi \left( {}^i/_N \right)$, being $N$ the number of acquired projections.

# Appendix B

# Variance of the Alignment Error

## B.1  Mean Rotation Matrix

Given a set of $N$ rotation matrices $\mathscr{R} = \{R_1, \ldots, R_N\}$ and its arithmetic mean matrix

$$M = \frac{1}{N} \sum_{k=1}^{N} R_k,$$

the mean rotation matrix could be computed as [Moa02]

$$\bar{R} = UV^T \tag{B.1}$$

where $U \Sigma V^T$ is the singular value decomposition of $M$. In [Moa02] the author shows that the matrix $\bar{R}$ is $M$'s projection into $SO(3)$ and is the solution to

$$\bar{R} = \arg\min_R \sum_{k=1}^{N} \|R_k - R\|_F.$$

## B.2  Measuring the alignment error

If the groundtruth orientation angles are known, it is possible to have a qualitative measure of the errors in the alignments. If the groundtruth's orientation angles and the ones obtained by CSP and USP are computed in the same references systems it is possible to directly compare them, but in our case, due to the reconstruction procedure, the obtained models, thus the alignment angles, have a global rotation avoiding this direct comparison.

Applying a 3D global rotation to the set of particle projections implies a rotation in the set of Euler angles for each particle projection. When there is no alignment error all the angles are affected by the same rotation. If the mean error in the alignments is zero, then averaging the orientations gives a good approximation to the global rotation. Being $R_G$ the rotation matrix

## Appendix B. Variance of the Alignment Error

associated with the original Euler angles, $R_C$ the rotation matrix associated with the rotated Euler angles and $R_\Theta$ the global 3D rotation, the following holds

$$R_C = R_G R_\Theta \quad \Rightarrow \quad R_\Theta = R_G^T R_C. \tag{B.2}$$

Therefore, if we have the groundtruth set of Euler angles represented by their rotation matrices $\mathscr{R}_G = \{R_G^1, \ldots, R_G^N\}$ and the corresponding set of Euler Angles estimated by CSP (or USP) represented as matrices, too, $\mathscr{R}_C = \{R_C^1, \ldots, R_C^N\}$, it is possible to estimate the rotation matrix between each correspondent pair of rotations using equation B.2,

$$R_C^k = R_G^k R_\Theta^k \quad \Rightarrow \quad R_\Theta^k = (R_G^k)^T R_C^k.$$

An estimation of the global rotation in the reconstructed density map can be computed using equation (B.1)

$$\bar{R}_C = U_C V_C^T \quad \text{where} \quad M_C = \frac{1}{N} \sum_{k=1}^{N} R_C^k = U_C \Sigma V_C^T.$$

The alignment error is computed averaging the difference in the Frobenius norm between each rotation matrix $R_C^k$ and $\bar{R}_C$

$$\varepsilon_C = \frac{1}{N} \sum_{k=1}^{N} \|R_C^k - \bar{R}_C\|_F. \tag{B.3}$$

# Appendix C

# (Simplified) Image Formation in a TEM and CTF correction

The process of image formation in the TEM is based in the interaction between the electron planar wave and the object, i.e., the incoming plane wave $\Psi_0$ is modified according to a phase shift proportional to the integral of the 3D Coulomb potential $C(r)$ of the object in the direction of propagation

$$\Psi(r) = \Psi_0 \exp\left(j\Phi(r)\right), \quad \text{where} \quad \Phi(r) = \int C(r,z)\mathrm{d}z.$$

Under some approximations, fulfilled in the data acquisition with a TEM, the wave function in the "back focal plane" $\Psi_{bf}(k)$ is the Fourier Transform of $\Psi(r)$. The wave is also affected by the lens aberrations and the defocusing, shifting the phase by an amount proportional to the "wave aberration function"

$$\chi(k,\phi) = \frac{1}{4}C_s\lambda^3 k^4 - \frac{1}{2}\Delta z \lambda k^2 - \frac{z_a}{4}\sin\left(2(\phi - \phi_0)\right)\lambda k^2,$$

where $C_s$ is the spherical aberration coefficient, $\lambda$ is the electron wavelength, $\Delta z$ is the defocus, and $z_a$ is the focal difference due to astigmatism.

The wave function in the image plane is the wave in the back focal plane affected by the aperture function $A(k)$. The aperture affects the contrast in the image, since, reducing the aperture only electrons deflected at lower angles are transferred and the contrast is enhanced. There is also an attenuation of the wave for higher spatial frequencies, modeled by another term $E(k)$ known as "compound envelope function". Finally, the intensity distribution in the image plane is

$$I(r) = |\psi(r)|^2 \quad \text{where} \quad \psi(r) = \mathscr{F}^{-1}\left\{\mathscr{F}\left\{\Psi_{bf}(r)\right\}A(k)E(k)\exp\left(j2\pi\chi(k)\right)\right\}.$$

Assuming that $\Phi(r)$ is real, the Weak-Phase Object Approximation ($\Phi(r) \ll 1$) is valid, and the axial astigmatism is fully compensated (i.e., $Z_a = 0$), the

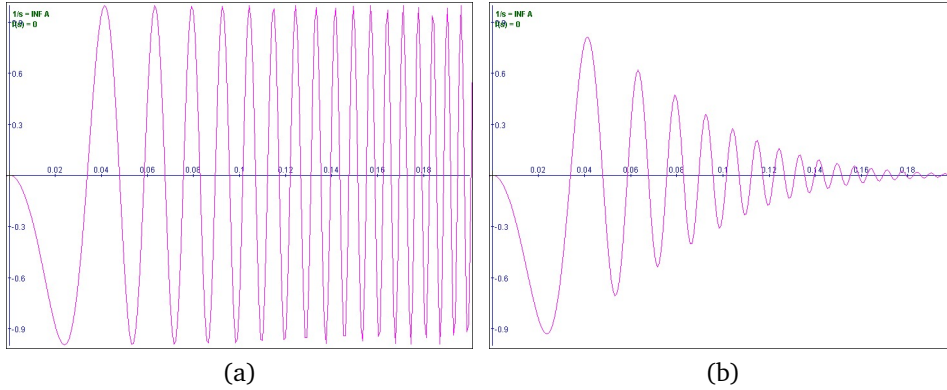# Appendix C. (Simplified) Image Formation in a TEM and CTF correction



Figure C.1: (a) A typical CTF profile for a TEM in EM. (b) Effect of the envelope function in the CTF shown in (a).

Fourier Transform of the image contrast $\mathscr{F}\{I(r)\}$ has a linear relationship with $O(k) = \mathscr{F}\{\Phi(r)\}$

$$\mathscr{F}\{I(r)\} = O(k)A(k)E(k)\sin\left(2\pi\chi(k)\right),$$

where $\sin\left(2\pi\chi(k)\right)$ is called the Contrast Transfer Function (CTF) [Fra06b].

The negative staining procedure to create contrasted images has some drawbacks, causing for example deformations in the specimens by the physical attachment to the staining substance. Thus, the unstained procedure by embedding the specimen in glucose or vitreous ice represents a major advantage in data collection. However, for unstained specimens with atomic masses under 400 kDa, the image formation procedure in a TEM needs to use relatively high defocuses in order to create a well contrasted image [Cro04, Fra06b]. Therefore, although the spherical aberration of the lenses could be reduced in the lenses preparation, the necessary defocus will add an oscillatory behavior to the CTF.

A typical profile of a CTF plot is shown in figure C.1a. The CTF starts close to zero, presenting a bandpass shape until the first zero crossing, then an oscillatory pattern attenuates the amplitudes and also adds a phase of $\pi$ to many of the frequencies. But the critical aspect of this behavior are the zero crossing frequencies, where all the information is suppressed.

There are many ways to deal with this effect. One way is to take more than one image of the specimen at different defocuses, then the zero crossings are at different frequencies and supplementary information is given by both images. Other approach only corrects the phase of the contrast, inverting the sign for the frequencies with phase $\pi$. The most complete correction involves the correction of the amplitudes and phases, generally via Wiener filtering to avoid the division by zero. See for example [MG03] and references therein.

# Appendix D

# Resolution measurement

Several measures to estimate the resolution of the obtained density map are presented in the literature, most of them based in the spectral components of the Fourier transform of the map. None of them has become a standard, however the FSC (see below) is widely used in the consulted references. Clearly, the evidence of structural features, such as secondary structure of proteins ($\alpha$-helices or $\beta$-sheets), gives the best idea of the order of the resolution obtained. For example, a resolution close to 5Å is needed to discriminate these structures [SG04]. But if the resolution is far from this no hints can be obtained directly from the density map.

The *resolution* of an imaging system is the minimum distance between two different points to be distinguishable (individually identified) in the formed image. Usually it is measured by the *angular resolution* being the smallest angle formed between the points and the image sensor, and transformed to a *spatial resolution* multiplying by the distance to the points. Even considering an imaging device with perfect lenses, the diffraction of the wave through the aperture limits the resolution of the system; when imperfect lenses are used their aberration also affects the limits of the resolution, being characterized by the Point Spread Function (PSF) or its Fourier transform, the Contrast Transfer Function (CTF).

As a first classification of the resolution measures, these can be separated into two groups depending if the statistical analysis is performed in the whole set of data or is splitted in two equally large *independent* subsets. The methods that split the data set in two halves, obtain two different reconstructions of the map and generate an estimation of the resolution comparing some features of the Fourier Transform of both maps. The main disadvantage of these methods is that the averaged maps are not completely independent since they are usually computed using the same model; this effect can be overcome if two different models are used or a reference free alignment is perform.

# D.1 Half-set methods

The Differential Phase Residual (DPR) and the Fourier Shell Correlation (FSC) are two representative methods considering two subsets of the data. Let us call $\mathscr{F}_1$ and $\mathscr{F}_2$ the Fourier Transform of the map computed from each subset. Both methods compute a frequency dependent factor evaluated in shells in Fourier space generating a curve for the range of working frequencies.

## D.1.1 Differential Phase Residual

The Differential Phase Residual (DPR) is defined for a shell of frequencies $S_{f,\delta} = (f - \delta, f + \delta)$ as

$$\text{DPR}(f, \delta) = \sqrt{\left| \frac{\sum_{|k| \in S_{f,\delta}} \Delta\phi(k)^2 \left( |\mathscr{F}_1(k)| + |\mathscr{F}_2(k)| \right)}{\sum_{|k| \in S_{f,\delta}} \left( |\mathscr{F}_1(k)| + |\mathscr{F}_2(k)| \right)} \right|} \qquad \text{(D.1)}$$

where $\Delta\phi(k)$ is the phase difference between $\mathscr{F}_1$ and $\mathscr{F}_2$ at frequency $k$.

The analysis of the curve of the DPR computed for the range of working frequencies is needed to characterize the consistency between the two maps. This curve starts from zero at low spatial frequencies and gradually grows, indicating inconsistencies at high frequencies between both maps. In order to have a single value for quantification of the resolution, the frequency $f_{45}$ where the DPR is 45° is selected as the representative resolution. The justification for this value comes from the argument that, if two sines waves are superimposed, they enforce each other if the phase differential is at least 45°, and if it is greater they interfere in a destructive way.

