Universidad Nacional de Colombia

Facultad de Ingeniería
Grupo de Investigación Bioingenium
Universidad Nacional de Colombia

EFFICIENT NAVIGATION OF MEDICAL MEGA-IMAGES

Francisco Albeiro Gómez Jaramillo

Mentor: Prof. Eduardo Romero Castro

Jury composition:
Prof. Marcela Iregui (UMNG, Colombia)
Prof. Norberto Malpica (URJ, Spain)
Prof. Luis Niño (UNC, Colombia)
Dr. Javier Pascau (HGM, Spain)
Dr. Andrea Soddu (ULG, Belgium)

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor en Ingeniería - Sistemas y Computación

July 2010

Abstract

Recent advances in imaging technology have provided large volumes of image data. This has imposed novel interaction requirements to medical doctors. Development of these interaction mechanisms is a challenging unexplored research problem. This thesis was devoted to the study of the complex processes underlying the events produced when a medical user interacts with a large volume of medical data, and how can be improved interaction by using this information. Acquirement of medical knowledge from these data implies intensive interaction. Our aim was to provide mechanisms for obtaining significant information during this interaction in a minimal time. To achieve this, we proposed original strategies for inferring and synthesizing medical knowledge from these data. Our inspiration was the biological mechanism that controls the visual data exploration and the complex movements involved in the navigation task. Proposed methods were successfully used to improve navigation velocities in virtual microscopy and to increase the tracking capabilities of the complex human motion.

Acknowledgements

Esta tesis significa un paso más en mi formación como investigador. Ahora es el momento parar y agradecer a todas las personas/entidades que hicieron posible este proyecto:

- A mi familia (Laureano, Flora, Betsy, Ignacio, Nico y Laura), quienes desde hace muchos años me brindan todo su apoyo, amor y compresión. A María Nury Escobar por su apoyo incondicional durante la ejecución del proyecto y por toda su paciencia. A mis amigos de toda la vida Wilson, Wilson, Wilfredo, Fernando y Juan Carlos gracias por confiar en mí. A los patas por mostrarme mi país y cambiar mi concepción de la vida.

- A mi director el profesor Eduardo Romero por su confianza, paciencia, colaboración, ejemplo y sobre todo por todo su tiempo. A Marcela Iregui por su invaluable guía en la concepción y desarrollo de este proyecto. A Fabio González por su colaboración y ejemplo. A Lucia Roa por introducirme al mundo de los patólogos. Al departamento de Patología de la Universidad Nacional por su colaboración. A todos mis coautores: Eduardo Romero, Marcela Iregui, Fabio Martínez, Ricardo Gutiérrez, Diana Marín, Lucia Roa y Cristina Santa-Marta, muchísimas gracias por su colaboración y su tiempo. A mis compañeros de doctorado Gloria, Juan K y Andrea por todo el tiempo compartido y toda su colaboración. A Norberto, Cris, Daniel, Quentin y Andreas gracias por mostrarme otra forma de hacer investigación. A mis jurados gracias por su valiosa colaboración y retroalimentación. A toda la gente del grupo Bioingenium por su apoyo y colaboración.

- Este proyecto no hubiera sido posible sin la financiación, soporte y colaboración de la Universidad Nacional de Colombia, y en especial, gracias al programa de doctorado en Ingeniería Sistemas y Computación por su colaboración y soporte.

Y a todas aquellas personas que de una u otra forma, colaboraron en la realización de este proyecto, hago extensivo mi más sincero agradecimiento.

Contents

Acronyms List

ANOVA Analysis of Variance

bpp bits per pixel

DWT Discrete Wavelet Transform

EBCOT Embedded Block Coding with Optimal Truncation

EWMA Exponential Weighted Moving Average

FFT Fast Fourier Transform

FOV Field of View

GUI Graphic User Interface

HVS Human Visual System

J2K JPEG2000 standard

JPIP JPEG2000 Interactive Protocol

LFU Last Frequently Used

LRU Last Recently Used

ML4UM Machine Learning for User Modeling

MSE Mean Square Error

pdf probability distribution function

RCT Reversible Color Transform

RD Rate-Distortion

RMSE Root Mean Square Error

RoI Region of Interest

SNR Signal to Noise Ratio

TGP Time Gain Percentage

VM Virtual Microscopy

VMV Virtual Microscopy Viewer

VS Virtual Slide

WoI Window of Interest

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Nowadays, mega-images or images composed of a larger number of pixels than those allowed by conventional capturing devices, are used in very different applications such as satellite, astronomic or medical images [22, 84, 147], among others. Large medical images comprise barely every diagnostic modality because of the amount of information that each generates. Over the last few years, medical specialties have been rapidly evolving from poorly detailed human body observations to very accurate pictures about the internal biological structures and body functions. These advances have been possible by recent developments in imaging technology that provide non-invasive high resolution 3D imaging. Others specialties such as pathology, are in a state of transition, from classical instruments of visualization, such as light microscopy, to digital tools as virtual microscopy. Virtual microscopy is a last technological innovation, consisting in merging digital imaging with conventional light microscopy to enable exploration through high resolution digital images of glass microscope slides, emulating tissue visualization, as in a real microscope [104].

All these novel medical image applications have imposed high demanding requirements for archiving, transmission, accessing and navigation. Since the last decade, on new standards for high resolution image store and transmission have been proposed [131]. However, many of their functionalities have not been exploited in the medical image domain. Additionally, the access velocity to these data is insufficient to have a seamless navigation. There are several methods which allow improved navigation such as highly scalable storage, prefetching and cache strategies, but their performance is highly dependent on the degree of knowledge of the expert navigation pat-

terns. Unfortunately, the underlying mechanisms which control the medical image exploration have been poorly studied. A deep understanding of these mechanisms is at the base of the fundamental question of how to store and to deliver visual data according to the medical doctor expectations.

This thesis is devoted to the study of the complex process underlying the events produced when a medical user interacts with a mega-image, and how we can improve the adaptive delivery of visual data using this information. In general, this type of processes are the result of the interaction between two complementary factors: on the one hand, top-down mechanisms such as the user knowledge or motor skills that govern many decisions and actions and, on other hand, bottom-up information such as the image low-level content that triggers particular types of associations. These mechanisms have been identified in many different interaction scenarios [60]. They are very robust to noise and allow to find a maximal amount of information with a minimum effort [104]. In several applications, these mechanisms may improve the way we use resources and the design of adaptable systems [60].

In this thesis we were focusing upon the particular domain of navigation in virtual microscopy for histopathology. This is one of the most promising technologies to be integrated in the clinical routines of pathology [104], but also one of the less studied problems in the medical imaging domain [13]. One of the first challenges that we found was that simple request and send operations, provided by the state-of-the-art image codification methods, were really limited for a fluid navigation. For this reason, we developed a set of efficient techniques for representing and accessing digital slides, such as highly scalable storage, prefetching and cache. Several tests were carried out in real navigations resulting in improvements of the exploration times. These first approaches were based on simple interaction mechanisms that nevertheless, resulted effective to accelerate the exploration. Later the research work was oriented to the identification of exploration patterns in the pathologist's navigation. An experimental study was carried out to recognize the more relevant characteristics of the navigation for diagnosis tasks in histopathology. Finally, several computational models for predicting these patterns and delivering the image data adaptively were developed.

1.2 Goals and Main Contributions

The main goal of this work was to reduce the navigation times in exploration of medical mega-image through the use of several strategies for storing and adaptive delivering the visual data. To meet this challenge we had explored the image compression techniques, the communications protocol, the graphical user interface, the navigation patterns and in general every part involved in the exploration of medical mega-images. During this work several original contributions in the area of virtual microscopy for histopathology have been produced and are summed up here after:

- A new strategy for efficiently browsing of histopathological images was proposed. The whole proposal is supported on the JPEG2000 standard [131] and strategies for adaptive delivering of image data, namely cache and prefetching. Main contribution was a complete analysis of the virtual microscopy relevant aspects so that multiple recommendations were presented on using each of the virtual microscopy modules, based on a detailed study from different perspectives by exploiting up-to-date tools. The strategy was tested in a real scenario in pathology, resulting in considerable reductions of the navigation times. An initial version of this work titled *Virtual microscopy using JPEG2000* was presented in the *12th Conference on Computer Analysis of Images and Patterns (CAIP2007)* [54]. The complete contribution was published in the specialized microscopy journal *Micron* as a research article titled *Strategies for efficient virtual microscopy in pathological samples using JPEG2000* [68].

- An original soft-cache strategy to improve the navigation times in virtual microscopy was formulated and tested. The main contribution was the design of an adaptable task oriented model which takes advantage from the JPEG2000 standard scalability by working at the minimal JPEG2000 information unit: the packet. Evaluations in very variable navigation patterns showed considerable reductions in the navigation times when was compared with state-of-the-art approaches. Preliminary results of this work were presented in the peer-reviewed conference *V international seminar of medical image processing - SIPAIM*. The complete contribution was accepted for publication in the specialized microscopy journal *Microscopy Research and Technique* under

the title *A Soft-cache Strategy for Pathologist's Navigation in Virtual Microscopy*.

- A novel pathologist's exploration pattern was identified and explained. The pattern was observed in the velocity navigation during explorations between Windows of Interest and results of the interaction between the neuromuscular user system and the tool used for the image exploration. Evidence about the occurrence of this pattern was presented in the *Third IASTED International Conference on Human-Computer Interaction* as part of an article titled *Prediction of Pathologist Navigation Patterns In Virtual Microscopy Based on a Soft-Computing Approach* [55]. More experimental evidence was presented in the research article *A Model for Predicting Pathologist's Velocity Profiles When Navigating Virtual Slides* [57] published in the specialized microscopy journal *Microscopy Research and Technique*.

- A computational model for tracking and predicting biological complex movements was formulated and evaluated. The model was based on a two step Bayesian machine learning strategy. An offline step, where the parameters of a hidden dynamical system which control the movement are learned. Followed by an online tracking phase, where the optimal trajectories are recursively approached from simple movement observations. The model was tested by predicting velocity patterns registered in the pathologist's navigations. The results showed that the proposed model outperforms the state-of-the-art proposals. This contribution was published in a chapter titled *Predicting Complex Patterns of Human Movements Using Bayesian Online Learning in Medical Imaging Applications* [56] as part of the book *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques*. The application of this model to the velocity prediction in virtual microscopy was presented in the research article *A Model for Predicting Pathologist's Velocity Profiles When Navigating Virtual Slides* [57].

- A computational model for prediction of Regions of Interest in histopathology images was formulated and tested. The model emulates the first slide examination phase, which consists on a coarse search of structures to separate image in relevant regions. The computational strategy is supported on visual attention model theory, the grouping Gestalt laws and machine learning. The main contribution of this work

was to model the visual attention process involved in histopathological image exploration by bringing together bottom-up and top-down information. The model was evaluated on manually segmented images resulting in more consisting Regions of Interest compared to classical visual attention models. This contribution was presented in the conference *SPIE Medical Imaging 2009* as a research article under the title *Finding Regions of Interest in Pathological Images: An Attentional Model Approach* [59]. The complete contribution was submitted for publication consideration to the specialized publication *Journal of Visual Communication and Image Representation* as a research article titled *A Supervised Attentional Model for Finding Regions of Interest in Basal Cell Carcionoma Images* [68].

During the development of this work we found interesting problems directly related with our original research question. Some methodological issues involved in the development of this work are applicable to these problems. Additional efforts were devoted to explore these problems and a series of original contributions were proposed in these areas:

- The Center of Mass trajectory is of paramount importance for the description of the human movements, because provides measures about the energy consumptions and their trajectories can be associated to several neuromuscular diseases. However, the tracking of the Center of Mass is a complicated problem since it is located inside the body. We approached this problem by means of the proposed computational model for predicting biological complex movements. Main contribution was to use a very simple methodology to follow the complex non linear dynamic associated to the Center of Mass Movement. This contribution was accepted for publication in the biomechanics specialized journal *Computer Methods in Biomechanics* under the title *A Kinematic Method for Computing the Motion of the Body Center-of-Mass (CoM) During Walking: A Bayesian Approach.*

- The texture is a fundamental characteristic for the histopathology image representation. Some of the most commonly used texture features are based on the application of bank filters, followed by a description of the corresponding probability distribution function. These representations result in highly discriminative for classification and image retrieval problems. However, these strategies ignore the fundamental fact

that many natural images are composed of directional information. In this thesis, we proposed and tested a novel low level texture feature based on sparse directional representations. The feature is based on Curvelets and generalized Gaussian distributions. The main contribution was the construction of a highly discriminative, precise and simple descriptor of natural textures. We demonstrated the effectiveness of the proposed descriptor for classification and retrieval tasks in natural images, obtaining significant improvements. An initial version of this contribution was published in *the 14th Iberoamerican Congress on Pattern Recognition (CIARP 2009)* as a research article titled *Texture Characterization using a Curvelet Based Descriptor* [58]. An extended version of this work titled *Rotation Invariant Texture Characterization using a Curvelet Based Descriptor* was invited to publish in a special issue in the *Pattern Recognition Letters* journal.

We encourage the lector to review these references that has been annexed at the end of this dissertation.

1.3 Structure of this Thesis

As we said previously, this thesis is focused on the following research goal:

To reduce the navigation times in exploration of medical mega-image through the use of several strategies for storing and adaptive delivering the visual data.

In order to meet this goal we propose to understand the dynamic generated by navigation of these images, and therefore to construct strategies that take advantage of this knowledge to accelerate exploration. In this context, we consider the development of a system that allows exploration of high resolution medical images. In this thesis we investigate three aspects of such system:

1. the design of a computational tool for exploring these images, it includes the representation used to efficiently store these visual data,

2. the navigation patterns that arise during the medical doctor explorations when he/she is using this tool,

3. the computational model that would be used to accelerate the exploration by taking advantage of the previously identified navigation patterns.

We explicitly considered histopathological images (also called *virtual-slides*), but it is not a restrictive choice. Navigation in other medical image domains could also be addressed by similar strategies. For instance, in neuroimage domain where data dimensionality is too high to explore with classical strategies, visual attention models as presented in chapter 7 would be used to select relevant regions for diagnosis.

Regarding the image representation, the JPEG2000 algorithm has been chosen. This choice was motivated by several reasons: their compression efficiency and the growing interest that it has been generated in the medical image domain [131]. But most importantly, this algorithm is highly scalable, a quite general concept that meets naturally the way experts exploring medical images. We believe that this concept is going to remain valid in novel compression algorithms and therefore strategies developed over this general idea are going to be valid also in future. As we demostrated in this investigation, navigation patterns arise from a complex interaction between image, graphical user interface and image content. Their understanding is of paramount importance not only for the acceleration of navigation problem, but also to answer a more relevant question: *How medical doctors explore visual data to devise a diagnosis?*. The answer of this question is out of scope of this thesis, however we identified *exploration patterns* that would be of interest for the image perception and medical image communities. For instance, patterns related to the use of the neuromotor system during exploration with mouse. In addition, we worked in the development of computational models to predict and take advantage of these patterns. These models were constructed based on our same observations of the exploration nature, and prior biological knowledge about the neuromotor system and the human visual system parts involved in exploration tasks. All these knowledge were integrated over a general probabilistic setting that fits well with machine learning, i.e., we were able to construct computational models that learn from actual navigation. It provides flexibility for adaptation of the proposed strategies to other domains. Finally, prediction strategies were carefully designed to meet requirements of faster exploration, i.e., precision and computational efficiency.

In the following, each of the three aspects enumerated above are investigated in a separate part, which we now briefly introduce. We invite to the reader to refer to Figure 1.1 to get a visual illustration of the thesis organization.

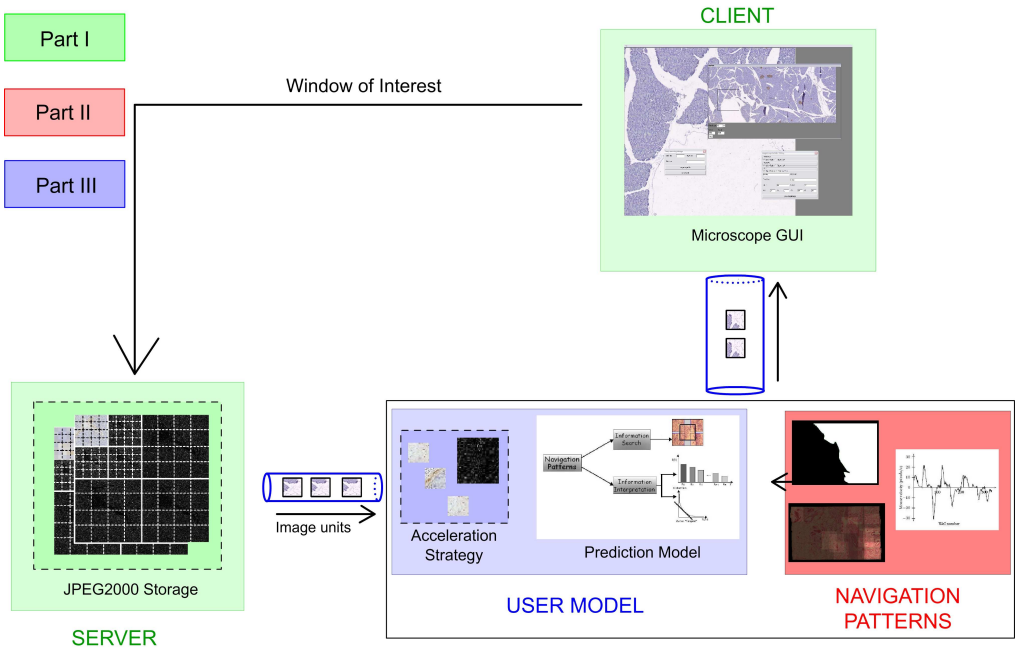


Figure 1.1: Thesis organization. Three aspects of a system for exploring high resolution medical images are studied: the computational tool for navigating mega-images that also includes image storage, the exploration patterns generated during navigation, and finally, strategies for accelerating the exploration taking advantage of these patterns.

As complement to this introduction, the first chapter presents a state of the art in virtual microscopy. In particular, we review the computational pipeline used for virtual microscopy: stitching, storage and navigation. Different stitching approaches presented in literature are described. We review main advantages and drawbacks of different methods used for high resolution medical images storage. Later, we detail precedent works in navigation of high resolution images, we are mainly concerned in strategies for accelerating navigation of JPEG2000 high resolution images. We report also some ele-

ments of the Graphical User Interface used to explore this images. Finally, we conclude with some future works of interests for the area.

Part I is devoted to the design of a tool for exploring high resolution medical images. Specifically, we detail the JPEG2000 functional and technical adaptations required to construct a digital tool to explore images in histopathological domain. These adaptations are not limited to the standard itself, but also include several strategies to address the problem of accelerating the navigations. These strategies take advantage of the dynamic nature of the image exploration. In this part we demonstrate improvements in navigation velocities by a careful combination of JPEG2000 features and strategies that exploit the exploration dynamic. The main conclusion of this part was the relevance of a good exploration model, i.e., a model to predict the navigation behavior in future.

In **Part II** we present an study carried out to identify pathologists navigation patterns when exploring virtual microscopy slides using the previously designed tool. A set of *similar* pathologists explored a series of high resolution histopathological images. Different issues of these explorations were evaluated, namely, the percentage of common visited image regions, the time spent at each and its coincidence level, that is to say, the region of interest location. In addition, navigation patterns were also assessed, i.e., mouse movement velocities and linearity of the diagnostic paths. Results suggest that RoIs are determined by a complex combination of the region visited, the time spent at each visit and the coincidence level among pathologists. Additionally, linear trajectories and particular velocity patterns were found for the registered diagnostic paths. Presence of these patterns is quite relevant, because it is an indicator of general behaviors that would be *learned* by a computational system in order to construct predictions about navigation future.

Part III deals with the construction of computational models that take advantage of patterns identified in the navigations. The first strategy is a new optimal cache strategy that improves the navigation times in medical image exploration. The entire method includes an optimal soft-cache policy and a dynamical probabilistic model of a pathologist's navigation. The whole model is based on a general exploration pattern identified in Part II. Results show improvements in velocity compared with state of the art approaches. We presented also in this part, two computational models to predict navigation in histopathological domain. Firstly, a soft computing model that permits to anticipate the pathologist trajectories in diagnosis

tasks when he/she is exploring virtual slides. The model combines in a very general Bayesian framework both: an offline model of a baseline pathologist knowledge (the prior) and a prediction online module (the likelihood) which captures a particular pathologist navigation pattern. The prior knowledge is learned from actual navigations performed by several pathologists in different virtual slides. Results indicate high improvements in trajectories prediction by using the proposed approach. Later, it is introduced a novel "bottomp-up" and "top-down" visual attention model for finding diagnostic regions of interest in histopathological images. The method is based on the cognitive process of visual selective attention that arises during a pathologist's image examination. The method was evaluated on a set of histopathological images and compared with regions selected by an expert pathologists finding considerable quality gains compared with state of the art methods.

Finally, in the **last chapter**, we summarize the contributions of each chapter and present several future research directions. In particular, importance of image quality in image exploration, development of automatic image navigation systems and the use of the semantic knowledge associated to images for improving navigation experience.

1.4 Publications

1.4.1 International Journals, Book Chapters and Peer-Reviewed Conferences

- *Romero E., Gómez F., Iregui M., Virtual Microscopy in Medical Images: a Survey, in Mendez-Vilas A. Diaz J. (eds.) Modern Research and Educational Topics in Microscopy, Formatex. pages 996-1006, 2007.*

- *Gómez F., Iregui M., Romero E., Virtual microscopy using JPEG2000, 12th Conference on Computer Analysis of Images and Patterns. Lectures Notes in Computer Science series. Springer Verlang. LNCS 4673, pages 181-188,2007*

- *Iregui M., Gómez F., Romero E., Strategies for efficient virtual microscopy in pathological samples using JPEG2000, Micron Vol 38 pag 700-713, 2007.*

- *Gómez F., Iregui M., Romero E., Prediction of Pathologist Navigation Patterns In Virtual Microscopy Based on a Soft-Computing Approach. Third IASTED International Conference on Human-Computer Interaction, Innsbruck, Mars 17-19, pages 150-155, 2008.*

- *Gómez F., Romero E., A Model for Predicting Pathologist's Velocity Profiles When Navigating Virtual Slides. Microscopy Research and Technique. Volume 73, pages 85-98, 2009.*

- *Gómez F., Martínez F., Romero E. Predicting Complex Patterns of Human Movements Using Bayesian Online Learning in Medical Imaging Applications. Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques. Eduardo Romero and Fabio González (eds), Medical Information Science Reference, pages 283-306, 2009.*

- *Gómez F., Villalón J, Gutiérrez R., Romero E., Finding Regions of Interest in Pathological images: an attentional model approach. SPIE Medical Images, Proceedings of SPIE Volume 7260, 2009.*

- *Gómez F, Romero E., Texture Characterization using a Curvelet Based Descriptor. Proceedings of the 14th Iberoamerican Conference on Pattern Recognition: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Lecture Notes In Computer Science; Vol. 5856, Pages: 113 - 120, 2009.*

- *Martínez F., Gómez F., Romero E., A Kinematic Method for Computing the Motion of the Body Center-of-Mass (CoM) During Walking: A Bayesian Approach. Computer Methods in Biomechanics. Accepted. 2010.*

- *Gómez F., Marín D., Romero E., A Soft-cache Strategy for Pathologist's Navigation in Virtual Microscopy. Microscopy Research and Technique. Accepted. 2010.*

- *Roa-Peña L., Gómez F., Romero E., An experimental study of pathologist's navigation patterns in virtual microscopy, Submitted to Human Pathology.*

- *Gómez F, Romero E., Rotation Invariant Texture Characterization using a Curvelet Based Descriptor. Submitted to Pattern Recognition Letters.*

- *Gómez F., Gutiérrez R., Romero E.. A Supervised Attentional Model for Finding Regions of Interest in Basal Cell Carcionoma Images. Submitted to Journal of Visual Communication and Image Representation.*

1.4.2 Local Journals, Book Chapters and Peer-Reviewed Conferences

- *Marín D, Gómez F, Romero E. Virtual microscopy based on a prefetched and soft cache strategy using the JPIP protocol. SIB-SIPAIM: V Seminario Internacional de Procesamiento y Anlisis de Imágenes Médicas, 2009.*

- *Martínez F, Gómez F, Romero E. Video analysis for human motion estimation: a review. MED Vol 17 95-106, 2009.*

2

Virtual Microscopy in Medical Images: a Survey

Romero E., Gómez F., Iregui M., Virtual Microscopy in Medical Images: a Survey, Modern Research and Educational Topics in Microscopy, Formatex, 2007.

Abstract *A major objective of the present chapter is to provide a comprehensive vision of the most recent methods for the construction of a Virtual Microscopy Viewer. The Virtual Microscopy is the microscopy area that provides a realistic digital emulation of a conventional light microscope and the VMV is the software tool that provides such emulation. Construction of a mega-image by stitching a sequential set of microscopic fields of view is the first step toward a useful VMV development. Once the mega-image is assembled, this huge amount of generated data should be compressed and stored in hard disk for later recovery. Finally, whether this mega-image is compressed or not, data should be suitably accessed for navigation. The reviewed approaches are herein classified according to their role in VM, that is to say, at the stitching, storage and navigation phases. Main contributions, advantages, and drawbacks of the current navigation methods are presented and discussed, as well as the outlook for future research.*

2.1 Introduction

Microscopic pathology is a subdiscipline of pathology, devoted to the study of the disease manifestations which are visible only at the histological level. The main core of the present disease knowledge comes up with the His-

tology development, a period spanning the nineteenth century and part of the twenty. The former light microscope is nowadays part of a large set of observation tools which together with the available number of dyes, allows precise exploration of cellular and sub-cellular microstructures. Microscopical examination of tissues requires a biological sample is cut into very thin slices which are deposited on glass slides. Once this tissue lies down onto the slide, a staining process highlights the relevant cellular information either by specifically staining nuclei or cytoplasm organelle. A very recent field, known as Virtual Microscopy (VM), makes possible digital exploration of these histological slides [104] as well as information archiving, tissue or cellular quantification and easy access to the information using modern communication resources. This technology shall be in the very near future a useful tool in most routine microscopical applications. Theses mega-images are constructed by a sequential capturing process which results in the called virtual slide which makes possible a so far unknown information availability for image retrieval in case of latter studies, medical training, distribution by electronic media, image exchange between pathologists, annotation capabilities and morphometrical measurements [21, 48, 104, 81, 117]. Overall, these virtual slides are high resolution images whose visualization requires specialized software: the Virtual Microscopy Viewer (VMV), a specific tool devised for running over images composed of thousands of microscopical fields of view (FOVs). Efficient navigation strategies within such virtual slides should take into account the multiple disk accesses and network latencies for locating, extracting and processing the requested information. Minimal requirements for this kind of viewer are: spatial random access to image information, representation of the different magnification levels and an adaptable interface to the user needs.

Construction of such a VMV implies to solve different kind of problems, whose nature depends on the different involved steps: assembling the mega-image, efficient storage and rapid information availability for navigation. These three general requirements define three complementary processes:

- Stitching. The different microscopical FOVs, result of the digitalization process, must be assembled together into a single high resolution image [9], a process which involves registering strategies for finding overlaps between neighboring FOVs.

- Storage. Clever storage strategies are needed since the virtual slide demands a huge amount of memory space. For this reason it is required

an intensive use of image compression methods [128], but subjected to the restriction that such methods must also allow an efficient access to information when required. Additionally, compression in medical images must be lossless since minimal distortions may lead to a false diagnosis [106].

- Navigation. This is the process which permits a user to carry out a microscopical examination of a particular sample as it would be possible using a conventional microscope. This virtual microscope must allow sequential and random translational movements at any of the xy-axes or zooms when moving along the z-axis [51].

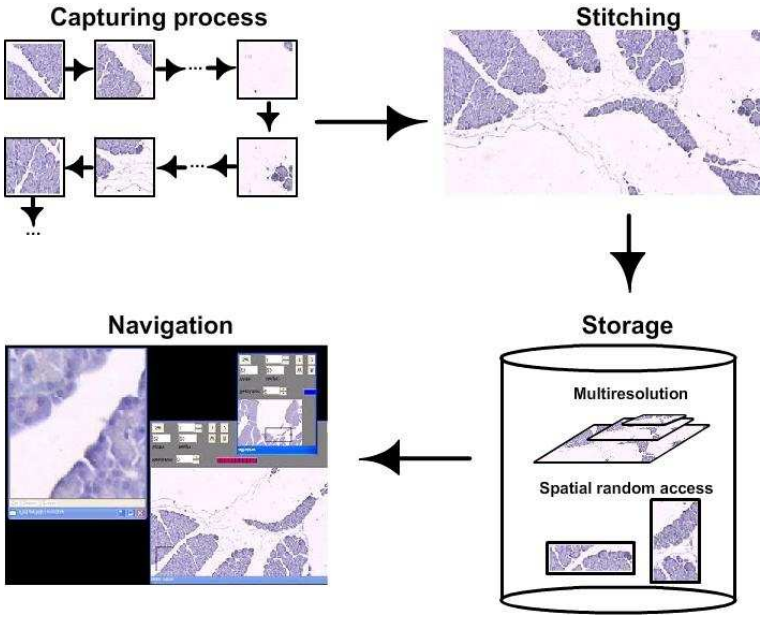


Figure 2.1: The whole scheme illustrates the Virtual Microscopy process: first a sequence of microscopical FOV is assembled into a mega-image, which must be stored because of its large size. Finally, the navigator must facilitate visualization of different windows of interest at different enlargements and with variable spatial displacements.

In conclusion, virtual microscopy requires a VMV which is able to integrate these three different processes. The main objective of this work is to give a brief introduction to some of the different strategies used in VMVs.

2.2 Stitching

A complete digitalization of histological slides is reached using a Whole-Slide-Imaging (WSI) device [116]. These systems must be capable of digitizing slides at any magnification and any desired resolution. The whole system is usually composed of an optical microscopy system, an acquisition system and a software that controls the scan process, as illustrated in Figure 2.2.

Figure 2.2: A Whole-Slide-Imaging is basically composed of a conventional optical system to which a motorized device is somehow adapted for automatic control of the stage. Likewise, a digital imaging system allows capturing the observed microscopical FOV.

The acquisition system is constituted of a digital camera provided with a good charged coupled device (CCD) sensor, a motorized stage which is controlled by some electronic device and a high resolution monitor for visualization of the digitalized FOV. Acquisition of microscopical FOV is usually performed frame per frame, following a particular order in the slide, i.e., most cases from upper left to the lower right corners (Figure 2.1), a frame of coordinates which is usually set on the computer screen for lower magnifications of the microscopical image to be captured. As a general rule, the capturing frame is overlaid with its neighbors in order to avoid possible information

losses.

The output of this process is a set of images with overlapping frames, corresponding to the virtual slide. Although the digitalization process is simple, several sources of errors may come out in the process such as variable illumination conditions between different FOV, geometric deformations due to the radial camera distortion and aligning errors because of the microscope stage backlashes [119], resulting in variable seams between neighboring FOVs.

Adequate microscopic virtual reconstruction of a desired part of a biological sample is achieved using image registration and stitching. In the VM context, registration is the process of finding the amount of overlay between two neighbor frames by maximizing a particular similarity measure between them. Two kinds of similarity measures have been used in VM systems:

- Area based methods. These measures are based on the similarity of intensities between the two neighbors FOV, using their intersected regions. They are based on low-level image intrinsic properties and therefore they are very sensible to the type of noise.

  - Sum of squared differences. Thévenaz et. al. have used the sum of squared differences as similarity measure in a VM system [136]. This measure has shown to be appropriate in many applications since it is simple and optimal under controlled conditions, i.e., when differences between images are exclusively caused by Gaussian noise [144]. However, inter-image intensity variations are mostly linear in histological applications and constitute the major source of noise [9], together with the unavoidable biological variability and the technical difficulties of any histological procedure.
  - Correlation. In routine microscopy, illumination settings are controlled in such a way that most changes regarding intensities between neighboring FOVs can be modeled as linear [68]. Therefore, similarity measures based on correlation such as the normalized cross correlation or phase correlation [120, 14], result more robust and become also more general. They are remarkably less sensible to noise than simple measures at the level of pixel differences such as the sum of squared differences and they are also robust to image displacement or rotations produced by microscope stage instabilities.

- Feature based methods. These approaches are based on the detection of salient features in the image intersection which can be used in a general manner [158].

  - Corners. Sun et al. [129] have proposed a method which finds a set of corners in the overlapping region of each field of view, based on the Harris corner detector method [65], which is followed by a match of the corresponding features. In this case the similarity measure is the Euclidean distance between the corresponding features. Although the method is rapid, this is not general or robust since there are no guarantees for the corners to exist in every microscopical image.

After a similarity measure has been set, the registering phase consists in finding the optimal transformation function which maximizes that similarity between neighboring FOV. Notice that a large image must be generated by registering hundred or thousand FOVs, whereby optimal registration schemes are required. A simple naive registration of one image with its neighbors can result expensive from a computational standpoint because of the number of needed registrations. An efficient strategy consists in registering a couple of neighboring images, which then form a new image. This new image is then registered against a third consecutive image to construct a new image and so on along the selected row. The process is performed in parallel until rows of images constitute individual images which are then registered to generate the final mega-image. Rankov et al. [112] proposed to start at the center of the digitalization framework and to follow a spiral-like pattern, under the hypothesis that the image at the center contains the higher information, as illustrates Figure 2.3. Appleton et al. [9] aligned simultaneously rows of FOV, while images associated to each row are firstly stitched into one single image using an efficient dynamic programming algorithm for solving the optimization problem. Thévenaz [136] developed a method for refining a rough stitching, starting with an initial conventional stitching. During this phase, images are ordered after the intersection surface between them and then aligned following this order.

After an optimal displacement is found, it is quite frequent that visible seams persist between two neighboring FOV. The stitching process must then correct the seam between two FOV by modifying the intensity values within the boundary of the overlap region. For so doing, Iregui et al. [68] proposed

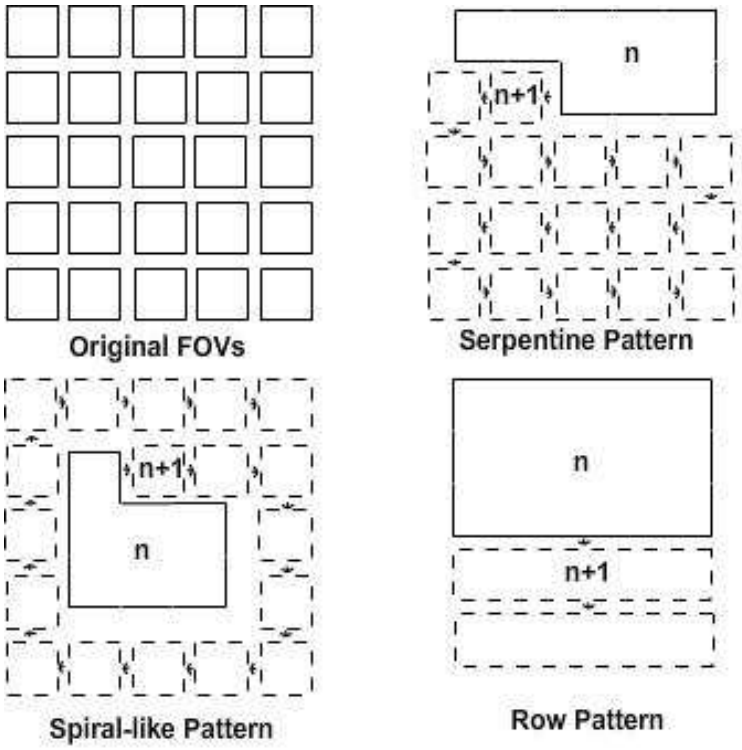


Figure 2.3: A sequence of microscopical fields of view is assembled together into a single large image, as illustrated in upper left panel. Different stitching schemes are showed in the other panels: a serpentine pattern in the upper right panel, a spiral-like pattern in the lower left panel and a structural generation for which the large image is constructed by aligning rows of FOV.

to apply a Gaussian filter on the two neighboring rows of pixels to smooth them out. Rankov et al. [112] and Thévenaz et al. [136] calculated a weighted bilinear interpolation of the two overlaid images. A pixel value is calculated as an average of its values at both overlaid images, weighted by their distances to their closest image edge. Correction of the illumination differences in different FOV is achieved in Sun [129] by weighting the intensity values with a second order polynomial which attempts to compensate the intensity differences between every pair of neighboring images and approximate the corrected illumination to the mean intensity value. Although differences are corrected using this low pass filtering scheme, it is very difficult to ensure that also

relevant information could not be hidden.