## D.1.2 Fourier Shell Correlation

The Fourier Shell Correlation (FSC) is also defined for a shell of frequencies $S_{f,\delta}$ as

$$\text{FSC}(f, \delta) = \frac{\text{Re} \left( \sum_{|k| \in S_{f,\delta}} \mathscr{F}_1(k) \mathscr{F}_2^{*}(k) \right)}{\sqrt{\sum_{|k| \in S_{f,\delta}} |\mathscr{F}_1(k)|^2 \sum_{|k| \in S_{f,\delta}} |\mathscr{F}_2(k)|^2}} \qquad \text{(D.2)}$$

The FSC curve starts from one at low spatial frequencies and gradually falls to zero. As in the case of the DPR, a single value of resolution is desired. In this case, at what value of the FSC curve takes the representative frequency is not so clear, or at least has various proposed values. Naming

$$\sigma = \frac{1}{\sqrt{N}} \qquad \text{(D.3)}$$

the estimation of the correlation for pure noise in a shell with $N$ points, the criterion of $2\sigma$, $3\sigma$ and $5\sigma$ where used by different authors (see the discussion at [Fra06b, chap.5, §8.2.1]). Also the 0.5 and 0.143 cutoffs were used. Comparing the reconstructions of the same structure, obtained with different techniques (X-ray crystallography and Cryo-EM) has been shown that the cutoff should be placed somewhere between 0.5 and $3\sigma$. Empirically the 0.5 cutoff has raised as a good compromise not too optimistic neither pessimistic and is quite consistent with the DPR=45° criterion [Fra09].

## D.2 Whole set methods

### D.2.1 Q-Factor

The Q-Factor is computed for the $j$-th component in the Fourier space as

$$Q(j) = \frac{\left|\sum_{i=1}^{M} \mathscr{F}_i(j)\right|}{\sum_{i=1}^{M} \left|\mathscr{F}_i(j)\right|} \tag{D.4}$$

where $\mathscr{F}_i(j)$ is the $j$-th Fourier component for the $i$-th image in the data set. Considering each Fourier component as a vector in the complex plane, the Q-Factor is the ratio between the length of the sum vector and the length of the path from the origin to the end of this vector following the directions given by each image component. Thus, if there is no noise and all the image components are correlated and aligned, the Q-Factor is one. In the case of pure uncorrelated noise, then the numerator should be $M$ times the noise mean and the Q-Factor is zero.

In the 3D case, each voxel of the Q-Factor is computed using Equation (D.4). Averaging the components in shells $S_{f,\delta}$, a curve is obtained for the range of working frequencies. In this case the cutoff with a frequency dependent critical Q-Factor curve is defined. This critical Q-Factor is a proportional to the correlation for pure noise given by Equation (D.3). Frank [PRF92] suggests using $3\sigma$ as the critical Q-Factor, Grigorieff [Gri98] uses $\sigma$.

Concerning the number of images, Grigorieff [Gri00] shows that for $M$ large, the Q-Factor is essentially independent of this number and depends only on the mutual correlation between the images.

### D.2.2 Spectral Signal to Noise Ratio

The Spectral Signal to Noise Ratio (SSNR) was introduced by Unser *et al.* [UTS87] and revisited by Grigorieff [Gri00] and Penczek [Pen02] in the 3D case. The SSNR is based on estimations of the SNR in consecutive shells in the Fourier space.

Let $X_i = S_i + N_i$ be the Fourier Transform of the $i$-th image in the data set, being $S_i$ and $N_i$ the FT of the signal and noise images respectively. An estimation of the energy of the signal present in a region R of the Fourier space can be obtained as

$$\hat{\sigma}_{Rs}^2 = \frac{\sum_R \left\| \hat{S} \right\|^2}{n_R} \quad \text{and} \quad \hat{S} = \frac{1}{N} \sum_{1=1}^{N} S_i$$

where $n_R$ is the number of Fourier components in $R$, and $\hat{S}$ is an estimation of the signal component obtained by averaging $N$ images. On the other hand, an estimation of the energy of the noise present in the same region is obtained as

$$\hat{\sigma}_{Rn}^2 = \frac{\sum_R \sum_{1=1}^{N} \left\| X_i - \hat{S} \right\|^2}{(N-1)n_R}$$

The authors show that the ratio $F_R = N\hat{\sigma}_{Rs}^2/\hat{\sigma}_{Rn}^2$ is a biased estimate of the SNR of the averaged image. Compensating this bias the SSNR is defined as

$$\text{SSNR} = \begin{cases} F_R - 1 & \text{if } F_R \geq 1 \\ 0 & \text{if } F_R < 1 \end{cases} \tag{D.5}$$

Taking the regions $R$ as consecutive shells a frequency dependent curve is calculated. The authors consider two thresholds $f_4$ and $f_0$ where the SSNR equals 4 and 0 respectively. The first threshold defines an empirical "minimum acceptable threshold" considered adequate for "unambiguous visual interpretation". The second threshold defines the region where no signal can be detected.

Compared with the DPR and the FSC, the SSNR is statistically more precise, being similar only if all the possible half-set partitions are considered. It is also shown that a DPR $= 45°$ corresponds to a SSNR $\approx 3$, being an optimistic estimation.

## D.2.3   RMEASURE: Fourier Neighbor Correlation

Following the analysis [Gri00, SG04] of partial noise alignment to the references in the refinement process (model bias) in the resolution estimation using FSC, Sousa and Grigorieff [SG07] propose to compute the Fourier Neighbor Correlation (FNC) based in correlation between neighboring Fourier terms.

The FNC idea is based on the observation that the Fourier terms of a volume filled with zero mean random noise are uncorrelated, but if a 3D binary mask is applied, the convolution introduces correlation between the Fourier terms. This correlation depends on the Fourier Transform of the mask, $G_{hp}$ following the notation in [SG07].

FNC is defined for a thin shell $S_R$ at resolution $R$ and a neighborhood $V_p$, consisting of the six Fourier terms connected with $p$, as

$$\text{FNC}(R) = \frac{\sum_{p \in S_R} \sum_{q \in V_p} \mathscr{F}_p \mathscr{F}_q{}^*}{\sqrt{\sum_{p \in S_R} \sum_{q \in V_p} |\mathscr{F}_p|^2 \sum_{p \in S_R} \sum_{q \in V_p} |\mathscr{F}_q|^2}}. \tag{D.6}$$

If only noise is present in the volume the FNC is approximately constant through the frequencies but with a large variance in the lower frequencies where the number of terms in the shells is reduced. The constant depends on the mask $G_{hp}$. When a macromolecular structure is added in the volume, an additional density (another source of noise) affects the computation of the FNC. In this case the term $\mathscr{F}_p$ is replaced by $S_p + N_p$ reflecting the two "signal" sources, where $S_p$ is the contribution from the macromolecule and $N_p$ the background noise; similarly with $\mathscr{F}_q$. Manipulation from Equation (D.6) gives

$$\text{FNC}(R) = \frac{FNC_F \, \text{SNR}(R) + FNC_N}{\text{SNR}(R) + 1} \quad \text{or} \quad \text{SNR}(R) = \frac{FNC_N - \text{FNC}(R)}{\text{FNC}(R) - FNC_F}, \tag{D.7}$$

where $FNC_F$ and $FNC_C$ are the constants describing the FNC arising from the structure and the background noise, respectively. These constant values are estimated masking the macromolecular structure and computing the FNC when filling the mask with noise. Any density outside the mask counts as noise and any density inside the mask counts as signal *plus* noise. To obtain the structure mask a low-pass filtered version of the volume is computed retaining sufficient information about the shape of the structure, but without finer and noisier details.

Equation (D.7) is valid when the FNC of the signals is approximately constant, but at very low resolutions the FNC value depends strongly on the details in the Fourier space of the molecular envelope. Following this result and the relation between the SNR and the FSC [Gri00] the FSC($R$) could be compared with the *real* FSC computed splitting the data set is halves.

The FNC measure seems to be less affected by model bias or over-fitting to the noise in the refinement compared to the FSC measure.

# References

[BDA+07]  Daniel R Beniac, Shauna L DeVarennes, Anton Andonov, Runtao He, and Tim F Booth. Conformational reorganization of the sars coronavirus spike following receptor binding: Implications for membrane fusion. *PLoS ONE*, 2(10):7, 2007.

[Ber99]  Dimitri Bertsekas. *Nonlinear Programming (scanned)*. Athena Scientific, 1999.

[BS09]  Alberto Bartesaghi and Sriram Subramaniam. Membrane protein structure determination using cryo-electron tomography and 3d image averaging. *Current Opinion in Structural Biology*, 19(4):402 – 407, 2009.

[BSL+08]  A. Bartesaghi, P. Sprechmann, J. Liu, G. Randall, G. Sapiro, and S. Subramaniam. Classification and 3d averaging with missing wedge correction in biological electron tomography. *Journal of Structural Biology*, 162(3):436 – 450, 2008.

[BSZM04]  Mario J Borgnia, Dan Shi, Peijun Zhang, and Jacqueline L S Milne. Visualization of alpha-helical features in a density map constructed using 9 molecular images of the 1.8 mda icosahedral core of pyruvate dehydrogenase. *Journal of Structural Biology*, 147(2):136–145, 2004.

[CHS96]  R A Crowther, R Henderson, and J M Smith. Mrc image processing programs. *Journal of Structural Biology*, 116(1):9–16, 1996.

[Cro04]  R A Crowther. Viruses and the development of quantitative biological electron microscopy. *IUBMB Life*, 56(5):239–248, 2004.

[Fau93]  Olivier Faugeras. *Three-Dimensional Computer Vision : a geometric viewpoint*. Artificial Intelligence. MIT Press, Cambrigde Mass., 1993.

[FLC06]  J J Fernández, S Li, and R A Crowther. Ctf determination and correction in electron cryotomography. *Ultramicroscopy*, 106(7):587–596, 2006.

References

[Fra06a]     Joachim Frank. *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*. Springer, 2 edition, 2006.

[Fra06b]     Joachim Frank. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford University Press US, 2 edition, 2006.

[Fra09]      Joachim Frank. Single-particle reconstruction of biological macromolecules in electron microscopy - 30 years. *Quarterly Reviews of Biophysics*, pages 1–20, 2009.

[FRP+96]     Joachim Frank, Michael Radermacher, Pawel Penczek, Jun Zhu, Yanhong Li, Mahieddine Ladjadj, and Ardean Leith. Spider and web: Processing and visualization of images in 3d electron microscopy and related fields. *Journal of Structural Biology*, 116(1):190 – 199, 1996.

[GMFH11]     R M Glaeser, G McMullan, A R Faruqi, and R Henderson. Images of paraffin monolayer crystals with perfect contrast: minimization of beam-induced specimen motion. *Ultramicroscopy*, 111(2):90–100, 2011.

[Gri98]      N. Grigorieff. Three-dimensional structure of bovine nadh:ubiquinone oxidoreductase (complex i) at 22 a in ice. *J Mol Biol*, 277(5):1033–46, Apr 17 1998.

[Gri00]      N. Grigorieff. Resolution measurement in structures derived from single particles. *Acta Crystallogr D Biol Crystallogr*, 56(Pt 10):1270–7, Oct 2000.

[Gri07]      Nikolaus Grigorieff. Frealign: High-resolution refinement of single particle structures. *Journal of Structural Biology*, 157(1):117 – 125, 2007. Software tools for macromolecular microscopy.

[Hor11]      Arthur L Horwich. Protein folding in the cell: an inside story. *Nature Medicine*, 17(10):1211–1216, 2011.

[HZ04]       R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[JMK+08]     Liang Jin, Anna-Clare Milazzo, Stuart Kleinfelder, Shengdong Li, Philippe Leblanc, Fred Duttweiler, James C Bouwer, Steven T Peltier, Mark H Ellisman, and Nguyen-Huu Xuong. Applications of direct detection device in transmission electron microscopy. *Journal of Structural Biology*, 161(3):352–358, 2008.

[JSB08]      S Jonić, C O S Sorzano, and N Boisset.   Comparison of single-particle analysis and electron tomography approaches: an overview. *Journal of Microscopy*, 232(3):562–579, 2008.

[KLP11]      Philip D. Kiser, George H. Lorimer, and Krzysztof Palczewski. Crystal structure of wild-type groel with bound thallium ions. `http://www.pdb.org/pdb/explore/explore.do? structureId=3e76`, November 2011.

[LBB+08]     Jun Liu, Alberto Bartesaghi, Mario J. Borgnia, Guillermo Sapiro, and Sriram Subramaniam. Molecular architecture of native hiv-1 gp120 trimers. *Nature*, 455(7209):109–113, Sep 2008.

[LBC99]      S. J. Ludtke, P. R. Baldwin, and W. Chiu. Eman: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol*, 128(1):82–97, Dec 1999.

[LBC+08]     Steven J Ludtke, Matthew L Baker, Dong-Hua Chen, Jiu-Li Song, David T Chuang, and Wah Chiu.  De novo backbone trace of groel from single particle electron cryomicroscopy. *Structure*, 16(3):441–448, Mar 2008.

[LBW+10]     Xin Liu, David A Bushnell, Dong Wang, Guillermo Calero, and Roger D Kornberg.  Structure of an rna polymerase ii-tfiib complex and the transcription initiation mechanism. *Science*, 327(5962):206–209, 2010.

[LBW+11]     Xin Liu, David A. Bushnell, Dong Wang, Guillermo Calero, and Roger D. Kornberg.  Structure of an rna polymerase ii-tfiib complex and the transcription initiation mechanism. `http://www.pdb.org/pdb/explore/explore.do? structureId=3k7a`, November 2011.

[MCM+11]     Anna-Clare Milazzo, Anchi Cheng, Arne Moeller, Dmitry Lyumkis, Erica Jacovetty, James Polukas, Mark H Ellisman, Nguyen-Huu Xuong, Bridget Carragher, and Clinton S Potter. Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. *J Struct Biol*, 176(3):404–408, Dec 2011.

[MG03]       Joseph A Mindell and Nikolaus Grigorieff. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol*, 142(3):334–347, Jun 2003.

[MML+10]     Anna-Clare Milazzo, Grigore Moldovan, Jason Lanman, Liang Jin, James C Bouwer, Stuart Klienfelder, Steven T Peltier, Mark H

## References

Ellisman, Angus I Kirkland, and Nguyen-Huu Xuong. Characterization of a direct detection device imaging camera for transmission electron microscopy. *Ultramicroscopy*, 110(7):744–747, 2010.

[Moa02]    Maher Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24:1–16, 2002.

[Pen02]    Pawel A Penczek. Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. *Journal of Structural Biology*, 138:34–46, 2002.

[PRF92]    P. Penczek, M. Radermacher, and J. Frank. Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, 40(1):33–53, Jan 1992.

[Rad88]    M Radermacher. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *Journal of Electron Microscopy Technique*, 9(4):359–394, 1988.

[SBL+07]   Sriram Subramaniam, Alberto Bartesaghi, Jun Liu, Adam E Bennett, and Rachid Sougrat. Electron tomography of viruses. *Current Opinion in Structural Biology*, 17(5):596 – 602, 2007.

[SC09]     Sjors H W Scheres and José María Carazo. Introducing robustness to maximum-likelihood refinement of electron-microscopy data. *Acta Crystallographica Section D Biological Crystallography*, 65(Pt 7):672–678, 2009.

[SFF+03]   Daniel Stoffler, Bernhard Feja, Birthe Fahrenkrog, Jochen Walz, Dieter Typke, and Ueli Aebi. Cryo-electron tomography provides novel insights into nuclear pore architecture: Implications for nucleocytoplasmic transport. *Journal of Molecular Biology*, 328(1):119–130, 2003.

[SG04]     Alex Stewart and Nikolaus Grigorieff. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy*, 102(1):67–84, Dec 2004.

[SG07]     Duncan Sousa and Nikolaus Grigorieff. Ab initio resolution measurement for single particle structures. *J Struct Biol*, 157(1):201–210, Jan 2007.

[Sig98]    F. J. Sigworth. A maximum-likelihood approach to single-particle image refinement. *Journal of Structural Biology*, 122(3):328 – 339, 1998.

[SMVC09]  Sjors H W Scheres, Roberto Melero, Mikel Valle, and Jose-Maria Carazo. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. *Structure*, 17(12):1563–1572, Dec 2009.

[Sub05]  Sriram Subramaniam. Bridging the imaging gap: visualizing subcellular architecture with electron tomography. *Current Opinion in Microbiology*, 8(3):316 – 322, 2005.

[UTS87]  M. Unser, B.L. Trus, and A.C. Steven. A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy*, 23(1):39–51, 1987.

[vH87]  Marin van Heel. Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2):111–123, 1987.

[vHGM⁺00]  Marin van Heel, Brent Gowen, Rishi Matadeen, Elena V. Orlova, Robert Finn, Tillmann Pape, Dana Cohen, Holger Stark, Ralf Schmidt, Michael Schatz, and Ardan Patwardhan. Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics*, 33(04):307–369, 2000.

[vHHO⁺96]  M van Heel, G Harauz, E V Orlova, R Schmidt, and M Schatz. A new generation of the imagic image processing system. *Journal of Structural Biology*, 116(1):17–24, 1996.

[WTN⁺97]  J Walz, D Typke, M Nitsch, Aj Koster, R Hegerl, and W Baumeister. Electron tomography of single ice-embedded macromolecules: Three-dimensional alignment and classification. *Journal of Structural Biology*, 120(3):387–95, 1997.

# Part II

# Simultaneous Object Classification and Segmentation with High-Order Multiple Shape Models

# Chapter 5

# Object segmentation

Object segmentation is one of the most fundamental tasks in image processing, still lacking a completely automatic solution. The main idea is to find a set of features that describes and discriminates the object of interest from the rest of the image. Object color is a low level feature that can be used as such descriptor, although its discrimination capacity is often insufficient in real images. Using shape as a high level feature is a common approach to augment such low level features.

The shape of the desired object is added as a descriptor, constraining the set of possible solutions to regions of the image that simultaneously "match" this shape and the low level features (color, edges, etc.). The most common way to add this shape information is in the form of a weighted linear combination of functionals addressing, on one hand, the low level features and, on the other hand, the shape priors or models. This leads to a minimization problem where the solution is a compromise between the shape of the final contour and the information given by the image. The minimization techniques used in the literature, include, among others, gradient descent methods [LGF00, COS06, TYWW+01, CFK07, CSS06] and graph-cuts [VM08]. The used shape representations can be signed distance functions (SDF) [LGF00, COS06, TYWW+01, VM08, CSS06, RJZ06], quadratic splines [CKS03], characteristic functions [CFK07], and landmark points [CTCG95].

When $M$ different objects (or object classes) can appear in an image, a single shape prior (model) is not sufficient, and multiple shape priors must be considered. A possible, but not elegant, approach is to run the process with each one of the shape priors separately, and then choose the best solution. Vu and Manjunath [VM08] and Cremers *et al.* [CSS06] define $M$ possible labels for each pixel on the image, and then propose a segmentation energy that includes the optimization of these labels in order to determine where to apply each prior. In a different work, Cremers *et al.* [CKS03] perform density estimation in a non-linear feature space, where different objects are separable. The proposed energy is then minimized considering both the curve's control

points and the image.

Considering the natural deformations and the variability of objects within a class, high-order shape models (SMs) should be included in the segmentation. Leventon *et al.* [LGF00] compute PCA on a set of registered shapes, fitting a Gaussian probability distribution to the coefficients of the reconstruction. This allows to include the probability of a certain shape, in traditional geodesic active contours for low level features, and a MAP estimation of the object in the image. Tsai *et al.* [TYWW$^+$01] also use PCA to model shape variations, defining an energy for the aligning of the binary shape, and formulate a segmentation functional optimizing the parameters of the representation with the first deformation modes. Cootes and Taylor [CTCG95] and Kim and Lee [KL05], among other authors, compute, using PCA, a point distribution model of landmarks points defining a shape, based on the active shape model framework. More recently, Charpiat *et al.* [CFK07] proposed a framework to compute non-linear shape statistics based on the Hausdorff distance between shapes, and then model distributions similarly to [LGF00].

In this work, a new framework for image segmentation with multiple high-order shape models is introduced, addressing at the same time the selection of the model and its image-driven positioning and adjustment to the modeled deformations. Invariance is included as part of the framework as well. In particular, the high-order SMs are computed using PCA in a similar way as [LGF00, TYWW$^+$01], obtaining a set of eigenmodes of variations. In the case of dynamic shapes with large, non-linear deformations, a method to obtain a lineal approximation of the shape space is described using a dimensionality reduction algorithm. The selection of the model is obtained with a binary selection coefficient, on-line learned based on a similarity measure between shapes. The proposed framework follows from a functional that combines two terms. The first one is a region-based segmentation term [CV01]. The second term is a combination of the multiple high-order SMs, addressing the model selection and constraining the solution to the high-order shape information coming from the on-line selected model. While the framework is presented for planar curves, it can be easily extended to data in higher dimensions.

The remainder of this part is organized as following. Chapter 6 reviews briefly the definition and properties of shapes models. Chapter 7 describes the proposed framework. Chapter 8 presents experiments testing the ideas and theirs discussion. Chapter 9 proposes an invariance to translation extension of the framework. Finally, Chapter 10 concludes this part.

# Chapter 6

# High-order multiple shapes models

Consider $M$ sets $\Phi_k$, $k = 1,\ldots,M$, each with $N_k$ registered shapes $\Phi_k = \{\phi_k^1,\ldots,\phi_k^{N_k}\}$, where each $\phi_k^j$ is a signed distance function (SDF), whose zero level-set curve, $\mathscr{C}_k^j$, represents a shape from the $k$-th class of objects. Let $\mathscr{M}_k^d$ be a $d$-order model that captures the intrinsic deformations of the training set $\Phi_k$ for the class $k$. In this work, $\mathscr{M}_k^d$ is derived from a PCA decomposition of the training set $\Phi_k$ (all the shapes $\phi_k$ are represented as vectors in $\mathbb{R}^D$, $D$ being the size of the range of the corresponding SDFs),

$$\mathscr{M}_k^d := \{\mu_k, U_k^d\}, \tag{6.1}$$

where $\mu_k \in \mathbb{R}^D$ is the mean shape of $\Phi_k$, $U_k^d \in \mathbb{R}^{D \times d}$ is a matrix containing the first $d$ modes of variation (eigenmodes), $U_k^d = [\{u_k^i\}_{i=1}^d], u_k^i \in \mathbb{R}^D$.

A model $\mathscr{M}_k^d$ generates a representation of a new incoming shape $\hat{\phi}$ by the $d$-projection $\mathscr{P}_k^d \hat{\phi}$

$$\mathscr{P}_k^d \hat{\phi} = \mu_k + U_k^d \alpha_k, \tag{6.2}$$

where $\alpha_k = (U_k^d)^T (\hat{\phi} - \mu_k) \in \mathbb{R}^d$ are the corresponding reconstruction coefficients, computed directly, in closed form, from the shape $\hat{\phi}$ given the model $\mathscr{M}_k^d$ (see for example [LGF00, section 2.1] for details).

The accuracy of the representation depends on the similarity between $\hat{\phi}$ and the shapes in $\Phi_k$. Constraining small shape variations in the class $\Phi_k$ (compared with the deformations across different classes $k$) allows to obtain accurate representations using a linear approximation like PCA.