2.3 Storage

A virtual slide is a high resolution image, for instance a typical digitalization of a 1 cm^2 glass slide using 20× objective, results in 64×64 FOVs [68]. Provided that FOV are digitized to a resolution of 720×520 pixels, it is produced a reconstructed image of 45000×32000 pixels and 4.3 Gbytes. The demanded high resolution of the virtual slide leads to several problems:

- Large storage requirements. Although the cost of devices for storage is lately falling while their capacity is increasing, the massive application of VM would require an unthinkable amount of storage. A specialized hospital produces between 100.000 - 500.000 histological slides every year [122]. The storage of only a 10% of these preparations, suppose the hospital can count on at least 50 Petabytes [122]. It is then mandatory to incorporate efficient compression techniques into the storage procedures.

- Lossless compression. Image compression can solve the problem of efficient storage, but it is important to take into account some particularities of medical images. Lossless compression techniques allow exact reconstruction of the original image and avoid annoying distortions introduced by the broadly used loss compression approaches. Overall, in virtual slides of pathology the lossless compression is preferred by several reasons: firstly, it is not easy to reach a consensus about an acceptable quality loss since this is based on exclusively subjective criteria. Second, loss compression can lead to legal disagreements as for instance a controversial diagnosis which could be based on artifacts produced by the compression, or in case of a malpractice suit, if only the compressed version of the image is available [106].

The storage problem for VM has been approached using two different strategies: uni-resolution and multi-resolution formats.

2.3.1 Uni-resolution image storage

The virtual slide or mega-image is subdivided into smaller images called tiles [21] and each is stored in a classical format, such as JPEG, TIFF, or RAW

[90], as is illustrated in Figure 2.4. All these sub-images or tiles are stored at the original virtual slide resolution. This virtual slide partition facilitates spatial random access to certain regions of the virtual slide. However, generation of an entire virtual slide thumbnail requires both access to every tile and data down sampling. Chang et al. [21] developed a uni-resolution VM system which runs in parallel architectures. This system spends between 1s - 8s to perform extraction of different Windows of Interest (WoIs) at different resolutions. When the requested WoI intersects two or more tiles, the size of the initial tiles becomes definitive, i.e., the largest the selected tile the more time the system spends. Fontelo et al. [48] compared a real microscopy system with a uni-resolution VM system using surgical pathology specimens commonly encountered in a University Hospital. This work shows an 88% agreement level when comparing both systems. However, this study reports low satisfaction rates when pathologists navigates using low band channels (modem channel - 56 kbps).

2.3.2 Multi-resolution formats

The multi-resolution formats [88] are characterized by access to different resolutions of the image which is compressed in a unique file.

Pyramidal formats

In these formats, the high resolution image is subdivided into spatial tiles which will be used for generating the whole image at different resolutions (see Figure 2.4). After an initial tiling, image versions of multiple resolutions (different levels of enlargement) are obtained from each tile and each one is stored in JPEG or TIFF formats, for lossy or lossless compression, respectively. FlashPix [66] is the typical example, this format provides spatial random access to the image data using the tile as the information unit and multiple zoom levels, including the thumbnail. However, storage requirements are very high since multiple versions of the high resolution image are independently stored into the same file. This type of information management is not optimal in cases in which the requested region overlaps four tiles and therefore the four tiles must be decompressed. In this kind of applications, a compromise must be reached between the size of the tile and the amount of available information. Fred et al. [30] evaluated the utility of the VM as a complementary learning tool during a Cancer Workshop using pyra-

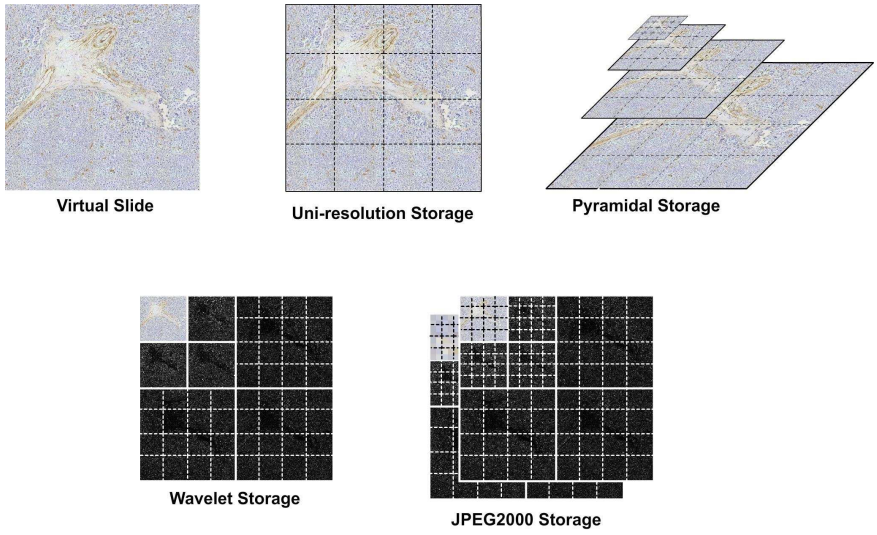


Figure 2.4: Different storage strategies from the original virtual slide. Thick-line corresponds to the spatial subdivision into sub-images or tiles. The Figure illustrates also different available storage formats. Uni-resolution allows only spatial random access. Pyramidal storage provides spatial random access at several magnifications, but multiple versions of the image are needed to be stored for reconstruction. Storage based on wavelets needs a unique version of the image for spatial random access at various magnifications. Finally, J2K uses all wavelet advantages and introduces the additional property of progression in quality.

midal formats. Results show that VM is most effective when compared with traditional microscopy, but this study fails to present storage results. Mikula et al. [95] developed a VM system for display of high resolution brain maps and atlases. The store strategy supports up to 35 Terabytes of information using a standard lossless JPEG compression.

Wavelet based formats

Zhang et al. [145, 156] used the Haar wavelet [142] for efficiently storing large resolution images. The proposed algorithm starts by separately processing each component of the RGB image and is as follows:

1. Computes the average of 22 non-overlapping blocks.

2. Differences between these averages and the original 22 blocks is calculated and stored.

3. A new image with every calculated average is constructed.

4. Repeat steps 1-2 on the new image of step 2 until the size of the low resolution version achieves a desired dimension.

5. Store the last low resolution version in a separated file.

The average values correspond to a low frequency version of the image while the differences keep information of details, that is to say, image high frequencies. Every frequency sub-band is stored in a separate file. Spatial random access is reached by splitting the larger high frequency files into several files, each corresponding to spatial blocks of wavelet coefficients, as shown in Figure 2.4. After the image is decorrelated using this wavelet transformation, every file is codified using a Huffman coder. Lossless compression is achieved using integer arithmetic. The Haar transform allows multiple enlargement versions of the mega-image from a unique representation. Besides, provided that wavelets are local operators, the use of the spatial sub-division of the coefficients guarantees a spatial random access. Yet this approach is suitable for many VM applications, a great limitation is that there is no a progressive quality representation which then results in a major drawback for image transmission and seamless navigations.

JPEG2000 (J2K)

J2K [131, 111] is a compression standard developed by the Joint Photographic Expert Group and is based on the Discrete Wavelet Transform (DWT) and the Embedded Block Coding with Optimized Truncation (EBCOT) [130]. The compression algorithm works in different complementary steps as detailed in Chapter 3.

Wildermoth et al. [15] proposed a VM system with an extended depth of field, which for the registering phase uses a normalized cross correlation similarity measurement and J2K for storing. However, the least information units for decompression are the tiles, which correspond to the primary J2K image partition that could not be optimal for navigation since they need to be large enough for avoiding distortion effects at low bit rates [111]. This study fails to exploit the minimal coding unit allowed by the standard: the precinct [150].

Iregui et al. [68] developed a VM system which allows progression by quality and uses strategies such as cache and pre-fetching for accelerating navigation. The storage uses the J2K algorithm in mode lossless, while extraction time amounts to 500 ms in average for real pathology navigations, using the precinct as the basic information unit. Additionally, this study reports that a correct tuning of the compressor parameters reduces the size of the file up to a 20%.

2.4 Virtual Microscopy and Modern Communication Scenarios

Conventionally, virtual microscopy systems are implemented as a client-server architecture [116], basically because this configuration is based on a central processing-storage and therefore clients with low computational resources can access to such level of information. Unfortunately, current implementations hardly will meet modern client requirements [95], mainly because the communication velocity results limited for the high spatial resolution of these images. In client-server architectures, the client runs on an end-user's laptop which allows exploration over the VS while the server provides storage, retrieving and processing of the huge VS data. Many client-server architectures have been proposed in the literature for VM, its goes from simple storage strategies, based on a unique server [100], to parallel computers with high computational resources [8]. Most of these approaches have been proposed for a low number of pathologists, navigating a reduced number digital slides, i.e., they do not take into account the use of network resources. Lately, there has appeared a growing trend for digitizing the whole tissue slide as part of the clinical routine in pathology [148]. Moreover, many novel educational scenarios require a concurrent VS exploration. In conclusion most actual scenarios require an intensive use of the available network resources. This limitation would be addressed by using the J2K high scalability [150]. It allows flexible access to the virtual slide in several image resolutions, spatial regions and quality layers [150]. The J2K standard defines a representation composed by a wavelet transform followed by a scalable codification using EBCOT to achieve this flexibility. Unfortunately, this standard does not specify how to access to J2K data stored in a remote way. A major question that must be addressed in order to have real and useful applications, such as

virtual microscopy over network environments.

Several computational architectures had been proposed to access high resolution J2K images in a remote way [3, 36, 102]. Most of them aim to exploit the scalability property to consume less bandwidth, by sending small image units such as tiles, packets or precincts, rather than the complete image. The JPIP protocol is defined in the part 9 of the J2K standard is an evolution of the JPIK protocol proposed by Tubman [133]. The associated J2K image communication protocol (JPIP) was devised to save work at the client side by concentrating the maximum processing at the server side [3]. Under this protocol, the client requests only WoIs and the server brings back large structural units such as tiles or precincts, who are finally decompressed. Microscopical navigations demand more flexible information representations because of the complex exploration paths, namely, series of back and forth direction changes, sideway movements, different magnification requests, all at different speeds and times. Therefore, any design must be as adaptable as possible to these complex requirements. Recently, an extension to the JPIP protocol, called JPIP-W, has been proposed [102]. This extension improves the J2K image transmission by means of a proxy cache that provides faster delivery by exploring the content redundancy of the navigated areas. Deshpande et al. [36] proposed a Web oriented architecture based on the HTTP protocol. Their solution consists in creating an index file, associated to every J2K image file. This file contains the byte range corresponding to a J2K image unit. This file can be downloaded from a client, which can use it for accessing the appropriate image parts by using the byte-ranging feature offered by HTTP/1.1. This approach is very flexible, requires a simple implementation in the clients and provides a full integration in Web Systems. The major limitation of this proposal is that clients must to have previously the complete index before accessing data, i.e., it is required to wait before to show any reconstruction. All these architectures are suitable to remote access to the image data, however the J2K decompression process introduces time latencies, because the standard complexity, that would affect the navigation velocities. For this reason many research efforts have been oriented to deal with this problem [67, 33, 18, 151, 41]. Two interesting approaches [67, 33], which are closely related to the proposals herein presented were developed in the UCL /TELE laboratory, these two doctoral works exploit the dynamic nature of navigation in high resolution images to propose novel cache and prefetching approaches in J2K adapted to the problem of navigation of natural images [134]. The work herein developed would be interpreted as a

natural application and extension of these techniques to the specific domain of virtual microscopy in pathology.

2.4.1 Cache Strategies in Virtual Microscopy

The cache problem in virtual microscopy can be twofold addressed: by data granularity, i.e., the level flexibility of data representation in the cache, and the cache replacement policy, which should specify how data must be replaced when a new image portion arrives. In terms of granularity, the simpler strategy may be to store the lower image resolution levels [8, 68] so that a fast access to WoIs at specific zoom levels is reached, but at a high memory cost. Besides, the size of the cache information units to be stored is important for devising a preference model. If these units are small, a large image WoI can be more probable than another with less elements of higher size [151]. Furthermore, cache elements with larger size, associated to relevant information [134], can yield smaller distortion results when they are queried [134]. Another factor, which influences the weight given to this basic cache unit, is its associated cost [8], expressed in terms of processing or transmission time. Therefore, a proper balance for a particular application should be maintained. An alternative to this choice may be to store complete portions of the coded image, whereby the granularity level is determined by the compression format. Overall, JPEG is the broadly used format in many virtual microscopy applications [18, 151, 41]. Nevertheless, this format does not provide access by resolution and introduces quality losses in the final result, an unacceptable issue in most diagnostic tasks [106]. In contrast, J2K is a more flexible alternative [68, 102], provided with random spatial access at different levels of resolution, quality, and lossless codification.

The cache replacement policies should be based on the user preferences, that is to say, elements with lower preference levels should be firstly removed, but only when the cache is full. The most popular cache replacement policy in virtual microscopy has been the Last Recently Used (LRU) [8, 68], for which the user preference is modeled as a function that decreases with the element age. The underlying hypothesis in this model is that the pathologist will revisit, with higher probability, the image zones that she/he just visited, a navigation pattern that very rarely is observed in actual navigations [79], above all when the pathologist reaches a particular degree of expertise. Another broadly used strategy is the Last Frequently Used (LFU) [102, 151, 8, 68], whose fundamental assumption is that higher probabilities

are given to those regions with larger number of visits. These two strategies turn out to be quite general for a very oriented domain as virtual microscopy, for which times associated to navigations are straightforwardly related to expertise. In the context of the present investigation, both strategies can be considerate as comparable since experimental evidence in multiclient environments reports similar performances with LRU and LFU, for queries from multiple applications and variable input/output requirements [64]. Finally, Iregui et al. proposed [67] a packet importance by using information, such as their location and resolution. It is interesting to point out the importance of prefetching strategies for accelerating the navigation. In this type of strategies the times in which user does explore the image are exploited to preload image data. Two main proposals had been proposed. Firstly, a dynamical rate-distortion scheduling [132] that maximizes the received image quality within the Window of Interest, at each point in the transmission. Later, the anticipated prefetching proposed by Descampe et al. [35] where is demonstrated that priori knowledge about future navigation patterns may help to improve the overall reactivity of the navigation system.

2.5 Navigator

Even in the larger monitor screens it is impossible a complete display of the virtual slide since its resolution is much higher than typical resolutions supported by conventional display devices, below 2000×2000 pixels. In consequence, it is essential a design of methods for efficient access to the image data regarding the different dimensions of the problem: spatial displacements and enlargement representations. A Graphic User Interface (GUI) in a VM system should emulate a microscopical examination performed by an expert in a real microscope. In general, this design should exploit the importance of low magnification for exploration and high enlargements for navigation. Typically, the navigator consists of two windows: (i) a thumbnail version of the mega-image, in which a re-sizable window is displaced and used to define a particular WoI and (ii) the window in which the larger resolution is displayed. Several approaches have been introduced in VM systems. A rough approximation has been to bring together this WoI into the whole system by a second re-sizable window with the original resolution [122]. Iregui et al. [68] proposed the use of a second window for displaying intermediate resolutions, while higher resolutions are displayed in a third window. This

WoI can be configured by size, resolution and quality [21, 68]. The expert can then request a rectangular area at any resolution and any quality while the WoI can be displaced for exploration. Mikula [95] proposed an interface for faster visualization of stacks of virtual slides by pyramidal resolution representations which are stored using a quadtree structure [46]. Quadrants are related by their parent tile. This method displays with higher resolution the quadtree nodes whose pixels are closest to the screen.

2.6 Acceleration of Navigation

Spatial locations, decompression and visualization unavoidably introduce considerable response delays, which make impossible interactive and fluid navigations [116]. Strategies such as the cache or the pre-fetching have been developed for decreasing the latency times and therefore to permit a fluid navigation. Cache is a rapid access to a space of memory in which it is stored the portions of the image that shall be visited in the future [29]. Prefetching is the anticipated uploading of those parts of the mega-image to which the navigation will be addressed in the future [34]. Those techniques have shown to highly improve navigation times [19].

Spatial cache [68] is a reserved part of memory, which is set to store visited pixels. When a WoI is requested, the algorithm calculates the intersection between what is stored in the cache and what the WoI is demanding, and this intersection is displayed directly from the cache. In zoom-in operations, Catalayeuk [18] used information of the thumbnail for a temporal display while the rest of data are being loaded. Iregui et al. [68] used the multi-resolution nature of the DWT while the cache was dynamically constructed by storing the wavelet coefficients of lower resolutions, which had been already visited. This strategy is a soft cache [141], which maintains image wavelets coefficients of low resolution versions in memory for reusing them to construct high resolution versions of the image [141]. Using these two strategies, navigation velocities grows up to a 30 %.

2.7 Future Works

VM is an incipient area with multiple open problems and a great variety of applications. So far the state-of-the-art technology has allowed development

of useful prototypes with some critical limitations, which have restricted a broader use. Future navigators would require faster registering methods for handling the huge amount of data generated from each particular application. These navigators also require flexible storage methodologies with optimal compromise between a quantity of stored data and image processing for reconstructing requested pieces of images. Acceleration of the navigation turns out to be a critical factor in seamless navigation either when images are locally stored or must be remotely accessed from a local client. VM can not be a real option until it is fully reliable, efficient and easy to use when accessing sets of mega-images. Main open problems which need new approaches are a more flexible access to image data, efficient indexing of data for rapid and opportune retrieval, compression strategies able to adapt to this kind of applications and seamless navigation methods. J2K has introduced a preliminary step toward novel navigation proposals; the paradigm has been changed into a fast, easy and optimal access to the image data rather than simple compression policies. New techniques must easily allow adaptation of interfaces for the use of experts and analysis of the image contents for selective compression or selection of a particular multi-dimensional representation which can be set after a matching pursuit [89] assessment. These new approaches must also include decompression parameters which must permit integral tunings for increasing benefit of information packets from quality and utility standpoints. Furthermore, in cases in which VM is used in server-client systems, it is fundamental to count on optimal strategies for managing information when the band channel is narrow as it will always be when one considers the amount of data that may be generated. Cache and prefetching have shown to be efficient strategies for accelerating navigations in VM, but more sophisticated methods are nevertheless needed for reaching real time performances in those navigations.

Part I

JPEG2000 for Navigation of Large Medical Images

3

Strategies for efficient virtual microscopy using JPEG2000

Iregui M., Gómez F., Romero E., Strategies for efficient virtual microscopy using JPEG2000, Micron, Elsevier. Vol 38, 2007.

Abstract *This chapter describes the design, implementation and validation of a new strategy for efficiently browsing large microscopical images (mega-images). The mega-image is constructed by registration a sequential set of microscopic fields of view, compressed and stored in hard disk using the J2K standard. Navigation is accelerated by fully exploiting J2K properties through the introduction of a cache strategy and an optimal delivering of quality information. Cache is introduced at the level of the spatial and resolution dimensions while optimal delivering is implemented on the organization of minimal information units. Navigation with the conventional use of J2K results in extraction times of about* 500 *ms. We show that these strategies can improve navigation velocities up to a* 30 %*, while we can efficiently represent high-quality and high-resolution colour images of microscopic specimens.*

3.1 Introduction

In this chapter an innovative J2K-based virtual microscope was designed, developed and evaluated. Virtual microscopy was here three-fold approached by developing registration, store and navigation modules. Firstly, mega-images are automatically constructed with a standard registration method, using similarity measures such as the normalized cross correlation, phase registration and mutual information. These mega-images are then compressed and

stored, using the J2K standard with an appropriate setting of spatial, resolution and quality parameterizations. For navigation, the structure of the J2K codestream is conveniently accessed for obtaining specific regions of the image. Besides, a cache strategy is designed and implemented for improving navigation velocity. The entire system is a user friendly interface which allows the pathologist to navigate through mega-images as he would by using a microscope. Main contribution of this work is an exhaustive analysis of the virtual microscopy relevant aspects so that multiple recommendations are herein presented on using each of the virtual microscopy modules, based on a detailed study from different perspectives by exploiting up-to-date tools. This chapter is organized as follows. In Section 3.2 we describe the problem which is approached from the standpoint of these three issues: registration, storage and navigation. Section 3.3 describes a general methodology for constructing a mega-image, as well as a complete study of all the compression, storage, displaying and navigation issues, focusing on the capabilities of the J2K standard. Section 3.4 presents the results and finally Section 3.5 is dedicated to discussion and conclusions.

3.2 The Problem

A virtual microscopy system comprises four components: registration, storing, displaying and navigation. A mega-image construction implies a registration problem since digitized images are not perfectly aligned [15]. In addition, this virtual representation demands huge amounts of information, which must be navigated at different resolutions, displayed for different WoIs, efficiently stored and easily accessed. Finally, navigation and visualization should manage adaptable, scalable and friendly interfaces.

3.2.1 Registration

Under the assumption that the capturing process introduces only translational shifts, the problem is to find the optimal spatial transformation so that images match with a minimal registration error. For so doing, a small window of one image is used for feature matching into the other, using similarity measures between spatial regions. Amongst the most common similarity measurements for registration are normalized cross correlation [120], phase correlation [14] or mutual information [144]. In normalized cross correlation

Figure 3.1: Some properties of the J2K.

one small image $T \subseteq I_2$, called the template is matched to a larger image I_1. The cross correlation is then defined as:

$$C(u,v) = \frac{\sum_x \sum_y T(x,y) I_1(x-u, y-v)}{\sqrt{\sum_x \sum_y I_1(x-u, y-v)^2}}$$

Phase correlation is based on the Fourier Theorem Shift, computes the cross-power spectrum of I_1 and I_2 and find the peak of its inverse.

$$\mathcal{Q}(u,v) = \frac{\mathcal{F}(I_1)(u,v)\mathcal{F}(I_2)(u,v)^*}{|\mathcal{F}(I_1)(u,v)\mathcal{F}(I_2)(u,v)^*|} = e^{2\pi i(u\delta_x + v\delta_y)} \tag{3.1}$$

where $*$ is the complex conjugate and $\mathcal{F}(I)$ is the Fourier transform of I. Mutual information, coming from information theory, is likely the more frequently similarity measure used in multimodality problems. The mutual information between two random variables X and Y is given by

$$MI(X,Y) = H(X) + H(Y) - H(X,Y)$$

where H is the entropy of the intensity distributions [144].

3.2.2 Compression and Storage

J2K [2, 111] is a compression standard developed by the Joint Photographic Expert Group and is based on the DWT and the EBCOT [130].

This standard supports lossy and lossless compression, progressive recovery by quality and resolution as well as random access to specific regions of any size, as illustrated in Figure 3.1. It is then reasonable to exploit such versatility for storage and navigation of large images, in particular of histological images.

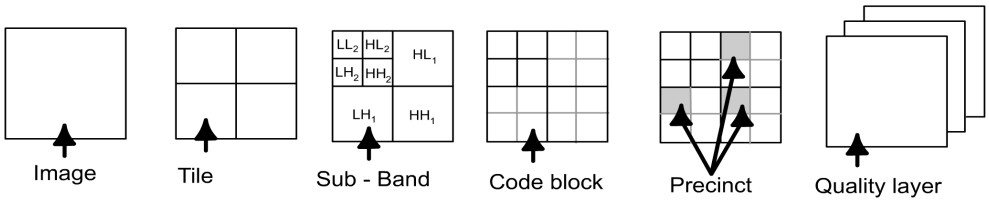


Figure 3.2: Structure of J2K, L and H are the low and high pass filters, respectively.

Previous to compression there is a pre-processing phase, which prepares the image for transformation. Firstly, the image is split into rectangular tiles or image regions, particularly useful for memory management. The sample values are then level shifted to make its value symmetric around zero and simplify implementation issues. After a level-shifting on the image values, each tile - usually in RGB format - is decorrelated into luminance and chrominance components (YUV) by means of a reversible or irreversible color transform. This phase allows an independent access to each component of the image. Afterwards, a DWT decomposes the input signal into frequency bands called *subbands* (see Figure 3.2), providing a multi-resolution image representation. The bank of filters L and H splits the signals into two levels of resolutions, each with low and high frequencies. Two filters can be chosen for transformation: either the Daubechies 9-7 which is adopted for lossy compression or the reversible Daubechies 5-3 for lossless compression. After an optional quantization, the next step is the entropy coding of each tile with the EBCOT. The DWT coefficients of each subband are subdivided into small blocks called *code-blocks* (codeblocks). Each codeblock is composed of bit-planes ordered by significance levels, obtained from the coefficient binary representation, as illustrated in Figure 3.3.

These bit-planes are encoded in significance, magnitude and clean up coding passes, with an arithmetic encoder which provides the final compression the so-called MQ coder [131]. The coding passes provide several truncation points of the bit-stream. The encoded codeblocks of each tile-component are then assigned to different *quality layers* by using a Post-Compression Rate-Distortion Optimisation (PSRD-opt [130]).

Briefly, each bit-plane is encoded in three passes introducing independent bit-streams. A convex hull analysis [131] permits to obtain the N_i optimal

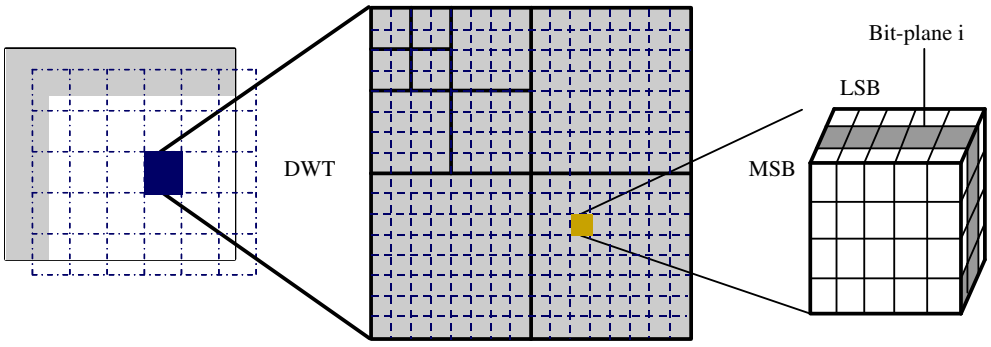


Figure 3.3: Codeblock structure by layers of bits.

truncation points for each block β_i, each entailed with a slope $\frac{\Delta D_i^{(t)}}{\Delta L_i^{(t)}}$, where $\Delta D_i^{(t)}$ defines the distortion gain when coding a particular bitstream associated to one truncation point t and $\Delta L_i^{(t)}$ is the subsequent increase in bytes. Distortion is here proportional to the cumulated Mean Square Error (MSE), calculated from the coefficients of the original wavelet and those obtained when decoding until a particular truncation point. The monotonically decreasing organization of the overall truncation points in the image defines an optimal bit-stream partition for threshold rates or quality increments: the layers of quality. Each of the quality layers q is formed by including incremental contributions of coded bytes from different codeblocks and is defined by distortion-length slope thresholds T_q. This layering strategy has the property that truncating the code-stream to any whole layer boundary results in a compressed representation, which is optimal in the rate-distortion sense.

Once EBCOT is applied, each block-stream stands for a spatial region in a particular subband and is represented in several quality layers. In order to obtain a codestream with embedded information about resolution, region and quality layer (see Figure 3.1), the algorithm introduces the precinct and packet concepts. A precinct can be defined as a block grouping from different sub-bands (subbands) with the same resolution that is mapped to a specific region of the image (see Figure 3.2). A precinct from a specific tile, component, quality layer and resolution that appears in codestream units, is called a packet, which is the basic unit of the J2K codestream. The J2K structure is shown in Figure 3.2.

Finally, the reconstruction problem can be defined in terms of packets.

Respecting the constraints imposed by the standard, these packets are puzzle pieces that are dynamically used to attend navigation requests.

3.2.3 Displaying and Navigation

Navigation can be considered as a dynamic interaction between three complementary variables: quality[1], WoI (Windows of Interest) displacements and zooms (resolution changes). When a particular WoI is requested, the navigator starts to sequentially uploading quality layers for the pathologist sees how the image is progressively improving. This uploading is stopped when the expert decides to change the WoI. Unlike the quality progression, both resolution changes and space displacements are triggered by specific requests. Navigations in virtual microscopy are characterized by either examinations at different magnifications or by spatial explorations, issues provided by the J2K characteristics concerning flexibility and granular decoding. A natural navigation should permit a dynamic selection of image regions by panning, i.e., a selection of adjacent regions, while the system should also allow flexible resolution changes (zoom-in and zoom-out). In terms of J2K it is equivalent to progressively decode J2K packets for upgrading resolutions. In general, when pathologists start a navigation, they select a WoI at a low resolution and then they move onto this region at higher resolutions [18]. Very rarely, they have a random access to regions in the image but rather they use their expertise in order to have an oriented search of information. Resolution in navigation is introduced by the frequency decomposition given by the DWT.

3.3 Methodology

3.3.1 Image Acquisition

Three histological specimen were digitized for evaluation. The first specimen was a normal mouse pancreas which was fixed, embedded into paraffin and inmunostained as described in [124]. Microscopical fields were digitized through a Zeiss microscope coupled with a JVC KY-F58 color digital camera (Victor Company of Japan, Japan Ltd). Microscope overlap was set to a 5 % of the image dimensions. A captured grid of 64×64 microscopical fields

[1]Recall that quality is herein understood as a measurement of how similar are the displayed and the original images.

(752×560 pixels) represented an effective area of $13.114 \times 9.641\, mm$. A grid of 160×159 microscopical fields was captured from a second specimen of a neuroendocrine tumor of thyroid stained with Hematoxylin-Eosin, but overlay was to a 10%. Finally, a grid of 91×123 microscopical fields was captured from a third specimen, an atypical thyroidian adenoma marked with thyroid peroxidase, were digitized using the same microscopical settings.

3.3.2 Registration

Yet microscope overlapping is a pre-defined parameter in any capturing system, it is very likely that mechanical backlashes or motorization non-linearities result in inaccurate microscopical movements so that this overlapping is usually spilled over. Three registration algorithms were implemented in C++: cross normalized correlation, phase correlation and mutual information. A predefined region of the image, near to the border is used as a template for registration with cross correlation and mutual information. In the present work, a 5 % of the image size in either dimension was used to define overlap. Phase correlation was calculated by previously transforming the two images with a standard fast Fourier transform (FFT) algorithm and finding the maximum spectrum displacement, as defined in 3.1. After registration any pair of images, a Gaussian filter is applied on the two neighboring rows of pixels to smooth them out, and a new image is constructed by combining them.

A set of 270 digitized images (752×560 image size in pixels) were chosen in each mega-image for evaluation. This set of images corresponded to a $8.33\, mm^2$ area. The three registration methods herein implemented were twofold assessed: time performance and accuracy, using as Gold Standard the pre-defined microscope overlapping set to 10 % of the image size in either dimension. Time performance was defined as the time needed to find the Gold Standard displacement with an error less than 1 % of the image size in either dimension. Accuracy was estimated on the same set of images by calculating the absolute difference between the obtained registration and the Gold Standard microscope overlapping. The registration algorithm was allowed to run until convergence. Accuracy and time performance were also statistically evaluated: the null hypothesis was that the three registration approaches produced no difference among the means. This assertion was assessed with a two-way Analysis of Variance (ANOVA) test, using image as a random factor and the method as a fixed factor. In all tests, differences

were considered to be significant when $p < 0.05$.

3.3.3 Compression and Storage

The Jasper Library [5], an implementation of the J2K handling standard, was adapted to the virtual microscope navigation problem. Jasper is a general use library for the J2K standard and it is neither provided with particular modules for codification of mega-images nor the needed tools for continuously extracting pieces of information with different sizes (windows of interest), qualities or resolutions. Some new functionalities were in consequence added:

- **Compression of mega-images:** A driver module for coding a mega-image file was implemented. This file contains references to any image which is part of the set of digitized images as well as the capturing method used, i.e., serpentine, rows, columns, etc.. The module uploads the desired WoI for displaying and also registers the entire set of images for generating the mega-image. Finally this module also codes the mega-image.

- **Specific packet localization:** For a rapid extraction of a specific image region a J2K index construction is useful. This index contains information about image dimensions, number of components and tile size. These three values are drawn from the size image marker, which is a specific marker defined in the J2K standard [2]. This index also stores coding parameters such as the number of decomposition levels, the number of layers and the progression order, information located at the coding style default marker which is also defined in the J2K standard [2]. Finally, the packet size used for faster packet localization, is extracted from the packet header information coding [54].

- **Soft cache for inverse wavelet transform:** An additional modification was introduced to manage different resolutions by storing the wavelet coefficients of lower resolutions into the cache memory. When a particular pixel is required, if visited before at any lower resolution, it is not needed to fully decode this pixel but the stored coefficients at lower resolution are used to reconstruct it.

Compression efficiency was evaluated for different tile sizes, resolution levels and different precincts and codeblock sizes. For so doing, the compressed

file size is calculated as $8 \times width \times length \times$ number of components, where width and length are the image dimensions, the number of components stand for the number of represented channels (RGB) and 8 is the number of bits per channel. The efficiency is defined as the ratio between the compressed and non compressed file. Data are shown by calculating the number of bits which represent a pixel, that is to say, in bits per pixel (*bpp*). Three different mega-images were built up and used for this evaluation.

3.3.4 Displaying

Graphical User Interface GUI

The graphical user interface is written in C++ using the wxwindow open source library [1].

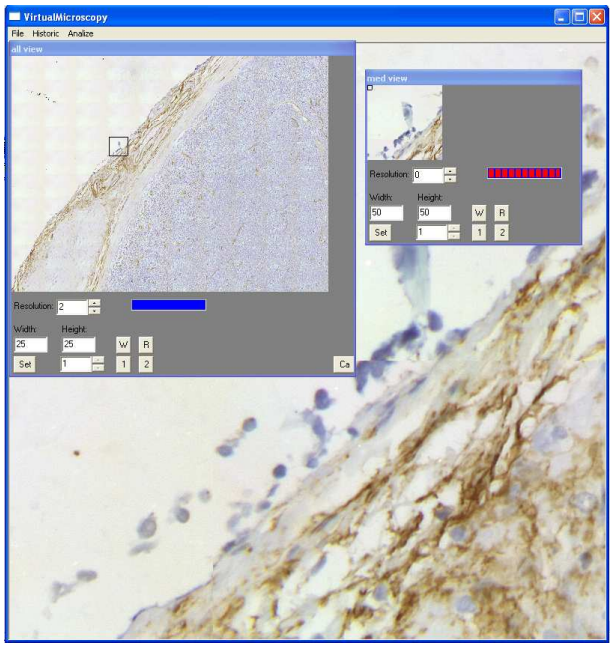


Figure 3.4: Virtual microscope prototype. This illustration reproduces a small microscope magnification (4×) in the coarse image while the high enlargement displays a 20× microscope magnification.

The GUI is devised for facilitating expert visualization and follows, as

close as possible, a routine expert examination [53]. This design exploits the importance of low magnification for exploration and high enlargements for diagnosis. Likewise, this interface pays attention to navigation into areas of the specimen with different magnifications at the same time. The interface consists of three windows: a coarse mega-image version in which a re-sizable window is displaced and used to define the particular window of interest. This WoI is shown in a second re-sizable window with intermediate resolution and quality, actually an operation which models the type of navigation performed by a pathologist. Finally, the third window shows the higher resolution and quality, i.e., the enlarged version. The WoI window is parameterizable for size, resolution and quality (see Figure 3.4). The expert can then request a rectangular WoI at any resolution, with any quality and can displace the window on the coarse image to explore the whole specimen. All algorithms were written in C++ and run under windows in a Intel Centrino processor of 1.7 GHz with 1 GB in RAM with a 80 GB hard disk.

3.3.5 Navigation Methods

Given the particular navigation characteristics and provided that there exists a high probability that a packet is required by different WoIs, we propose in this chapter a cache strategy capable to optimize the decoding phase by minimizing the data to process for a specific WoI.

The Cache Strategy

Virtual seamless navigations are nearly impossible because of both the huge image sizes and the limited machine resources. This combination unavoidably leads to annoying delays and hindering of normal operation of the computing system. Delay problems depend on several factors: image characteristics, physical resources, the used codec and the GUI. In this work, we propose a method to improve the decoding process performance by applying a cache strategy.

A cache strategy is a temporal store of data in the faster memory for future retrieval [29]. A cache mechanism is useful since this reduces direct hard disk access, bandwidth consumption, computational cost and the time for accessing relevant information. A spatial cache is used for reducing the amount of data to decode each time a new request involves adjacent WoIs, given that it is likely that adjacent WoIs overlap. On the other hand, a soft

cache [73] improves navigation through several resolutions by storing coefficients into the allocated space. The principle of this technique is based on a separate delay analysis for J2K coding and transformation algorithms. These implementation issues are not included in the decoding processes defined in the J2K standard, yet they are herein fully justified since the purpose of this technique is an efficient information handling during navigation.