Finally, let

$$\mathfrak{M} = \{\mathscr{M}_1,\ldots,\mathscr{M}_M\}$$

be a set of SMs for the $M$ different classes of objects. For simplicity, the order $d$ in the notation is omitted from now on.

Figure 6.1 shows SMs for a walking person. Figure 6.1a shows four different shapes from one of the classes of shapes, note the similarity between
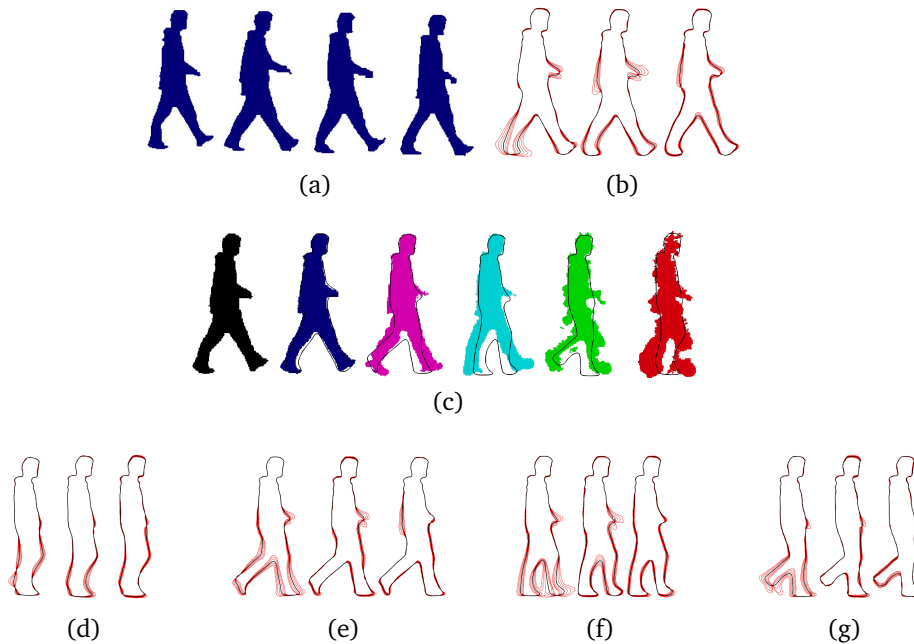
Figure 6.1: (a) Four shapes from one of the classes, and (b) the first three modes of variation of the corresponding model in the walking sequence. The thick black line is the mean shape, the red lines are obtained varying the amplitude (see text). (c) Original shape (in black) and its projections $\mathscr{P}_k \phi$ (colored as in figure 6.2) to $M = 5$ different models in the walking sequence, one for each cluster. The mean shape of the corresponding model is plotted too (black curve). The projections are ordered based in the measure given by Equation (7.6). Note how the projection is completely deformed when using the wrong shape model. (d) - (g) First three modes of variation for four different shape models.

them. Figure 6.1b shows the first three modes of variation of the corresponding model in a walking sequence. The data was obtained filming a single person walking with a static background [ARS07]. The thick black line is the zero level set of the mean shape. The red lines are the zero level sets of the addition of the mean shape and a constant times the first, second, or third eigenmode respectively, varying the amplitude. Figure 6.1c shows an original shape from the set and its projections (with $d = 21$) to $M = 5$ different models in a walking sequence. The mean shape of the corresponding model is plotted too (black curve). The projections are ordered based in the measure given later by Equation (7.6). Note how the projection is completely deformed when using the wrong model, clearly illustrating the importance of selecting the correct shape model (prior). Figures 6.1d, 6.1e, 6.1f and 6.1g show the first three modes of variation of the other four models obtained from the same walking sequence. The procedure to obtain these models is explained in the next section.

|(a)| | |(b)|
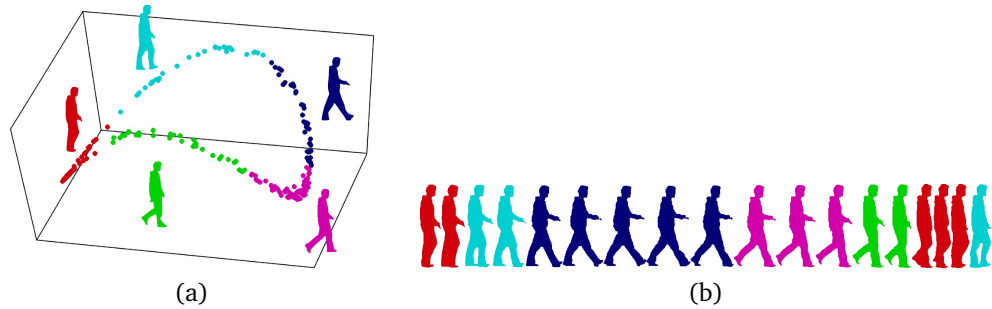
Figure 6.2: (a) Low dimensional embedding. Each point correspond to the first three coordinates of the mapping obtained with DM. The colors correspond to the $M = 5$ obtained cluster. One sample shape from each cluster (walking-cycle position) is shown. (b) Eighteen samples from the walking sequence colored based on the obtained clusters.

## 6.1 Clustering a set of shapes

One of the used datasets of shapes was taken from the sequence of a walking person [ARS07]. Considered as a unique deformable object, this shape has large, non-linear deformations, invalidating the hypothesis of small (and linear) shape variations for this set. To alleviate this, a set of clusters may be considered. In this way linear approximations can be used to approximate shape deformations within each cluster to obtain $\mathcal{M}_k$. In order to obtain the clusters, in this work a non-linear mapping to an Euclidean space is performed based on Diffusion Maps (DM) [CL06]. DM is a general framework for data analysis based on a diffusion process over an undirected weighted graph, defining a new metric on the data called Diffusion Distance. Two properties of this metric are important in the present work. First, as a consequence of the density renormalized kernel defined to build the graph, the graph-Laplacian (see Von-Luxburg's tutorial [vL07] for definition and properties of the graph Laplacian) is an approximation of the Laplace-Beltrami operator on the underlying manifold, allowing to recover the Riemannian geometry of the data set regardless the distribution of the points in the underlying manifold. Second, the Diffusion Distance is equivalent to the Euclidean distance in the space with coordinates given by the mapping function. This allows to simply compute K-means in the corresponding Euclidean space in order to group the shapes into $M$ clusters and then obtain a local model in each cluster.

To recapitulate, the clusters are obtained by mapping into a new space via DM (a kernel method) and then applying K-means on this space. Note that the subsequent PCA could actually be performed in this space as well (using Kernel PCA [SS02, DRT06]), though the clustering makes the inner class variations already well approximated by ordinary PCA.

Figure 6.2 shows the clustering result. Figure 6.2a shows the low dimensional embedding manifold. Each point correspond to the first three coordi-

nates of the mapping colored based on the obtained clusters. One sample shape from each cluster (walking-cycle position) is shown too. Figure 6.2b shows eighteen consecutive samples from the walking sequence colored based on the obtained clusters.

# Chapter 7

# Proposed variational framework

Given an input image $\mathscr{I} : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ containing a shape generated by one of the shape models $\mathscr{M}_k \in \mathfrak{M}$, an energy $E$ is defined to simultaneously select the best model and obtain a segmentation of the corresponding object in $\mathscr{I}$ (we will discuss the issue of multiple objects later),

$$(\mathscr{M}^*, \phi^*) := \arg \min_{\phi, \mathscr{M} \in \mathfrak{M}} E(\mathscr{I}, \phi, \mathfrak{M}). \qquad (7.1)$$

This energy includes two terms linearly combined with the constant $\lambda$,

$$E(\mathscr{I}, \phi \mathfrak{M}) = E_{\text{IM}}(\mathscr{I}, \phi) + \lambda E_{\text{SM}}(\phi, \mathfrak{M}). \qquad (7.2)$$

The $E_{\text{IM}}$ term is related to any low level image descriptor such as gray level, color, edges, texture, etc. For the examples in this work, $E_{\text{IM}} = E_{\text{CV}}$ is the energy introduced by Chan and Vese [CV01],

$$
\begin{aligned}
E_{\text{CV}}(\mathscr{I}, \phi, c_+, c_-) \;\; = \;\; & \int_\Omega |\mathscr{I}(x) - c_+|^2 H(\phi(x)) \mathrm{d}x + \\
& \int_\Omega |\mathscr{I}(x) - c_-|^2 (1 - H(\phi(x))) \mathrm{d}x + \\
& \mu \int_\Omega \delta(\phi(x)) |\nabla \phi(x)| \mathrm{d}x, \qquad (7.3)
\end{aligned}
$$

where $c_+$ and $c_-$ are the averages of the input data inside and outside the curve $\mathscr{C}$ (the zero level set of $\phi$), respectively, $H(\cdot)$ is the Heaviside function, and $\delta(\cdot)$ is the Dirac function. This energy attempts to split the input data into two different regions of approximately piecewise constant color or gray level values ($c_+$ and $c_-$). Other low level descriptors could be used for a better discrimination, for example texture [PD99] or edges [SKC95].

The term $E_{\text{SM}}$ adds an additional force aiming at maximizing the similarity between the evolving shape $\phi$ and its projection onto *only one* of the $d$-order

models from $\mathfrak{M}$. Which one of the $M$ models is used depends on the evolution of the shape and its projection to each model. The proposed term is

$$E_{\text{SM}}(\phi, \mathfrak{M}) = \sum_{k=1}^{M} \beta_k E_k(\phi, \mathcal{M}_k), \tag{7.4}$$

defining

$$E_k(\phi, \mathcal{M}_k) = \int_{\Omega} \left\| H(\phi(p)) - H(\mathscr{P}_k \phi(p)) \right\|^2 dp, \tag{7.5}$$

where again $H(\cdot)$ is the Heaviside function, $\beta_k$ is a binary coefficient that (on-line) selects which of the $M$ models is used, and $\mathscr{P}_k \phi$ is the projection of $\phi$ onto the model $\mathcal{M}_k$, given by Equation (6.2). Only one of the $\beta_k$ must be different from zero in (7.4), since it is not fair to penalize for models that do not correspond to the object in the image. This is detailed next.

## 7.1 Shape dissimilarity measure and model selection

Which is the non-zero $\beta_k$ in Equation (7.4) is computed based on a shape dissimilarity measure ($\Upsilon$) between two shapes $\phi_1$ and $\phi_2$,

$$\Upsilon(\phi_1, \phi_2) = \int_{\Omega} \frac{|\phi_1(p)| \delta(\phi_2(p))}{\text{length}(\mathscr{C}_2)} dp + \int_{\Omega} \frac{|\phi_2(p)| \delta(\phi_1(p))}{\text{length}(\mathscr{C}_1)} dp. \tag{7.6}$$

This is a length-normalized variation of the measure introduced by Funkhouser *et al.* [FKS$^+$04]. This measure evaluates the sum of Euclidean distances corresponding to moving the contour of the first shape to points in the contour of the second shape, and viceversa, scaled by the curves lengths. In figure 6.1c, the projected shapes are ordered according to increasing values of $\Upsilon(\phi, \mathscr{P}_k \phi)$. These ordered values are 1.35, 2.83, 3.59, 5.87, and 7.83 respectively.

Based on (7.6), a normalized shape similarity measure $\bar{\xi}_k(\phi)$ between a shape $\phi$ and its projection $\mathscr{P}_k \phi$ to the $d$-order $k$-th model is computed as

$$\bar{\xi}_k(\phi) = \frac{\exp\left(-\Upsilon(\phi, \mathscr{P}_k \phi)\right)}{\sum_{l=1}^{M} \exp\left(-\Upsilon(\phi, \mathscr{P}_l \phi)\right)}. \tag{7.7}$$

This normalized similarity measure $\bar{\xi}_k(\phi)$ is close to one for the model that better represents $\phi$. Finally to force the binary value in $\beta_k$, soft thresholding, based on a sigmoid function, is performed. Note that a unique coefficient is used as model selector, instead of one coefficient in each pixel as in [CSS06, VM08]. This encourages shape consistency and significantly simplifies the optimization.

This procedure can be viewed as a case of an analysis-by-synthesis framework for model selection, similar to the one introduced by Srivastava and Jermyn [SJ09]. Instead of comparing the current curve with high-probability sampled contours from the models, in our case, its best representation, given by its projection to the model, is used for comparison.

In the particular case where two models represent objects with too similar features, the proposed model selection framework might fail if both models obtain high $\beta_k$s (a human observer might fail as well in such scenario). This could be resolved adding a "multiple hypothesis" procedure, selecting these two models and first evolving two curves "in parallel" under different model hypothesis, and then confirming one of these hypothesis with the dissimilarity measure or the validation procedure proposed in Equation (7.8) in the next section.

If the typical features of the object are defined and captured with the curve, the model selection procedure does present a stable behavior, also allowing model switching to recover from errors in the selection. To achieve this, the $E_{\mathrm{CV}}$ term plays an important role. An example of model switching and recovering from a potential error is shown in Chapter 8 (Figure 8.1).

## 7.2 Shape validation

With the proposed method, one model is always selected and a segmentation is obtained, even if the shape in the image has no appropriate model in $\mathfrak{M}$ that provides a good representation. The validation of the final segmentation cannot be directly compared to the original non-occluded shape in all the cases, since there is no way to "create" the particular features or attributes that are occluded in it. Instead of this, the resulting segmentation is evaluated taking into account the fact that the shape is generated by a model $\mathcal{M}_k$ in $\mathfrak{M}$ and the solution should then be a "valid" shape generated by this model. The following measure permits to discard a segmentation $\tilde{\phi}$ given the selected model. First, the mean $\bar{\Upsilon}_k$ and variance $\sigma^2_{\Upsilon_k}$ of $\Upsilon(\phi_k^j, \mu_k)$ are computed $\forall \phi_k^j \in \Phi_k$. If

$$\Upsilon(\tilde{\phi}, \mu_k) > \bar{\Upsilon}_k + \tau \sigma^2_{\Upsilon_k}, \tag{7.8}$$

then the segmentation is discarded, and the shape can not be recognized. Empirically it was found that a value of $\tau = 1.5$ is useful for the acceptance of shapes from the correspondent model and the rejection of shapes from other models. Moreover, a model-dependent $\tau_k$ could be used, where the particular variations of the set of shapes are considered.

## 7.3 Energy minimization

The proposed energy is minimized using a classical gradient descent method. The first variation of Equation (7.2) becomes

$$\frac{\partial E}{\partial \phi} = \frac{\partial E_{\text{CV}}}{\partial \phi} + \lambda \frac{\partial E_{\text{SM}}}{\partial \phi}. \tag{7.9}$$

For the gradient descent of $E_{\text{CV}}$, the expression is given in [CV01, Equation (9)]

$$\frac{\partial E_{\text{CV}}}{\partial \phi} = \delta(\phi) \left( (\mathscr{I} - c_+)^2 - (\mathscr{I} - c_-)^2 - \mu \nabla \cdot \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right).$$

For $E_{\text{SM}}$, defining $W = U_k U_k^T$, the obtained expression is

$$\frac{\partial E_{\text{SM}}}{\partial \phi} = -2 \sum_{k=1}^{M} \beta_k \left\| H(\phi) - H(\mathscr{P}_k \phi) \right\| \left( \delta(\phi) - \delta(\mathscr{P}_k \phi) W \right).$$

Although the model selector $\beta_k$ depends on $\phi$, is treated as static, as a first order approximation for the gradient descent, since it affects the model selection and only indirectly the evolution of the curve. This model selector coefficient is computed/updated at every iteration $r$ of the gradient descent, i.e., given a new updated curve $\mathscr{C}^{(r)}$ at iteration $r$, its projection to the models in $\mathfrak{M}$ are computed and the binary selector $\beta_k^{(r+1)}$ is obtained selecting the model for iteration $(r+1)$.

## 7.4 Prior activation

The first steps of the optimization are performed without SMs information ($\lambda = 0$), until stationarity, then the "prior is activated" adding $E_{\text{SM}}$ with $\lambda \neq 0$ (manually selected) until a new stationary point is reached[1], now combining the image and the shape information. This helps to determine the object in the image without affecting the initial steps of the evolution with the projections to the models of the initial curve used in the minimization, which in the general has no similarity with the shapes in the models. A similar idea of "prior activation" is considered by Vu and Manjunath [VM08] using shape prior templates instead of SMs.

The Gestalt Principles [Kan79] can give some intuition to this initial step. The "Principle of Similarity" states that people try to organize visual elements into groups based on the similarity of certain features (shape, intensity, texture, etc.). This gives an additional argument for trying to start grouping

---

[1]Stationary points are determined taking into account the variation of the area of the evolving curve $\phi$, i.e., if there is no substantial variation in this area the iterations are stopped. Also note that $\lambda$ does not affect the first stationary point, previous to the "prior activation."

regions of similar intensity and use the results as an initial point or "primary units" for helping the minimization process. After the identification of these "primary units," the addition of priors is used for a better interpretation of the object or scene.

The framework, as is presented, obtains the segmentation of only one object (at a time) in the image. If more than one object is present in an image, the segmentation will depend on the initial curve that roughly selects the object and hence the segmentation at the "prior activation" step. This will be exemplified in Figure 9.1b, where five different objects are present and always only one is selected and segmented, even if the initial curve contains more than one object in the image.

In order to be able to segment an image with more than one object, a possible solution, as a direct extension of the proposed framework, is to consider a multiphase level set framework [VC02], where each phase has its own model selector. Alternatively an analysis of the connected components of $\phi$ at the "prior activation" step could be performed, and then use the multiphase level set framework from this step. Lastly, we could also consider multiple steps, and once an object has been detected and segmented, its region is eliminated/ignored and the rest of the image is further studied for additional objects.

Finally, the selection of $\lambda$ and its effects on the balance of the energy terms is, as in a general framework of active contours, important in determining the final segmentation. A detailed discussion of this point is presented in Sections 8.1 and 8.2.

# Chapter 8

# Experimental results

## 8.1   Models of ellipses

The first example is a "toy example," though illustrative and challenging, with two models of ellipses, where the only difference is that the first (and only beyond the mean shape) eigenmode is rotated $\frac{\pi}{2}$ (this already exemplifies the importance of high-order models). Let us name $\mathcal{M}_{\mathrm{V}}^1$ the model with vertical deformations and $\mathcal{M}_{\mathrm{H}}^1$ the model with horizontal deformations. Figure 8.1a shows the mode of variation for both models, in green for $\mathcal{M}_{\mathrm{V}}^1$ and in red for $\mathcal{M}_{\mathrm{H}}^1$.

   The input image contains an occluded vertical ellipse, not present in the training set. Two different experiments are presented, varying the order $d$ of the model $\mathcal{M}_{\mathrm{V}}^d$ while maintaining the highest dimension for the model that does not represent the input shape, $\mathcal{M}_{\mathrm{H}}^1$. With $d = 0$, only the mean shape is considered in the shape prior (no deformations), with $d = 1$ the vertical deformations are considered. All the parameters are the same in both experiments. Figures 8.1b and 8.1d show some steps in the minimization, and Figures 8.1c and 8.1d show the evolution of the shape dissimilarity measure, for both experiments, respectively. Steps ① and ⑤ show an intermediate curve in the evolution with $\lambda = 0$, and the projections, $\mathcal{P}_{\mathrm{H}}^1 \phi$ and $\mathcal{P}_{\mathrm{V}}^d \phi$, to both models, dashed colored lines. The initial curve (in yellow) is also shown. Note that $\mathcal{P}_{\mathrm{V}}^0 \phi$ has no vertical deformations. The following steps (②,③ and ⑥,⑦) show the evolution after the "prior activation" adding the $E_{\mathrm{SM}}$ term ($\lambda = 1.1$), and the obtained segmentation (④ and ⑧).

   In the first experiment (Figure 8.1b), the projections to both models end in the same shape, the mean shape. This is reflected also in the graph of dissimilarity measure (Figure 8.1c) by the overlapping of the green and red curves. In the second experiment (Figure 8.1d), $\mathcal{M}_{\mathrm{V}}^1$ captures the variation of the input shape, as reflected in the obtained segmentation. In this case there is also a model switching around iteration 200 (step ⑥), where the $\mathcal{M}_{\mathrm{H}}^1$ is selected
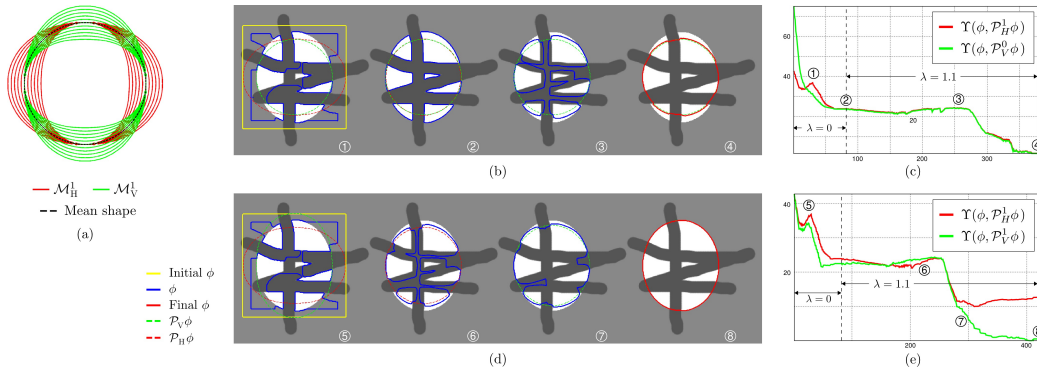
Figure 8.1: (a) Mode of variation for the two ellipses models ($\mathcal{M}_v^1$ in green and $\mathcal{M}_h^1$ in red), the mean shape of both models is the same and is plotted in black dash line. (b) Results for the experiments with $\mathcal{M}_h^1$ and $\mathcal{M}_v^0$ (only mean shape). Some steps in the segmentation (see text) (c) Evolution of the shape dissimilarity measure for the experiments with $\mathcal{M}_h^1$ and $\mathcal{M}_v^0$. (d) Results for experiments with $\mathcal{M}_h^1$ and $\mathcal{M}_v^1$ (complete model). (e) Evolution of the shape dissimilarity measure for the experiments with $\mathcal{M}_h^1$ and $\mathcal{M}_v^1$.

while the occlusions are being filled. After this point, the vertical deformation determines the selection of $\mathcal{M}_v^1$ for the rest of the evolution, ending with an accurate segmentation. Clearly, the high-order model and the automatic model selection are critical to obtain the correct segmentation.

## 8.2 Models from the walking sequence

Five high-dimensional models of a walking person cycle $\mathcal{M}_{w_k}^{21}$ ($k = 1, \ldots, 5$, $d = 21$) were obtained with the procedure explained in Section 6.1. The first three modes of variation for each model are shown in Figure 6.1. These are the models in the set of models $\mathfrak{M} = \{\mathcal{M}_{w_k}\}_{k=1}^5$ for the next experiment. This set of models is particularly challenging for model selection since they are different deformations of the "same object."