Spatial Cache

Algorithms used for optimizing information exploit the fact that pathological images are redundant and that once a query is performed on a particular spatial location, there exists a high probability of querying spatially adjacent data in the next future. Adjacency is herein defined in terms of logical proximity between image pixels. Spatial cache is a part of memory which is set to store visited WoI pixels within a spatial exploration at a particular resolution. When a WoI is requested, the algorithm makes the difference between what is stored in the cache and what the WoI is demanding so that it assesses if some information is missing. Should it be the case, the algorithm extracts what is needed from the bitstream, otherwise information is directly drawn from the cache. Finally, the substitution strategy, the Last Recently Used (LRU), replaces the oldest data that has not been reused. In other words, the decoder decompresses only those precincts which are not included in the intersection between both WoIs.

Soft Cache

A common slide microscopical exploration demands navigation through different resolutions, which is a difficult issue to mimic in any virtual microscopy design. Provided that the DWT is by nature a multi-resolution decomposition, it turns out reasonable to devise a cache strategy based on storing the wavelet coefficients of lower resolutions. This strategy is a soft cache [73], which maintains image wavelets coefficients of low resolution versions in memory for reusing them to make up a high resolution version [73]. The soft cache herein designed is located between the arithmetic decoder phase and the Inverse Wavelet Transform. From the J2K perspective this makes sense since the major complexity and time consumption are focused on the EBCOT because of the bitplane and coding pass strategy [134]. Many resolution levels can reduce the potential benefits of soft caching since a WoI

can remain in memory for a long time. In contrast, few resolution levels would get rid of the unneeded WoI quicker but it reduces the quality of the useful WoI. Therefore, the number of resolution cache levels should coincide with the number of resolution levels to which the image is compressed [130], this results in a good compromise between compression rate and quality (see some data in Section 3.4.3). Finally, the replacement strategy, as for spatial cache, is the LRU.

Quality Maximization of the Retrieved Image

Yet the cache strategy is a central issue toward fluent navigation, there exist several visual factors which may be improved for the pathologist to quickly interact with the displayed information. In other words, the objective is to uploading information scheduled by quality optimization, so that pathologists will have more elements to make faster decisions as for instance, a change of the WoI location or the WoI current resolution. Quality may be obtained either by objective - such as peak signal-to-noise ratio (PSNR) or mean square error (MSE) - or perceptual measurements, which could include a-priori information concerning this kind of images. In this chapter we address the problem by analysis based on objective measurements.

As explained in Section 3.2.2, the J2K codestream layers are defined by the ordering of the slopes of truncation points. This particular ordering is carried out on the whole image, so that when a particular WoI is requested, the implicit WoI sequencing of the codestream is not optimal any more. This may be overcome giving priority to packets by using a rate-distortion optimization approach (RD) to maximize the quality of the rendered WoI, as Taubman and Rosenbaum propose [134]. On the other hand, the Post-Compression Rate-distortion Optimization (see Section 3.2.2) serves to determine the distribution of the bit-streams in the quality layers. This is obtained by organizing the slopes of truncation points for the whole image in decreasing order. This distribution reflects the fact that some layers may contain several associated truncation points. A selection of packets for maximizing a WoI quality would require to keep the slopes at each truncation point after compression. This is herein solved by coding the image in several layers in such a way that it can be stated that all layers have approximately one truncation point per block. In this way, the re-ordering of packets could be performed with the layer available information, as Taubman proposes [134].

These strategies finally attempt that the pathologist feels comfortable

during navigation, but also that he/she may rapidly focus on the relevant parts. In this case, a compression scheme based on layers is useful since quality is progressively increased in function of needs. In any case, regions with larger contents of information can be fully and lossless reconstructed, offering the specialist the possibility of a diagnosis based on reliable information.

Navigation Assessment

A natural measure for efficiency assessment of the navigator is the time that spends the GUI for decoding and displaying a WoI. In this work we are not evaluating the efficiency of the GUI, but the gain in time obtained by introducing some techniques - cache and WoI's quality optimization - embedded in the navigator. Some authors prefer a measurement on the number of decoding bytes for making the results independent of the platform. However, in the case of the soft cache, the number of decoding bytes is not proportional to the spent time since the decoding time is different due to the cache contents. On the other hand, it is supposed that when a pathologist changes from a current WoI, it is because the image has reached enough quality as to decide a navigation move. The underlying fundamental hypothesis is that the attained WoI quality is the one required by the expert, Q_{WoI}.

Three different pathologists used the three constructed mega-images for evaluation. In general, pathologists are not quite used to computer, so for avoiding any navigation bias because of an inappropriate use of the GUI, before the first navigation each pathologist was instructed on this interface with a test image. Each of the navigation operations were then previously assessed by them, the different zooms, the re-sizing operations as well as the spatial jumps. For each pathologist's navigation, we store the WoI parameters of location, resolution and quality along with the span in time until the next WoI request. This quality was measured and defined as the quality threshold for the WoI (Q_{WoI}). Thus, three pathologists evaluated the system using the three built mega-images and the whole navigation sequence was recorded for further analysis. The reference for comparisons was the raw navigation, i.e., navigation without the cache and quality maximization strategies. Results are shown as the calculated difference between the reference and the improved navigation with any strategy together with the percentage improvement for any situation.

Evaluation of the Quality Maximization of the Retrieved Image

As seen in Section 3.3.5, the navigation time delays can be shortened by using an *RD* strategy. As suggested by Taubman and Rosenbaum [134], a re-sequencing of packets of a particular WoI can improve the quality for the same quantity of bytes, when comparing with a progressive sequencing of layers. On the other way around, setting a required quality to the navigation requested quality Q_{WoI} (as defined in Section 3.4.3), allows to measure the gain in number of bytes and thereby to reduce the navigation time. For so doing, the PSNR was calculated for both the strategy of re-sequencing packets (*RD*) as well as for re-sequencing of layers. The experiment was performed for a random selected WoI and for a continuous navigation. Once a quality threshold was set[2], the quality gain can be calculated as the difference in the number of bytes between these two strategies.

3.4 Results

3.4.1 Registration

A classical registration method is basically composed of a type of transformation, a similarity measurement and an optimization strategy. In the present work, affine transformations were used along with a descent gradient method for setting the best match between a template and a floating image. Descent gradient algorithm stops when it finds a displacement error compared with the Gold Standard displacement of less than 1 % of the image size. When automatically capturing, only simple displacements are expected. However, these displacements frequently result in a change of the focus plane so that the scale may change. The template was a small strip of pixels (image width/heigth 10 pixels) extracted from the border of one image which was utilized for searching its better match on a second image, i.e., the microscope overlay between a particular pair of images. In addition, three different similarity measurements were evaluated regarding their time performance and accuracy: mutual information, phase correlation and normalized cross correlation.

Figure 3.5 shows the accumulated time results when constructing one mega-image with 270 images, using a serpentine scanning pattern, which

[2]the quality requested from the actual navigation

correspond to 8.13 mm^2. As observed, computation time was higher when using mutual information compared with the two other similarity measurements. Overall, these results seem to be independent for each image so that differences are linear and become larger with the number of processed images.

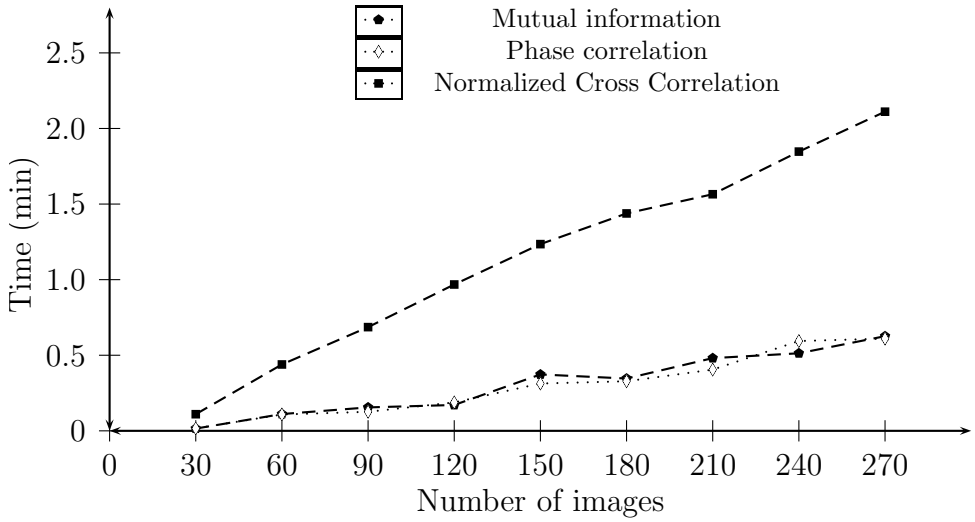


Figure 3.5: Cumulated times in registration for the three different similarity measurements. Provided that main hypothesis was that microscope produced only space displacement, affine transformations were used for transformation together with a conventional gradient method for optimization. The set of evaluation consisted in 270 images.

The observed difference in time for the three techniques may be attributed to the different calculation methods, since cross correlation and phase correlation are implemented using the FFT algorithm, which results in a smaller processing time. In general the time for processing one image was about 30ms for the registration techniques using phase correlation or normalized cross correlation as similarity measures, while the method with mutual information took about 200 ms.

On the other hand accuracy was assessed on the same set of images as the absolute difference between the obtained displacement with any of the three techniques and the Gold Standard overlapping displacements. Figures in Table 3.1 show no major difference, error mean is less than 1 % while

Method	$Mean \pm SD(\%)$
Normalized Cross Correlation	0.13 ± 0.037
Phase Correlation	0.12 ± 0.028
Mutual Information	0.22 ± 0.052

Table 3.1: Accuracy was measured as the difference between the set microscope overlay and the obtained with any of the three methods over a set of 270 images

standard deviation is below 0.25 %.

Usual image dimensions herein used were 752×560 and for this size the described error mean amounts to 2 ± 1 pixels. An observation that was confirmed through the ANOVA evaluation since differences were found to be not significant ($p > 0.05$).

3.4.2 Lossless Compression

J2K compression assessment was performed on the three different mega-images. The reversible color transform (RCT) and the Daubechies 5-3 wavelet transform were adopted to allow lossless compression. The influence of main parameters was investigated on the compression rate, i.e., the tile size, the resolution level and the size of different precincts and codeblocks. From the original three mega-images, three sets of 20×20 digitized images were used for constructing three 14288×11200 mega-images, which were used in this Section.

Compression by Tile Size

A first phase in the sequence of J2K steps prepares the image for further processing. The division of the image in rectangular tiles is useful not only for memory management but also for gaining access to different regions of interest in the image.

Table 3.2 compares the performance of the J2K algorithm for different tile sizes while resolution was set to 3, codeblock and precinct sizes to 64, respectively. Latter values correspond to default Jasper parameters.

Data show that each image follows a different compression pattern. However, these figures indicate that there is a decreasing of the compression rate when increasing the tile size, which is different for the three images, i.e., while

tile size	Image 1	Image 2	Image 3
64 × 64	13.35 ± 0.31	14.57 ± 1.26	14 ± 1.08
128 × 128	13.21 ± 0.26	14.28 ± 1	13.72 ± 0.9
256 × 256	13.19 ± 0.24	14.1 ± 0.96	13.54 ± 0.74
512 × 512	13.16 ± 0.13	14.09 ± 1.1	13.52 ± 0.8

Table 3.2: Evaluation of the tile size in the compression rate. Data show the mean and standard deviation in ($Mean \pm SD(bpp)$) for the three different mega-images (The original pixel size was 24 *bpp*)

for the first image a better compression gain is observed at the maximum tile size, for the other two images very little is gained after a 256 × 256 tile size.

Compression by Resolution Level

Compression was performed for six different resolution levels (see Table 3.3), the tile size was set to 512 × 512 and the above described parameters to the default Jasper values.

Parameter	Image 1	Image 2	Image 3
Level 0	14.1 ± 0.41	15.84 ± 0.84	14.49 ± 0.44
Level 1	13.4 ± 0.25	14.6 ± 0.85	13.97 ± 0.68
Level 2	13.19 ± 0.12	14.51 ± 0.97	13.94 ± 0.85
Level 3	13.16 ± 0.13	14.57 ± 1.04	14 ± 0.94
Level 4	13.17 ± 0.13	14.54 ± 1.06	13.98 ± 0.98
Level 5	13.17 ± 0.17	14.66 ± 1.13	14.1 ± 1.04

Table 3.3: Influence of the number of resolution levels on the compression rate. Figures correspond to the mean and standard deviation of the three images in (*bpp*).

Table 3.3 summarizes results. Data show that after two or three passes of the filter bank, the image is sufficiently decorrelated and very little is gained with new passes.

Compression by Precinct and Codeblock

A precinct, as shown in Figure 3.2, is the combined information of the different frequency subbands for a particular resolution. Data in Table 3.4 suggest

Pc	**Cds**	Image 1	Image 2	Image 3
32 × 32	16 × 16	13.17 ± 0.13	14.59 ± 1.06	14.02 ± 0.96
32 × 32	32 × 32	13.16 ± 0.13	14.57 ± 1.04	14 ± 0.94
64 × 64	16 × 16	13.11 ± 0.13	14.5 ± 1.03	13.92 ± 0.91
64 × 64	32 × 32	12.57 ± 0.14	13.79 ± 0.86	13.21 ± 0.74
64 × 64	64 × 64	12.56 ± 0.14	13.8 ± 0.87	13.23 ± 0.75
128 × 128	32 × 32	12.44 ± 0.07	13.77 ± 0.86	13.19 ± 0.73
128 × 128	64 × 64	12.38 ± 0.14	13.42 ± 0.7	12.83 ± 0.55

Table 3.4: Precincts and codeblocks represent the least information units in the standard. Table shows the influence of these parameters by varying their size and calculating the effect on the compression rate for the three mega-images. $Mean \pm SD(bpp)$ for different precincts (Pc) and codeblocks (Cd). Other parameters were set to the default Jasper values, i.e., tile size to 512×512 and resolution to 3.

a general trend: increasing the precinct size results in a continuous decrement in the compression rate. This statement turns out to be independent of the image contents and can be regarded from different points of view:

1. A large codeblock is obtained from a large pixel neighboring and the compression rate improves, however with too larger codeblocks the codestream loss granularity and the rate-distortion optimization is lost.

2. A precinct which is larger than the codeblock, slightly improves the compression rate because less information can be included in the packet header.

3.4.3 Navigation

Although J2K has many advantages regarding the mega-image handling, the time for accessing information is a cost issue which can be improved by using a store of objects in memory. Our basic strategy is to make relevant information available before the expert requests it.

Figure 3.6 illustrates a common navigation upon the pancreas mega-image with and without a cache strategy. The x-axis shows the sequence of consecutive WoIs and the y-axis the time needed for displaying each. Overall, the time required for extracting, building and displaying a WoI is lower than 600 ms, which in most applications can be considered as adequate for a

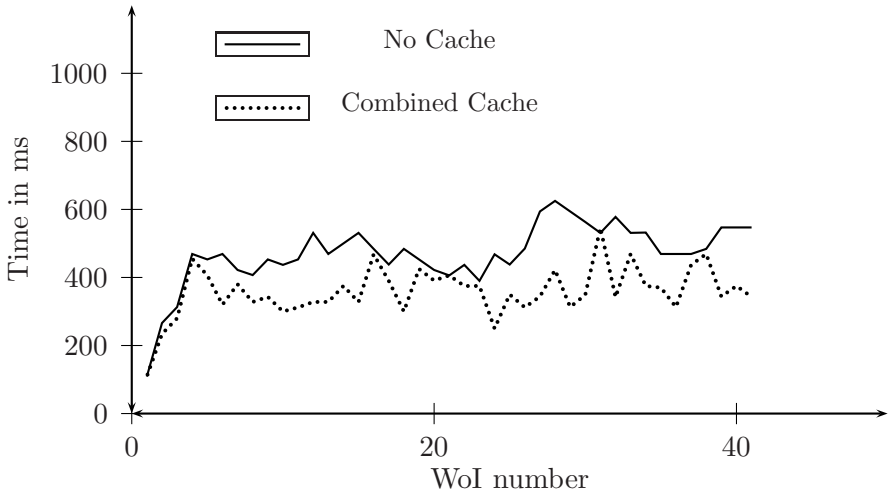
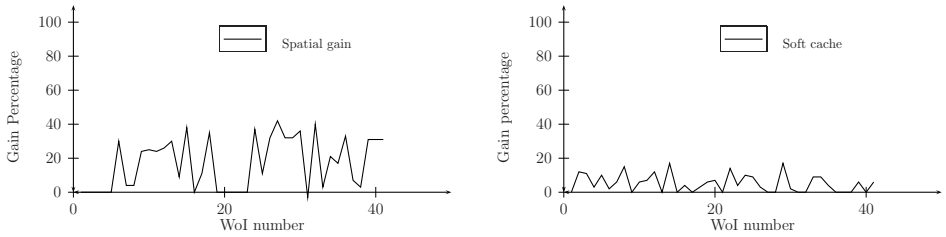


Figure 3.6: A typical navigation, plotted with and without a cache strategy in dotted and thick lines, respectively. $x-$axis shows the sequence of the different WoIs requested in this navigation, which corresponds to a total of 222 s, while the $y-$axis shows in ms the required time for displaying the WoI. The default exploration window was set to a 200×200 pixels, but it was changed once in this navigation by re-sizing (WoI 27).

continuous navigation. However, when a pathology expert changes rapidly the navigation focus, a slight delay appears and depending on the resolution, this can be a quite annoying effect. In this navigation the pathologist rapidly found an object of interest, the islets of Langerhans, and focused the navigation onto this region by zooming-in. In the graph this is observed as a rapid slope increase for the first four WoIs, which corresponds to a change from 109 ms to 469 ms, from the lower to the intermediate resolution. Afterward, the pathologist run over the image at this resolution: a spatial examination. This spatial navigation presents little variations, as observed in the plot from WoI 5 to WoI 25, with a base line on about 430 ms and a standard deviation on 40 ms. In one occasion, pathologist re-sized the WoI from the default 200×200 to 350×350 since the field of view was very small, this is also seen in this graph as the slight increase of time reconstruction for the WoI variation (WoIs 27 and 28). Likewise, the cache benefit can be observed in the superimposed dashed curve. The plot shows a combined strategy, including the spatial design for examination at the same resolution together with the soft cache when zooming in. Overall, for the whole set of naviga-



(a) Gain in percentage for the spatial cache

(b) Gain in percentage for the soft cache

Figure 3.7: The cache gain has been divided into the corresponding spatial and soft contributions for a separate analysis.

tions, the cache strategy is always below the raw navigation curve in about a 15 %, except on those points in which the expert changed abruptly the focus on navigation and re-initialize the entire memory cache space. For the case shown in Figure 3.6 the gain is about 23 %.

A separate analysis on the two complementary cache strategies is hereafter presented, starting by the spatial cache. This graphic displays in two panels the gain time in percentage for uploading information of a particular WoI quality requirement, Q_{WoI}, when using the spatial cache to the left and soft cache to the right.

Figure 3.7 presents the obtained percentage gain when using a cache strategy, either a spatial or a soft method. This gain was calculated as the time difference between the raw and the cache navigations related to the raw navigation and expressed in percentage. The left panel shows the spatial cache gain, in which figures present a high variability, between 4 % and 42 %. This pattern overall reflects the fact that this spatial cache is important when the degree of overlay between consecutive requested WoIs is large. In cases where overlay amounts to 85 %, a gain close to a 50 % is seen, while gain falls down to 5 % when this intersection is barely a 10 %. In addition, the plot shows occasionally negative peaks or null gains, which stand for explorations in which the expert changed completely the WoI, there was no intersection at all, and so an extra time was needed for filling again the cache memory space within the new context. These results show how a spatial cache can effectively improve navigation by reducing latency times, but clearly this is more effective when the pathologist is moving in a region

of high interest or is close to the target information, i.e., once the resolution navigation level is set. On the other hand, graph in right panel shows the relative gain of the soft cache. In general, results evidence lower levels of gain when comparing with the spatial cache, although the soft cache strategy maintains a positive gain through the entire navigation. In the best scenario, benefits reach up to a 17% time performance improvement, related to the conventional navigator performance. Again, null gains are observed when the expert rapidly displaces the navigation focus. Yet this gain presents small variations, no systematic trend is observed, an observation which is consistent with the fact that navigation is essentially a random phenomenon. These figures also show that the soft cache effect is lower than the spatial, but the fact that this strategy looks steadier makes that in zones in which the spatial cache shows no gain at all, the soft cache provides a quite constant level of gain. It should be pointed out that all these evaluations are limited by the fact that the expert leaps out the field of view with certain frequency for exploring new image zones, but in general these experiments demonstrate the benefit when using a cache strategy.

The above navigation experiment was repeated by the three pathologists on the three constructed mega-images and data gain is expressed in percentage. Results in Table 3.5, calculated as a mean and standard deviation of the time gain percentage by using a cache strategy, show an average of the benefits of the spatial and soft caches.

Overall, these figures show a great gain variability for each navigation, a result which comes from the pathologist multiple navigation patterns as well as for the variable image information contents. Figures in Table 3.5 confirms what was illustrated in the particular navigation before, that is to say, gains are overall higher for the spatial than for the soft cache. In despite of the different navigation approaches and the particular preferences of each pathologist, results show no major differences concerning the cache gain percentages. Average, figures are also alike and are around a gain percentage of 15 % for the spatial cache while the soft cache reaches about an additional gain percentage of 5 %. This can be attributed to the particular methodology a pathologist follows for searching information in a virtual slide, i.e., a pathologist explores a slide on an actual microscope by a spatial search of information while they usually require one or two resolution changes. Besides, most navigation patterns suggest that a combination of these two strategies is a good alternative for reducing latency times. More specifically, these results also show that for some WoIs, these caching procedures reach a 32 % gain

E - N	Spacial Cache Improvement (TGP %)	Soft caching improvement (TGP %)
$E_1 - N_1$	17 ± 15	4 ± 3
$E_1 - N_2$	15 ± 13	4 ± 3
$E_1 - N_3$	16 ± 13	4 ± 3
$E_2 - N_1$	13 ± 10	4 ± 3
$E_2 - N_2$	14 ± 9	2 ± 1
$E_2 - N_3$	14 ± 12	4 ± 2
$E_3 - N_1$	15 ± 12	4 ± 3
$E_3 - N_2$	14 ± 12	3 ± 2
$E_3 - N_3$	12 ± 10	3 ± 2

Table 3.5: Each expert carried out three navigations on the three built mega-images. First column stands for the particular expert along with any of the corresponding navigations, i.e., each expert makes three navigation and this is notes as E for expert and N for navigation. Second and third column present the mean and standard deviation, respectively, of the time gain percentage (TGP %) by using a cache strategy for the three different navigations using the spatial and the soft cache versions.

for a pure spatial cache and up to 37 % when the strategy is combined. Yet in very few occasions figures show negative gains because of the uploading charge time of the cache when the expert abruptly changes the image focus, in average these negative or null gains are more than compensated thanks to the cache in every navigation.

Finally, last part of evaluation corresponded to the quality maximization of the retrieved image. Provided that quality is essentially a subjective process, it is quite hard to have an approximated estimation of quality. In other words, the weight of every packet is not the same in terms of quality and then it makes sense that the order in which they are stored is important in terms of availability and therefore of navigation velocity. For doing so, evaluation was twofold performed: on the one hand, as illustrated in the left panel of Figure 3.8, the PSNR was calculated for one randomly selected WoI for the two sequencing strategies. Then, a desired quality was fixed as the last one constructed before the expert changed the navigation WoI. The quality performance for the two sequencing strategies can thus be calculated as the difference in number of bytes, needed to attain the same quality. The effect of the quality maximization was also evaluated on a whole navigation sequence of 157 s on the pancreas image, which corresponded to a set of 37 different WoIs, on which the analysis was performed.

Figure 3.8 shows the obtained gain when using the RD strategy. This is firstly illustrated in the left panel for a particular WoI. In this plot a level of quality is fixed by the expert[3] and permits to compare both strategies simply by the number of bytes needed to reach such desired level of quality. Using the difference in the number of bytes for both strategies, the right panel shows the pattern for the entire navigation. These results demonstrate that the number of bytes for RD is generally smaller. This represents a significant save of time in terms of navigation since less bytes are needed for reaching the desired quality. However, RD strategy is not always optimal in actual navigations, for instance observe that in the left panel the RD line slope is not always concave as it should be since this ordering is optimal. This can be attributed to the fact that the WoI is not aligned with every precinct for every case so that these WoIs are generally constructed by combining different packets from different precincts, i.e., it is needed to search information in different branches of the structure of the former defined compression tree with

[3]This level corresponds to the quality WoI reached in the navigation before the user makes a new request

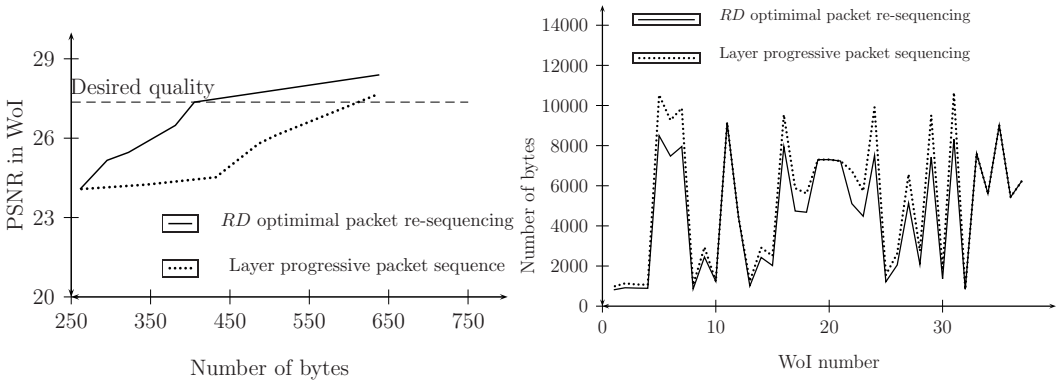


Figure 3.8: Left panel plots the PSNR in dB against the number of processed bytes for a selected WoI. The thick line stands for the number of bytes of the re-sequenced packets over the specific WoI, which are progressively uploaded by means of an RD strategy. The dotted line corresponds to the number of bytes calculated from the default distribution of layers in the codestream. Right panel shows the same measurements in a continuous navigation.

a resulting sub-optimal pattern. However, the RD strategy for re-sequencing the needed packets for a particular request over the WoI shows higher gains, which means that it needs a smaller number of bytes for reaching a desired level of quality. The right panel (b) corresponds to the same kind of analysis but extended to an entire navigation. Overall, the pattern is consistent with what was observed in one WoI, that is to say, the RD strategy requires a smaller number of bytes, except in cases in which the expert requires a full quality for visualization. In these cases, no strategy can be devised for improving navigation velocities and the expert merely decompresses the whole bitstream in order to obtain the original image.

3.5 Discussion

This study describes an entire system for virtual microscopy of large microscopical images. The microscopic navigation scheme herein presented starts with a correction of the acquisition overlap for generating one single mega-image. Adequate compression hints are then determined for this kind of images together with an efficient navigation strategy, using the J2K stan-

dard. The system is divided into three components: a registration module, which automatically constructs a mega-image for a given set of microscope parameters. This mega-image is generated by using known registration techniques for a set of digitized images from a particular specimen and determines the best overlapping between any two sequentially acquired images. A second module tackles with an efficient implementation of the J2K standard and allows adequate file compression rates along with a versatile information handling. Finally, an adapted navigation design permits a pathologist to easily use this virtual microscope for routine examinations or teaching activities. The entire system allows navigation through WoIs, resolution and qualities with a friendly interface.

Some previous works have focused on specific steps of the virtual microscopy process, either at the mosaicking [9, 25], the compression [22] phases or both [15], but they have missed microscopic image particularities as an important part of the problem. A basic concept in understanding histology is that there are four basic types of tissue: epithelial tissue, connective tissue, muscle tissue, and nerve tissue [72]. With very few exceptions, all organs contain a different proportion of these four basic tissues. In general, histological techniques highlight these tissues with few colors since dyes are designed to specifically arise a particular tissue feature. Color variability stems mainly from a large intensity range as well as dye deterioration. In terms of image processing, histological images are distinguished by having more or less homogeneous textures or repeated patterns, which may be used to characterize the image and to decide a particular strategy for compression.

An extensive use of registration techniques is on the base of histopathology slide assembling and stitching [9, 25]. In this kind of methods, windows of a predefined size are used for feature matching using similarity measures between spatial regions. A displaced overlay between two digitized images is used as an estimation of the superposition between the two images [158]. Overall, simpler similarity measurements are functions of the intensity differences between the corresponding pixels of the two images. This kind of measures, such as *the sum of squared differences*, is optimal when differences between images are exclusively caused by a Gaussian noise [144]. However, virtual microscopy is the result of automated capturing mechanisms, which introduce variable errors that alter image intensities (among 10 and 50 pixels) [9] so that this kind of measures may result inappropriate for a general stitching strategy. When this change is linear, similarity measures based on the *correlation* are more indicated [115]. A more robust approach is phase

correlation [80], in which the correlation is based only on phase information that is insensitive to intensity changes. Likewise, the *mutual information*, which has been successfully used in multimodality registration problems [87, 144], assumes a statistical relationship between the images to register. This measure is robust, but intensities are uniquely processed in a qualitative manner, with no consideration at all of the spatial information which may exist between near intensity values [114]. It is important to recall that differences in intensities between neighboring Fields of View are approximately linear, for this reason the more appropriated measurements are those based on correlation. In addition, these intensity differences are not too high. In consequence, it is not necessary to use time consuming measurements, such as mutual information. Yet the main focus of the present work was not on the mosaicking problem, we did explore this topic by evaluating an automated strategy for generation of mega-images. A simple image stitching algorithm was implemented for determining a set of optimal image translations, using a greedy optimization scheme as in [15]. In addition, several similarity measurements were assessed, finding that their performance are comparable in terms of accuracy and different in terms of time. Time performance becomes an irrelevant factor in actual applications because mosaicking and stitching are off-line operations. In contrast, the mega-image assembling depends on the proper overlay approximation. Accuracy was equivalent for the three evaluated measures since it is very likely intensity changes were herein linear. However, a more general perspective indicate a use of more robust measures, case in which non linear measures may result more appropriate. Finally, through visual evaluation the three experts assessed the quality of the images in large computer screens (5 megapixels Kodak Direct View Monochrome Display) as adequate for diagnosis, indicating that such simple greedy strategy is sufficient for this kind of applications and that more complex methods such as dynamic programming [9] are not needed.

In microscopical applications, recent advances in capturing devices and improvement in computer capacities have facilitated microscopic acquisition. It has been nowadays envisaged to make 3D acquisition in some morphometrical applications [110]. Yet larger computer resources are presently available, they still remain limited for handling such volume of data. Data must thus be reduced with the demand to deliver image information [53]. Done with JPEG compression, a $1cm^2$ microscopical sample, even to a high compression factor, would still result in several gigabytes. However, a pathologist will never need all of the data at the same time. Besides, main current trend goes toward

the use of distributed systems in which Doctors may access microscopical images from any place and appropriately navigate on them without a complete downloading. All these issues may be approached using J2K standard, which allows scalability, error resilience and progressive access in quality and resolution. This progressive access to image in J2K is based on the concept of packet. An image is represented as an ordered set of these packets. Each contains a piece of image in its spatial, resolution and quality components. When a specific region of the image is requested, a subset of these packets is obtained and decompressed. A usual slide examination requires a small set of packets so that a specific extraction of them highly improves navigation times. Once the requested packets are located, decoding time is divided between the decoding phase and the inverse wavelet transform. In the present work we designed and developed a combined strategy for each phase and concluded that a simple cache strategy such as the Last Recently used, produces a considerable improvement in the navigation velocity only because of reusing the wavelet previously coefficients. Additionally, an RD optimized resequencing of packets was explored over a WoI and a complete navigation. This strategy clearly reduces the time required to reach a fixed quality with the consequent save of time since a smaller number of bytes is needed to upload.

As already said, it is very likely that computational techniques in the near future may improve so that the decoding time becomes negligible. In this scenario, the spatial and soft caches may become mainly a buffering of J2K packets, preselected to be decoded for avoiding disk access. However, the medical devices are also changing and access to Medical Information Systems is more and more achieved through devices with little capacities such as mobile phones or pocket devices that Doctors can easily bring to any medical situation in hospitals. On the other hand, yet the storage capacity in modern machines is highly growing and codification techniques are also more powerful, it is also true that the amount of data is exponentially increasing in any modern hospital and also the requirements for current applications. For instance, the Radiology Department of the University Hospital of Geneva (2000 beds, 40000 in-patients and 400000 out-patients) produced more than 12000 images by day in 2002, and the total amount of cardiologic image data produced in the same hospital for the same year was around 1 TB [97], a figure that is difficult to handle for a conventional desk machine. It is worthy to strengthen out that this number of images does not include potential applications as virtual microscopy. As a conclusion, it is reasonable to assert

that these cache strategies will be at the stake in actual medical applications for the next decade.

Finally, yet this work is addressed to virtual microscopy, it can be easily adapted to any type of application for navigation on large images. Within the J2K context the present work points out on the utility of including, inside the decoder, several techniques which permit random access to the images and take advantage of its scalability. In despite of J2K being by nature a very useful technique for scalability during navigation, there exist few works in the literature that have explored this issue. On the other hand, our results in terms of navigation are encouraging and can improve as long as optimized implementations are available and include perceptive measurements of quality and *a priori* informations regarding the type of images. Our results suggest that the technique can be improved to the point of a seamless navigation and an actual tool on the routines pathologist work. Future works should include faster methods for registration as well as use of optimal replacement and store strategies of the soft cache.

Part II

Pathologist's Navigation Patterns

4

An study of pathologists navigation patterns in virtual microscopy

Roa-Peña L., Gómez F., Romero E., Submitted to Human Pathology.

Abstract *The aim of this chapter was to establish main pathologists navigation patterns when exploring virtual microscopy slides, using a GUI adapted to the pathologist's workflow. Four pathologists with a similar level of experience, graduated from the same pathology program, navigated six virtual slides. Different issues were evaluated, namely, the percentage of common visited image regions, the time spent at each and its coincidence level, that is to say, the region of interest location. In addition, navigation patterns were also assessed, i.e., mouse movement velocities and linearity of the diagnostic paths. Results suggest that RoIs are determined by a complex combination of the region visited, the time spent at each visit and the coincidence level among pathologists. Additionally, linear trajectories and particular velocity patterns were found for the registered diagnostic paths.*

4.1 Introduction

As we discussed previously, the bottle neck for extending the use of this technology is very likely the large volume of data, which must be stored in dedicated machines and the limitation to transmit through fixed bandwidth channels. Caching and prefetching may speed up these processes, when associated to a proper knowledge of the pathologist navigation patterns, which surely are very variable. A complete automated navigation would eliminate that variability, whereby it is required a deep knowledge of the two main

issues of this problem: how to find the relevant information or Regions of Interest (RoIs) in virtual slides and then, how an expert interacts with these regions. Regarding the first issue, a RoI has been traditionally considered as a particular image area in which the level of information from a user stand point is higher [79]. However, this is still an incomplete definition since relevant information can also depends on the time spent in these regions and then this time should also be considered to define RoIs.

4.1.1 Pathologist Navigation Patterns

At a first naive glance, a pathologist navigation turns out to be an intricate (and intriguing) series of back and forth changes, sideway movements, all at different speeds and tempos. However, such pattern is not at all a random movement since complex mechanisms, guided by a long training period, are triggered by the content of the slide and the spatial organization of the information therein [79]. Overall, every pathologist follows a standard training addressed to strength out both the diagnosis precision and time [105]. Coarsely, diagnosis in pathology can be considered as a process composed of four sequential steps: look, see, recognize and understand [16]. A definitive diagnosis is achieved by following a standard methodology with two coarse phases: first examination is carried out at the lower magnification (panoramic) in order to locate relevant information in terms of a spatial organization of the histological sample (scanning), while the second and further examination is conducted for analysis of the slide contents which implies changing magnification (zoom) [16, 28, 140]. This analysis is performed through navigation of the zoomed areas [28], on which gentle movements are generally required. This learned strategy has been observed in multiple studies in which it has been possible to determine the existence of these two phases. Tsuchihashi et. al. studied one pathologist exploring twenty different slides in telepathology. This investigation identified two patterns: exploration at low magnification and analysis at higher magnifications [140]. Crowley et. al. recorded on videotape diagnoses performed on four histological slides by fifteen pathologists, distributed into three different categories: novices, intermediates and experts [28]. Results showed that intermediates and experts exhibited patterns very similar to the two described before, i.e., to apply a general search strategy first, such as examining the entire slide at low power, and to select areas to revisit them at higher power.