The input image in this experiment contains a new occluded shape $\hat{\phi}_1$ (Figure 8.2a) that belongs to the model $\mathcal{M}_{w_1}$ and is not in its training set $\Phi_1$. Figure 8.2 shows details about the segmentation of $\hat{\phi}_1$. Figure 8.2c shows four steps after the "prior activation" (steps ① to ④) in the evolution of $\phi$ (blue curve) and their projections onto the automatically on-line selected model $\mathcal{M}_{w_1}$ (green curve). Also the obtained segmentation $\tilde{\phi}_1$ (red curve) and its projection are shown. Figure 8.2b shows the projections of $\phi$ (blue curve) in the "prior activation" iteration onto the five models, ordered based on $\Upsilon(\phi, \mathscr{P}_{w_k}\phi)$ for this iteration. The mean shape of the corresponding model is plotted too (black curve). Compare the projections of the occluded shapes (Figure 8.2b) with those of a similar non-occluded shape in Figure 6.1c. Note how the projections onto the incorrect models are not too different, but the

Figure 8.2: (a) Input image with an occluded shape $\hat{\phi}_1$ in gray levels. (b) Projections of $\phi$ in the "prior activation" iteration (blue curve in step ① in Figure 8.2c) onto the five models, ordered based on $\Upsilon(\phi, \mathscr{P}_{w_k}\phi)$. The mean shape of the corresponding model is plotted too (black curve). (c) Steps ① to ④ in the evolution of $\phi$ (blue curve) and its projections onto the selected model $\mathscr{M}_{w_1}$ (green curve). The obtained segmentation is the red curve. (d) Evolution of the shape dissimilarity measure, $\Upsilon(\phi, \mathscr{P}_{w_k}\phi)$ with the iterations. The curves in steps ①-④ are shown in Figure 8.2c. (This figure is in colors.)

projections onto the correct model have significative differences. Figure 8.2d plots the evolution of the shape dissimilarity measure, $\Upsilon(\phi, \mathscr{P}_{w_k}\phi)$, for all the iterations and the five models. Note how the correct model is the one with lowest dissimilarity measure.

This experiment is repeated four times, maintaining the same set of models $\mathfrak{M} = \{\mathscr{M}_{w_k}\}_{k=1}^{5}$ and changing the input image. In each repetition, the input image contains a new occluded shape $\hat{\phi}_k$, $k = 2, 3, 4, 5$, belonging to the models $\mathscr{M}_{w_k}$, $k = 2, 3, 4, 5$, respectively. These four images are shown in Figure 8.3a with the corresponding obtained segmentations $\tilde{\phi}_k$ (red curves) and the projections onto the corresponding selected model (dashed green curves).

The results in Figure 8.2 show a number of important characteristics of the proposed framework that are consistent for all the presented experiments. First, in all the examples the selected model is the one to which the input shape belongs and the obtained segmentation is accurate to the data given by the image. Also, for all the experiments, during the minimization iterations the model selection is stable and there is no switch between the models once the shape prior is activated.

Second, it is relatively easy to follow the variations of the projection in the shape dissimilarity measure graph as the curve evolves. When the occlusions are being filled, the projection gets more similar to the shapes in $\mathscr{M}_{w_1}$ and the dissimilarity $\Upsilon(\phi, \mathscr{P}_{w_1}\phi)$ reduces. This is due to the force generated by

(a)                (b)
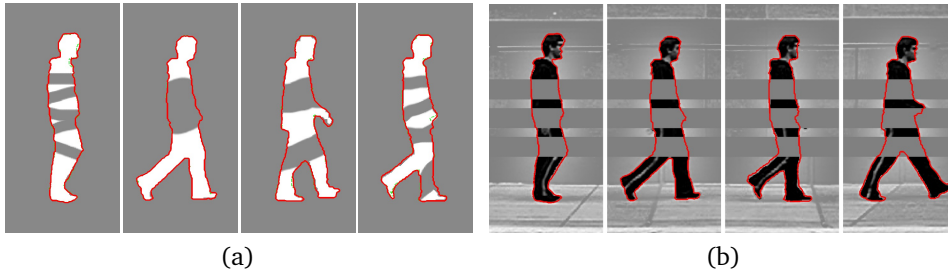
Figure 8.3: Segmentations obtained with the proposed framework with the set of models $\mathfrak{M} = \{\mathscr{M}_{w_k}\}_{k=1}^5$ for different input images. (a) Segmentations for the binary occluded shapes $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ and $\hat{\phi}_5$ belonging to different models $\mathscr{M}_{w_k}$, $k = 2, 3, 4, 5$, respectively. (b) Segmentations of the gray level images with added occlusions. The shapes in these images also belong to different models $\mathscr{M}_{w_k}$, which are all correctly selected by the proposed framework.

the shape term $E_{\mathrm{SM}}$, and as the curve gets closer to its projection this term attracts the curve strongly. Although this behavior is due to the $E_{\mathrm{SM}}$ term, the competition of both energy terms in areas of the shape where there is no occlusion preserves the curve close to the contour of the original shape, preventing to locally follow the projection, meaning that the $E_{\mathrm{CV}}$ energy term is stronger than the prior in this area of the image. This can be seen in the final segmentation (red curve in Figure 8.2c), where the projection (dashed green curve) in the hand goes through the original shape but the curve respects the gray level information. Similar details can be seen in Figure 8.3a for a different example. This example shows how the two energy terms collaborate to obtain a good segmentation of the occluded shape and each term attempts to define the curve in the regions where it better describes the solution. Where there is an occlusion the shape prior term takes control of the curve and where there is information of the actual shape (given by the intensity of the pixels), the data term controls the curve. This is done in a collaborative way, there is no discontinuity in the curve and it remains smooth. In order to achieve this, the projection onto the proper model is critical. Also the selection of

| $k(\mathscr{M}_{w_k})$ | $\Upsilon(\tilde{\phi}_k, \mu_k)$ | $\bar{\Upsilon}$ | $\sigma_{\Upsilon}^2$ | $\Upsilon(\tilde{\phi}_k, \phi_k)$ |
|---|---|---|---|---|
| 1 | 4.53 | 3.85 | 0.91 | 0.86 |
| 2 | 3.39 | 2.66 | 0.68 | 1.05 |
| 3 | 2.60 | 3.42 | 1.31 | 0.59 |
| 4 | 4.14 | 3.75 | 1.37 | 1.44 |
| 5 | 2.99 | 3.18 | 1.17 | 2.18 |

Table 8.1: Numerical validation results (Equation (7.8)) and dissimilarity measures (Equation (7.6)) for the obtained segmentations $\hat{\phi}_k$ for the experiment in Figures 8.2 and 8.3.

the parameter $\lambda$ is important, determining these collaboration/competition between both energy terms. In this work, as often done in the literature, $\lambda$ is manually obtained. As a rule of thumb $\lambda \in (1.0, 1.2)$ was found to be a good initial estimation.

Table 8.1 shows numerical results for the validation of the obtained segmentation (see Equation (7.8)). The dissimilarity between the obtained shape and the original non-occluded shape $\Upsilon(\tilde{\phi}_k, \phi_k)$ (possible in these experiments since it is accessible) is shown in the last column. Note that these last measures are, in general, significantly smaller than the mean dissimilarity between the shapes in the training set and the mean shape, $\bar{\Upsilon}_k$. This indicates the high accuracy of the proposed framework for this data.

Table 8.2 shows the dissimilarity measures between the obtained segmentations and the mean shapes for all the models in $\mathfrak{M}$, for the five different images and shapes in Figures 8.2 and 8.3, the minimum for each is in bold in each row, and obtained in the diagonal as expected from a correct model selection. Note that the difference between the minima and the next greater value in each row are considerable. This further indicates how the automatic model selection is correctly performed. Taking into account that the obtained segmentations correspond to shapes generated by the selected models, in all the experiments, the results were validated with the proposed measure, Equation (7.8), and were not validated by the other four used models. This further supports the validity of the proposed framework in general and the on-line automatic selection of the correct model in particular.

Figure 8.3b shows the obtained segmentations with the proposed framework for four different gray level images. The configuration of the framework is the same as in the previous examples, using the set of models $\mathfrak{M} = \{\mathcal{M}_{w_k}\}_{k=1}^5$. The automatically selected models, as well as the obtained segmentations, are also correct and accurate.

|  | $\Upsilon(\tilde{\phi}_k, \mu_1)$ | $\Upsilon(\tilde{\phi}_k, \mu_2)$ | $\Upsilon(\tilde{\phi}_k, \mu_3)$ | $\Upsilon(\tilde{\phi}_k, \mu_4)$ | $\Upsilon(\tilde{\phi}_k, \mu_5)$ |
|---|---|---|---|---|---|
| $\tilde{\phi}_1$ | **4.53** | 13.16 | 6.99 | 9.08 | 11.54 |
| $\tilde{\phi}_2$ | 12.99 | **3.39** | 11.05 | 7.65 | 6.05 |
| $\tilde{\phi}_3$ | 6.87 | 9.14 | **2.60** | 6.34 | 7.45 |
| $\tilde{\phi}_4$ | 6.89 | 8.43 | 7.48 | **4.14** | 7.97 |
| $\tilde{\phi}_5$ | 8.81 | 6.64 | 7.01 | 6.54 | **2.99** |

Table 8.2: Dissimilarities measures (Equation (7.6)) between the obtained segmentations $\tilde{\phi}_k$ and the mean shapes $\mu_k$ of the models in $\mathfrak{M}$ for the experiment in Figures 8.2 and 8.3.

|   |   |   |   |
|---|---|---|---|
| (a) | (b) | (c) | (d) |
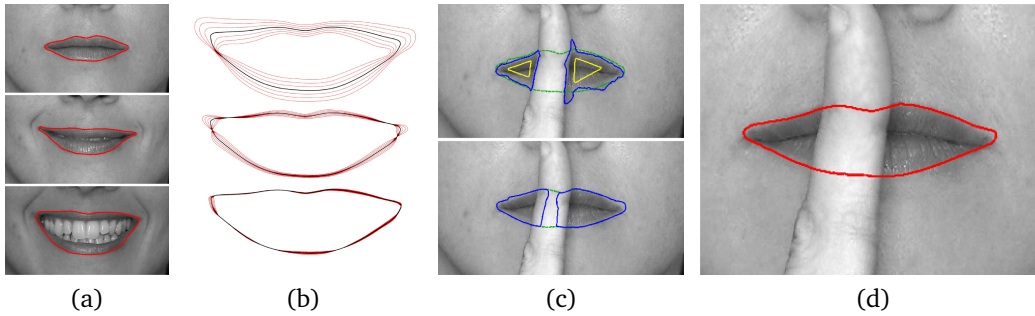
Figure 8.4: Segmentation obtained with the model of lips and mouth $\mathcal{M}_l$. (a) Three images and the corresponding contour (in red) of the defined shape in the set of training shapes. (b) First three modes of variation of the model $\mathcal{M}_l$. (c) A new image with an occluded shape and the evolution of the curve $\phi$. Top: Initial curve (in yellow), curve in the "prior activation" step (in blue) and its projection onto the model (dashed green curve). Bottom: curve (in blue) at an intermediate step with the $E_{sm}$ term added, and its projection onto the model (dashed green curve). (d) Obtained segmentation (zoom).