Tiersma et. al. investigated visual exploration patterns in pathology

using an eye tracker mechanism. Results showed two patterns: scanning (saccadic eye movements) and selective (eye fixation over specific points for further exploration) [137]. Finally, Krupinski et. al., using an eye tracker system, demonstrated that visual pathologist exploration is characterized by a rapid determination of RoIs, which likely contain diagnostic information [79]. It is worthy to strengthen out that the experimental setup of that study allowed only scanning patterns, i.e., magnification changes were not available. Furthermore, a recent study [151] has shown that a pathologist exploration is composed of one or two magnification changes, at most, and two or three movements through the slide which are linear. Moreover, other works [151] have shown that these movements are linear. This evidence points out to a twofold underlying mechanism of pathologists' navigation which should be integrated into a navigation system.

Overall, the experience with images in virtual microscopy should be very similar to an actual optical examination. Design of friendly and useful graphical user interfaces is a fundamental issue, i.e., with a conventional mouse, pathologists navigate to any region within a slide as well as they change to any desired magnification (zooming in and out). Indeed, navigation patterns arise from two intermixed processes: the motor control associated to some movement automations and a refined search information process, which reflects the level of expertise. Yet patterns may be different, the more expert is the group of pathologist the more similar are the locations they visit when exploring a histological slide [79]. It is important to keep in mind that the virtual microscopy tool does not imitate exactly a conventional light microscope, but rather it has the purpose to allow a pathologist to navigate at any resolution, while the VS is always available at the lowest resolution so the expert always conserves a thorough panorama of the histological image.

This study aims to determine main factors involved in the genesis of the navigation patterns from a particular diagnostic path. These patterns are the result of the interaction between the image contents and the expert experience. Other studies have focused before on studying either the attentional mechanisms which guide trajectories in this type of images or the general mechanisms at the very base of the interaction of an expert with an image. On the contrary, our study integrates these two visions and also introduces new elements as the magnification changes, which minimize the navigation times dedicated to search information. Finally, as far as we know, this is the first study which actually evaluates how a pathologist "moves the stage", in other words, we dedicated our endeavours to figure out the influence of both

the image contents and the pathologist methodology in an actual interaction context.

The rest of the chapter is organized as follows: Section Materials and Methods the experimental assesment proposed, Section Results presents evidence about new Pathologist Navigation Patterns in virtual microscopy. Finally, the last Section concludes with a discussion.

4.2 Materials and Methods

4.2.1 Virtual Slides

A total of four randomly selected histological specimen were digitized and six VS were assembled using an acquisition system consisted of a Sony high resolution digital video camera Handycam DCR-HC85 (640×480 pixels) coupled to a Carl Zeiss Axiostar Plus microscope, provided with Carl Zeiss 426126 and 456006 adapters (Carl Zeiss, Light Microscopy, Gottingen, Germany). Hematoxylin-eosin tissue samples from endomyometrium, gallbladder, prostate and a uterus leiomyoma were used for this study. Histopathological slides were selected from a set of representative routine cases from the Pathology Department at the National University of Colombia, a pathology laboratory of medium complexity. These samples were selected by an expert pathologist. The endometrium sample was obtained by a legrade so that the biological specimen was composed of irregular hemorrhagic fragments, corresponding to epithelial tissue. The gallbladder specimen, characterized by homogeneous distributions of transitional epithelium, came from a cistocopy. The prostate sample was obtained by transrectal resection, case in which the tissue is usually characterized by the presence of glands supported by stroma. Finally, the uterus leyomioma, obtained from a myomectomy, is basically composed of muscular tissue whose distribution is quite homogeneous. They were digitized and six different images were assembled from these acquisitions whose sizes in pixels were 53280×39360, 42480×15840, 33840×21600, 53280×39360, 42480×15840 and 49680×28320 which stand for an effective area of 11.97×8.84, 9.54×3.56, 11.16×6.36, 11.97×8.84, 9.54×3.56, $11.16 \times 6.36\, mm^2$ respectively (pixel size of $1.98\, \mu m^2$). Mega-images were stitched using automatic registration with cross correlation as the similarity measure and were stored in JPEG2000 format for latter access and navigation [68].

4.2.2 Virtual microscopy GUI

Pathologist navigation patterns were recorded using a virtual microscope prototype and a GUI adapted to virtual microscopy in pathology [68]. This design exploits the importance of low magnifications for exploration and analysis at high resolutions for diagnosis. The GUI is composed of a thumbnail and an auxiliary window. The former displays the lowest resolution thumbnail image, in which a rectangular re-sizable window allows a required selection. The thumbnail window is set to a particular selection size at the beginning, while the auxiliary window is constantly varying with the level of magnification of the selected RoI in the thumbnail window. Displacements of a particular RoI were only allowed in the thumbnail window through drag and drop operations. Finally, for each requested RoI, its position, size, resolution and time were recorded for later analysis. Figure 4.1 shows the virtual microscope GUI. Navigation in the developed prototype is carried out through a conventional mouse and consists of two processes: first, a window picking at low resolution image, followed by a displacement of this window to an interest point, proportional to the mouse movement. This prototype was aimed at achieving integration of this kind of tools with a routine pathologist's work, whose main advantage is a design that permits a simultaneous display of different magnifications.

4.2.3 Observers and tasks

The aim of the present study was to compare navigation patterns when expert pathologists are exploring virtual slides. Four expert pathologists participated in this study, all of them had similar years of experience (about five years), and were graduated from the same pathology school program. Each pathologist was previously trained on virtual microscopy using two test virtual slides. Average, the spent time for training was 20 minutes so that at the end of this time they were free to navigate the slides at will, until they could reach a probable diagnosis and organ identification. Six virtual slides were chosen, as parts of full histological slides, with a relative size which varied between 10 % to a 30 % of the whole histological sample. An experienced pathologist, with at least five years of experience, selected the digitized area. For the sake of the experiment, images shown to the pathologist belonged to areas in which it was difficult to determine both the organ and the pathological entity so they were forced to spend more time,

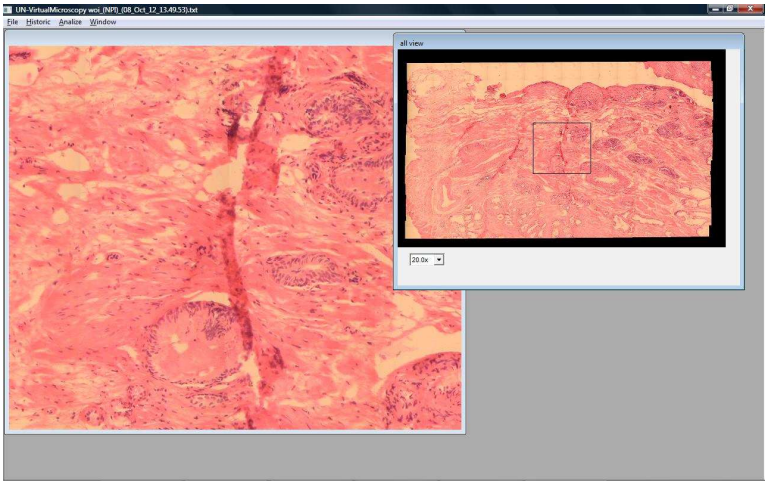


Figure 4.1: GUI of the virtual microscopy prototype: the Figure shows the whole Virtual slide at the right panel (small magnification) while a large resolution of the black square in the VS enlargement is displayed in the auxiliary window at the left panel. Panning was onely allowed in the VS window.

exploiting the navigation tool [17]. The six VS were randomly displayed for each of the pathologists so that the examination order was always different. Each pathologist was asked to run over the virtual slide, up to a diagnosis was set, using the same screen monitor they use in their routine computer work (CCFL (220 *nits*) WXGA (1280×800 of $13,3$"$)$). During examinations, every pathologist action was recorded for later analysis, namely, we recorded the RoI location, the RoI size related to the thumbnail image, the time any action (drop or drag) was carried out and the level of magnification.

4.2.4 Evaluation Issues

We claim that from a navigation point of view an image can be thought of as an ordered partition of spatial locations with different levels of relevance associated to each. Therefore, an image is composed of regions with different levels of interest and pathologists shall visit a minimum number of regions, whereby they would gain a maximal amount of information spend-

ing a minimal time. Evaluation was then addressed to verify these two main issues: images are composed by pieces of information with different levels of relevance (RoIs) and pathologists will use a minimum time exploring them (Navigation Patterns). The first item was assessed as follows:

- The percentage of visited image area was calculated for the group of pathologists.
- The percentage of coincidence among the different visited areas, i.e., at least two pathologists examined the same image region within the navigation.
- For each pathologist, the spent time for region was also computed.
- Coincidence between RoIs, defined by visited areas, and RoIs defined by time.

The second item was assessed as follows:

- Mouse movement velocities were registered and analyzed.
- Trajectories among RoIs might be linear so that their Euclidian distance was computed and compared against the actual distance, i.e., a ratio between the two distances was calculated.

4.3 Results

4.3.1 Regions of Interests

As previously discussed, we suppose that RoIs are defined following a combination of the three criteria presented hereafter:

Percentage of visited image area

The percentage of visited area by at least one pathologist was computed for each of the virtual slides (Table 4.1). These values varied between 44 % and 91 % with an average of 66 %, indicating that the amount of tissue explored is highly dependent on the image contents, that is to say, some virtual slides were little-explored because relevant information was easily available.

Image	Percentage of visited area per image
image 1	91%
image 2	48%
image 3	67%
image 4	86%
image 5	44%
image 6	60%

Table 4.1: Percentage of visited total area by at least one pathologist per image.

Pathologists were forced to further explore the image as to obtain a maximum level of information, but in general this was hardly established since information was not enough as to consolidate a diagnosis. Results support this statement since the percentage of explored image was larger than a 50 %.

Percentage of coincidence

The percentage of coincidence of the visited areas among pathologists of each VS was computed (Table 4.2). These values varied between 41 % and 97 % with an average of 69 %, indicating that the explored areas were quite similar, though the virtual slide content is entirely different. The coincidence level turns out to be dependent on the kind of information present in the virtual slide and located in specific regions. For example, Table 4.1 shows the pathologists visited a 48 % of the 2th VS, while its level of coincidence was 97 %. In this virtual slide it is observed that there is no tissue in about a 30 % of the entire VS. Interestingly, the histological sample corresponds to an endomyometrium, in which the glands are the fundamental part of any diagnosis and in the virtual slide they are located in specific areas. The diagnostic path, in this case, searched for these structures all over the virtual slide and overall, the four pathologists run over the same parts of the VS.

As expected, samples where information is located show higher levels of coincidence. The first virtual slide corresponds to prostatic tissue in which there is a large number of glands and pathologists dedicated most of the navigation exploring them so the coincidence level is high (larger than 50 %). The second virtual slide corresponds to an endomyometrial sample, its epithelial component is quite located and constitutes a very small area of the sample so the coincidence level is shown to be high. The third virtual slide corresponds

Image	Percentage of coincidence in visited areas
image 1	63%
image 2	97%
image 3	95%
image 4	69%
image 5	41%
image 6	58%

Table 4.2: Influence of the image contents in the navigation pattern observed from our experiments was assessed by measuring the coincidence level in the visited regions, namely, the area percentage which was visited by more of one pathologist. Coincidence average was 70.5 %, which demonstrates that there are relevant information areas.

to a fragment of a leiomyoma with a predominat stromal component. The areas visited by the pathologists corresponded to structures with a luminal space, seen at the low magnification. Interestingly, as it was not clear if they corresponded to glands or vessels, the resultant coincidence level was high. The fourth virtual slide was clearly a leiomyoma and yet the visited area was large (86 %), the coincidence level was only 69 %. The fifth virtual slide was a gallbladder, the epithelial component is minimum but scattered and then the area to explore large, the coincidence level was not as higher as in the other images. Finally, the sixth virtual slide corresponded to a prostate, in which the epithelial component was sparse and hence also the coincidence level and the area to explore.

Spent time

A potential RoI could arise either when every pathologist stops in particular image locations and therefore information therein is relevant, or when one pathologist spans a longer period in a precise area so that even if the interest in the region is not shared among the group of experts, there exists a potential source of knowledge. In consequence, we also evaluated the time every pathologist required for examining regions as the total time of visit for each pixel in the thumbnail window, which was estimated by accumulating the set of visit times and compute their average. A RoI (in the time sense) was then defined as the region composed of those pixels for which this quantity was larger than the mean. Once these regions were set for every pathologist, the

time at each of these regions was calculated and regions where pathologists spent times larger than the average were highlighted.

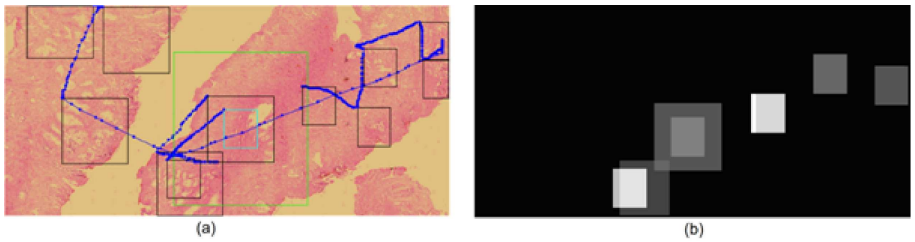


Figure 4.2: Left panel a whole diagnostic path, composed of multiples jumps among different image locations. This diagnostic path shows not only scanning patterns at the beginning, but also magnification changes, highlighted in the Figure as the green and blue squares. Once the magnification changes are established, navigation keeps under scanning patterns and the observation window is smaller since resolution is higher. At the right panel it is displayed the image locations with higher time rates. The white squares correspond to longer times while grey ones stand for smaller. Note that longer times are spent in the part of the diagnostic path which was conducted at the higher magnifications.

The analysis in this Section was pointed out to determine whether or not there exists any pattern regarding the time used for analysis. Therefore, a diagnostic path could be set not only in terms of the image contents, but also according to the time a pathologist needs to explore the VS. Figure 4.2 illustrates a thorough diagnostic path in a VS, a scanning pattern with two different magnifications. The magnification changes are highlighted in the image as the green and blue squares. There are two scanning patterns, each at a different resolution. Interestingly, the scanning pattern at higher magnification required also higher time rates, indicating that once the search has been established, these experts devote their efforts to analysis and diagnosis on regions in the image that contain relevant information. This analysis was extended to the entire set of pathologists upon the available VS.

Figure 4.3 shows the percentage time every pathologist spent over the RoIs previously established at any of the six images. Overall, pathologists spent at least a 50 % of the average of the navigation time on these RoIs, most of them detected at the larger magnification. The plot shows that two and five VSs were less explored regarding these regions. Overall, in despite of the different contents in these images, these results indicate that

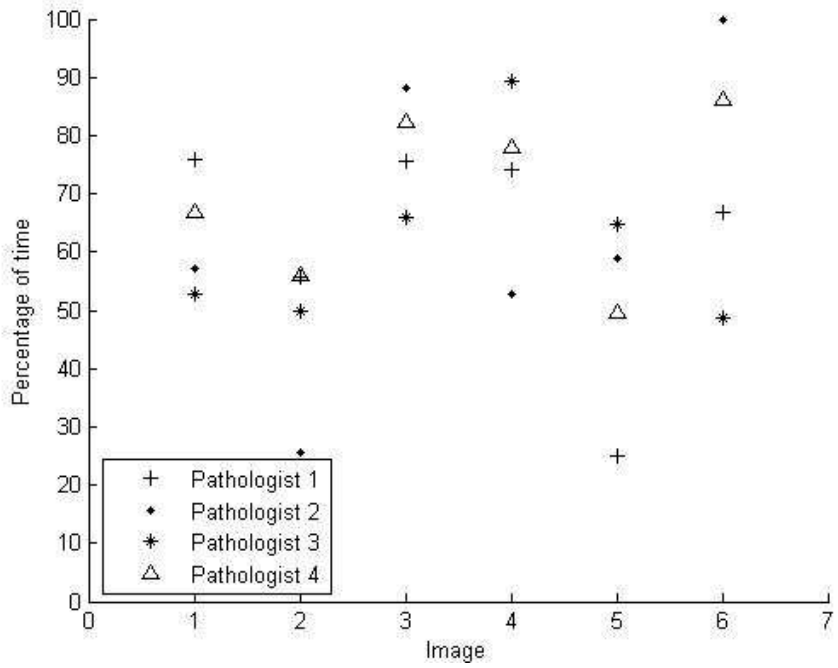


Figure 4.3: It is displayed the average time percentage spent among the different RoIs for each of the pathologists and through the set of six images. Total times were normalized by the maximum visited time for comparing navigations with different durations. The x axis corresponds to the set of available images while the y axis stands for the average of percentage of time spent in the previously determined RoIs. Note that in general pathologists spend more than a 50 % of the navigation time exploring these regions.

pathologists spend most of the navigation in regions where information is more relevant and the scanning process turns out to be dedicated to search such information.

Coincidence in RoIs

One important question we addressed consisted in determining, whether or not the RoIs are defined, by spatial preferences or spent time, would coincide. For measuring so, a standard measure of the degree of intersection between regions was used: the Jaccard coefficient, a measurement of the similarity between sample sets, defined as the intersection divided by the union of the

sample sets. This coefficient has a maximum value of 1 when there is total agreement and zero when there is none.

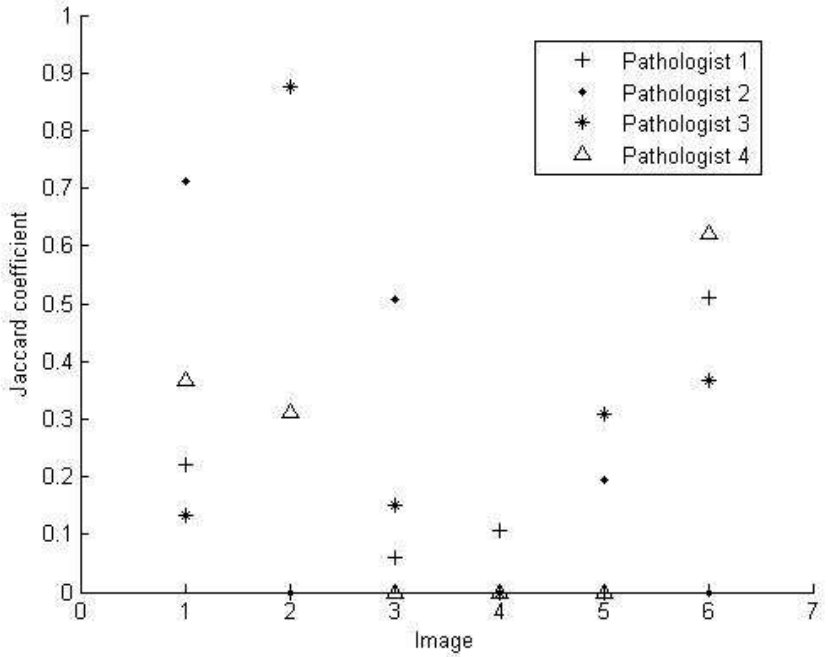


Figure 4.4: Jaccard coefficient is shown in the y-axis against the particular VS in the x-axis. This coefficient measures the level of coincidence between the two sort of RoIs, namely, the ones determined by the number of visits and those established when a pathologist spent a significant time exploring them. In general, there is not a systematic trend and pathologists use different navigation patterns regarding times and preferred locations. These results indicate that RoIs may be defined, depending on the application, by two different sets of features.

The coefficient was thus calculated for the six VS, showing different degree of overlapping, from a 0.9 of the third pathologist for the second image to a zero overlap coefficient for at least one of the pathologists along the whole sets of images. Overall, Figure 4.4 show coincidence levels under a 0.5 in most images, indicating that effectively it turns out there exist different RoIs, defined by the spent time or by the number of visits along the group of pathologists, that is to say the level of coincidence is low. There is no a specific pattern, for instance the fourth pathologist (triangle) has no coincidence

level in images three, four and five, while in images one, two the Jaccard coefficient is between 0.3 and 0.4 and 0.6 for images six. This statement can be also extended to the other pathologists, each using different preferences when searching further information, either by time or preferred location.

4.3.2 Navigation patterns

Mouse Movements

By default, all navigations start at the upper-left corner using a standard $15 \times 15\ \mu m$ microscopical field of view, which corresponded to a 100×100 pixel window, within the thumbnail virtual slide. In the first part of the navigation, the WoI is displaced through the virtual slide under a drop-drag-drop paradigm, constituted as the basic operation so that differences are mainly observed in the velocity profiles with which these navigations are carried out. In addition, pathologists could change magnification during exploration, either zooming in or out. An intermediate operation is an adjustment of the field of view when changing magnifications, i.e., a window re-sizing which allows covering the same area when resolution changed. Overall, once this new size was set, the pathologist continued the spatial exploration using the same magnification, as observed in Figure 4.5. So far our observations indicate what has been described in the literature, that is to say, navigation is composed of two complementary strategies: scanning and magnification.

When comparing the navigations available over the same image, a main conclusion is that every pathologist always uses both scanning and zooming operations. Interestingly, the coincidence level in the image in Figure 4.6 was 63 %, a fact that definitely suggests that the image contents steers the resultant navigation profile. The virtual slide corresponds to a prostate sample, with different cuts and RoIs defined by the loci with high gland density. Figure 4.6 shows the four different navigation profiles, with very different diagnostic paths and observation strategies. Overall, pathologists explored a variable virtual slide surface, with different levels of interest with two main navigation patterns: in the former case, three of the pathologists used the default window and run over the virtual slide with occasional magnification changes while in the latter case, pathologist enlarged the initial window to cover the maximum surface while running over the slide. The second strategy wastes much more computational and network resources because when using bigger areas the system has to load more information and in consequence it

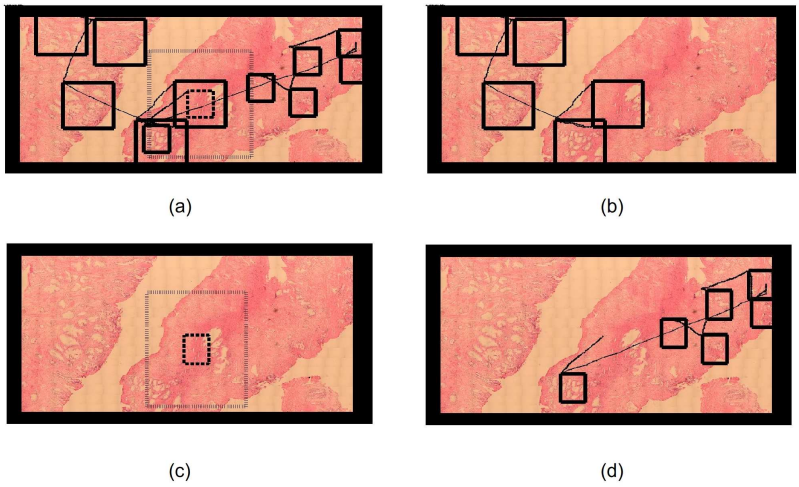


Figure 4.5: Complete navigation, split into four sequential panels (a, b, c and d) for the sake of understanding. Panel (a) shows the whole picture, which obviously looks very jammed by the number of windows and window movements. Hence navigation was split into a sequence shown in the next three panels: panel b starts by depicting a classical scanning pattern, composed of multiple jumps between RoIs. Panel (c) shows the magnification change from the dotted window to the thick one. Finally, the expert uses this magnification for exploring the rest of virtual slide in panel (d).

takes more time and nothing can ensure that the level of interactivity may lead to diagnosis in minimal time.

Figure 4.6 shows different navigations of each pathologist over the same virtual slide. In all of them the navigation pattern consists in both scanning and zooming in different RoIs. Interestingly, three pathologists used the default window (a, b, c) while one of them enlarged the window to cover a maximum area. Finally, the velocity profiles corresponding to the navigation displacements of all the pathologists were observed showing that although every pathologist has different navigation patterns, exists a common velocity profile, i.e., velocity rapidly increases up to a certain level and then it decays with lower slopes. This profile is likely a complex mix of associated factors such as the microscopical magnification, the neuromuscular mechanics and

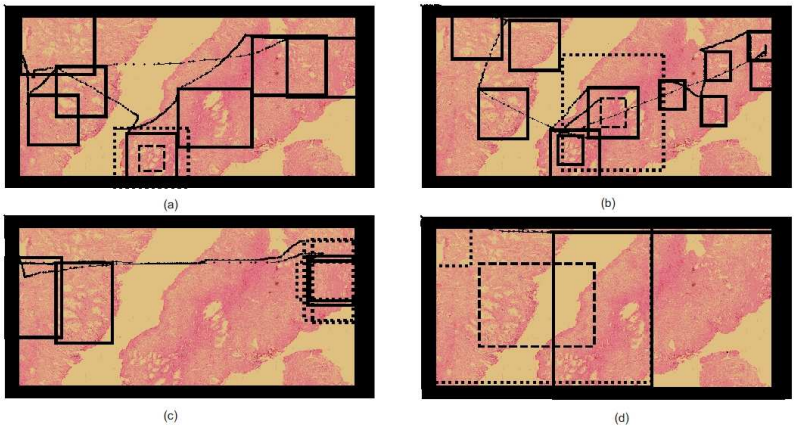


Figure 4.6: Different navigations of each pathology over the same virtual slide. In all of them the navigation pattern consists in both scanning and zooming in different RoIs. Interestingly, that 3 of the pathologists used the default window (a, b, c) while one of them enlarged the window to cover a maximum surface.

the type of restriction demanded by the developed GUI, i.e., a drop-drag-drop sequence (screen and mouse). As shown in Figure 4.7, explorations show high velocity profiles when experts are moving between RoIs, and lower velocity when approaching them. The GUI design allows to easily jumping from one information zone to a next in the smaller resolution windows, an effect observed in terms of velocity as the increasing part of the peak while a new zone is reached and a decreasing velocity profile since this zone deserves a certain amount of time for examination.

Path linearity among RoIs

The average coefficient of linearity is herein defined as the distance between two RoIs divided by the actual run distance. This coefficient has a maximum value of 1 when there is a linear trajectory and lower when is not linear. This coefficient was calculated for any trajectory between the previously determined RoIs and for every available navigation. Results are depicted in Figure 4.8, in which the percentage of the entire set of trajectories (157) is drawn in y-axis and the different linear intervals in x-axis. The coefficient average is about 0.8 and its standard deviation of $\pm 18\%$, thereby one can

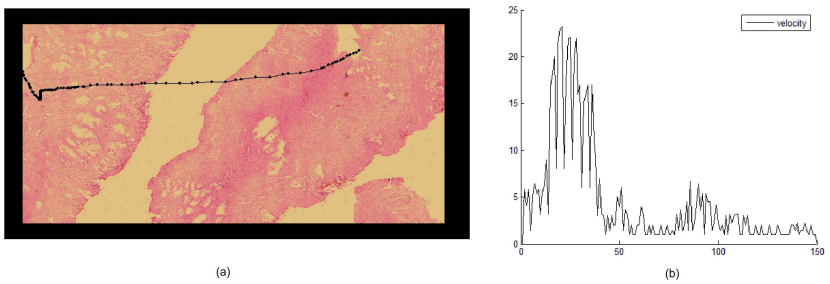


Figure 4.7: Example of a characteristic velocity pattern generated by a left to right movement.

conclude that movement between RoIs are basically linear.

4.4 Discussion

The present investigation was addressed to determine main pathologists navigation patterns when using virtual microscopy slides, using a GUI, adapted to the pathologist's workflow. Four pathologists with a similar level of experience, graduated from the same pathology program, navigated six virtual slides. Provided that a microscopical navigation is an interactive process, this study was devised to establish the relative importance of both image contents and navigation patterns of pathologists with high degree of expertise. A contribution of this study is that our GUI allowed to study not only scanning patterns, as described in the literature [79], but also magnification changes, a scenario really close to what pathologists are doing in their daily routine. Likewise, we explored the concept of RoI from different perspectives, either by analyzing the number of visits to a particular VS location as described in other previous investigations [79], or more importantly, by taking into account two new issues, i.e., the time a pathologist dedicates to explore this particular location and the coincidence level among pathologists, determinant factors which have not been evaluated so far. In addition, we also assessed the different paths described during these navigations, namely, mouse movement velocities and linearity of the diagnostic paths.

In previous studies, RoIs have been defined as areas to which the examiner is rapidly attracted to look and that may contain diagnostic information

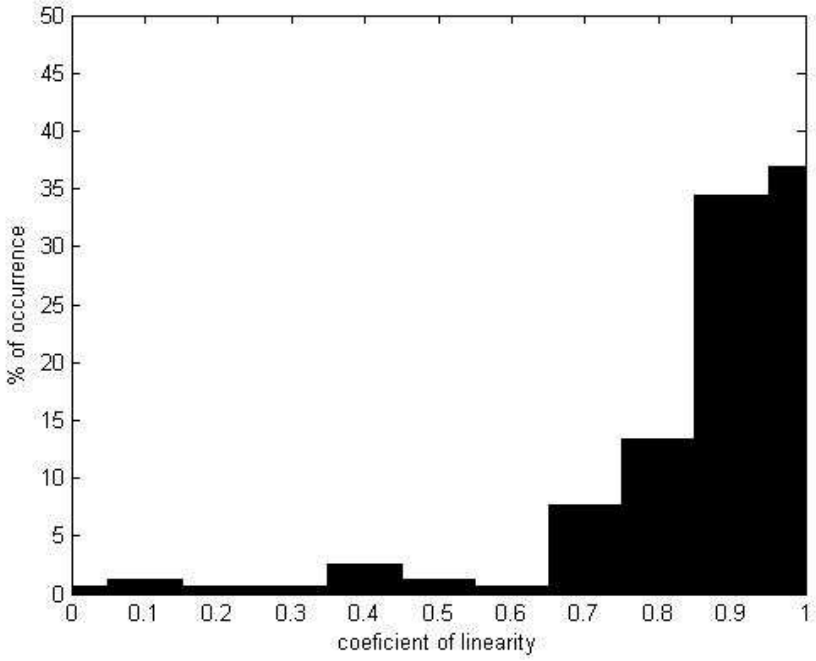


Figure 4.8: y-axis shows the percentage of occurrence for each coefficient of linearity in all the pathologist navigating the same image. The x-axis represents the different ranges of linearity, as observed in the Figure, most of the movements are done in a linear pattern.

[79], where visual stimulus is analyzed in detail [17] or by regions that are prioritized by the image content (first criterion) [4]. However, such definitions are still incomplete since a particular examiner can be rapidly drawn to regions which require a further exploration for classifying the type of information. Such pattern can be easily observed in difficult cases in which relevant information is hidden or information associated to the case is not enough as to establish an objective judgment. This leads us to acknowledge that the first path approximation is insufficient and that other criteria should be included. Classical psychophysical theory claims that the time spent in any particular task is directly related to the degree an examiner is familiar with a particular pattern [43]. We decided then to include the time spent at examining, as an evaluation criterion (second criterion) of what a RoI means in pathology, because the common workflow consists in developing technical

skills as to search abnormalities, a very harsh picture in many pathologies. In the present investigation, this factor resulted crucial since our GUI permitted to zoom in at any of the available preparations and so we could evaluate the importance of a particular image locus, not only because the examiner stopped there, but basically because information was valuable and required actual analysis. This fact could be established because we could compare the time spent at any of these image loci. Finally, another analysis direction (third criterion) could be the coincidence area visited during navigations, a factor which can be much more objectively included as a criterion, even though its inclusion in clinical routine is very difficult because the number of pathologists examining the same slide is very rarely larger than two. Overall, our results have supported the importance of simultaneously taking into account the three issues, mentioned before, as the base of an actual RoI definition because: (a) Pathologists are effectively attracted by some regions, as inferred from Table 4.1, with percentages of visited areas from 44 % to 91 %, a variable Figure which directly depends on diagnostic difficulty. Recall that the present study was devised for studying navigation patterns and therefore information related to the case was not available, this factor did increase the navigation time but even in these hard conditions, pathologists do not need to visit the entire virtual slide. Images associated to larger visited areas are consistent with the ones in which it was more difficult to determine the organ and/or the pathologic entity. (b) The interest for particular image loci is shared by most of the pathologists, a claim inferred from a coincidence level nearby to 70.5 %, as observed in Table 4.2. Finally, we found that the time spent at examining each of these regions was at least a 50 % of their navigation time, on specific regions previously defined, whereby some of the original regions were ruled out when considering the time factor. This evidence suggests that none of these factors could be considered as the base for defining the RoI and probably complex combinations of them are required for specific applications.

Classically, exploration patterns in medical imaging have been studied by both tracking the visual system or analyzing the diagnostic path complexity [123]. However, due to the restrictions of visual systems for acquiring and interpreting high resolution images and the limited display capabilities of the computational resources [153], new alternatives for exploring virtual slides have been recently proposed [68]. In actual virtual microscopy scenarios, any diagnostic task requires the use of a device (a kind of joystick), mainly mouse devices, to point out relevant information over a low resolution version

of the VS. The kind of patterns generated by the interaction of an expert and a virtual slide is obviously related to the type of interface. Overall, navigation patterns in virtual microscopy have been recently studied [151], and the few reported studies use very complicated interfaces such as eye trackers, which may bias the observed patterns. As far as we know, there exists only another study that has recorded diagnostic paths during actual navigations [151] but its analysis is totally different since therein, authors do not analyze velocities and time is not included as a criterion at defining what a RoI is. Such study is focused on exploring prefetching and caching as possibilities to reduce navigation and transfer times. This study reported partial image covering, as observed in the present investigation, and an average linearity coefficient of 0.41 for the complete diagnostic path. Interestingly, they also found a 0.85 linearity coefficient when the analysis was carried out on three consecutive steps of the diagnostic path. On the contrary, we assessed linearity for every single step which composes the diagnostic path, when using virtual slides in a conventional laptop and mostly under the scanning phase. Our results showed a 0.8 linearity coefficient, a quite coherent figure when comparing with the study mentioned before. In addition, we identified velocity patterns in scanning tasks, consisting in a rapid increase of velocity when pathologists leave the RoI and a decreasing profile velocity when the expert is nearby to a new RoI.

So far an optimal GUI design in virtual microscopy is still an open problem. Many virtual microscopes try to emulate the experience of navigating a real microscope, that is to say, to move a microscopical stage while zooming in and out. The simpler exploration strategy consists in using a unique window, which stands for the microscopical objective and provides interactions with a virtual stage by means of mouse panning while the zoom operations are simulated by clicking. This strategy is of course closely related to a real microscope exploration, however if an expert might leap between two high magnification regions, that expert must zoom out from the first region, displaces the field of view to the second region and then zoom in. This pattern constitutes a natural movement with any actual microscope, but for a virtual device, it ignores main display capacities of a virtual interface. This complex set of operations can be drastically reduced by taking advantage of both digital storage and display potentialities. The GUI herein presented approached this problem by "focus & context" [78], a set of techniques that combines a "focus view" (the auxiliary window), i.e., the GUI part charged of displaying the high degree of detail, and a "context view" (thumbnail window), charged

of presenting the VS at low resolution.

A fundamental hypothesis in the present work for the presented design is that an expert filters information out using the "context view" and then switches to the "focus view", on which the process of information refinement and diagnostic, is achieved. Therefore, this focus view occupies most of the available area, while the smaller part for the context provides orientation during interaction. For the sake of interaction, this focus view is placed within the context view, allowing a scanning-like display as well as an additional view of the whole contents. This design compensates many disadvantages of ordinary scanning because every interaction can then be executed using only the content view so that leaps between high magnification regions become simple displacements of this focus view within the context view.

Part III

Pathologist's Navigation Model

5

Soft-cache Strategy for Pathologist's Navigation in Virtual Microscopy

Gómez F., Marín D., Romero E.. A Soft-cache Strategy for Pathologist's Navigation in Virtual Microscopy. Accepted for publication in Microscopy Research and Technique.

Abstract *This chapter presents an optimal soft-cache strategy, which improves the navigation times in virtual microscopy. The entire method includes an optimal soft-cache strategy and a dynamical probabilistic model of a pathologist's navigation. This strategy was implemented as a Client-Server application, using the JPEG2000-JPIP standard and evaluated using different navigation patterns, namely, four different pathologists exploring ten VS, stained with different dyes. The present approach was compared with a conventional soft-cache method and the cache performance improved, in average, in about a* 10%.

5.1 Introduction

A virtual microscopy is a system which allows efficient storage of large microscopical images while displays different Windows of Interest (WoIs) at any resolution and desired quality [48]. Such microscopical emulator should be adaptable, scalable and friendly so that it effectively makes possible to explore actual histological slides. Overall, the size of these microscopical images introduces considerable delays, which results in difficult interactive and fluid navigations [85]. Strategies such as caching or prefetching aim to decrease the latency times and therefore to allow seamless navigations. Cache is a rapid

access to a space of memory in which it is stored the portions of the VS that shall be visited in the future [29], while prefetching consists in anticipating user requirements. These techniques have shown to highly improve navigation times [68, 135], but their performance depends on the prediction capacity of the system [35, 113]. Recent investigations [35] have demonstrated that seamless navigations can be reached using cache and prefetching [68]. Furthermore, processing times have improved up to a 30 % using a naive cache strategy [68]. Nevertheless, the success of these strategies is highly dependent on the computational architecture used for exploration [102, 151, 68] and on the degree of knowledgement one may have about how a user runs over the virtual slide [35]. This chapter presents a prediction strategy which is adapted to navigating pathological images, permitting to prefetch pieces of images while also stores blocks which very likely will be used in the future. The main contribution of the present work has been the design of a task oriented model which takes advantage from the JPEG2000 (J2K) [24] scalability by working at the minimal J2K information unit: the packet. The model uses a soft-cache, adapted from [101], for which the access probabilities are generated by a navigation model of the pathologist diagnostic pathways. This strategy was assessed with a very variable set of navigation patterns, generated by different pathologists and several test specimens, stained with two dyes. Results show a systematic gain when comparing this strategy with a traditional soft-cache method.

This chapter adapts a modern communication architecture, specifically designed to support the efficient access to high resolution images, for virtual microscopy applications. The soft-cache strategy herein presented was designed for the client-server architecture (See Section 2.4), similar to JPIP-W [102] which is illustrated in figure 5.1 .

At the beginning, the client requests a Window of Interest to the JPIP server, which then figures out the list of packets that belong to that WoI. This list is then sent to the proxy cache which checks the required packets out in the cache and sends them to the decoder. Missing packets, in the proxy cache, are requested to the JPIP server. The cache is updated using a Cache Model that privileges packets with high probability to be used in the future. Finally, the complete set of packets is sent to the decoder and the image is displayed at the client side. As illustrated in figure 5.1, this architecture can save network resources because only packets not available in the proxy cache are required. This time gain highly depends on how the cache model is devised.

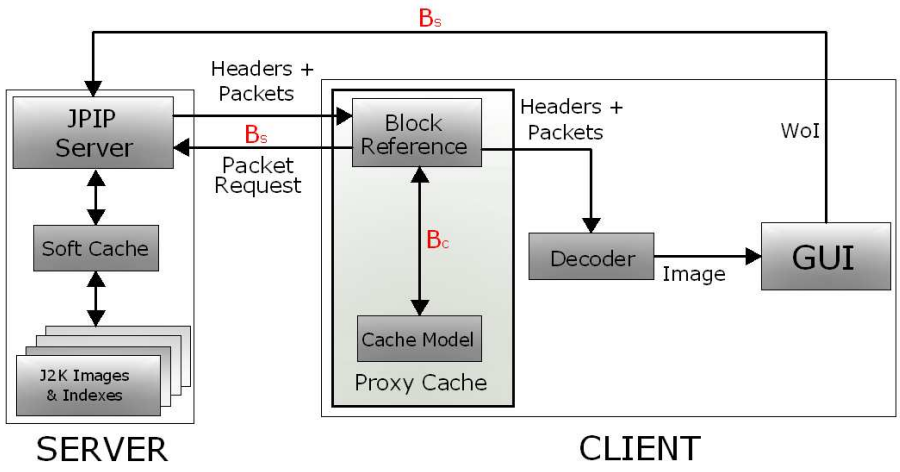


Figure 5.1: The soft cache architecture: when a WoI is requested, the server sends to the client the list of packets that belong to that WoI, the client then checks out which packets in the list are in the proxy cache, and queries the missing ones from the server.

The chapter is organized as follows: this Section presents the cache problem and a brief discussion about how related problems have been approached, Section Materials and Methods introduces our cache policy and the proposed navigation model, Section Results demonstrates the effectiveness of the cache method. Finally the last Section concludes with a discussion.

5.1.1 Cache Strategies in Virtual Microscopy

The cache problem in virtual microscopy can be twofold addressed: by data granularity, i.e., the level flexibility of data representation in the cache, and the cache replacement policy, which should specify how data must be replaced when a new image portion arrives. In terms of granularity, the simpler strategy may be to store the lower image resolution levels [8, 68] so that a fast access to WoIs at specific zoom levels is reached, but at a high memory cost. Besides, the size of the cache information units to be stored is important for devising a preference model. If these units are small, a large image WoI can be more probable than another with less elements of higher size [151]. Furthermore, cache elements with larger size, associated to relevant information [134], can yield smaller distortion results when they are queried [134].

Another factor, which influences the weight given to this basic cache unit, is its associated cost [8], expressed in terms of processing or transmission time. Therefore, a proper balance for a particular application should be maintained. An alternative to this choice may be to store complete portions of the coded image, whereby the granularity level is determined by the compression format. Overall, JPEG is the broadly used format in many virtual microscopy applications [18, 151, 41]. Nevertheless, this format does not provide access by resolution and introduces quality losses in the final result, an unacceptable issue in most diagnostic tasks [106]. In contrast, J2K is a more flexible alternative [68, 102], provided with random spatial access at different levels of resolution, quality, and lossless codification.

The cache replacement policies should be based on the user preferences, that is to say, elements with lower preference levels should be firstly removed, but only when the cache is full. The most popular cache replacement policy in virtual microscopy has been the Last Recently Used (LRU) [8, 68], for which the user preference is modeled as a function that decreases with the element age. The underlying hypothesis in this model is that the pathologist will revisit, with higher probability, the image zones that she/he just visited, a navigation pattern that very rarely is observed in actual navigations [79], above all when the pathologist reaches a particular degree of expertise. Another broadly used strategy is the Last Frequently Used (LFU) [102, 151, 8, 68], whose fundamental assumption is that higher probabilities are given to those regions with larger number of visits. These two strategies turn out to be quite general for a very oriented domain as virtual microscopy, for which times associated to navigations are straightforwardly related to expertise. In the context of the present investigation, both strategies can be considerate as comparable since experimental evidence in multiclient environments reports similar performances with LRU and LFU, for queries from multiple applications and variable input/output requirements [64].

5.1.2 Pathologist Navigation Patterns

The success of a virtual microscopy system depends totally on the system ability to fulfill the user expectations so that the system must be devised to easily meet different expert needs. An obvious question is then how to adapt this kind of systems to processes which turn out to be stochastic by nature, at least at a first sight. In fact, pathologist navigation appears to be a complicated series of back and forth changes, sideway movements, all at

different speeds and tempos. However, experimental evidence [79, 28, 137] obtained during the last decade, indicates that such patterns are not random at all but rather complex movements, guided by a long training period which minimizes the interaction time between the expert and the slide contents in terms of the spatial information organization [79]. Section 4.1.1 disccused evidence about how pathologists develop particular skills for exploring slides, following two complementary patterns: scanning (panning) and interpretation [16, 140, 28]. In general, at the scanning phase, examination is carried out at the lower magnification because pathologists attempt to locate relevant information as some particular spatial arrangements of the histological sample. This phase spans most of the navigation time, while a second examination phase (interpretation), is conducted when a further information analysis is required. In this case, navigation patterns consist in spatial exploration of small zones [28] of the VS, upon which only gentle movements are required. During this phase pathologists change magnifications, interpret information and identify structures for diagnosis [28, 137, 79]. All this evidence contradicts then the hypothesis that exploration patterns are unpredictable.

5.2 Materials and Methods

The pathologists navigations used in the present investigation were described in Sections 3.3 and 4.2.

5.2.1 The Fundamental Cache Unit

A microscopical exploration of an actual slide is composed of a complex sequence of different events, basically of two types: magnification changes (zoom in and out) and spatial displacements at different velocities, times and directions. A virtual microscope should be able to emulate such heterogeneous sequence on a VS. Actually, depending on both the image contents and its size, this process can result so variable that information has to be constantly reconstructed from basic units, in terms of different magnifications, qualities and random accesses to the image. This makes that flexible image data representations such as the J2K standard [2] results adequate to meet such requirements. The advantage with this standard is that it allows data representation with high granularity and loose coupling, using basic in-

formation units, namely the packets. A packet corresponds to a flexible J2K unit, which contains data from a resolution level (magnification), precinct (a particular image region at different resolutions or magnifications) and quality[1], all of them independently stored. Hence, several VS versions with different zoom levels or image regions can be constructed when decoding packets of a specific image region with a particular setting of magnification and quality. An actual clinical scenario requires not only such data representation (J2K) but also adaptable communication protocols [131, 2].

5.2.2 The Soft-cache Strategy

A J2K image corresponds to a set $D = \{k_1, k_2, k_3, \ldots, k_n\}$ of n packets. Each packet k_i contains a portion of image data, indexed by precinct pr, resolution r and quality l, i.e., each packet is associated to a (pr, r, l) vector quantity. Two types of access can be implemented over the J2K packets [73]: hard and soft. In the former case, the user always requires the entire set of packets, that is to say, the maximum resolution and quality levels. In the latter case, intermediate resolutions and qualities are demanded by the user and therefore a variable subset of packets is required. The strategy herein proposed is then adapted to having a soft access because it is the natural request in virtual microscopy. For doing so, let us be B_s and B_c the server-client and Proxy cache-client bandwidth connections, respectively (see Figure 5.1). A navigation is composed of a sequence of WoIs so that at each time t, a WoI can be considered as a set of packets, i.e., this WoI is also a vector quantity with a spatial associated region pr, a given resolution r and a quality level l. Optimal navigations are reached when a set of packets A_t, $A_t \subseteq D$ associated to the current WoI, is retrieved in minimum time. According to [73], the delay $\delta(A_t, X_{(pr,r,l)})$ associated to this WoI can be expressed as:

$$\delta(A_t, X_{(pr,r,l)}) = \sum_{k(pr,r,l)\in A_t} (X_{(pr,r,l)} S_{(pr,r,l)}) B_c^{-1} + ((1 - X_{(pr,r,l)}) S_{(pr,r,l)}) B_s^{-1} \quad (5.1)$$

[1]The abstraction in the standard allows to navigate along different quality levels which amount to splitting quality information by independent layers that incrementally contribute to the observed image quality.

where $S_{(pr,r,l)}$ corresponds to the size of each packet $k(pr,r,l)$ and $X_{(pr,r,l)}$ is a binary variable that indicates whether or not a packet is in the cache, that is to say:

$$X_{(pr,r,l)} = \begin{cases} 1, \text{ if } k(pr,r,l) \in Q_t \\ 0, \text{ in otherwise} \end{cases}$$

with Q_t the set of packets stored in cache at the time t. The first term in Equation 5.1 corresponds to the delay associated to uploading packets from the proxy cache while the second term amounts to the delay when uploading packets from the server.

An optimal bandwidth use is then reached when the expected total uploading time over every possible user request is minimized, that is to say:

$$\min_X \sum_{A_t \in C} P_{A_t} \delta(A_t, X_{(pr,r,l)}) \tag{5.2}$$

$$\text{subject to} \sum_{k(pr,r,l) \in D} X_{(pr,r,l)} S_{(pr,r,l)} \leq S_Q$$

where P_{A_t} is the probability of requesting A_t, C is the set of every possible request and the restriction is related to the maximum cache size S_Q. Note that if this problem were solved in the C space, which corresponds to every possible WoI, the problem would result in a very difficult combinatorial problem. Hence, this formulation was simplified as the problem of estimating the expected values associated to the delays produced by a WoI, and reformulated as the problem of finding the set of packets, in terms of a query, which maximally save time. Equation 5.2 is then re-written as the following knapsack problem [92]:

$$\max_X \sum_{k(pr,r,l) \in D} X_{(pr,r,l)} \left(S_{(pr,r,l)} B_a^{-1} \sum_{A_t \in \Omega_{(pr,r,l)}} P_{A_t} \right) \tag{5.3}$$

$$\text{subject to} \sum_{k(pr,r,l) \in D} X_{(pr,r,l)} S_{(pr,r,l)} \leq S_Q$$

where $\Omega_{(pr,r,l)} = \{A_i \mid A_i \in C,\ k(pr,r,l) \in A_i\}$ is the set of requests which contains the packet $k(pr,r,l)$ and $B_a^{-1} = B_s^{-1} - B_c^{-1}$ is an acceleration factor given by the cache use. The term in parenthesis stands for the relevance

level of a packet, which in terms of a knapsack problem is a consequence of a proper balance among the probability of accessing a packet, its size and the bandwidth gain. In this expression we have to estimate $\sum_{A_t \in \Omega_{(pr,r,l)}} P_{A_t}$, a quantity that corresponds to the sum over every possible WoI containing the packet $k(pr, r, l)$, and which is approximated by the probability of the packets belonging to the intersection (see Figure 5.2), namely the packet probability itself $P_{(pr,r,l)}$.

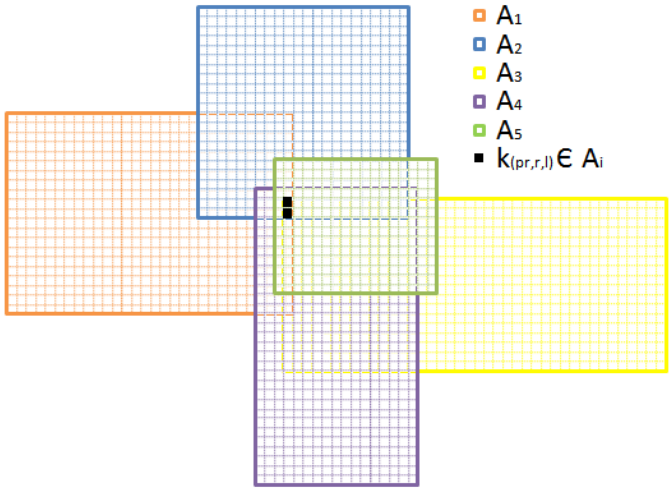


Figure 5.2: The probability of the intersection of different overlapping WoIs (A_i) is approximated by the packet probability located in this intersection.

Finally, the soft-cache problem for J2K packets in the JPIP-W architecture reads as:

$$\max_{X} \sum_{k(pr,r,l) \in D} X_{(pr,r,l)} S_{(pr,r,l)} P_{(pr,r,l)} B_a^{-1} \qquad (5.4)$$

$$\text{subject to} \sum_{k(pr,r,l) \in D} X_{(pr,r,l)} \cdot S_{(pr,r,l)} \leq S_Q$$

In this expression the term to be calculated is $P_{(pr,r,l)}$ since the other terms, $S_{(pr,r,l)}$ and B_a^{-1}, are already known. In the next subsection we introduce a novel probability model for $P_{(pr,r,l)}$, adapted to the process of navigating VS.

5.2.3 The Pathologist's Navigation Model

Overall, a diagnostic path is defined by a process of information search within a VS. In Figure 5.3 an actual navigation, spanning $113s$, illustrates the whole process: the pathologist scans the VS looking for some specific patterns which, once located, are further explored using a higher magnification. The whole sequence has been split into the four panels as (a), (b), (c) and (d). Panel (a) shows the beginning, the pathologist is forced to start at the upper left corner and displaces the WoI looking for a region of interest. Once this one is located, panel (b) shows the magnification change from the thick to the dotted window. Panel (c) displays a re-sizing operation of the auxiliary window. Finally, in panel (d), the pathologist scans the rest of the VS, using the same magnification.

Provided that a diagnosis is set when certain histopathological patterns are well established, and that these patterns are usually located in specific places of the VS, a pathologist's navigation can be though of as a sequence of jumps between WoIs [28]. A GUI of any virtual microscope should then be devised as to easily allow jumps from one WoI to any other location, dragging a desired WoI within the thumbnail window. Once this one is reached, a decreasing velocity profile is observed since this new WoI deserves a certain amount of time for examination [79]. When this level of detail is not enough for taking a diagnostic decision, the pathologist will go into a further information level by increasing magnification. In consequence, the navigation process can be seen as a search of information through two basic operations: scanning and zooming.

In general, every pathologist follows a standard methodology with two coarse phases: first examination is carried out at the lower magnification in order to locate relevant information, while the second and further examination is conducted for analysis of the slide contents [79]. This last task is typically carried out by zooming-in, as in the VM case, as well as by increasing the quality level. This experimental evidence leads us to consider that magnification changes and quality levels may be both handled together. Based on this, our fundamental hypothesis is that a navigation is a process composed of two complementary strategies: a spatial scanning for classifying information and a further exploration for image interpretation, i.e., higher magnifications and quality levels (Figure 5.4). In consequence, the probability of requesting a particular packet should depend on the particular navigation phase: information classification or image interpretation. In these

terms, we propose to model a quantifiable interest on the packet $P_{(t,pr,r,l)}$ at time t as the weighted average among three sources of preference: spatial $P_{t,pr}$, resolution $P_{t,r}$ and quality $P_{t,l}$, as follows:

$$P_{(t,pr,r,l)} \propto \alpha_t P_{t,pr} + (1-\alpha_t)(P_{t,r} + P_{t,l})$$

the ($\alpha_t \in [0,1]$) factor changes during the navigation and determines the importance of each of the exploration phases, in other words, this parameter switches between either scanning the slide or looking for further information. This interaction is then modeled by supposing that the last recorded user event is used for approximating the near future as a first order, i.e., the α_t weight takes its value from what happened in the very recent navigation history. When a transition between the navigation states is registered (scanning to interpretation or viceversa), the α_{t+1} value changes to $1-\alpha_t$ for the current operation is privileged. An estimated value for this parameter was obtained by a sensitivity test using the whole set of navigations, a result which is presented in Section 7.3.

By simplicity, we are going to consider that each preference source can be then modeled as a first order discrete stochastic process, under the assumption that a navigation decision is made in terms of the current navigation state, plus a short history. Yet a higher order model could be used, its order selection is not a trivial issue, while a high computational cost should surely have to be paid. The initial conditions for the model are set as follows:

- *Spatial preference*, at the beginning, there is no reason to privilege a specific spatial region and therefore every spatial region is equally probable. In this phase (scanning), specific image WoI information obtained from segmentation algorithms [99, 59] or visiting logs [123] can be used for initializing the probability distribution.

- *Resolution preference*, the choice of a resolution r is independent of the sequential selection of previous resolutions $0, \ldots, r-1$. This statement is consistent with the fact that pathologists know at which particular level of detail they have to look for information. Hence, the $0, \ldots, r-1$ sequence can be considered as failures while the r resolution selection as a success. If each resolution selection is independent and the probability of success is λ, then the probability of a resolution r can be modeled by a Bernoulli distribution $(1-\lambda)^r\lambda$.

- *Quality preference*, the initial probability state for the quality dimension was calculated as the rate-distortion ratio, defined in the J2K standard and associated to each packet [134]. This value provides an approximation to an optimal compromise between the packet size and its inherent visual quality, as previously demonstrated in previous works [35].

Once these initial conditions are set, probabilities are updated after the pathologist carries out a particular action. Probabilities are then updated as follows:

- *Spatial preference*, when a pathologist is spatially moving or scanning the VS, its current position defines a neighborhood of influence in which the probability should be higher, that is:

$$P_{t,pr} = P_{t-1,pr} + \sum_{pr' \in N(pr)} P_{t-1,pr'}$$

  where pr is a precinct and $N(pr)$ corresponds to its neighborhood (8-SED). This probability update is carried out for every precinct within the current navigated area.

- *Resolution preference*, is updated by taking into account the sequence of previously used resolutions:

$$P_{t,r} = \begin{cases} P_{t-1,r}, \text{ if there are no zoom changes} \\ P_{t-1,r} + \sum_{r \leq R(t)} P_{t-1,r}, \text{ in zoom-in operations} \\ \max\{0, P_{t-1,r} - \sum_{r > R(t)} P_{t-1,r}\}, \text{ in zoom-out operations} \end{cases}$$

  where $R(t)$ is the zoom-level at time t, the zoom-in operation corresponds to $R(t) > R(t-1)$ and zoom-out to $R(t) < R(t-1)$.

- *Quality preference*, the probability associated to a quality should increase when the pathologist stops at any VS location, looking for further information:

$$P_{t,l} = P_{t-1,l} + \sum_{l \leq L} P_{t-1,l}$$

  with L the current level of quality, note that only increasing probability levels are possible.

Once each preference source is updated, each of these values are normalized and the probability packet $P_{t,pr,r,l}$ is computed.

5.2.4 Evaluation

Assessment of the proposed strategy was carried out using recorded navigations, performed by the group of pathologists on the set of available Virtual Slides [57]. The recorded queries -different spatial regions, magnifications and quality levels- were then expressed in terms of the required packets. The performance of the presented strategy was measured as the percentage of packets, in bytes, which were within the cache when they were required. For so doing, the percentage of hits and misses in bytes was determined for different cache sizes and compared with a baseline strategy, i.e., a LRU cache[2]. LRU was selected as the baseline because this is a well known replacement strategy which is independent of the architecture. The cache size was calculated as a variable percentage of the image at the lower resolution, that is to say, the sum over the size of all packets at lower resolution $\sum S_{(pr,0,l)}$. Several cache sizes were then evaluated, the 50%, 100%, 105%, 110%, 120% and 200% of the lower resolution size. The soft-cache problem (equation 5.4) was solved through a greedy solution [92], the cache space was filled with packets $k(pr, r, l)$ for which the product $P_{(pr,r,l)}S_{(pr,r,l)}B_a^{-1}/S_{(pr,r,l)} = P_{(pr,r,l)}B_a^{-1}$ is the highest, until the available space was completely full. Some basic system characteristics were fixed as constants, such as B_c and B_s, as to simplify the analysis. The exact value of these parameters is not really critical because the client cache system applies over a wide range of the system parameters. Additionally, the model parameters α and λ were tuned, searching at the best model performance, by varying both values in the interval $[0, 1]$.

5.3 Results

The four pathologists navigated through the ten VS so that a total of 40 navigations were available either for training or assessing. Overall, the relative size of these VS varied between 10 % to a 30 % of the whole histological sample so information was limited or partial. This strategy pursuited to accomplish a maximal navigation time since pathologists tried to exhaustively

[2]Our LRU replacement policy consisted in replacing the packets with no use which longer had lasted within the cache

search the needed information as to reach a diagnosis. The patterns generated by this type of navigation resulted then ideal to assess this cache strategy. Each pathologist was previously trained on virtual microscopy using two test images, which were natural mega-images with similar sizes to the ones previously described. Finally, the time used for navigating varied between 35 s and 398 s, depending on the image contents. In general, navigations took shorter intervals of time for images stained with immunofluorescent techniques, for which the islets of Langerhans constitute the main information source. On the contrary, inflammatory pathologies stained with Hematoxylin-Eosin spanned larger exploration intervals and very rarely they were able to accomplish a correct organ diagnosis, but they could repair the most important histological structures, achieving a maximum and comparable level of description.

5.3.1 The Cache Strategy

The Figure 5.5 shows, through the four panels, the cache performance of the same navigation illustrated in Figure 5.3. For so doing, the navigation operations have been superimposed to the percentage of hits that were within a cache of a 110 % of the lower VS resolution size. In this case both α and λ model parameters were set to 0.8.

Each of these plots depicts the percentage of hits in the cache space in kbytes, the y-axis represents the percentage of packets within the cache space, while the x-axis stands for the WoI number. The two different cache assessed policies are drawn at each of the four plots, the thick line corresponds to the cache strategy herein presented while the dotted one stands for the LRU cache procedure. The first panel (left upper panel) shows the very first part of this navigation, corresponding to a exclusively scanning pattern. This navigation part spans the interval between the first WoI and the WoI defined by the number one, also superimposed to this graph. Within this interval, the percentage of hits is quite high for both strategies, above a 60 %. However, our method outperforms always the LRU in this phase, between 10 % to 20 %. Once the expert located the desired WoI defined in panel a, a higher magnification level was set in panel b. This navigation part corresponds to the interval between numbers one and two. The number of hits decreases dramatically for both strategies, from a 90 % to a 30 %, but again the proposed method is slightly better, with a performance which is higher in a 5 % to 10 %. The change shown in panel (c) corresponds to an enlargement

of the window size, increasing the available field of view, that is to say, the pathologist remains at the same resolution but searching further information. The interval defined by this operation is between numbers two and three. The performance decreases even more during this interval, yet the reduction in the number of packets within the cache is nearly constant. This pattern indicates that both strategies adapts very poorly to this type of change since their values are alike, even though ours is slightly better. Finally, in panel (d), the expert continues to scan the VS, an interval defined by numbers three and four. The performance improves for both cache strategies, up to reach the initial level, but our strategy reaches much more rapidly these levels, indicating a better adaptation capacity to these discontinuous jumps, and again outperforms the LRU cache.

5.3.2 Adjustment of the Parameters α and λ

The α parameter maps what is happening within an actual navigation to a particular pattern, i.e., it controls the balance between scanning the slide or looking for further information. An optimal value for this α parameter was selected by a sensibility analysis using the whole set of navigations. Table 5.1 shows the percentage of hits when α and λ were set to five and three different values, respectively. Figures indicate that an α selection of 0 (interpretation pattern) results in a performance of 36 %, and of about 42 % when this value is set to 1 (scanning pattern). Intermediate α values result in higher performance percentages, with a peak when $\alpha = 0.8$. These results suggest that navigations are actually a complex mix of the two exploration patterns and that the α parameter allows to follow their importance. A higher importance of the scanning pattern was herein observed, probably because the scanning phase spanned larger times at interacting during a navigation.

α	λ	% Hits	α	λ	% Hits	α	λ	% Hits
1	0.8	42.59 ± 21.31	1	0.6	42.53 ± 21.38	1	0.3	42.49 ± 22.14
0.8	0.8	44.18 ± 19.83	0.8	0.6	44.15 ± 20.30	0.8	0.3	43.98 ± 19.93
0.5	0.8	41.22 ± 24.17	0.5	0.6	41.16 ± 23.32	0.5	0.3	40.94 ± 24.38
0.2	0.8	38.83 ± 21.15	0.2	0.6	38.29 ± 20.50	0.2	0.3	38.07 ± 20.36
0	0.8	36.56 ± 24.10	0	0.6	36.22 ± 24.22	0	0.3	36.01 ± 24.73

Table 5.1: Percentage of cache hits for different α and λ values.

Likewise, λ is related to the probability of making zoom operations.

Higher values correspond to larger probabilities of zooming into high resolutions and viceversa. A λ selection of 0.8 privileges zoom-in operations, resulting in improvements of nearly a 1%, a gain which is non significant (t-student test, $p < 0.05$). These results indicate a low sensibility to the λ parameter for fixed α values (see figures per rows), a fact that may be attributed to the low number of zoom operations compared to the total number of user interactions.

5.3.3 Proposed Cache Strategy vs LRU

Figure 5.6 shows the averaged percentage of hits in the cache, calculated for the whole set of navigations, for different cache sizes. The baseline size or the basic cache size was set as the size occupied by the packets needed to reconstruct the lower resolution VS, because these packets are the ones with the minimal information at the reconstruction time and then any cache system should start from this store capacity [85], at least.

For small cache sizes, near to a 50% of the basic cache size, the difference between policies is neglible and their performace reaches approximately a 20%. Once the cache capacity is enlarged to 100%, the proposed strategy outperforms in about a 10% the LRU, a gain which is maintained nearly constant up to a size of 200 %. Interestingly, both strategies improve the cache performance which, in the current experiment, is between 30% and 45%, a result which illustrates the importance of using cache strategies for improving the level of interaction during a navigation.

On the other hand, as the gain introduced by the proposed cache is dependent on the particular temporal patterns, which are completely hidden when one computes means and standard deviations, it is more suitable as a global performance evaluation, a difference of temporal series. This was herein computed as a point-to-point difference between the two strategies during the entire set of navigations. For doing so, we calculated the differences between both hit ratios at each time of each of the navigations. This set of differences were then expressed as a mean and a standard deviation so that a benefit is stable if this difference is positive and its standard deviation is small, i.e., our method outperforms LRU. The benefit obtained was evaluated by subtracting both performances, i.e., the proposed strategy and LRU policy, with an α value set to 0.8. Computations were carried out for the complete set of navigations under different cache sizes, namely 50%, 100%, 105%, 110%, 120% and 200%.

Cache Size	**Pth 1** Gain	**Pth 2** Gain	**Pth 3** Gain	**Pth 4** Gain
100%	5.34 ± 2.77	9.02 ± 7.56	8.37 ± 3.37	2.02 ± 1.10
105%	5.48 ± 2.86	9.36 ± 7.58	8.62 ± 3.62	2.04 ± 0.98
110%	5.54 ± 2.88	9.41 ± 7.51	8.78 ± 3.78	2.06 ± 1.05
120%	6.31 ± 3.02	10.71 ± 7.64	9.27 ± 4.46	2.11 ± 1.01

Table 5.2: Benefit of the proposed method over LRU, expressed in mean and variance of the difference between cache hits for different cache sizes.

Given that different experts cope with different patterns, this analysis was carried out individually for each of the four pathologists, as observed in table 5.2. It should be strengthen out that all figures in the table are positive so that it can be globally inferred that our method always outperforms the LRU, disregarding the particular pathologist and the different cache sizes used in this evaluation. Table 5.2 shows that this gain is systematically higher for the first three pathologists and smaller for the last one. This particular case corresponded to a pathologist who simply enlarged the auxiliary window as to cover a maximum slide surface, and then displaced it all over the image. Yet this pattern is valid and this is why we decided to include it in this study, it does not use most of the flexibility and potentialities that the GUI offers and then our model could not capture the observed pattern. At any case, there exists a gain using the proposed strategy on the total cache performance.

Finally, as the navigation is composed of two basic patterns, namely scanning and interpretation, a part of the analysis was dedicated to figure out in which phase the cache policy is more important. It should be strengthen out that from the entire set of navigations, the scanning pattern stands for a 92 % of the total number of interactions with the GUI.

Patterns	**Pth 1** Gain	**Pth 2** Gain	**Pth 3** Gain	**Pth 4** Gain
Scanning	5.53 ± 2.96	10.04 ± 6.84	8.33 ± 4.33	2.07 ± 1.14
Interpretation	11.70 ± 4.40	7.18 ± 2.29	7.21 ± 2.07	1.83 ± 0.59

Table 5.3: Benefit of the proposed method over LRU for different navigation phases, expressed in mean and variance of the difference between cache hits for a cache size of 120% of the size of the packets in the lower resolution for the complete set of navigations.

Table 5.3 shows the benefit obtained for each of the navigation phases: scanning and interpretation, when the cache corresponds to a 110% of the basic cache size. As we can see our method provides a better performance for the scanning phase in the three last pathologists. Note that in the case of the first pathologist, much of the gain comes from the interpretation phase, an 11% compared with the 5% of the scanning phase. This pathologist used during navigations a larger number of zoom changes, when compared to the others. This pattern allowed that the model could better follow the interpretation phase when the pathologist was actually switching from zooming-in to zooming-out. The pattern obtained from the fourth pathologist produced a smaller performance, an issue already discussed.

In this section we have demonstrated that the presented cache policy provides a robust strategy to accelerate the navigation through various experiments. The results show better cache hit performance for the proposed method, when compared with the LRU replacement policy. A consistent benefit associated to the strategy was observed in three of the four subjects, as well as a marginal gain for the four pathologist. This benefit was also observed in both navigation phases: scanning and interpretation. The results also indicate that the main limitation of our strategy is the poor adaptation to resize WoIs without scanning.

5.4 Discussion

The present investigation has introduced a new cache strategy for accelerating actual pathologist navigations in virtual microscopy. The whole method is based on both, an optimal allocation strategy of J2K packets in the cache space and a new dynamical probabilistic navigation model of the slide examination. The allocation strategy results in a minimization of the pathologist navigation times upon server-client cache architectures. The pathologist navigation model is supported on two prior factors: the pathologist moves locally when spatially exploring the VS and the only magnification changes are given towards having larger information versions, i.e., the expert uses mostly zoom-in patterns during the exploration. Finally, a simple and fast updating strategy of the exploration probabilities, allows the strategy to adapt to both global and non-stationary navigation patterns, such as the observed in VM navigations.

A virtual microscopical exploration, performed by an expert pathologist,

is generally carried out in client-server architectures [151, 95, 6]. Such network structures allow remote access to large images, stored using several image formats [21, 151, 135, 102]. The proposal herein implemented was J2K [68] because of its large flexibility and adaptability to different navigation conditions, i.e., granularity. Unlike text or conventional images, for which computational resources, bandwidth and storage capacities are sufficient, in VM any of these factors will always be very limited because of the huge image requirements [118]. Moreover, it results fundamental for such an application, the development of strategies to handle the image data according to the specific exploration architecture. The classical approach to deal with this problem is the design of a cache policy, which in its classical version is very limited since it requires the storage of the complete image in memory (hard access) [73]. In the present work, the cache was designed, keeping in mind two fundamental issues: the image is stored in a granular format and exploration in virtual microscopy is soft by nature. The access to the image is then thought of as a soft one since the basic operations, zooming and scanning, have been already identified as local accesses [16, 140, 28].

Different soft-cache strategies have been already proposed in the literature, for exploring natural images [73], under the hypothesis of an underlying static navigation model, i.e., the user always visits the same regions in the image. Later, Li and Zheng [85] proposed an extended version of the previous work at including non-stationary patterns through a LRU replacement policy when locally navigating. In this work, an optimal cache strategy was used upon a time dependent probability distribution. The model could then directly code both the stationary and non-stationary components. Given that the cache solution is directly related to the probability values, it results of paramount importance to have a good model that captures the dynamics of the specific task, which corresponds in this case to navigation of pathological images. In the specific case of virtual microscopy, these strategies could hardly afford the complex requirements of an expert exploring a pathological image. In particular, it is nearly impossible to determine beforehand which regions the expert would visit in the case of the static cache, and for doing so it should be needed to construct a probability distribution function using either previous navigations -which obviously leads us to thousands and thousands of navigations- or explicit image information. In the case of the non-stationary cache policy, the drawback comes from the LRU foundation itself since this strategy supposes that the more important object within the cache is the one more recently visited, that is to say, that the underlying

hypothesis is that any user will revisit certain image regions, a statement which is hardly tenable in virtual microscopy since experts use a minimal number of regions with maximal information [151, 79]. On the contrary, our model is totally adapted to the specific task of exploring this VS and thereby is adaptable to different navigation circumstances.

In general, every pathologist follows a standard methodology with two coarse phases: first examination is carried out at the lower magnification in order to locate relevant information in terms of a spatial organization of the histological sample while the second and further examination is conducted for analysis of the slide contents [103]. Several studies have identified two complementary patterns for microscopical exploration: zoom-in and scanning (panning) operations [16, 140, 28]. Krupinsky has demonstrated [79], using an eye tracking system, that pathologists highly coincide on the regions they visit when navigating virtual slides. Although this information could be used for constructing a probability map associated to the VS, by storing the history of visits, this strategy results limited in clinical environments in which microscopical slides are explored by a reduced number of pathologists. In the present work this frequentist approach is rather substituted by a dynamical probabilistic model that captures most of the navigation patterns an expert follows. The proposed model is based on two intermingled phases of the navigation: scanning and interpretation. Both phases are closely related to the Information-processing theory which provides elements for current interpretations of visual search data in pathology [28, 79]. This theory establishes two main processes: firstly, a global impression of image content -such as symmetry, color, and gray scale content- is constructed by exploration. Then, the collected information is compared with information contained in the long-term memory that forms the viewer's cognitive schema (or expectations) about what information actually is, in the pathologic image. In low resolution medical images, such as radiology, the global impression is constructed by the Human Visual System (HVS) through saccadic and fixation movements [79]. These mechanisms are nevertheless insufficient to construct a general view of the VS because of the limitation for displaying such VS in conventional devices, and moreover, limitations of the HVS for capturing information in high resolution images [153]. In these terms, additional interactions are essential in order to locate relevant areas and construct the global impression. The model herein developed copes with these issues through a simplification of the navigation task, an expert is either scanning or looking for information of some interest or interpreting, which in terms of the naviga-

tion corresponds to the periods in which the expert has located information and is zooming-in or demanding a better quality representation.

Navigation patterns arise when a pathologist interacts with a particular VS, a complex relationship which is somehow regulated by articulating the two types of prior knowledge: the Pathologist's expertise and the information available on the particular clinical case. In the former case novices tend to wander around the slide while experts' observation paths are better located and shorter. In the latter case, additional information influences the kind of objects the expert is searching so that patterns change and again, experts' performance is neatly superior. In both scenarios, the model nicely adapts to the interaction rather than to a particular generated pattern, influence by any of these two situations.