## 8.3 Model of lips and mouth

A model of the shape variation of a mouth and lips, $\mathcal{M}_L$, of a person while smiling is obtained from a set of 50 images. A new image where the mouth and lips are occluded by a finger is then segmented using the proposed framework. Figure 8.4a shows three images and the manually segmented contour for three different elements in the set of training shapes. Figure 8.4b shows the first three modes of variation of the obtained shape model. At the top of Figure 8.4c the new testing image is shown with the curve in two different steps before the $E_{SM}$ is added. The initial curve is shown in yellow and the curve in the "prior activation" step is in blue, its projection to the model is also shown as a dashed green curve. At the bottom of Figure 8.4c an intermediate step after the "prior activation" is shown, the curve is in blue and its projection is the dashed green curve. Figure 8.4d shows the obtained segmentation (curve in red). The obtained segmentation is accurate and the occlusion is filled properly with the information given by the model.

## 8.4 Varying the order of the models

This section further analyzes the segmentations when the order $d$ of the model varies. The shapes used in this test are shown in Figure 8.5a. They are fifteen shapes of sharks taken from the SQUID database [AMK96]. Two different sets of shapes are defined, $\Phi_{s_1}$ and $\Phi_{s_2}$. $\Phi_{s_1}$ has fourteen shapes, leaving out the shape marked with a box in Figure 8.5a, while $\Phi_{s_2}$ uses the fifteen shapes. Two different models were created, $\mathcal{M}_{s_1}^d$ from $\Phi_{s_1}$ and $\mathcal{M}_{s_2}^d$ from $\Phi_{s_2}$. Since $\Phi_{s_2}$ is

| $\mathcal{M}^d_{s_1}$ | | | $\mathcal{M}^d_{s_2}$ | | |
|---|---|---|---|---|---|
| $d$ | $\Upsilon(\tilde{\phi}_{s_d}, \mu_{s_1})$ | $\Upsilon(\tilde{\phi}_{s_d}, \phi_s)$ | $d$ | $\Upsilon(\tilde{\phi}_{s_d}, \mu_{s_2})$ | $\Upsilon(\tilde{\phi}_{s_d}, \phi_s)$ |
| 3 | 3.59 | 1.46 | 3 | 5.68 | 6.34 |
| 7 | 4.15 | 0.97 | 10 | 5.49 | 2.60 |
| 10 | 4.25 | 0.95 | 13 | 3.69 | 1.63 |
| 13 | 3.96 | 1.38 | 14 | 4.00 | 0.63 |
| (a) $\bar{\Upsilon} = 6.11$, $\sigma^2_\Upsilon = 1.42$ | | | (b) $\bar{\Upsilon} = 5.93$, $\sigma^2_\Upsilon = 1.48$ | | |

Table 8.3: Numerical validation results (Equation (7.8)) and dissimilarity measures (Equation (7.6)) for the obtained segmentations $\tilde{\phi}_{s_d}$ shown in Figure 8.5.

larger than $\Phi_{s_1}$, $\mathcal{M}^d_{s_2}$ might have more maximal modes of variation than $\mathcal{M}^d_{s_1}$, this happen in this case being $d = 14$ the number of modes of variation for $\mathcal{M}^d_{s_2}$ and $d = 13$ for $\mathcal{M}^d_{s_1}$. Figure 8.5b shows the first three modes of variation of $\mathcal{M}^d_{s_1}$. The modes of variation of $\mathcal{M}^d_{s_1}$ are very similar to those of $\mathcal{M}^d_{s_2}$.

Inspite of the "visual" similarity of the shapes in the set, their variations are larger than in the previous examples. For instance, they are not just a sampling of the deformation of an object like the walking sequence or the ellipses. This can be observed from the mean dissimilarity measure between the shapes in the training set and the mean shape, $\bar{\Upsilon}$. For the models $\mathcal{M}_{s_k}$ this value is close to 6 (see Tables 8.3a and 8.3b) whereas for the models $\mathcal{M}_{w_k}$ is smaller than 4 (see Table 8.1). This is a significant difference for this dissimilarity measure. (Note that the dissimilarities can be compared since they are normalized by their corresponding curve length.)

In order to analyze the segmentations varying the order of the model, at first a single model is used in $\mathfrak{M}$ at a time, without the influence of the model selection component of the framework.

The first experiment consists of the segmentation of an input image with an occluded version of the boxed shape in Figure 8.5a, with the model $\mathcal{M}^d_{s_1}$ with different order $d$ (number of modes of variation). Let $\phi_s$ be the original shape, $\hat{\phi}_s$ its occluded version, and $\tilde{\phi}_{s_d}$ the obtained segmentation with the $d$-order model. Figure 8.5c shows the obtained segmentations $\tilde{\phi}_{s_d}$ (red curves) for $d = [3, 7, 10, 13]$. The projection to the model is also plotted (dashed green curves).

This experiment is repeated using the model $\mathcal{M}^d_{s_2}$. Figure 8.5d shows the obtained segmentations for $d = [3, 10, 13, 14]$ and the corresponding projections onto the model.

Table 8.3a shows, for the model $\mathcal{M}_{s_1}$, the dissimilarity measure between the obtained segmentations $\tilde{\phi}_{s_d}$ for different $d$ and the original non-occluded shape, $\Upsilon(\tilde{\phi}_{s_d}, \phi_s)$, and the dissimilarity measure with respect to the mean
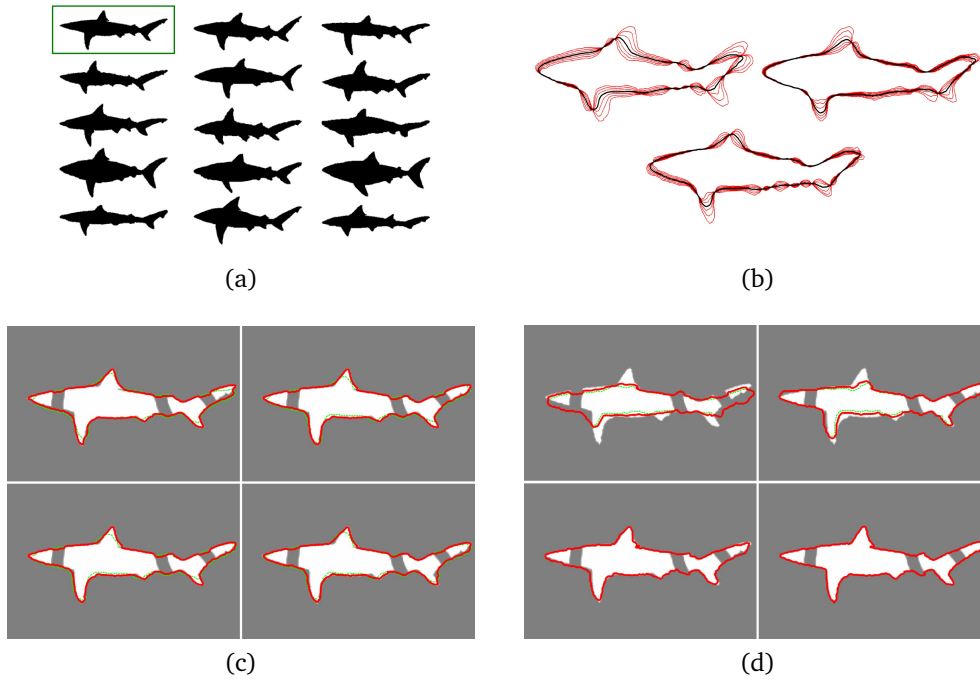
Figure 8.5: Obtained segmentations with different orders of the shape model. (a) Set of shapes used in the experiment, the boxed shape is $\phi_s$. (b) First three modes of variation of the model $\mathcal{M}_{s_1}$. (c) Segmentation of $\hat{\phi}_s$ (an occluded version of $\phi_s$) when $\phi_s$ is not in the training set (left to right and top to bottom $d = [3, 7, 10, 13]$). (d) Segmentation of $\hat{\phi}_s$ when $\phi_s$ is in the training set (left to right and top to bottom $d = [3, 10, 13, 14]$).

shape $\mu_{s_1}$, $\Upsilon(\tilde{\phi}_{s_d}, \mu_{s_1})$. Table 8.3b shows the same results for the model $\mathcal{M}_{s_2}^d$ and the mean shape $\mu_{s_2}$.

As can be observed in both experiments, the projection better represents the shape as the order increases.

In the first experiment (Figure 8.5c), the obtained segmentation improves the adjustment to the shape as the model has more details to represent. This can be seen for example in the pectoral and tail fin and under the head. However, since the projection does not perfectly adjust to the object, there is a competition between both energy terms, generating an intermediate curve that does not completely fit the present object. If more weight is added to the $E_{SM}$ term, other regions of the curve, in non-occluded areas of the object, will follow a less accurate approximation of the projection and lead to a worst segmentation, for example in the belly of the shark. Trying to choose the best segmentation from this four cases of $d$, the curve obtained with $d = 10$ seems to be the slightly more accurate than the curve obtained with $d = 13$, for example, analyzing the adjustment in the tail and the pectoral and pelvic fins. This is also supported by the dissimilarity between the obtained segmentations $\tilde{\phi}_{s_d}$ and the original non-occluded shape, last column of Table 8.3a. This pro-
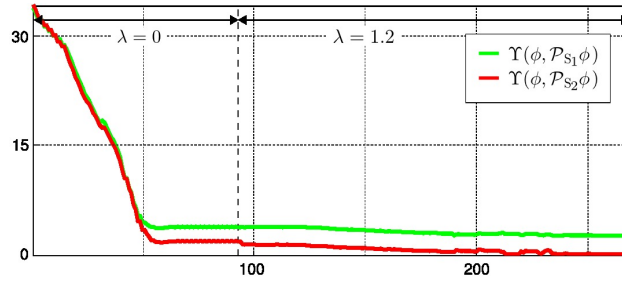
Figure 8.6: Evolution of the shape dissimilarity measure, $\Upsilon(\phi, \mathscr{P}_{S_k}\phi)$, with the minimization iterations.

vides an example of a kind of over-fitting of the model to the shapes in the training set, capturing features that are too specific in the higher eigenmodes.

On the other hand, with the model $\mathscr{M}_{S_2}^d$ that includes the non-occluded shape in the training set (Figure 8.5c), the $E_{\text{SM}}$ shape term has more relevance in the segmentation. When the order of the model increases, the projection gets more accurate and the segmentation improves. When the order is low and the projection is not accurate, the segmentation again is a compromise between the two energy terms, being an intermediate curve between the projection and the edges of the gray level information. Finally, the main difference between the segmentations with $d = 13$ and $d = 14$ are the fine details like high curvature points, see the extreme points in the tail and the pectoral fin.

The last experiment of this section is done using both models in the set $\mathfrak{M} = \{\mathscr{M}_{S_1}^{13}, \mathscr{M}_{S_2}^{14}\}$, and the same occluded shape in the input image. Figure 8.6 plots the dissimilarity measure for this example. The selected model is $\mathscr{M}_{S_2}^{14}$ which has one additional eigenmode and obtains a better description. The obtained segmentation is, of course, the same segmentation shown in Figure 8.5d with $d = 14$. This further supports the necessity of high-order models in order to obtain accurate segmentations, in particular when the different object classes are relatively similar.

These obtained segmentations are also validated by the proposed validation process, (Equation (7.8)).

# Chapter 9

# Invariance to translation

Invariance to geometric transformations (such as translations, rotations and scaling) is a desirable property in a general framework for segmentation. One way to do this is to substitute $H(\phi)$ by

$$H(\sigma R_\theta(\phi(x - x_0)))$$

in Equation (7.5), as in the work of Cremers *et al.* [COS06, section 5.1]. Here, $\sigma$ is a scale factor, $R_\theta$ a rotation matrix of a given angle $\theta$, and $x_0$ a translation vector.