Our results illustrate many of these issues. It can be inferred from Figure 5.3 that effectively the use of a LRU policy results in acceleration of the pathologists navigations up to 30%. Also, from the same figure, our task oriented model provides a better perfomance, resulting in additional acceleration values of about a 10%. These results confirm the previously formuled hypothesis about the utility of a task oriented model and recall the importance of a proper undestanding of the exploration process in virtual microscopy [151, 57]. Likewise, as observed in table 5.3, the method gain remains systematically higher when comparing with the LRU, in despite of the fact that for certain navigation patterns, highly deviated of our priors, the obtained gain is not so high. It should be strengthen out that these patterns were observed in one pathologist who did not exploit adequately the GUI potentialities. These results lead us to conclude that this oriented task model may speed up actual navigations in virtual microscopy [151]. On the other hand, the sensibility analysis table 5.3 of the α parameter shows that both phases, scanning and interpretation, are the basic patterns of an actual exploration. An optimal cache hit performance is reached in the present investigation when α is set to 0.8, indicating the importance of simultaneously taking into account both navigation phases. As expected, the proposed method provides different levels of performance for different subjects and navigation phases see table 5.3, however, its benefit remains higher at any test. This model was evaluated by tracking different patterns in very different experimental conditions, namely four pathologists performing fourty navigations in ten Virtual slides, stained with two different dyes. Yet it is frequent in psycophysic studies the use of virtual slides stained by a single method, we used VS with two different colorations, demonstrating

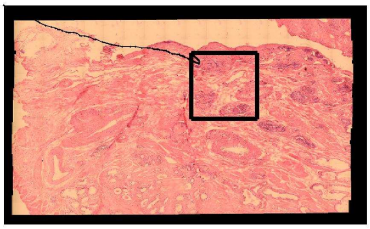
comparable results for each and indicating that the entire strategy is highly adaptable to the particular pathologist's navigation and therefore it results to be much more general.

Some recent works have found that it is possible to automatically determine RoIs [99, 57] so that a probabilistic map can be associated to the image. This kind of maps could be included as a prior knowledge in the probabilistic strategy herein presented and the initial conditions for the model could be drawn from these maps. This would accelerate the performance of the complete strategy, but then it would be necessary a preprocessing phase. In addition, observation paths should follow optimal sampling strategies [77] which improve the diagnosis times, that is to say, a minimal number of RoIs would drive the navigation. In that scenario, a pathologist would achieve a diagnosis spending a minimal time, so the pathologist workflow could be highly improved, basically because it would be possible to prefetch the whole examination path at the very navigation beginning and little additional information would be needed. However, observation paths are dependent on the image contents, on the pathologist's experience and the image quality, and therefore highly variable. These probability maps should then change along the navigation, depending on the expert needs. The advantage with the model proposed here is that it relatively independent of these variable factors and constructs a future diagnostic path using basic information of the mouse velocity and history of the navigation. It is worthy to point out here that our model was designed to cope with both phases, showing a consistent improvement during the entire navigation, when comparing with a baseline cache strategy. These predictions could be included into the whole strategy by using smart sampling procedures [77].

The main concern of this study was not the computational complexity so that the model was formulated basically searching to gain cache importance. However, the whole strategy can be considered as $O(n(\log(n))$, with n the number of image packets, corresponding to the probability updating step (Section 5.2.3) with $O(n)$ whilst the packet schedule procedure takes $O(n(\log(n))$. This last complexity is associated to the solution of the optimization problem that consists in sorting out packets by interest (see Section 5.2.4). In our simulations, for the image sizes described in Section 4.2.1, the proposed method takes less than $2s$ for each navigation step, when running on a Intel Xeon X5460 Quad-Core 3.16 GHz with 8 GB in RAM. In principle, these delays can of course impact negatively the possibility of seamless navigation. Nonetheless, again the model allows to deal with this problem

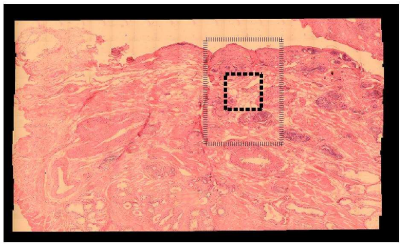
by updating the probability only on those packets which are susceptible to be changed, a strategy very similar to that used for the particle filter when estimating a *pdf*. This idea consists in representing a *pdf* with particles [10], which basically are subsets of a set of events Ω. In this context, a subset of events stands for the packets associated to a particular query, that is to say, a particular location, resolution and quality. Each particle is composed of a large set of packets and is entailed with an associated weight, calculated as the sum of probabilities of the packets within it [10]. This representation allows the probability updating step to be implemented as a simple change of the particle weights, while the optimization scheme consists in resampling these *pdf* weights, which means in terms of our simulations that the time falls from a couple of seconds to about $20\,ms$ with 1000 particles (data not shown).

There are several avenues for future research in this direction, namely, integration of more complex interaction patterns using Machine learning, characterization of navigation patterns with larger sets of experts and finally, acceleration of the image navigation by prefetching the useful image information during periods in which the expert is analyzing information [35]. Another research direction is the integration of automatically generated probability maps [59, 99] that would help to find, at any navigation time, the relevant information to be displayed so that maximum interaction can be reached with a minimal time.



(a) Scanning pattern

(b) Zoom in



(c) Window enlargement

(d) Multiple scanning patterns

Figure 5.3: This Figure illustrates a whole navigation, superimposed upon the thumbnail window. For the sake of the comprehension, this navigation has been split into the four panels and the sequence follows the order (a), (b), (c) and (d). Panel (a) shows the beginning of a navigation, the pathologist starts at the upper left corner and displaces the WoI. In panel (b) there is a change of magnification and therefore the WoI size changes from the thick WoI to the dotted one. Panel (c) displays the adaptation the pathologist performs for having the same visual field at the new resolution level, a re-size of the auxiliary window. Finally, in panel (d) multiple displacements are shown (a scanning profile) that are carried out at the new magnification and WoI size. Note that panels (a) and (d) corresponds to scanning phase and (b), (c) to the interpretation step.

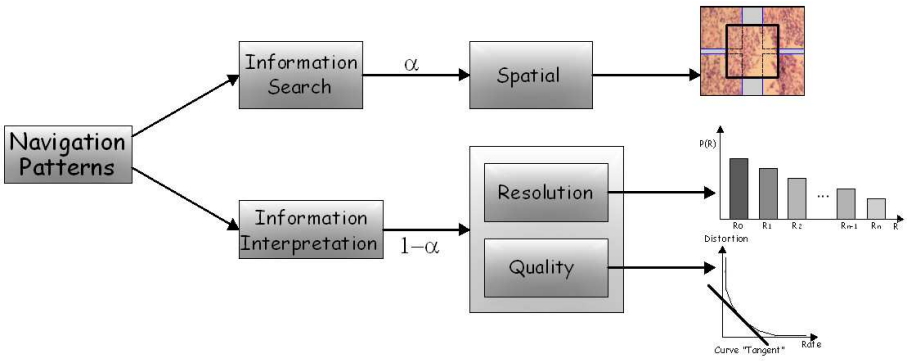


Figure 5.4: Navigation patterns in VM. A pathologist navigation is composed of two main tasks: information search and interpretation information. The former corresponds to spatial displacements, case in which the current navigation neighborhoods are relevant. The latter stands for every operation conducted for improving the information resolution, either zooming in or increasing the quality level. Note that the two operation are modeled by decreasing functions.

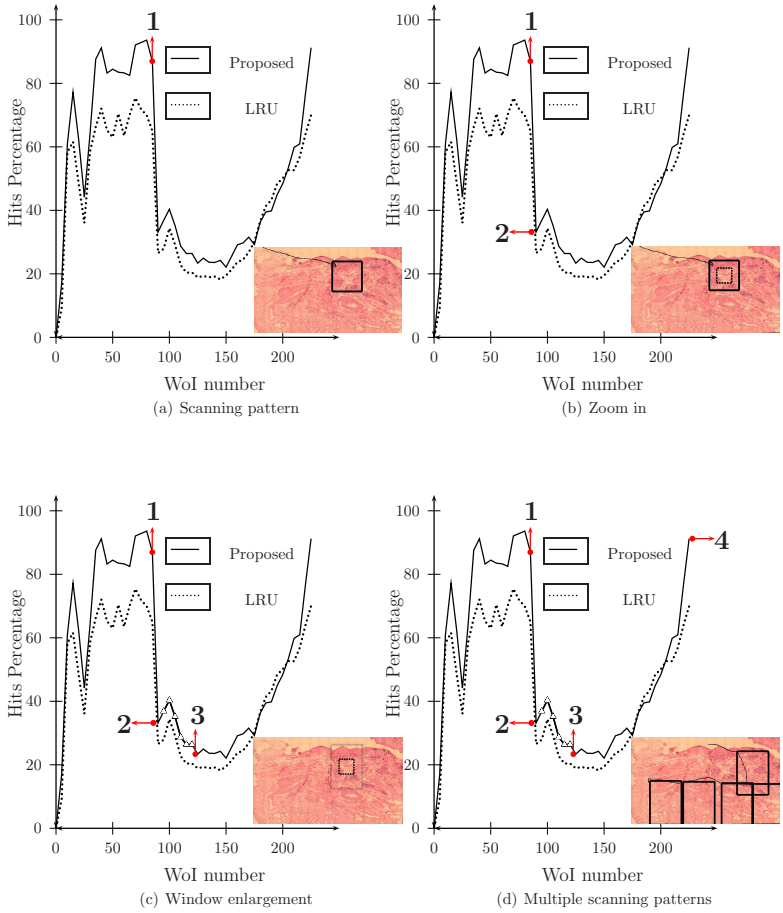


Figure 5.5: The four panels correspond to the same navigation illustrated in Figure 5.3, in the same consecutive order. Each panel displays the percentage of hits (y-axis) in a cache space of a 110 % for both strategies (ours and LRU), against the navigation time (WoI number in the x-axis). The selected parameters for our strategy were $\alpha = 0.8$ and $\lambda = 0.8$. The thumbnail has been superimposed upon each of these plots so that each part of the navigation can then be located as a percentage of hits in the cache. Panel (a) corresponds to the initial scanning task, from the beginning up to the arrow marked with one. Note that the proposed method outperforms the LRU in about a 20 %. Afterwards, the zoom-in operation displayed in panel (b) is observed in the graph a a rapid performance decrease for both methods, a period which corresponds to the points between arrows one and two. Interestingly, the proposed strategy always remains outperforming LRU. Panel (c), the time interval between arrows two and three, shows a WoI resizing operation. Note that both strategies hardly follow these changes but again, ours outperforms LRU (about 5 %). Finally, a series of scanning operations (panel (d)), at the same magnification, are observed in the performance graph as a rapid increasing which stabilizes around the same values of the navigation beginning.

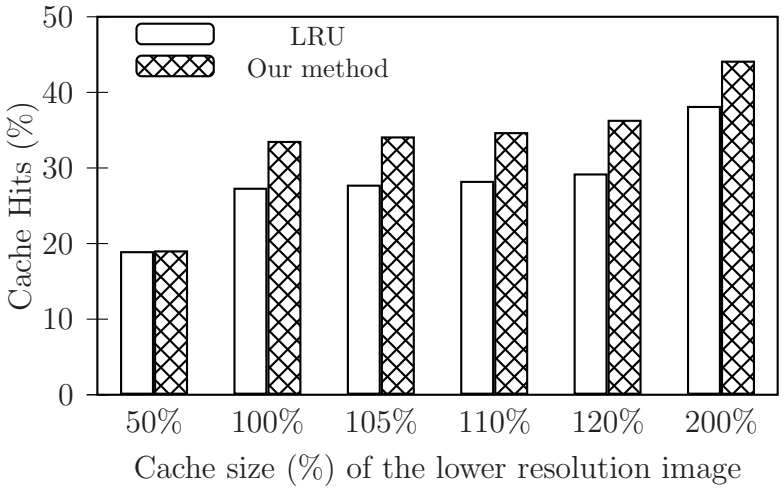


Figure 5.6: Cache performance measured in cache hits ratios for different cache sizes, corresponding to different percentages of the size of the packets in the lower resolution.

6

Predicting Pathologist's Velocity Profiles

Gómez F., Romero E. A Model for Predicting Pathologist's Velocity Profiles When Navigating Virtual Slides. Accepted for publication in Microscopy Research and Technique.

Abstract *This chapter presents a soft computing model which permits to anticipate the pathologist trajectories in diagnosis tasks when exploring virtual slides. The Bayesian strategy combines an offline model of a baseline pathologist knowledge (the prior) and a prediction online module (the likelihood) which captures a particular pathologist navigation pattern. While optimal parameters for the biologically inspired offline model, are calculated using an Expectation-Maximization strategy, prediction is carried out by a particle filter. Parameters are estimated from several series of actual navigations performed by several pathologists in different virtual slides. The present approach is compared with other conventional prediction methods and decreases the calculated MSE in about a* 50 % *for the entire group of pathologists.*

6.1 Introduction

Because of the size of these microscopical images, some image operation unavoidably introduces considerable response delays which result in difficult interactive and fluid navigations [85]. Strategies such as the cache or the prefetching have been developed for aiming at decreasing the latency times and therefore to permit fluid navigations. Cache is a rapid access to a space of memory in which it is stored the portions of the mega-image that shall be

visited in the future [29]. Prefetching is the anticipated uploading of those parts of the mega-image to which the navigation will be addressed in the future [34]. These techniques have shown to highly improve navigation times [34, 135], but their performance depends on the prediction capacity of the system [35, 113]. Available VMV software, such as ImageScope [39] or the John Hopkins' VMV [18], improves the navigation velocity using prefetching and cache strategies. In general, the used strategies are linear or second order approximations to the navigation velocity using considerations such as either the navigation velocity is constant [26] or the velocity is an Exponential Weighted Moving Average (EWMA) of the past velocities [23] or an auto regressive velocity model [12]. Even though these models improve navigation times, their performance has been lower compared with a specific prediction model oriented to the task [19]. The present work is focused on developing such a strategy when navigation is carried out by an expert pathologist.

Several studies have identified two complementary patterns for microscopy exploration when pathologists are examining a particular sample: zoom-in and scanning (panning) operations [16, 140, 28]. The scanning patterns span the largest interval of time in a pathologist navigation since a magnification change is more related to the level of visual detail rather than a search of information, i.e., pathologists identify structures very easily with little level of detail [103]. Indeed, navigation patterns arise from two inter-related processes: the motor control of the virtual microscope associated to some movement automations and a refined search information process, which reflects a level of expertise [28].

Yet patterns may be different, the more expert is the group of pathologist the more similar are the locations they visit when exploring a histological slide [28]. The present chapter formulates and validates a mathematical model which predicts the exploration velocity patterns of a group of four pathologists with similar expertise. The approach herein proposed takes into account the very basic influence of motor movements through a neuromuscular inspired model, coupled to a Bayesian learning strategy which attempts to capture the navigation pattern of a pathologist. The presented method approaches both processes through a hierarchical strategy, articulated in three phases: firstly a biologically inspired model [109] emulates non linear and linear velocity changes of actual microscopic navigations so that this model resulted suitable for tracking low and rapid velocity patterns. Second, a Bayesian learning strategy generates a navigation probability distribution function (pdf) from the actual pathologist navigation samples. Finally, the

navigation patterns are predicted by a Bayesian filter which weights the pdf values using real velocity samples. Two main contributions of this chapter are: identification of the isotonic velocity patterns in actual scanning movements of a computer mouse using a GUI designed for exploring virtual slides and a probabilistic model which accurately predicts such patterns.

The rest of this chapter is organized as follows: next section presents the velocity profiles of a group of four certified pathologists navigating ten megaimages, an analysis of these patterns and a brief discussion about how similar problems have been approached, section Materials and methods introduces our prediction method, section Results demonstrates the effectiveness of the predicting method. Finally the last section concludes with a discussion.

6.1.1 Modeling slow and fast human movements

Human movements have been studied under two different conditions: isotonic or isometric. A movement is known as isotonic if a muscular contraction results in a joint movement and the muscle force is maintained constant in despite of the change of the muscle length. Likewise, isometric conditions are present when the muscle is contracted in equilibrium with other forces, whereby the muscular attachments are always at the same distance. Most of the natural human movements are basically isotonic with some isometric components. The point is that these complex movements follow a precise sequence of patterns which are susceptible of modeling and have been used for prediction [107, 108, 109]. Experimentally, Gielen [50] demonstrated that velocity and acceleration patterns of aimed flexion arm movements, with equal duration but different amplitude, were basically equal - apart from a scaling factor, i.e., ratio between movement amplitudes -. Gottlieb [62] compared two subjects performing elbow flexions in a horizontal plane over different distances, from a stationary initial position to a visually defined stationary target. Joint angles and accelerations were measured from two agonist and antagonist muscles, finding that the initial rises of the acceleration (and therefore of the inertial torque) were all invariant over changes in the target distance. Likewise, several authors [11, 49, 96, 98, 127] have reported that the velocity profiles of rapid-aimed movements have a global "asymmetric bell-shape" which is invariant over a wide range of movement sizes and speeds. Plamondon [109] compared 23 different models that can be used to describe the asymmetric bell-shaped velocity profiles of rapid-aimed movements and found a clear superiority of the support-bounded lognormal

model to globally describe them. The very basic idea is that specific tasks involving repetitive isotonic movements can be completely described in the velocity domain, as a response to the synergistic action of a combination of an agonist and an antagonist neuromuscular network [107, 109]. Each network of this synergy is represented as a set of complex subsystems, which reacts to an input command with a lognormal impulse response.

The VMV developed within the present investigation uses a conventional computer mouse, whereby navigation consists of two processes: selection or drop of a WoI over the lower resolution image, followed by a window drag towards an interest point and a drop operation once the target is reached. In the particular case of a virtual navigation, an expert displaces a mouse whose resistance to the arm forces is zero or constant and is so defined as isotonic [154]. The important issue with a conventional mouse is that its relative movement is mapped to the cursor position, its control is intuitive and facilitates the use of ordinary motor skills. This conventional mouse allows small and large fine movements, yet a user must lift and reposition the device to move over large distances. In the present investigation, the GUI design forced the pathologist to move within the thumbnail image, for which small and large movements demanded fine control strategies. In the case of small movements, the pathologist is forced to maintain around a previously chosen RoI, while for large movements, the pathologist applies a foraging strategy which requires fine control in order to fit the Fitts's law [47]. We therefore claim that this movement can be considered as isotonic and can be modeled using the kinematic theory [109].

6.1.2 Navigation models

The problem of predicting navigation patterns has been already approached in other domains. Lui [86] proposed a method for predicting navigation patterns of user when browsing multiresolution images, the strategy is based on the hypothesis that a navigation is a sequence of multiple jumps among various spatial positions in the image. The image is divided into blocks at different resolutions, each with an associated fidelity information measure, defined as a measure of visual attention [69] weighted by the resolution. A navigation is so defined as a collection of jumps, each described by its beginning and end blocks, together with the navigation delay. This model supposes that navigations use a maximum benefit at a minimum cost policy, i.e., a foraging pattern. In similar problems, such as the prediction problem

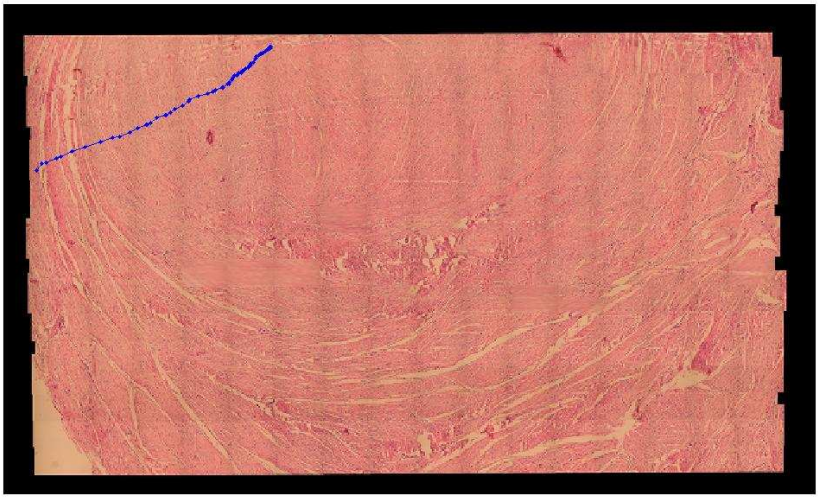
of the mouse position in distributed virtual environments [19], it has been proposed the use of a polynomial predictor, under the hypothesis that the mouse movement is a constant. This method is appropriate for slow velocity profiles but its delay is considerable in case of rapid movements [19]. Chan finds that navigation is composed of a series of velocity peaks and when studied in the phase space (velocity-acceleration), data may be approximated by a second order differential equation. Nonetheless, low velocity profiles can not be predicted by this model. For the problem of predicting navigation patterns of pathologists performing diagnosis tasks, as far as we know, there are no proposed models in the literature. In conclusion, the velocity profile of a user handing a conventional device as a computer mouse has been modeled using two different approaches, one for slow and another for rapid profiles. Unfortunately, integration of these two models into real applications is very difficult since it is nearly impossible to decide which should be used for complicated velocity profiles.

6.2 Materials and methods

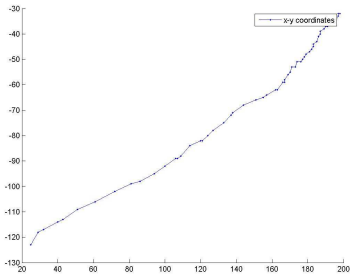
The pathologists navigations used in the present investigation were described in Sections 3.3 and 4.2.

6.2.1 The velocity profile

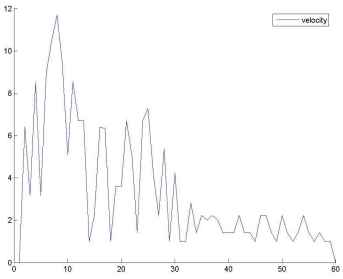
A portion of an actual navigation is also shown in Figure 6.1, where the upper panel shows the path defined by the pathologist. These mouse coordinates are better observed in the left-bottom panel, recall that the number of points depends on the velocity and the sampling frequency. Finally, the velocity modulus has been calculated from the plot shown in the left-bottom panel and displayed in the right-bottom panel. Note that the velocity agrees well with what has been described in the literature for isotonic movements, i.e., an asymmetric bell shaped curve which can be approximated by a log normal law.



(a)



(b)



(c)

Figure 6.1: Figure shows a drop-drag-drop operation superimposed upon the thumbnail image in the upper panel (panel *a*). The trajectory in the thumbnail image coordinates is plotted in the left bottom panel (panel *b*) while its velocity modulus is depicted in the right bottom panel (panel *c*)

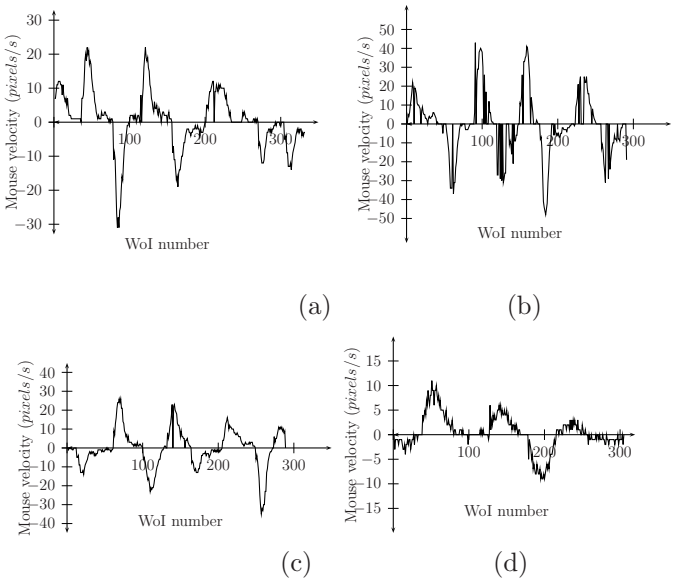


Figure 6.2: The four panels show uni-dimensional navigation segments of the four pathologists (x-axis). Explorations show high velocity profiles when experts are looking for information, In these four plots, the y-axis stands for the WoI velocity while the x-axis represents the sequence of requested WoIs, during the navigation. The GUI design allows to easily jump from one RoI to any location, dragging a desired WoI within the thumbnail windows, an effect observed in terms of velocity as the increasing part of the peak. When a new RoI is reached, a decreasing velocity profile is observed since this new RoI deserves a certain amount of time for examination.

Figure 6.2 shows typical velocity profiles corresponding to some navigation displacement of the four pathologists. Yet every pathologist presents different navigation patterns, the four panels evidence that there exists a common velocity profile, i.e., velocity rapidly increases up to a certain level and then it decays with lower slopes. This profile is likely a complex mix of associated factors such as the microscopical magnification, the neuromuscular mechanics and the type of restriction demanded by the developed GUI, i.e., a drop-drag-drop sequence (screen and mouse).

6.2.2 A machine learning approach for user modeling

The prediction problem is posed in terms of how to anticipate scanning patterns for a particular navigation. This problem has been here addressed as to forecasting velocities so that next positions can be predicted by a simple integration of these found velocities. For so doing, it is needed an analytical expression which approaches the observed scanning profiles in their peaks and valleys. This expression should be dependent on a set of parameters and the time so that it can be evaluated in the future. A convenient expression is a Δ-lognormal law, result of the competitive interaction between two log-Gaussians [109], each representing the fundamental parts of the neuromuscular system, i.e., the agonist and antagonist components.

$$v(t) = D_1\Delta(t; t_0, \mu_1, \sigma_1^2) - D_2\Delta(t; t_0, \mu_2, \sigma_2^2) \tag{6.1}$$

where D_1 and D_2 are Gaussian weights, which are the two input commands of the neuromuscular system, t_0 is a trigger time of the whole process and the the log-Gaussian is

$$\Delta(t; t_0, \mu_i, \sigma_i^2) = \frac{1}{\sigma_i\sqrt{2\pi}(t - t_0)} e^{-\left\{\frac{[ln(t-t_0)-\mu_i]^2}{2\sigma_i^2}\right\}} \tag{6.2}$$

here μ_i estimates the delay time of the two log-Gaussian and is associated with the overall reaction velocity of the neuromuscular system, while σ_i is related to how large and asymmetric the entire movement is. Plamondon demonstrated [109] that different velocity profiles under various experimental conditions are described by a Δ-lognormal law, whereby he was able to determine the adequate parameters when one individual was performing specific motor tasks. Kinematic theory is here applicable because the mouse movement is considered as isotonic with constant resistance [154], i.e., there exists a linear relationship between the mouse movement and the WoI displacements, using the developed GUI.

This formulation allows a great variety of scanning velocity profiles with different durations, slopes and heights. The associated parameters can be modeled as random variables of a unique pdf which can be evaluated in future times. The user model herein proposed is based on a machine learning strategy, an approach presently known as Machine Learning for user modeling ML4UM [56] and is divided in two phases:

1. An offline learning phase in which a general pdf is generated using a competitive Δ-lognormal law and a set of navigation examples

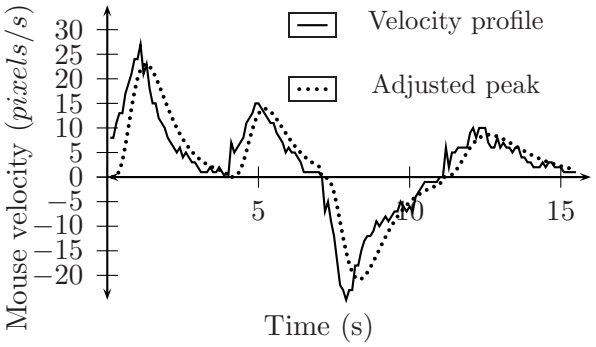


Figure 6.3: A segment of a typical navigation is shown (thick line) superimposed with the result obtained by adjusting a Δ-lognormal law (dotted line).

2. Tracking online phase in which the initial general pdf is modified estimating the best set of parameters for a current navigation of a determined specialist, using the collected data during this navigation. This function is thus adjusted for achieving prediction.

Offline learning

In this phase the algorithm estimates the optimal parameters that maximize the likelihood for a set of actual navigations using the available training set of mega-images. Parameters associated with each of the scanning profiles are determined by approximating the Δ-lognormal law (Equation 6.1) to the actual data using a simple non linear least square method [91]. The obtained set of parameters is then used as the input to an Expectation-Maximization algorithm, which solves the maximum likelihood problem (ML). Actual velocity profiles are shown in Figure 6.3, superimposed to the velocity peaks obtained using the Δ-lognormal law. The main objective of this part was to obtain a valid parametric expression for adjusting the shape of all these velocity profiles and use it for inferring a pathologist pattern from a training set of samples.

Figure 6.3 shows a typical velocity curve for a navigation performed by one of the pathologists. The thick line displays the navigation adjusted by a Δ-lognormal model. Among several advantages, this model allows in a natural way to include a proper response to rapid changes of acceleration while it is also capable to approach linear patterns. A Δ-lognormal was successfully used for modeling rapid responses to motor tasks [109], such

as it is the case when a pathologist uses a mouse to emulate microscopical routines. This parametric representation allows to introduce very powerful soft computing methods into this problem since these parameters are random variables with a specific pdf [109].

On the other hand, a Δ-lognormal constitutes a convenient description of navigation velocities but it is incapable in any case to capture the particularities corresponding to navigations of different experts. This makes necessary to find the set of Δ-lognormal parameters which better fits a particular pathologist navigation. Although the Δ-lognormal properly represents velocity profiles, parameter tuning is a difficult problem to deal with, provided the high parameter space dimensionality and their strong coupling [109]. The approach herein used is based on Bayesian learning, using navigation samples from the actual trajectories of real navigations. Each navigation peak is thus associated to an assembly of parameters $\gamma = (t_0, D_1, D_2, \mu_0, \mu_1, \sigma_1, \sigma_2)$, inferred from each velocity peak as follows: the agonist parameters are computed through a non-linear minimal square adjustment [91] of the velocity peak (Equation 6.1), $v(t) \approx D_1 \Delta(t; t_0, \mu_1, \sigma_1^2)$. The antagonist component is approximated by subtracting the observed peak velocity from the adjusted agonist component, i.e., $D_2 \approx D_1 \Delta(t; t_0, \mu_1, \sigma_1^2) - v(t)$ using non-linear minimal squares and the t_0 is calculated from the agonist component [63]. We assume that each single velocity peak is one instance of a true model, entailed with the parameter characteristics of an ideal navigation and which we wish to estimate. Given an observed set of parameters extracted from the pathologist navigation, the objective is to find the pdf for these data.

Let γ^j a random variable associated to the j_{th} peak-parameter (parameters defined in Equation 6.1) of the pathologist navigation, i.e:

$$(\gamma^1, \gamma^2, \ldots, \gamma^7) = (t_0, D_1, D_2, \mu_1, \mu_2, \sigma_1, \sigma_2)$$

Suppose that each parameter of the peak (Plamondon parameter) comes from a Mixture of Gaussians (MoG):

$$f(\gamma^j) = \sum_{m=1}^{M_j} \lambda_{m,j}\, G(\gamma^j \mid \mu_{m,j}, \sigma_{m,j}^2) \tag{6.3}$$

where $G(\gamma^j \mid \mu_{m,j}, \sigma_{m,j}^2)$ is a Gaussian distribution with mean $\mu_{m,j}$ and covariance $\sigma_{m,j}^2$ and $\lambda_{m,j}$ is the weight of the m^{th} Gaussian of the parameter j, M_j is the number of components of the MoG for the parameter j, $M_j = 2$

for $j = 2, 3$, i.e., for the D_1 and D_2 parameters. Recall that these parameters stand for the weights of the agonist and antagonist components of the movement that are herein modeled as a pair of log-Gaussians. Moreover, depending on the navigation direction a positive or negative sign is set to each of these D parameters. This adaptation of the Plamondon model allows to naturally deal with forward or backward movements and different velocity profiles.

The EM algorithm calculates a ML estimator of an incomplete set of data. For so doing, a new complete set is made up with the incomplete and hidden data so that calculation of the ML on this new set is trivial. In this case, the incomplete data are the set of parameters of the collected navigations $\Gamma^j = \{\gamma_i^j,\, i = 1, 2 \ldots, n\}$. These data are incomplete since it is impossible to determine from which Gaussian of the MoG observations come from, otherwise calculation would be trivial since estimation would be focused on this unique Gaussian. For this reason, observations Γ^j are completed using a variable m_i^j and $m_i^j \in \{1, 2 \ldots, M_j\}$ indicates which term of the mix of Gaussians generates the parameter γ_i^j.

Starting from the initial values of the ML estimator parameters, the EM algorithm iterates on two steps:

- The Expectation step (E-step [32]) calculates an estimator of the indicating variables m_i^j given the observations and parameters of the ML estimator. These estimators are used to complete the data set and the ML estimator parameters are calculated.

  $$P(m_i^j = m|\gamma_i^j, \theta_{(t)}^j) = \frac{P(m_i^j = m|\theta_{(t)}^j) G(\gamma_i^j|m_i^j = m, \theta_{(t)}^j)}{G(\gamma_i^j|\theta_{(t)}^j)}$$

  where

  $$P(m_i^j = m|\theta_{(t)}^j) = \frac{1}{n}\sum_{l=1}^{n} P(m_l^j = m|\gamma_l^j, \theta_{(t-1)}^j)$$

  here $\theta^j = \{\lambda_{m,j}, \mu_{m,j}, \sigma_{m,j}^2\}_{m=1}^{M_j}$, is the set of parameters of each MoG.

- In the Maximization step (M-step) the new parameters of the ML estimator are used for calculating a new value of the indicating variables (E-step). The new parameters of the ML estimator (M-step) are ob-

tained then from this new complete set

$$\lambda_{m,j} = \frac{1}{n}\sum_{l=1}^{n} P(m_l^j|\gamma_l^j, \theta_{(t-1)}^j)$$

$$\mu_{m,j} = \frac{\sum_{l=1}^{n} P(m_l^j|\gamma_l^j, \theta_{(t-1)}^j)\gamma_l^j}{\sum_{l=1}^{n} P(m_l^j|\gamma_l^j, \theta_{(t-1)}^j)}$$

$$\sigma_{m,j}^2 = \frac{\sum_{l=1}^{n} P(m_l^j|\gamma_l^j, \theta_{(t-1)}^j)(\gamma_l^j - \mu_{m,j})^2}{\sum_{l=1}^{n} P(m_l^j|\gamma_l^j, \theta_{(t-1)}^j)}$$

6.2.3 Tracking online learning via particle filtering

The offline phase provides a pdf to generate the entire set of pathologist's navigations, even though during navigation the belief must be adjusted to a particular pathologist's navigation. Adaptation of the found pdf to the current navigation is achieved through online learning by continuously adjusting the parameters to observations while predictions are performed. In this case the process is $x_t = (x_t^1, x_t^2, \ldots, x_t^7) = (t_{0,t}, D_{1,t}, D_{2,t}, \mu_{1,t}, \mu_{2,t}, \sigma_{1,t}, \sigma_{2,t})$, so that parameters of a velocity peak for a time t and observations correspond to the velocity profile for this time $z_t = (v_t)$.

The online adaptation problem consists in estimating the parameters of the velocity profile given a set of observed velocities, i.e., $p(x_t|z_1, z_2, \ldots, z_t)$. This estimation can be recursively found using a Bayesian filter [10] so that it is needed to define an estimation of the system states at the beginning of the whole process. The whole adaptation is governed by two equations: a dynamic equation which indicates how the system states are evolving and a equation of observations which relates the system states and observations. The recursive estimation requires two phases: a prediction step which uses the dynamic equation for estimating every possible state of the system in time t and an updating step which modifies those possible states for matching them to observations.