This chapter proposes an extension of the functional in Equation (7.5), adding invariance to translation (other invariances can be similarly added).

Consider all the shapes aligned with respect to their corresponding center of mass $p_\phi$, defined for a certain shape $\phi$ as

$$p_\phi = \frac{\int_\Omega p H(\phi(p)) \mathrm{d}p}{\int_\Omega H(\phi(p)) \mathrm{d}p}.$$

The shape models are built in the same way as in the previous chapter. Considering that all the shapes in $\Phi_k$ have the same center of the mass $p_k$, this point becomes the center of mass of the model. To obtain the projection of a new, not aligned, shape $\phi$ with center of mass $p_\phi$, first the shape is translated to $p_k$ and then projected. Mathematically, the projection to the translation invariant model $\mathcal{M}_k^{(\mathrm{IT})}$ centered at $p_k$, becomes $\mathscr{P}_k \phi(p + p_\phi - p_k)$, and lets call $p_\phi^k$ its center of mass. Finally, the projection needs to be translated back to the original center of mass. Defining $T(\phi(p), p_0) = \phi(p + p_0)$, the final projection to the translation invariant model is

$$\mathscr{P}_k^{(\mathrm{IT})} \phi = T(\mathscr{P}_k T(\phi(p), p_\phi - p_k), p_\phi^k - p_\phi). \tag{9.1}$$

Without loose of generality $p_k = 0$ is assumed from now on.
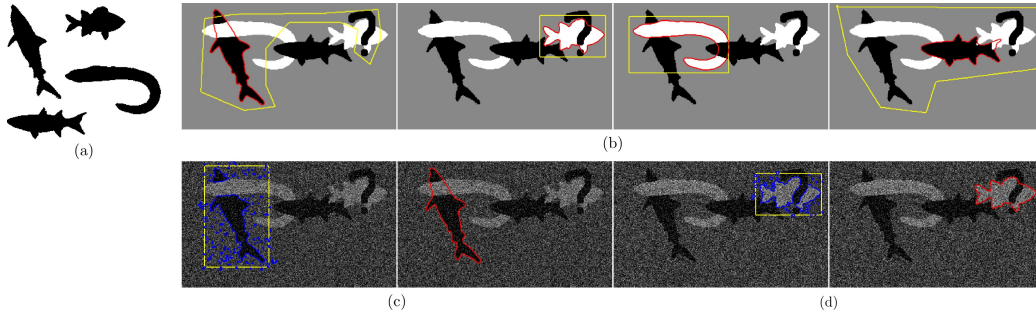
Figure 9.1: Segmentations obtained with the translation invariant energy (see Equation (9.2)). (a) Four zero-order models (shape prior templates). (b) Four different initial curves (yellow curves) and the obtained segmentations (red curves). (c)-(d) Two different initializations with Gaussian noise added. Initial curve (in yellow), last curve previous to the "prior activation" (in blue) and obtained accurate and valid segmentation (red curve).

In order to incorporate the invariance to translation in the original energy, the shape models terms $E_k(\phi, \mathcal{M}_k)$ become

$$E_k(\phi, \mathcal{M}_k) = \int_\Omega \left\| H\left(\phi(p + p_\phi)\right) - H\left(\mathcal{P}_k \phi(p + p_\phi)\right) \right\|^2 \mathrm{d}p. \qquad (9.2)$$

If the $k$th model provides a good representation of $\phi$, the corresponding centers of mass are close, $p_\phi^k \approx p_k = 0$. For the derivation of the corresponding gradient descent expression below, this approximation is assumed, simplifying the derivation here presented. However in the implementation the actual $p_\phi^k$ is used. The updated gradient descent expression is

$$\frac{\partial E_{\text{SP}}}{\partial \phi} = -2 \sum_{k=1}^{M} \beta_k \Big[ \Delta H(p) \Delta \delta(p) +$$

$$\frac{\delta(\phi(p))(p - p_\phi)^T}{\int_\Omega H(\phi(u))\mathrm{d}u} \int_\Omega \Delta H(z) \Delta \delta(z) \nabla \phi(z) \mathrm{d}z \Big], \quad (9.3)$$

where

$$\Delta H(p) = \left( H(\phi(p)) - H\left(\mathcal{P}_k^{(\text{IT})} \phi\right) \right)$$

and

$$\Delta \delta(p) = (\delta(\phi(p)) - \delta\left( T(W \mathcal{P}_k T(\phi(p), p_\phi), -p_\phi) \right)).$$

Again, $\beta_k$ is treated as static, as a first order approximation for the gradient descent.

## 9.1 Model selection with invariance to translation

An example of the model selection capabilities of the translation invariant framework is shown in Figure 9.1. In order to test only the model selection,

Figure 9.2: Segmentations obtained with the translation invariant energy (see Equation (9.2)). (a) Three steps in the segmentation of the ellipse from Figure 8.1. Note in the first image the projections translated to the center of mass of $\phi$. The second image shows the projection to the correct model and the curve filling the occlusions. The last image shows the obtained segmentation (red curve). (b) Obtained segmentation of a binary shape from the walking person cycle. (c) Obtained segmentation of a gray-valued shape from the walking person cycle.

without being influenced by the adjustment to the selected model, four zero-order models are used. A zero-order model obtains always the same synthesized shape for any input shape, this synthesized shape being the mean shape of the model in this work. The zero-order models are shown in Figure 9.1a. These four different shapes from the SQUID database are arranged in a single image with occlusions for each shape and this becomes the input image for testing the framework. Figure 9.1b shows four different initial curves (in yellow) and the segmentation (red curves) obtained with the proposed framework. Two examples with Gaussian noise added to the image are shown in Figures 9.1c and 9.1d.

In all the cases the segmentations are accurate which also implies that the selected model is the correct one. Note that these results are valid even when the initial curves are not clearly defining one object (following our definition of validity, Equation (7.8)).

## 9.2 Segmentations with invariance to translation

The last tests show the translation invariant framework working with the high-order models. The first experiment reproduces the test with the ellipses (Section 8.1), now with the addition of Gaussian noise and the translation of the ellipse. Figure 9.2a shows the details of the segmentation. The first subfigure shows the initial curve (in yellow), an intermediate step (blue curve), and its projections to both models (green and red dashed curves). Note the projections translated to the center of mass of $\phi$. The second subfigure shows the projection to the correct model and the curve filling the occlusions. The last subfigure shows the obtained segmentation (red curve).

Figure 9.2b shows the segmentation of an image with an occluded binary

shape from the walking sequence and Figure 9.2c shows the segmentation of an image with a gray level shape from the walking sequence, using the set of translation invariant models for this dataset. Again the result is an accurate segmentation with a valid shape from the correctly selected model.

# Chapter 10

# Concluding remarks

A framework for simultaneous and automatic model selection and object segmentation was introduced in this work. The proposed technique is based on a new energy that combines region based segmentation with on-line selection of the best model for the object present in the image, and an adjustment to the best description of the object given the selected model.

The segmentation is obtained via gradient descent energy minimization, and the model selection is automatic in each iteration, without the need to run the segmentation with all the models and then select the best solution. The on-line decision of best description is based on a shape dissimilarity measure between the curves. The selection is such that a unique model candidate is considered at each step of the minimization. Invariance to shape transformations are incorporated into the proposed framework as well.

Possible directions for further improvements include incorporating high-order modes in the validation step and considering going beyond PCA, as well as including class-dependent model orders ($d_k$). Results in these directions will be reported elsewhere.

# References

[AMK96]    Sadegh Abbasi, Farzin Mokhtarian, and Josef Kittler. Squid
           database (shape queries using image databases). WWW, 1996.
           Visited, April 26th. 2008.

[ARS07]    Pablo Arias, Gregory Randall, and Guillermo Sapiro. Connecting
           the out-of-sample and pre-image problems in kernel methods. In
           *CVPR*, pages 1–8, 2007.

[CFK07]    G. Charpiat, O. D. Faugeras, and R. Keriven. Shape statistics for
           image segmentation with prior. In *CVPR*, pages 1–6, 2007.

[CKS03]    D. Cremers, T. Kohlberger, and C. Schnorr. Shape statistics in
           kernel space for variational image segmentation. *Pattern Recog-
           nition*, 36(9):1929–1943, sep 2003.

[CL06]     R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Com-
           putational Harmonic Analysis*, 21(1):5–30, 2006.

[COS06]    D. Cremers, S. Osher, and S. Soatto. Kernel density estimation
           and intrinsic alignment for shape priors in level set segmenta-
           tion. *IJCV*, 69(3):335–351, 2006.

[CSS06]    D. Cremers, N. A. Sochen, and C. Schnörr. A multiphase dy-
           namic labeling model for variational recognition-driven image
           segmentation. *IJCV*, 66(1):67–81, 2006.

[CTCG95]   T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active
           shape models-their training and application. *Computer Vision
           and Image Understanding*, 61(1):38–59, 1995.

[CV01]     T. F. Chan and L. A. Vese. Active contours without edges. *IEEE
           Transactions on Image Processing*, 10(2):266–277, 2001.

[DRT06]    Samuel Dambreville, Yogesh Rathi, and Allen Tannenbaum.
           Shape-based approach to robust image segmentation using ker-
           nel pca. In *CVPR*, volume 1, pages 977–984, 2006.

# References

[FKS+04]  T. Funkhouser, M. Kazhdan, P. Shilane, P. Min, W.Kiefer, A. Tal, S. Rusinkiewicz, and D. Dobkin. Modeling by example. *ACM Transactions on Graphics*, 23(3):652–663, August 2004.

[Kan79]  G. Kanisza. *Organization in Vision*. Praeger, New York, NY, 1979.

[KL05]  Won Kim and Ju-Jang Lee. Object tracking based on the modular active shape model. *Mechatronics*, 15(3):371 – 402, 2005.

[LGF00]  M. E. Leventon, W. E. L. Grimson, and O. D. Faugeras. Statistical shape influence in geodesic active contours. In *CVPR*, volume 1, pages 1316–1323, 2000.

[PD99]  N. Paragios and R. Deriche. Geodesic active regions for supervised texture segmentation. In *ICCV*, pages 926–932, 1999.

[RJZ06]  Marie Rochery, Ian H. Jermyn, and Josiane Zerubia. Higher order active contours. *IJCV*, 69(1):27–42, 2006.

[SJ09]  Anuj Srivastava and Ian H. Jermyn. Looking for shapes in two-dimensional cluttered point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1616–1629, September 2009.

[SKC95]  G. Sapiro, R. Kimmel, and V. Caselles. Geodesic active contours. In *ICCV*, pages 694–699, 1995.

[SS02]  B. Schölkopf and A.J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, USA, 2002.

[TYWW+01]  A. Tsai, A. J. Yezzi, III W. Wells, C. Tempany, D. Tucker, A. C. Fan, W. E. L. Grimson, and A. S. Willsky. Model-based curve evolution technique for image segmentation. In *CVPR*, volume 1, pages 463–468, 2001.

[VC02]  L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *IJCV*, 50(3):271–293, 2002.

[vL07]  Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[VM08]  N. Vu and B. S. Manjunath. Shape prior segmentation of multiple objects with graph cuts. In *CVPR*, pages 1–8, 2008.

# List of Tables

# List of Figures

This is the last page.
`http://iie.fing.edu.uy/~fefo/thesis/`