The parameters of the initial pdf come from the offline learning phase. In the dynamic Equation 6.4, let us assume that before the pathologist moves the mouse, he/she has a pre-programmed desired path. The previous hypothesis is valid only between separated WoIs and indeed this has been already proved in eye tracking studies [79, 137]. Small adjustments to the system are

given by the visual feedback and are herein modeled by independent Brownian motions of each of the Plamondon parameters. Therefore, the hidden parameters defined in Equation 6.1 should be modified so that the predicted velocity can vary and properly approach the sequence of observations. For doing so, in the prediction phase, the hidden navigation parameters are modified by a Gaussian perturbation. The noise produced by the variability of the sensed mouse positions is assumed to be additive, as well as independent of the positions of the mouse neighbors, whereby it can be modeled as a Gaussian. The initial pdf parameters come from the offline learning phase (Equation 6.3). The dynamic system equation indicates how the parameters change within a velocity peak, a phenomenon modeled as a Gaussian perturbation of the hidden parameters

$$p(x_t^j|x_{t-1}^j) = G(x_t^j \,|\, x_{t-1}^j, \hat{\sigma_j}^2) \tag{6.4}$$

The updating equation is able to generate velocities using parameters of the velocity peak, this is modeled as the sum of the parametric expression (Δ-lognormal) and a Gaussian noise:

$$p(z_t|x_t) = N(z_t - D_1\Delta(t; t_0, \mu_1, \sigma_1^2) + D_2\Delta(t; t_0, \mu_2, \sigma_2^2), \Sigma^2)$$

where, x_t are the parameters for a navigation time t and $\hat{\sigma_j}^2$, Σ^2 are predefined parameters whose values were calculated by minimizing the error when predicting the velocity from observed velocities. The used loss function was:

$$\begin{aligned} f(x_t, \Sigma, \hat{\sigma_1}, \hat{\sigma_2}, \ldots, \hat{\sigma_7}) \quad &= \sum_{i=1}^{K} \frac{(z_t^i - D_{1,t}\Delta(t;t_{0,t},\mu_{1,t},\sigma_{1,t}^2) + D_{2,t}\Delta(t;t_{0,t},\mu_{2,t},\sigma_{2,t}^2))^2}{2\Sigma^2} \\ &\quad + \sum_{j=1}^{7}\sum_{i=1}^{K} \frac{(x_{t-1}^{j,i} - x_t^i)^2}{2\hat{\sigma_j}^2} \end{aligned}$$

where, K is the number of training samples, z_t^i and $x_{t-1}^{j,i}$ are the velocity samples and the navigation parameters used for training. In the loss function, the first term accounts for the error introduced by the mouse, while the second stands for the user adjustements during navigation. Minimization of equation was achiveve using the Levenberg−Marquardt method [91], with $K = 10$ training samples corresponding to velocity peaks randomly selected in the time t, while $x_{t-1}^{j,i}$ parameters were non-linear least square estimated, as mentioned before.

Because of the non linear nature of the model of observations, there is no analytical solution for the proposed Bayesian filter so that the estimation of

the state is achieved using a sequential importance resampling (SIR) filter [10]. This technique approximates a pdf with a random set of samples, each with an associated importance weight. Using this discretization, the prediction step applies the dynamic equation to each sample while the importance weights are modified in the updating step for the sequence of samples match observations. Resampling aims to concentrate the samples around the areas with a high importance ratio so that some samples will disappear. However, provided this a dynamic and stochastic process, importance can change within the navigation and some of the disappeared samples can become important but they result completely unretrievable. Therefore, the sequential importance resampling uses an additional resampling process which avoids the weight of some particles early vanishes.

6.2.4 Evaluation

To study the performance and the accuracy of the proposed prediction method, we have implemented two classical methods for forecasting: an expected weighted moving average (EWMA) and an elliptical predictor based on the Kalman filter (elliptical model) [19]. The algorithms were written in *C++* and run under windows with an Intel Centrino processor of 1.7 GHz and 1 GB in RAM. Time performance was assessed by calculating the mean running time for each algorithm. We also studied the prediction error of the three methods as follows: firstly, the prediction error for each method was computed by comparing with the recorded navigations. Secondly, we study the degree of dependence between the results provided by the proposed prediction method and the pathologist, i.e., whether or not the predicted results are independent of the pathologist. Statistical significance was determined using the Barlett's test [126]. The error for each navigation step was measured in pixels as the root mean squared error (RMSE) between the recorded position and the predicted position provided by each strategy. The predicted displacement vector was calculated as the velocity times the mouse sampling time ($0.1\ s$).

Parameter values were tuned for each of the three predictors using four different navigations, corresponding to two pathologists navigating on two different mega-images. The EWMA's time window size and the vector of weights were set as the least error found in the test images using an extensive numerical search, namely the time window size varied between 1 to 10, using a step increment of 1 while weights were varied between 0 and 1, with step

increments of 0.01. The parameters of the elliptical model were obtained from a non-linear optimization process, as previously described in [19].

6.3 Results

A total of 40 navigations were used either for training or assessing. Pathologists were instructed to navigate until they could achieve a diagnosis about the organ or the particular pathology. Overall, these mega-images were parts of full histological slides, a relative size which varied between 10 % to a 30 % of the whole histological sample. This means that pathologists never had the entire information as to have comparable levels of coincidence on the diagnosis and many times they just gave up. It is worthy to strength out that the main objective of this study was to determine the coincidence on the velocity patterns rather than the sequence of events for diagnosis or even the diagnosis. Finally, the time used for navigating varied between 20 s and 2 min, depending on the image contents. In general, navigations were shorter for images stained with inmunofluorescent techniques, for which the islets of Langerhans constitute the main information on which diagnosis lies. On the contrary, inflammatory pathologies stained with Hematoxylin-Eosin took larger exploration intervals and very rarely they were able to achieve a correct diagnosis.

6.3.1 Computational times

The prediction times for the three models are shown in Figure 6.4, the proposed method presents the highest computational load, near to 6 ms using 1000 particles. This is the result of the number of parameters of the dynamical model, which is 7, compared to the elliptical and EWMA models: 2 and 1 parameters, respectively.

These computational times are very low and, in any case, much smaller that the retrieval time of a WoI which is about 500 ms [68]. In addition, the time performance for the three methods is lower than 100 ms, which is the mouse time sampling. These results indicate that this method may be perfectly utilized for the prediction task in real time applications.

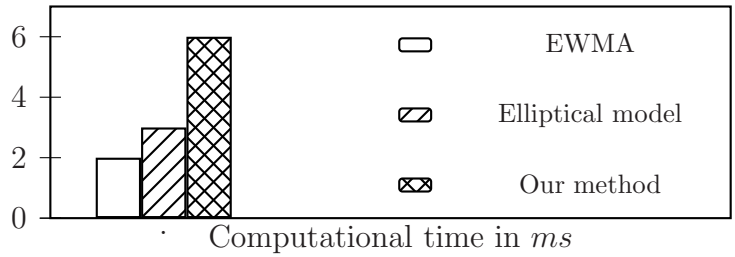


Figure 6.4: Comparison of the computational cost for the three predicting methods. Our model has a higher cost, but its time performance is in any case much lower than the sampling mouse period, i.e., the method can be used for virtual microscopy.

6.3.2 Prediction errors

Figure 6.5 shows a segment of an actual navigation (thick line), along with the predictions computed by two different strategies: an elliptical model [19] (dashed line) and the approach herein presented (dotted line).

The elliptical model shows a high variability which induces a series of peaks, quite far from the actual navigation. Very likely, this is due to the method itself since the switch from low to high velocities is controlled by a threshold which introduces a lot of noise in the prediction. In contrast, the presented method is steadier, more robust to the noise and shows a better approximation to the estimator in the sense of minimal squares. From a practical point of view, the mouse introduces a lot of noise so that having a robust-to-noise estimator improves precision.

Pathologist navigation error

The success of a preloading strategy, for the case of actual virtual microscopy applications, depends on the capacity of the navigation prediction for future times [35]. Therefore, the prediction in thumbnail coordinates was calculated up to 20 steps ahead with interval increments of 5. For this experiment, one step corresponds to the next time the mouse captures a new location. The whole set of navigations was split into each of the four pathologists. Only two navigations were used as training and the rest of individual navigations were used as a test set. Results for each pathologist were harvested independently

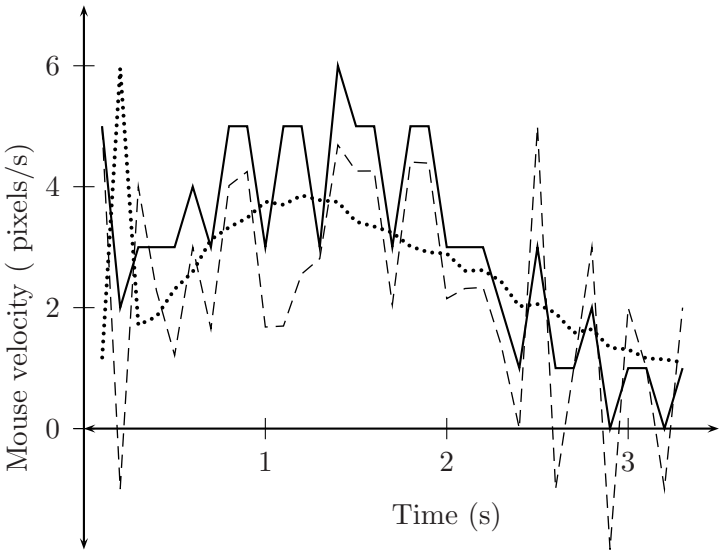


Figure 6.5: Thick line corresponds to a representative navigation segment. Observe the toothed saw-like pattern, mostly produced by the electrical noise from the mouse capture process. A good prediction strategy should approximate average values. Dashed line corresponds to the prediction obtained with the elliptical predictor while the dotted line depicts predictions obtained with our method.

and averaged for estimating means and variances.

Table 6.1 shows the RMSE of the prediction for the three methods for five different future steps. Overall, the error in pixel coordinates is very low for the five different intervals. However, the best approximation is achieved with our method, reaching high error reductions as the future step increases. As expected, the larger the prediction step the higher the prediction error, for the three methods. Interestingly, our method shows lower error rates, a pattern which results much more important as long as the step increases.

Prediction errors with different pathologists

The method was also assessed when different pathologists navigate one microscopical mega-image so that inter-pathologist prediction accuracy could be evaluated. For so doing, one image was randomly selected and predictions were estimated using the available navigations on this image. The RMSE

Future Steps	*EWMA*	*Elliptical Model*	*Our method*
1	1.3 ± 0.13	1.1 ± 0.1	0.9 ± 0.17
5	3.4 ± 0.14	2.1 ± 0.2	1.2 ± 0.19
10	7.9 ± 1.08	3.1 ± 0.3	1.8 ± 0.40
15	9.4 ± 1.4	4.4 ± 1.1	2.4 ± 0.80
20	11.4 ± 2.3	6.5 ± 1.5	3.2 ± 0.97

Table 6.1: RMSE for five different prediction steps, expressed as means and variances in pixels, for the three prediction methods, namely: EWMA, Elliptical model and the presented method.

variances were compared among the group of four pathologists for the next prediction step and results are shown in table 6.2.

pathologist	*Mean* ± SD	*Total Samples*
1	1.18 ± 0.09	180
2	1.36 ± 0.13	262
3	1.24 ± 0.14	283
4	1.46 ± 0.15	229

Table 6.2: RMSE means and variances in pixels for the next prediction step, calculated from the available navigations of the four pathologists

Table 6.2 shows the results as means ± the standard deviations. Prediction results were similar and differences were not significant under the Bartllet's test [126] (significance level set to $\alpha = 0.010$, Barlett's statistics to 51.7 and $\chi^2_{\alpha,k-1} = 50.8$), which suggests that prediction is independent of the particular set of training samples. This experiment was also carried out using the four next steps and results are alike.

In this section, we have demonstrated that the presented method is accurate and robust for predicting the mouse motion in virtual microscopy through various experiments. Our results show that the proposed method has lower prediction error compared with other popular prediction methods. Our results indicate that the prediction capacity is also independent of the pathologist. In contrast, a limitation of our method is their computational cost when comparing with other prediction methods, however this cost is sufficiently low as to use it in preloading strategies for virtual microscopy.

6.4 Discussion

The present investigation has demonstrated a common pattern in the velocity profiles obtained from actual pathologist navigations in virtual microscopy. In addition, a biologically inspired model served as prior knowledge of a machine learning strategy which could then be properly used for tracking and predicting these profiles. This velocity profile is the result of two different factors: firstly, the distribution of RoIs all over the virtual slide, and secondly, the movement restriction compelled by a particular device such as an ordinary computer mouse. Overall, the proposed model is based on a prior which takes into account main human movement restrictions when performing subtle tasks and learn the exploration patterns from a set of training navigations. The tracking and prediction strategies are both based on a nonlinear dynamic model which naturally adapts to any velocity profile, namely, to the slow and rapid movement at any desired direction.

Construction of a virtual slide can be thought of as three complementary processes: stitching, storage and efficient access [118]. The stitching aims to reconstruct a virtual slide (a digital high resolution image of the physical glass slide) as similar as possible to the real glass slide so that exploration may be carried out using conventional digital devices as a personal computer [6]. The storage process should improve main problems generated by the high resolution acquisition, namely large disk space and display limitation upon conventional visualization devices [6], using representation data which allow efficient access to the virtual slide at different magnifications or scales, at different RoIs and with any desired quality level [68]. Finally, the last objective, the efficient access, is the adaptive displaying of the virtual slide data or a seamless navigation which meets the pathologist requirements [68]. Recent investigations [35] have demonstrated that a seamless navigation can be reached using adaptive strategies such as cache and prefetching [68]. Processing times have improved up to a 30 % using the naviest cache strategies [68]. In any case, the success of these strategies is highly dependent on the degree of knowledge one has about how a user runs over the virtual slide, i.e., the prediction accuracy [35]. However, construction of any prediction method will depend on the identification of navigation profiles, if they exist. It is nearly impossible to determine such patterns for the general case, but oriented task application such as the virtual microscopy could take advantage that these patterns do exist.

CHAPTER 6. PREDICTING PATHOLOGIST'S VELOCITY PROFILES

A virtual microscopical exploration, performed by an expert pathologist, is composed of multiples jumps of spatial scanning or magnification changes so that information should be constantly reconstructed from minimal units. In general, every pathologist follows a standard methodology with two coarse phases: first examination is carried out at the lower magnification in order to locate relevant information in terms of a spatial organization of the histological sample while the second and further examination is conducted for analysis of the slide contents [103]. This analysis is performed through a spatial navigation of small zones of the whole slide [79, 137], upon which gentle movements are required. Several studies have identified two complementary patterns for microscopical exploration: zoom-in and scanning (panning) operations. Krupinsky has demonstrated [79], using an eye tracking system, that pathologists highly coincide on the regions they visit when navigating virtual slides. This fact could be used for constructing a probability map of the virtual slide by storing the history of visits and calculating the number of times a region is visited. This probability map would be then used for prediction after recording a significant number of navigations, i.e., a frequentists approach. However, this strategy results limited in clinical environments in which microscopical slides are explored by a reduced number of pathologists.

From the point of view of the image contents, it would be reasonable to characterize the image, to determine where the RoIs are located and use this information as a road map in the navigation process. A basic concept in understanding histology is that there are four basic types of tissue: epithelial tissue, connective tissue, muscle tissue, and nerve tissue [72]. With very few exceptions, all organs contain a different proportion of these four basic tissues. In general, histological techniques highlight these tissues with few colours since dyes are designed to specifically arise a particular tissue feature. Colour variability stems mainly from a large intensity range as well as dye deterioration. In terms of image processing, histological images are distinguished by having more or less homogeneous textures or repeated patterns, which may be used to characterize the image and to decide a particular strategy for compression. Nevertheless, very little information could be drawn from such representations because for the particular case of pathological images, these textures are highly complex stochastic systems, organized in a manner which is very badly understood. As far as we know, RoIs have been determined for specific types of images [71, 125] with some particular pathologies so that they could be hardly used for actual virtual microscopy applications.

The velocity profiles observed through this investigation follow what has been described in the literature [109] for gentle movements, that is to say, an asymmetric Gaussian [109] which is the result of the interaction between the agonist and antagonist mechanisms. Overall, once the expert leaves a particular RoI, the velocity modulus increases until the expert focuses on a new region. After the maximum is reached, an exponential decay is observed which follows the Fitts's law. This law formulates a model of human movement which predicts the time required to rapidly move to a target area, as a function of the distance to the target and the size of the target [47]. The maximal velocities and the time spanned for the whole phenomenon depend on the dimensions of the thumbnail image and the navigation device, i.e., the computer mouse, typically this peak magnitude varied between 0 and 60 $pixels/s$ while its duration varied 900 ms and 3 s. This work focuses on characterizing the trajectories between RoIs, using these velocity profiles. These profiles have been previously used in virtual navigation environments [19] and movement prediction in virtual reality [20]. This characterization is independent of the image contents, avoiding the variability problems and computational burden. Besides, thanks to the restrictions set by the exploration device (mouse in this case), it is possible to identify the velocity interval values. Finally, as the velocity is a two-dimensional variable, its tracking and prediction require low complexity calculation, such as it is shown in table 6.1. The work dedicated to these profiles lies on the hypothesis that exploration velocities are actually dynamical processes. In this sense, Chan initially pointed out that a linear model for velocity prediction was adequate. Later, he realized how high velocities followed Gaussian asymmetric patterns. Mixing up these two ideas, he formulated a strategy for prediction which has been successfully used in many scenarios [19, 20]. Nevertheless, adjustment of this model has not been carried out experimentally and besides requires heuristic parameters, such as the threshold level to which the model switches from low to high velocity profiles. The Plamondon approach [109], herein used, has many additional advantages and none of these defaults. Among the advantages, it is simple, based on physiological experimental knowledge and is adaptable to any velocity profile.

This chapter presents a novel method for predicting a pathologist's navigation path as well as its velocity profile in virtual microscopy, using a very simple soft computing strategy. Pathologist navigation profiles are found from the predicted velocities by a simple over-relaxation integration scheme. The method presents considerable reductions in the error rate, when com-

pared with previous approaches, and is also more robust and steadier for longer future prediction intervals. Future work includes integration of this method with a cache or prefetching strategy.

7

An Attentional Model for Finding Regions of Interest

Gómez F., Gutiérrez R., Romero E.. A Supervised Attentional Model for Finding Regions of Interest in Basal Cell Carcionoma Images. Submitted to Journal of Visual Communication and Image Representation.

Abstract *This chapter introduces a novel "bottomp-up" and "top-down" visual attetion model for finding diagnostic regions of interest in histopathological images. The method is based on the cognitive process of visual selective attention that arises during a pathologist's image examination. The selected bottom-up information includes local low level features such as intensity, color, orientation and texture information. Top-down information is related to the anatomical and pathological structures known by the expert. A coarse approximation to these structures is achieved by a still segmentation algorithm. The algorithm parameters are learned from an expert pathologist's segmentation. Top-down and bottom-up integration is achieved by calculating a unique index for each of the low level characteristics, inside the region. Relevancy is then estimated as a simple average of these low level indexes. Finally, a binary decision rule defines whether or not a region is interesting. The method was evaluated on a set of* 55 *images, using a perceptually-weighted evaluation criterion, finding a quality gain of 3dB when comparing to a classical attention bottom-up model.*

7.1 Introduction

A reliable determination of clinically meaningful RoIs in medical images is at the very base of strategies for adaptive delivering of image data and clever selective compression, which in turn have multiple applications in medical education, medical training and virtual microscopy [7]. The typical strategy so far has been to draw these regions by manual selection [28], but this method is time consuming and presents high inter-observer variability, in some studies of about 20% [40]. The automatic RoI extraction in histopathological images is a very challenging task because of their very complex color, shape and architectural variabilities [138]. This picture is even worst if one thinks that histological samples are randomly taken from a lesion and that the anatomical biopsy is cut at different orientations and locations [93]. The naive use of current low level-RoI-extraction methods for medical images [75] would probably fail since they disregard main histopathological issues such as particular tissue architecture and relations between different structures [157]. An expert nevertheless is capable to weight each of them and figure out a very precise diagnosis. Overall, a pathologist examination starts with a first search and identification of relevant areas at a low power, which will be later revisited at a higher power [28]. Attention processes are triggered in this phase to visualize, focus and identify these regions. Attention is herein thought of as the system capacity to select relevant information in function of a particular task. Some computational attention models rely on low-level image features to locate the relevant or conspicuous information within an image. One of these "bottom-up" models of attention, is the one proposed by Itti et al. [70]. Other theoretical and computational models of attention rely on "top-down" information, i.e., memory (semantic, episodic, declarative) and specific behavioral tasks [37].

The main contribution of this work is to model attention by bringing together bottom-up and top-down information. Specifically, this model emulates the pathologist's first examination step where she/he defines and separates high informative diagnostic regions [105]. Thus, the idea is to capture groupings, not necessarily neighbors, endowed with similar histopathological meaning. The method combines the advantages of a low level image characterization with a high discriminant power in terms of tissular properties and spatial grouping, information learned from the pathologists. This novel strategy was assessed in basal cell carcinoma images stained with Hematoxylin-

Eosin, but is extensible to other histopathological images since the methodological analysis is alike in many other medical entities. This carcinoma is a representative tumorous pathology constituted of abnormal epithelial and connective tissue arrangements, which are also found in many other pathologies [93]. Our results demonstrated more coherent RoI selections than those obtained with a classic strategy of visual attention.

This chapter is organized as follows: the problem and some previous works are introduced in this section, Section Materials and Methods is devoted to describe the proposed method for finding relevant information regions, evaluation and experimental results are presented in Section Results and some conclusions and perspectives are discussed in the last Section.

7.1.1 Related Work

The problem of selecting RoIs has been approached in several medical image modalities. For instance Karras et al. [75], using gray scale pictures from abdominal cancer, assumed that regions with high density of repetitive patterns were more relevant than others. A robust description was obtained by using a vector of texture characteristics like energy, correlation, inverse difference moments and entropy. These features were the input to a fuzzy c-means clustering algorithm that classified regions as important or non-important. Gokturk et al. [121] claimed that relevant information in CT colon images was mainly due to the boundaries, when they are separated by air from other tissues and are recognized as variations on the gray scale levels. This kind of strategies could not be straightforwardly applied to histopathology images because these techniques ignore information such as color, intensity or spatial correlation [152, 157], crucial in these images since they are basically characterized by a repetitive complex mix of these patterns. A classical approach, in natural images, has consisted in finding RoIs with high spatial edge density [146]. Again, this concept could hardly be applied to histopathological images because they contain regions with high edge concentration without clinical meaning [82] so that this approach would surely fail.

In the histopathological domain, a similar problem has been previously approached in automatic cancer diagnosis, case for which the aim was to automatically decide on the existence of cancer by examining the tissue properties [31]. These properties were characterized at two levels: cellular, focusing on cell abnormalities, [45, 38] and tissular, by description of changes in cell distributions [74]. The analysis in both cases was performed by low level

image characterization and a statistical analysis to discriminate normal from cancerous tissues. A large variety of low level image features have been used in histopathology: morphological, textural, fractal, topological and intensity based features [31]. These features are always computed at the pixel level, regardless the fundamental fact that histopahtological images are constituted by objects [105]. A recent work in this direction was proposed by Tosun et. al. [138]. In colon biopsy images, they approached the histopathological objects by circular primitives, upon which they computed an homogeneity measure. A growing and merging algorithm was used to segment cancerous tissues by minimizing these measures. Unfortunately, these algorithms highly depend on many non-intuitive parameters [74, 138], which must be manually tuned.

A pathologic diagnosis is the results of a complex series of activities mastered by the pathologist. Classical psychophysical theories suggest that complex visual tasks, such as histopathology examination, involve high degrees of visual attention [105]. There exists evidence showing that focal attention, displayed serially to different locations, integrates the constituting low level features of an object [139]. These findings have inspired several computational algorithms that somehow search to structure the low level features [69]. One of the most influential is the one proposed by Itti et al. [70], a pure bottom-up attention model that locates relevant foci, based on a conjoint map of three low level characteristics, i.e., color, intensity and orientation. Although this method has been successfully tested in natural images, primary results on histopathological ones were not (as it would be described later). The relevant semantic information of these images is mainly constituted by repetitive patterns, which cannot be linearly reconstructed from the three basic features used in Itti's model. These computational models have been used to characterize RoIs in natural images [42], however, their use in medical images is very limited.

7.2 Materials and Methods

7.2.1 Images and Ground Truth

A total of 55 histological microscopic fields of view of basal cell carcinoma, sampled from 25 randomly chosen patients, were chosen in this evaluation. Each biopsy was stained with the Hematoxylin-Eosin technique. Micros-

copical fields were digitized with a system Nikon eclipse E600 through a coupled to a Nikon DXM1200 camera at different magnifications (objective set to 20×) and stored in JPEG format at a 1280 × 1024 resolution. An experienced pathologist, with at least five years of experience, selected the digitized fields of view and manually segmented relevant regions. A manual segmentation, made by an expert pathologist, is shown in figure 7.1.

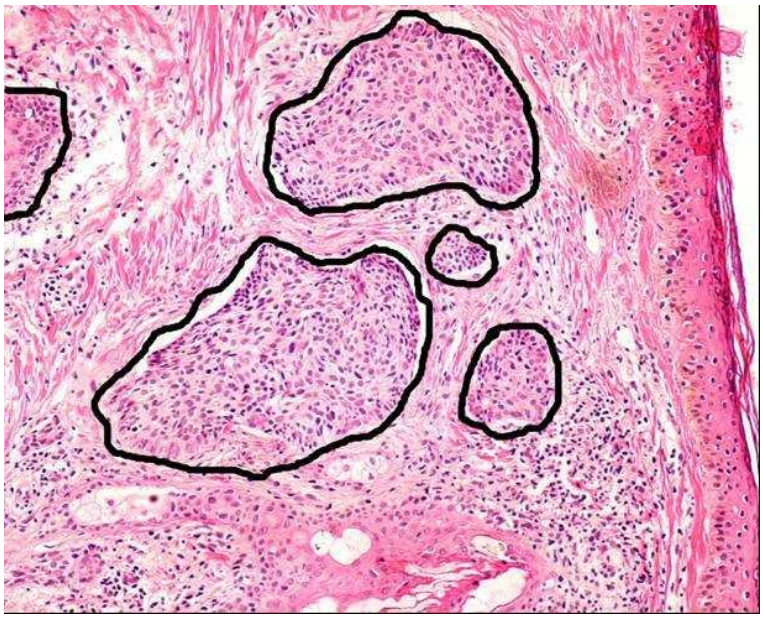


Figure 7.1: Illustration of a ground truth, drawn by the expert pathologist.

Note how difficult is to define a border, the tissue inside an islet is more cluttered than the outside, and the carcinoma is highlighted by Haematoxylin-Eosin. It should be strengthen out that in this kind of images, the color is very alike so that this characteristic has a low discriminative value.

7.2.2 Method Overview

A routine pathologist's visual examination is carried out in two sequential phases, when exploring a microscopic slide. An initial search for coarse tissue structures at a "low zoom" [28, 105] to separate the image into large regions, and a subsequent finer feature characterization at a "higher zoom" within

these regions to identify cellular structures [103]. The approach proposed herein attempts to emulate the pathologist's initial examination step where she/he defines the different regions of the image according to the inherent properties of each tissue type, such as level of visual attention, texture and homogeneity. Our approach tries to identify which of these regions are of diagnostic interest in a similar way as a pathologist decides where to look for finer details. The idea is to capture groupings, not necessarily neighbors, endowed with histopathological diagnostic meaning. These groups are determined by similarity relationships between the objects inside them. The groups compete at supposing that relevant regions will contain more relevant visual attention level.

The proposed strategy (figure 7.2) is composed of two parallel processes: one coarse and adjustable still-segmentation procedure [1] whose parameters are automatically extracted from a set of manually segmented images, and a modified version of Itti's attention model that runs over each. The goal of the still-segmentation process is to split the histopathological image into its constitutive objects, which are later characterized by their level of attention.

7.2.3 Splitting Histopathology Images in Regions

Visual attention is the ability of a biological or artificial system to find relevant region in an scene [69]. In the particular case of humans, they can not only find relevant regions, but also recognize complex structures in a scene. The Gestalt laws for proximity and resembling, illustrated in figure 7.3 have motivated the fundamental hypothesis of our model, i.e., a histological tissue is a grouping of objects which resemble in their very basic structural properties.

Visual systems reach grouping by clustering, proximity and resembling. Any strategy should at least look for any of these basic properties, just like carcinoma stained images have cluttered regions composed of simple structures with similar average intensity. The grouping characteristics defined before are herein used to segment, based on a comparison of the intraclass and interclass variances regarding the intensity value of each pixel. This strategy was implemented using the Felzenszwalb algorithm [44], for which firstly pixels are sorted out by similar intensity value and then neighborhoods are organized by grouping pixels with intensity values, which are previously de-

[1] A still-segmentation corresponds to an image partition in non-overlapping regions

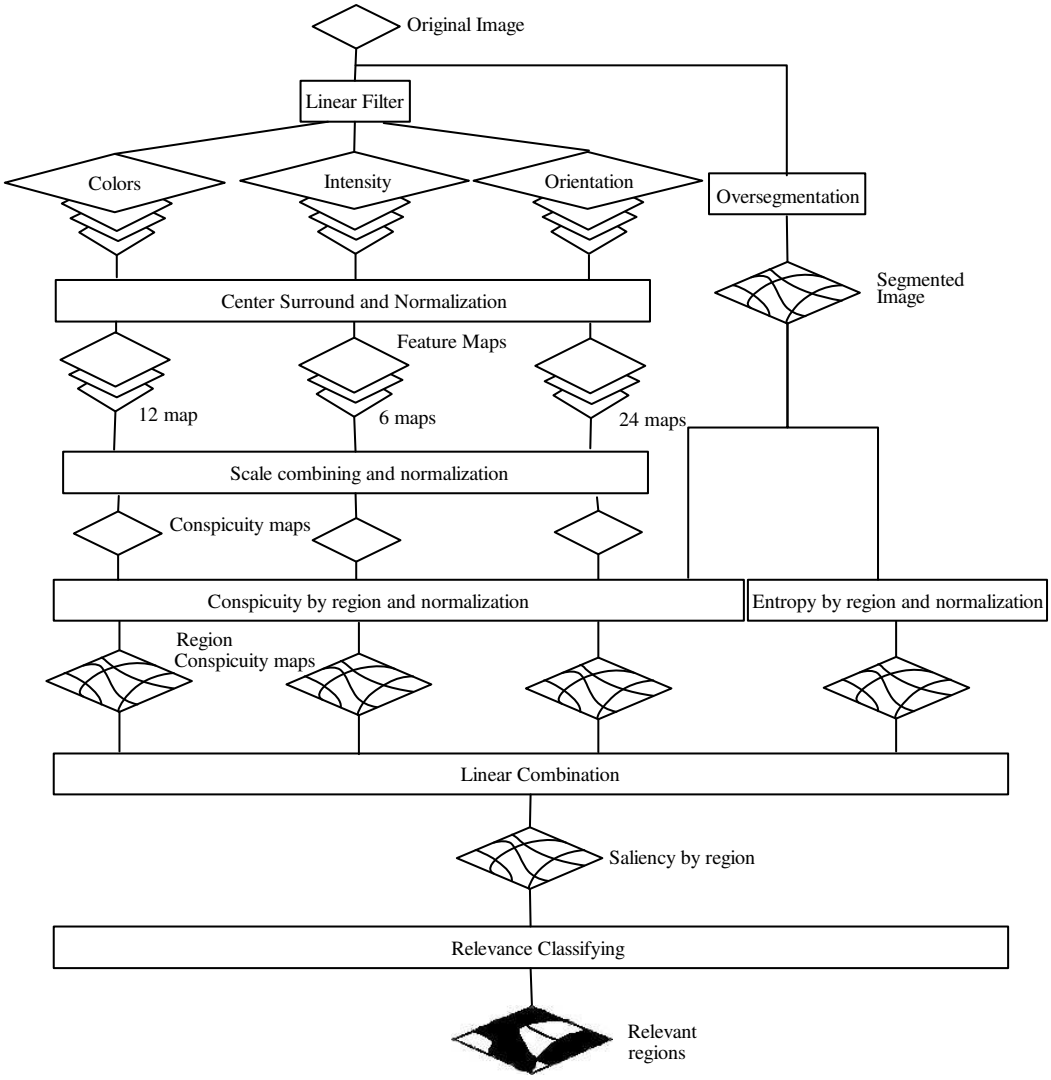


Figure 7.2: The proposed method: two independent processes run in parallel aiming at finding relevant regions. The first process utilizes low level information for determining a level of interest for each pixel, while the second process segments basic structures present in the image. Finally, the level of interest at each of the segmented regions is calculated from the average level of interest of the pixels inside the region.

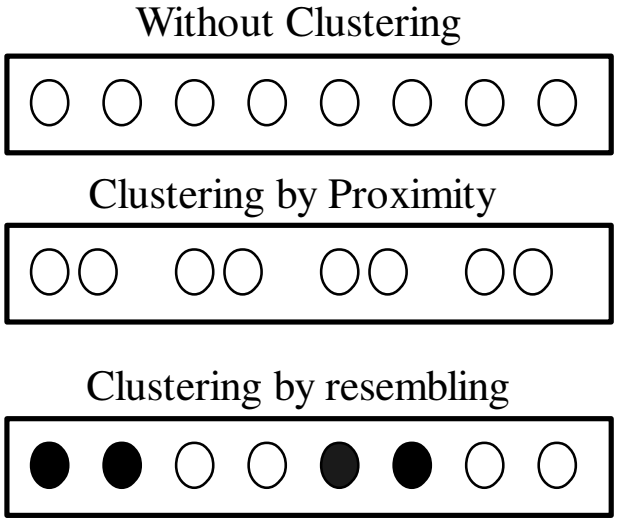


Figure 7.3: Gestalt laws for proximity and resembling.

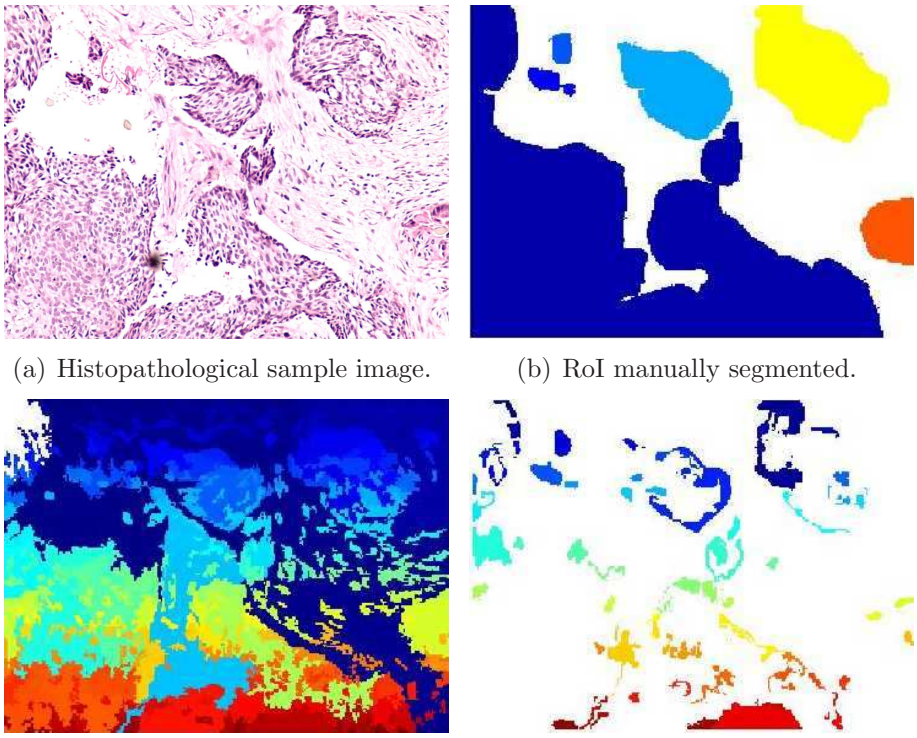
fined under a variance threshold. This method provides an still-segmentation strategy which is inspired from psychological grouping theories [149]. The main idea is that two regions are perceived as different if differences between them are larger than differences within them, according to a learned rule. The problem is defined in terms of a graph, where a non-linear decision function specifies if two elements c_1, c_2 in a graph partition should merge or not. The decision function reads as:

$$M(c_1, c_2) = \begin{cases} 1, & \text{if } \mathrm{Diff}_{bR}(c_1, c_2) < \min\{\mathrm{Diff}_{wR}(c_1) + \tau(c_1), \mathrm{Diff}_{wR}(c_2) + \tau(c_2)\} \\ 0, & \text{otherwise} \end{cases}$$

The two regions c_1 and c_2 are merged together when $M(c_1, c_2)$ is one, $\tau(c) = \frac{k}{|c|}$ depends on the size of c ($|c|$) and establishes an evidence for a boundary between two components, k is a scaling factor and sets preferences for specific component sizes, $\mathrm{Diff}_{wR}(c)$ corresponds to a within-region difference which stands for the largest difference inside the component, while $\mathrm{Diff}_{bR}(c_1, c_2)$ is a between-region difference that looks for evidence of a boundary between both components [44].

7.2.4 Automatic Still-Segmentation of Histopathology Images

The previous algorithm can be used to split the histopathology image into its constitutive tissue parts. As observed in figure 7.4, the quality partition is highly dependent on the segmentation parameters.



(a) Histopathological sample image.

(b) RoI manually segmented.

(c) Over-segmentation.

(d) Under-segmentation.

Figure 7.4: Results of a wrong selection of the segmentation parameters.

A manual selection of these parameters is always possible, but this is by no means an intuitive task for the expert. Therefore, we proposed an energy-based learning method for selecting an optimal set of segmentation parameters, based on manually segmented images. Let I a histopathological image, W a parameter vector and C the still-segmentation, resulting of running a segmentation algorithm over the image I with parameters W. Provided that it is possible to define an energy function $E(W, G, I)$ that quantifies the similarity between the still-segmentation C and a ground truth partition G,

then a set of training samples $S = \{(I^i, G^i),\ i = 1, 2, \ldots p\}$ corresponding to manually segmented images, will be used for finding the W^* optimal vector which solves the following optimization problem:

$$W^* = \min_{W \in \Omega} \frac{1}{p} \sum_{i=1}^{p} E(W, G^i, I^i) \tag{7.1}$$

with Ω the set containing any possible parameter vector. For this problem to be solved it is necessary to define the structure of the energy function $E(W, G, I)$.

Energy Function

The proposed energy function must quantify the similarity between two image partitions: the generated by the pathologist's selection and the produced by the segmentation method. This measure should cope with two different conditions: the perceptual relevance of the region center should be large and the measure should penalize miss-segmentations, i.e, classification is mainly addressed to regions rather than to pixels. We used the Mezaris metrics [94], an extension of another perceptual measure [143] which weights the visual relevance of any foreground-background segmentation:

$$q(g, c) = q_{MF}(g, c) + q_{AB}(g, c) \tag{7.2}$$

where $q_{MF}(g, c)$ amounts to the missing foreground pixels (MF) and $q_{AB}(g, c)$ to the false background pixels (AB), weighted by their distances to the closest region borders, as follows:

$$q_{MF}(g, c) = \sum_{i=1}^{D_{MF\max}} w_{MF}(i) \cdot |c_i \cap g^{\mathsf{c}}| \tag{7.3}$$

$$q_{AB}(g, c) = \sum_{i=1}^{D_{AB\max}} w_{AB}(i) \cdot |c_i^{\mathsf{c}} \cap g| \tag{7.4}$$

here c is the segmented RoI, g the ground truth, $\{\cdot\}^{\mathsf{c}}$ denotes complement, $c_i = \{\mathbf{x} | \mathbf{x} \in c, d(\mathbf{x}, c^{\mathsf{c}}) = i\}$ corresponds to the set of pixels inside the RoI at the same distance from its border, $d(\mathbf{x}, c)$ is the distance between the pixel $\mathbf{x}$ and the region c (in this case the Euclidean distance), $c_i^{\mathsf{c}} = \{\mathbf{x} | \mathbf{x} \in c^{\mathsf{c}}, d(\mathbf{x}, c) = i\}$ stands for the set of pixels outside the RoI at the same distance

from its closest border, $w_{MF}(i)$ and $w_{AB}(i)$ are the weighting functions for the missing foreground pixels and the false background pixels, both growing linearly, while $D_{MF\max}$ and $D_{AB\max}$ are the maximum permitted distance for the missing foreground pixels and the false background pixels, respectively.

Measure 7.2 was originally proposed for evaluating the segmentation quality in background-foreground segmentations, an approach which will fail in segmentations with multiple components. A multicomponent measurement was proposed by Mezaris et al. [94], exploring three error sources: inaccuracy of the region boundary location, under-segmentation and over-segmentation effects. For so doing, let $C = \{c_1, c_2, \ldots, c_K\}$ a still-segmentation composed of regions c_k and $G = \{g_1, g_2, \ldots, g_Q\}$ the ground truth partition. The inaccuracy is quantified by comparing the ground truth and the segmented images as corresponding region pairs. This correspondence is obtained by associating each ground truth region g_q to the still-segmentation region c_k with which the overlapped area is maximum. Once this association is established, the relationship is unique and unalterable. The inaccuracy e_{bl} for any pair of regions is computed as follows:

$$e_{bl}(A) = \sum_{(c_k, g_q) \in A} q(c_k, g_q)$$

where A is a set that contains the pairs of corresponding regions. Given that c_k and g_q constitute a unique couple and none of them can have a new link to another region, it is possible to obtain non coupled regions in C and G. A particular segmentation may result in a partition in which some regions have not a corresponding pair in two different situations: over and under segmentations, see Figure 7.4. When the actual region in the ground truth image corresponds to many regions in the segmented image, we are over-segmenting, case in which the measure penalize it by adding a term that takes into account the area defined by all these regions, as follows:

$$e_{ov}(A_c) = \sum_{c_i \in A_c} q_{MF}(B, c_i)$$

where e_{ov} is the over-segmentation error, A_c is the set of the c_i non coupled regions and B is a black image. Similarly, when there is a non coupled region in G and multiple regions in the still-segmentation image C, the under-segmentation error is calculated as follows:

$$e_{un}(A_g) = \sum_{g_i \in A_g} q_{MF}(B, g_i)$$

where e_{un} is the under-segmentation error and A_g is the set of the g_i non coupled regions. These error sources can be combined in a single quality segmentation measure, that can be used as energy function for the learning optimization problem 7.1:

$$E(W, G, I) = e_{bl}(A) + e_{ov}(A_c) + e_{un}(A_g) \tag{7.5}$$

The optimal segmentation parameters were found by a pattern search method [83], since that the energy function is not derivable.

7.2.5 Assigning Interest to Regions

A modified version of the Itti's algorithm was applied to the images. This model initially calculates multiscale "conspicuity" maps for three low level characteristics, i.e., intensity, color and orientation. These conspicuity maps are normalized and summed into the augmented conspicuity or "saliency" map, whose maximum defines the most relevant location. This normalization preserves information which is localized while other types of noise are ruled out. As mentioned, these low level features are not enough to conform the attention foci in the histopathological domain. Normal tissues appear as homogeneous architectures. Tumors and other pathologies introduce heterogeneous areas within this architecture, due to the presence of infiltrating, inflammatory and tumor cells, and the loss of marked boundaries [93]. Then, determining a measure of heterogeneity would be useful for locating the abnormal structures in the images. Heterogeneity might be understood as texture disorder that can be measured by entropy. Our approach adds the calculation of an additional conspicuity map for the intensity entropy.

Accordingly, the augmented saliency map is calculated by including intensity, color, orientation and entropy. The computed conspicuity maps, for the low level features and entropy, are considered as the bottom-up information. The segmentation, provided by the aforementioned algorithm, is considered as the top-down information. Top-down and bottom-up information are combined by firstly calculating an index for each low level feature from each region. This index is a pixel value average, inside the region conspicuity maps for intensity, color, orientation and entropy. Finally, the total region saliency is estimated by averaging its conspicuity map index.

7.3 Results

7.3.1 Evaluation Issues

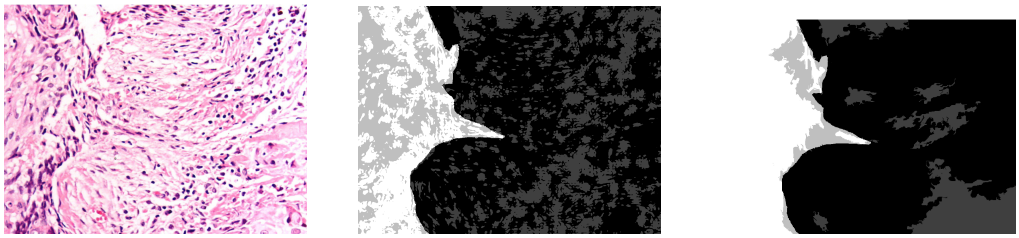
In this chapter two main issues were assessed, namely the accuracy of the proposed RoI extraction method and its generalization ability (Section 7.2.4), using a total of 55 manually segmented images (Section 7.2.1). Comparisons were performed among the three segmentations: manual, Itti's and ours. Itti's RoIs were computed by thresholding the resultant visual attention maps [155]. Likewise, the robustness of the automatic segmentation algorithm was evaluated by an 11-folding strategy, understanding this robustness as the method performance when the algorithm runs over a different set of data. A quality measurement was computed for each folding, namely the background-foreground quality measure $q(g,c) = q_{MF}(g,c) + q_{AB}(g,c)$, computed with the ground truth region and the extracted RoI, which was then normalized using $q(c,g)^{dB} = 10\log\frac{1}{1+q(c,g)}$. This measurement is a normalized accuracy measurement expressed in decibels (dB). The more negative these values are, the worse the RoI selection accuracy results.

7.3.2 RoI extraction

Figure 7.5 shows a visual illustration of the differences between the ground truth segmentation and the RoI obtained using the proposed method. Coincidences between RoIs are shown in white, method misses in gray and background coincidences in black.

As observed, the proposed method is able to capture different structures of interest, in spite of the complicated patterns present in the sample. The RoI computed by our method looks perceptually more similar to the ground truth, when compared to the RoI calculated using the Itti's model. While the Itti's RoI looks quite scattered, our method finds a more homogeneous region, clearly much more similar to the ground truth. Most misses are located near to the border, where we are supposing visual information is less important.

Figure 7.6 shows the original image in the first column, the ground truth in the second, Itti's RoI in the third and the RoI found by our method in the fourth. The four rows show different structures, as observed in the first column. Overall, these four original images show different configurations, with the carcinoma tissue in a darker violet color, which correspond to the zones highlighted in white since the expert considered them as the interest.



(a) Original image

(b) Coincidence level between the ground truth and and Itti's model

(c) Coincidence level between the ground truth and our approach

Figure 7.5: In panels (b) and (c) white and black stand for a perfect match while gray levels represent disagreement. Note a much smaller scattering level in (b).

Note that the level of structural organization is quite different so that it results impossible to determine RoIs by simply setting a set of parameters, i.e., structures present different sizes, shapes, colors and levels of hierarchy. As illustrated in the third column of figure 7.6, Itti's model misses important histological objects and instead highlights many small scattered regions. This can be attributed to the fact that this model performs a pixel-based analysis and therefore it finds interesting points rather than complete defined regions. From a semantic point of view, this is a great limitation because regions with some interest are distributed all over the image, following a complex mix of rules which are in general very variable. A clear advantage of the proposed strategy is that nearly every spatially coherent structure was found with different levels of noise. Interestingly, most relevant objects, within these RoIs, highly coincide with what the pathologist determined as important.

7.3.3 Automatic Segmentation

In this section we evaluated the robustness of the proposed algorithm, that is to say, how well this strategy performs when samples change. For doing so, the set of available images was split into 11 subsets of 5 images and a folding cross validation was applied, i.e., training with 10 subsets and test with the remaining one. Figure 7.7 shows the performance algorithm for the whole set of available image since each image has belonged at least once to a test subset. The graph plots at the x axis the available images and at the y

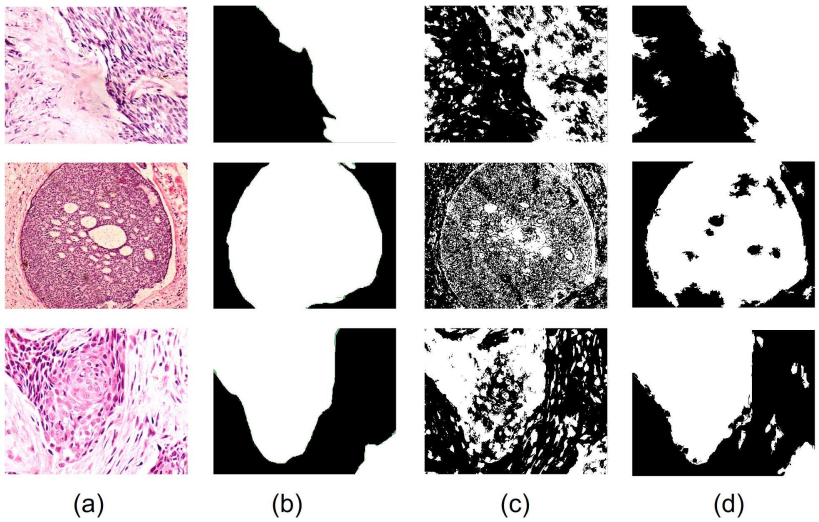


Figure 7.6: Column (a) shows examples of basal cell carcinoma images. Column (b) shows the corresponding RoIs selected by the pathologist. Column (c) shows results obtained by the Itti's model. Finally, column (d) shows results obtained by the proposed method. In the binarized images (b, c, d) the white color stands for the relevant diagnostic RoIs. The proposed method provides less scattered regions than those selected by the Itti's model. Note that our RoIs (column (c)) are more visually consistent with the expert selection (column (b)), than those computed by the Itti's model (column (d)).

axis the respective quality measurement for both strategies, namely Itti's and ours. As expected, the RoI quality measurements are different for each image and their values vary between -40dB and -63dB. Performance in general is quite acceptable around -50dB and only in two images, this performance decreased to -60dB. It is worthy to recall here that the more negative is this measurement the larger the number of both missing foreground and false background. The graph shows a systematic gain of our method performance: in most images, the proposed method provides better quality results and only in images 14,15 and 16 the Itti's model outperforms ours. These images correspond to samples in which there the objects are very small, they are not grouped and very scattered. Obviously, our strategy was devised for the opposite scenario.

The previous analysis was extended to find an estimation of the obtained

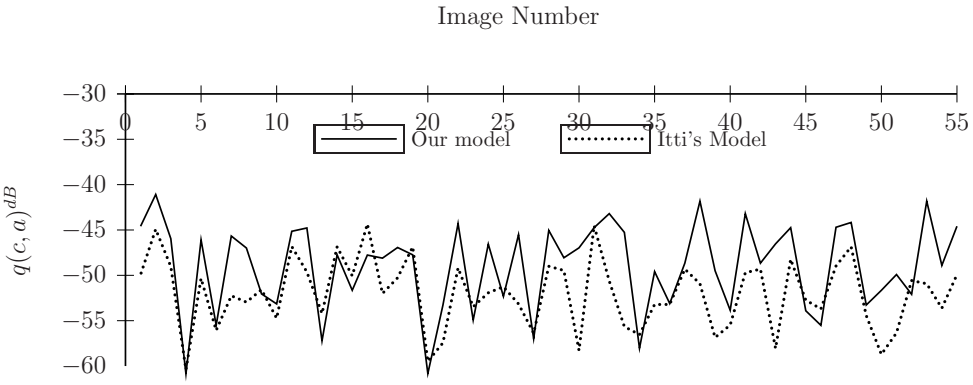


Figure 7.7: The x-axis represents the set of available images and the y-axis the quality measure in decibels (dB).

perceptual RoI quality. For doing so, a similar strategy to the 10-folding evaluation, commonly used in machine learning, was applied as follows: the complete data set was firstly partitioned into 11 subsets upon which one of the subsets was set aside for testing and the other ten were used for training. A total of 11 trainings were then carried out[2] and, for each training, averages of two perceptual quality measures were computed: one with the training images and the other with the test data. These calculations resulted therefore in an estimation of the segmentation quality in a particular subgroup and so, 11 quality measures correspond to the training sets, while the remaining 11 to the test sets. Values reported in table 7.1 are the averages and the standard deviations of these 11 quality measures for the missing foreground pixels, the false background pixels and the normalized estimation with the two last measures (total quality).

As observed in Table 7.1, RoI perceptual quality estimations (total quality) vary between -44dB and -50dB while its standard deviation is low, 1.3dB at most, indicating that the our strategy is stable in the sense that the strategy rarely selects regions with poor quality indexes. Overall, the proposed method outperforms the Itti's one when applied to the test set, in about 3dB, a figure which can be considered as a quite large difference. A

[2]each training implied a switch of the test set amongst the eleven possibilities

	Itti's Method		***Proposed Method***	
	Test (dB)	*Train (dB)*	*Test (dB)*	*Train (dB)*
False foreground	-48.2 ± 1.1	-44.2 ± 1.3	-46.6 ± 1.3	-42.8 ± 1.3
Added background	-47.3 ± 3.9	-43.3 ± 1.2	-41.1 ± 2.9	-37.7 ± 1.2
Total	-52.0 ± 1.2	-47.7 ± 1.4	-49.0 ± 1.2	-44.9 ± 1.3

Table 7.1: Average $\pm$ standard deviation of the perceptual quality in false foreground, added background and total quality for proposed and Itti's visual attention models obtained on the training and test sets.

Wilcoxon test ($p > 0.05$), applied on the two sets of eleven measurements from the Itti's method and ours, confirms that the perceptual quality improvements are statistically significant. As illustrated in figure 7.1, this level of difference is visually important, an issue considered in most image processing evaluations when comparing images with general measurements such as the SNR [143]. As expected, the perceptual quality performance for the training set is higher than for the test set. The quality measure, for this last set, is still high, about $49dB$, indicating that the strategy finds very similar regions to what the expert set. Importantly, the perceptual quality performance of our algorithm, for missed foreground and false background pixels, outperforms the Itti's method, very likely because interest is located in regions rather than in pixels.

7.4 Discussion

The present chapter has introduced a novel strategy, a complex mix of "bottomp-up" and "top-down" mechanisms, for setting RoIs in histopathological images. The model is inspired in the first phase of a pathological examination [16, 140, 28, 105], a process largely studied which starts by a scanning the slide at a low magnification zoom. This overview or first phase allows the expert a spatial pre-classification of areas of interest, upon which a further analysis shall be carried out. This first step is herein emulated as a two-fold strategy: a still-segmentation algorithm, which splits the image into several regions with potential interest, followed by a characterization step of the relevancy inside each region, using a linear combination of low level features (intensity, color opposition, orientation and entropy), weighed by parameters learned from the pathologist with a machine learning strategy.

So far the underlying mechanism that controls a RoI selection in histopathological samples has been poorly studied [13]. Recent studies suggest that some visual mechanisms, such as the one that allows to highlight an object from the background (figure-ground segmentation), and the visual attentional process, are connected [13]. The figure-ground segmentation models the process that occurs when an individual is exposed to a two-dimensional surface with some gentle structural differences, an she/he separates it into parts, one of which is consciously recognized as having a distinctive form whereas the region around is not [149]. This visual segmentation mechanism follows certain invariable rules that have been shown to be relevant in diagnosis of tissue sections [13]. These rules include convexity of contours, proximity of lines around it, closed contours, simple shapes, proximity and resembling of their components. The visual attention mechanism is related to the cognitive process of selectively concentrating on one aspect of the scene while ignoring others [69]. This fact suggests that the visual system is able to selectively focus on specific areas of the image, which besides are entailed with a high relevant meaning. Yet these ideas are far from being fully exploited, our approach has been able to capture the basic facts, that is to say, that relevancy is a global property somehow constructed by integrating local features.

Many endeavours have been dedicated to emulate these relationships for segmenting areas with cancer in histopathological images. the coarse structural recognition has been already implemented as an still-segmentation algorithm, using KKN and spectral clustering [138, 99], but these strategies only cope with local spatial relationships, and no perceptual meaning has to be assigned to each. One of the most challenging issues in histopathological images regarded the fact that semantic interest is related to similarity, no matter whether these regions are neighbors or not. This drawback was herein dealt with a graph-based image segmentation algorithm [44], which in contrast to previous approaches, is capable of capturing perceptually important regions such as tissue distribution. As illustrated in figure 7.6, regions obtained with the proposed strategy are perceptually more consistent and coherent with what the expert set. They are surrounded by closed contours and follow the proximity and resembling relationships, i.e., these regions satisfy the figure-ground segmentation rules. Interestingly, the ground truth also follows these figure-ground segmentation rules, as illustrated in figure 7.1, a finding that supports the choice of still-segmentation methods in this type of problems. The results herein presented support this selection, the perceptual quality

obtained with the proposed algorithm is around -49dB, an acceptable value for natural images. Likewise, table 7.1 shows the importance of the figure-ground segmentation mechanisms since a RoI selection strategy, focused on areas, shows a better performance, in near 3db of the quality measure herein used, when compared to a pixel based method, as reported in previous publications [155]. It is worthy to point out here, that the used figure-ground segmentation mechanishm may fail in some pathological entities, because it is strongly based on simple perceptual grouping rules that can not capture some ill-defined configurations, for example, concave-shaped structures, stromal changes in neoplasms, interstitial infiltrates and deposits. Importantly, these limitations also occur in humans, i.e., some cases are really hard to recognize [13].

Our method is based in two complementary strategies: the still segmentation that groups things and the relevance assignation by the attentional model. In general, the still-segmentation performance depends on many non-intuitive parameters. Figure 7.4 illustrates how incorrect the method performs for a wrong selection of parameters. In previous approaches these parameters must be exhaustively because of the intrinsec variability of these problems, for instance several type of tissues, different dyes or multiple acquisition magnifications [31]. In computer vision, this lack of flexibility has been aboarded via the use of meta-algorithms that look for the optimal set of parameters, i.e., the most similar segmentation to the manually segmented image, whereby a similarity measure is required. Previous approaches have used different quatifications of the overlapping area as similarity measurements, ignoring the perceptual coherence required for the segmented regions [27]. It should be strengthen out that perceptual meaning passes by a similarity measure which finds regions, but these regions are made up of pixels that share complex relationships so that we need a measurement that rewards strategies that find regions with perceptual meaning rather than pixels. In the present work, we proposed a novel meta-algorithm, based on a machine learning strategy which is trained using the regions set by an expert. The proposed measure is based on a functional that privileges the center region pixels and penalizes over and under segmentations, resulting in more coherent regions that should follow what the pathologist selected. The fact that this measurement uses a distance map allows to define an object from its to center to the bounday of interest. We recall that this work focused on the figure-ground segmentation process emulation, as discussed previously, this is a complex process influenced by the image content and the pathologist

experience. The proposed machine learning algorithm was developed for extracting in some way some the semantic of the selection task. The training samples used in this work were marked by an expert pathologist so that our machine learning method constrainted the segmentation parameter search to these ground truths. The results herein presented have demonstrated that the algorithm is very robust, that no signifcant change is observed when testing with completely new images, obtaining for them high perceptual qualities, near to -49dB.

Overall, previous works have used the typical approach of computer vision: image characterization followed by a statistical description. This strategy is limited by the fact that any algorithm should be re-trained with these two sequential steps for any new type of image i.e. this approach would easily fail with histopathological images. In contrast, we have proposed the use of attentional, much more robust to the image variability, introducing certain modifications for achieving saliency maps with semantic meaning. Yet the saliency elements in these images are very likely the same that are present in natural images, the way they are analyzed is totally different. While in any natural image there is practically no further analysis, pathologists must gather together areas which are highlighted by the classical mechanisms, but in many cases, areas which are hidden by these mechanisms. The present study has demonstrated that semantic searches require flexible methods that easily can include perceptual or prior knowledge, a statement supported herein by the difference between the quality of our segmentations and the scattered poorly defined regions found by the Itti's approach.

Part IV

Conclusions and Perspectives

8

Conclusions and Perspectives of this Work

8.1 Conclusions

This thesis has addressed the problem of navigation on medical mega-images from different perspectives. We mainly focused on the virtual microscopy domain for pathology (Chapter 2). The main contribution is the improvement in navigation velocities by combining several techniques such as highly scalable storage, caching, prefetching and user modeling.

In Chapter 3 we introduced a virtual microscopy framework which implements some techniques to make faster display of image portions by using soft cache and spatial cache to avoid redundancy when decoding J2K image packets. Additionally, a dynamically quality optimization strategy (prefetching) was designed to permit faster analysis of the images. Initially, both strategies were based on simple image exploration model: last recently used or last frequently used. Results were published and demonstrate the usefulness of the highly scalable storage and its associated clever strategies in the problem of navigation of medical mega-images.

In Chapter 4 an experimental study was carried out in order to establish main pathologists navigation patterns when exploring virtual microscopy slides. This study addressed two main issues: the RoI definition as primary concept of the navigation, and the existence of some navigation patterns generated by the interaction between expert, slide and graphical user interface. Results suggest that ROIs are determined by a complex combination of the region visited, the time spent at each visit and the coincidence level among pathologists. We found also evidence of the existence of linear trajectories

and particular velocity patterns in registered diagnostic paths. This lead us to conclude about the existence of a stochastic navigation process which probably is controlled by an underlying navigation model. We hypothesize that these patterns can be explained in terms of a complex interaction between the neuromuscular and visual systems, the navigation device and the image content.

In Chapter 5 an optimal soft-cache strategy for client-server architectures was proposed. The whole strategy is based on the natural concept of navigation by WoI, a complete mathematical derivation of the navigation problem in this scenario was developed. As result, a simple optimization problem was found; it includes as special cases the previous soft-cache and prefetching formulations. The strategy is strongly based on a navigation model on a dynamical probabilistic model of a pathologist's navigation. Results demonstrate an improvement in performance of the proposed method conventional approaches, as well as, the importance of a good selection of the navigation model. We also explore the possibility of determining a navigation model which permits to obtain benefits of the prefetching and cache in a optimal way.

In Chapter 6 an online Bayesian model which permits to anticipate the pathologist trajectories when exploring virtual slides was developed. The strategy was based on the intuition that when the exploration is carried out by mouse, the navigation velocity between two consecutive WoIs is related to an isotonic movement. The Bayesian strategy combines an offline learning model of velocity patterns described by a set of training pathologist's navigations and a prediction online module which predicts the velocity pattern for a particular pathologist. The proposed approach was evaluated, obtaining as result an increasing of the capacity prediction compared with state-of-the-art approaches used in similar problems. Finally, in Chapter 7 we introduced a supervised learning method for finding diagnostic RoIs in histopathological images. The method emulates the first examination phase, which consists in a coarse search for tissue structures at a "low zoom" to separate the image into relevant regions. In order to implement this approach we propose a novel visual attention model integrates bottom-up and top-down information. The first one, corresponds to local low level features such as intensity, color, orientation and texture information, and the second one, is related to the anatomical and pathological structures known by the expert, coarsely approached by an still-segmentation algorithm, inspired by psychological grouping theories. The method was evaluated on a set of

histopahtological images, using a perceptually-weighted evaluation criterion, finding a consistent quality gains when comparing to a classical attention models.

8.2 Perpectives

A virtual microscope accesses a Slide Image, upon which a pathologist navigates by locating RoIs which are then displayed at higher resolutions in auxiliary windows. A virtual navigation is thus composed of multiple spatial jumps and magnification changes. Actually, depending on both the image contents and its size, this process can result so variable that information has to be constantly reconstructed from basic units, in terms of different resolutions, qualities and random accesses to the image. This makes that flexible image data representations such as JPEG2000 result adequate to meet such requirements.

Overall, virtual slide images are usually of the order of thousands $\times$ thousands pixels, which must be stored in dedicated machines with large storage capacities. In actual applications, seamless navigation is dependent on a bandwidth which in general results insufficient for the huge volume of data to transfer. Caching and prefetching strategies may speed up image loading, but yet associated to pathologist navigation patterns, which are expected to be highly variable. A complete automated navigation would eliminate that variability, but in that case it is mandatory to establish a structured observation path for every WSI.

Some recent works have shown that it is possible to automatically determine RoIs [99, 59] so that a probabilistic map could be associated to the image. In addition, such observation path should follow optimal sampling strategies [77] which improve the diagnosis times, that is to say, a minimal number of RoIs would drive the navigation. In that scenario, a pathologist would achieve a diagnosis using this technology and spending a minimal time, so the pathologist workflow could be highly accelerated basically because it would be possible to prefetch the whole examination path at the very navigation beginning and little additional information would be needed in order to accomplish such diagnosis. However, observation paths are dependent on the image contents, on the pathologist's experience and the image quality, and therefore highly variable. These probability maps should then change along the navigation, depending on the expert needs. This model would need

then much more flexibility as to adapt the prior, defined by the observation paths, to the observations, such as the present investigation did at a local level. The work here presented was based on a local velocity model which was able to capture the expert velocity patterns within short time intervals and which adapted to the variability of the navigation. This local strategy could be used for defining observation paths, but predictions should be constrained by the probability map. It is worthy to point out here that our model was focused on the scanning phase of the microscopical exploration since the magnification changes take much less time in terms of navigation [79]. Moreover, these predictions could be included into the whole strategy by using smart sampling procedures [77].

On the other hand, the quality of the examined image is directly associated with the quality of the histological sample and the acquisition conditions, i.e., the standardization at different levels, namely, tissue fixation, embedding, cutting, dye preparation, stainning technique, microscope adjustment and acquisition procedures. The biological image contents and RoI selection strategies should also influence the design of protocols aimed at assuring the image quality [76]. Such protocols are embedded within the microscopy laboratory workflow [52, 61], but minor attention is paid to the factor user-image content interaction, i.e., the image expected contents should define the manner that many of these procedures are implemented. Development of new virtual navigation applications depends on the definition of image quality, a traditional concept which can be fed back by some evidence about how experts interact with images. This knowledge could be used for instance in compression applications, since the more knowledge of the exploration process is available the better is the compression, i.e., selective compression strategies [7]. Also, these observations or exploration paths may lead to new storage strategies in computational grids [52], as for example to improve the distribution algorithms by using more realistic exploration models.

Finally, modern diagnosis in pathology is a multimedia process in which the image findings are dictated and stored. In consequence, the only source of information is not visual but semantic knowledge associated to images [151]. This textual annotation should be integrated into all strategies previously discussed since they constitute a bridge between the low level image information and the semantic concept [123]. Nowadays, ontologies can be easily constructed from text documents and in virtual microscopy they could guide definition of RoIs, paths and velocities and obtain automatic navigations which maintain the diagnosis quality using a minimal time.

A

A Kinematic Method for Computing the Motion of the Body Center-of-Mass

Martínez F., Gómez F., Romero E. A Kinematic Method for Computing the Motion of the Body Center-of-Mass (CoM) During Walking: A Bayesian Approach. Accepted for publication in Computer Methods in Biomechanics.

B

Rotation Invariant Texture Characterization using a Curvelet Based Descriptor

Rotation Invariant Texture Characterization using a Curvelet Based Descriptor. Submitted by invitation to *Pattern Recognition Letters*. Extended version of *Texture Characterization using a Curvelet Based Descriptor. Published in Proceedings of the 14th Iberoamerican Conference on Pattern Recognition: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Lecture Notes In Computer Science; Vol. 5856, Pages: 113 - 120, 2009.*

Bibliography

[1] WxWindows open source library, http://www.wxwidgets.org.

[2] ISO/IEC 15444-1. Jpeg 2000 image coding system, 2000.

[3] ISO/IEC 15444-9. JPEG2000 image coding system part 9: Interactive tools, apis and protocols, 2003.

[4] S. Sridharan A. Nguyen, V. Chandran. Gaze tracking for region of interest coding in jpeg 2000. *Image Communication*, 21:359–377, 2006.

[5] M. Adams and F. Kossentini. JasPer: A Software-Based JPEG-2000 Codec Implementation.

[6] A. Afework, M.D. Beynon, F. Bustamante, A. Demarzo, R. Ferreira, R. Miller, M. Silberman, J. Saltz, A. Sussman, and H. Tsang. Digital dynamic telepathology – the virtual. In *AMIA Symp*, volume CS-TR-3892, pages 912–6, 1998.

[7] G. Anastassopoulos and A. Skodras. Jpeg 2000 roi coding in medical imaging applications. In *Proceedings of Visualization, imaging and image processing*, page 783788, Marbella, Spain, 2002. Acta Press.

[8] H. Andrade, T. Kurc, A. Sussman, and J. Saltz. Multiple query optimization support for the virtual microscope. *Archives of Pathology/ and Laboratory Medicine*, 2002.

[9] B. Appleton, A.P. Bradley, and M. Wildermoth. Towards optimal image stitching for virtual microscopy. In *DICTA*, page 44, 2005.

[10] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[11] W. Beggs and C. Howarth. The movement of the hand towards a target. *Quarterly Journal of Experimental Psychology*, 24:448–453, 1972.

[12] A. Blake and M. Isard. Active contours: The application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion. *Springer-Verlag New York*, 1998.

[13] A. Boer. Visual perception and consciousness in dermatopathology: Mechanisms of figure-ground segregation account for errors in diagnosis. *The American Journal of Dermatopathology*, 31(1):13–21, 2009.

[14] R. Bracewell. *The Fourier Transform and Its Applications*. McGraw-Hill, New York, 1965.

[15] A. Bradley, M. Wildermoth, and P. Mills. Virtual microscopy with extended depth of field. In *DICTA*, page 35, 2005.

[16] G. Bussolati. Dissecting the pathologists brain: mental processes that lead to pathological diagnoses. *Virchows Arch*, 448(6):739–743, 2006.

[17] L. Stark C. Privitera. Evaluating image processing algorithms that predict regions of interest. *Pattern Recognition Letters*, 19:1037–1043, 1998.

[18] U. Catalyurek, M. Beynon, C. Chang, T. Kurc, A. Sussman, and T. Saltz. The virtual microscope. *IEEE Transactions on Information Technology in Biomedicine*, 7(4):230–248, 2003.

[19] A. Chan, R. W. H. Lau, and B. Ng. Motion prediction for caching and prefetching in mouse-driven DVE navigation. *ACM Transactions on Internet Technology*, 5(1):70–91, 2005.

[20] A. Chan, R. W. H. L, , and Li. L. Hand motion prediction for distributed virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):146–159, 2008.

[21] C. Chang, T. Kurc, A. Sussman, U. Catalyurek, M. Beynon, and T. Saltz. The virtual microscope. *IEEE Transactions on Information Technology in Biomedicine*, 7(4):230–248, 2003.

[22] S. Chaudhuri. *Super-Resolution Imaging*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[23] J. Chim, M. Green, R.W.H. Lau, and H.V. A. Si. On caching and prefetching of virtual objects in distributed virtual environments. In *MULTIMEDIA 98: Proceedings of the sixth ACM international conference on Multimedia*, pages 171–180, 1998.

[24] C. Chistopoulos, A. Skodras, and T. Ebrahimi. The jpeg 2000 still image coding system: an overview. *IEEE transactions on Consumer Eletronics*, 46(4):1103–1127, November 2000.

[25] S. Chow, H. Hakozaki, D. Price, N. Mclean, T. Deerinck, J. Bouwer, M. Martone, S. Peltier, and M. Ellisman. Automated microscopy system for mosaic acquisition and processing. *Journal of Microscopy*, 222(2):76–S177, 2006.

[26] DIS COMMITTEE. IEEE Standard for distributed interactive simulation application protocols 1278, 1998.

[27] D. Crevier. Image segmentation algorithm development using ground truth image data sets. *Comput. Vis. Image Underst.*, 112(2):143–159, 2008.

[28] S. Crowley, G. Naus, and F. CP. Development of visual diagnostic expertise in pathology an information processing study. *Journal of the American Medical Informatics Association*, 10(1):30–51, 2003.

[29] B. Davison. A survey of proxy cache evaluation techniques. *Proceedings of the fourth international web caching workshop.*, 24(4):66–67, 1999.

[30] F. Dee, J. Lehman, D. Consoer, T. Leaven, and M. Cohen. Implementation of virtual microscope slides in the annual pathobiology of cancer workshop laboratory. *Human Pathology.*, 34(5):430–436, 2003.

[31] C. Demir and B. Yener. Automated cancer diagnosis based on histopathological images: a systematic survey. Technical report, Rensselaer Polytechnic Institute, Department of Computer Science,TR-05-09., 2005.

[32] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):138, 1977.

[33] A. Descampe. *Seamless Remote Browsing and Coarse-to-fine Compressed Retrieval using a Scalable Image Representation.* PhD thesis, Université catholique de Louvain, 2008.

[34] A. Descampe, J. Ou, P. Chevalier, and B. Macq. Data Prefetching for Smooth Navigation of Large Scale JPEG 2000 images. In *IEEE Conference on Multimedia and Exposition*, 2005.

[35] A. Descampe, C. De Vleeschouwer, M. Iregui, B. Macq, and F. Marques. Prefetching and caching strategies for remote and interactive browsing of JPEG2000 images. *IEEE transactions on Image processing*, 16:1339–1354, May 2007.

[36] S. Deshpande and W. Zeng. Http streaming of JPEG2000 images. In *ITCC*, pages 15–23, 2001.

[37] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18(1):193–222, 1995.

[38] G. Díaz, F.A. González, and E. Romero. A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. *J. of Biomedical Informatics*, 42(2):296–307, 2009.

[39] B. Eichhorn. Ole. system and method for viewing virtual slides, united states patent: 20040167806.

[40] J. Erasmus, G. Gladish, L. Broemeling, B. Sabloff, M. Truong, R. Herbst, and R. Munden. Interobserver and intraobserver variability in measurement of non small cell carcinoma lung lesions: Implications for assessment of tumor response. *Journal of Clinical Oncology*, 21:2574–2582, 2003.

[41] Bacus et al. US Patent 7596249 - Focusable virtual microscopy apparatus and method, September 29, 2009.

[42] X. Fan, H. Huang, D. Liang, and C. Qi. A hybrid parallel projection approach to object-based image restoration. *Pattern Recognition Letters*, 27(10):1045 – 1053, 2006.

[43] L. Fei-Fei, R. Van Rullen, C. Koch, and P. Perona. Why does natural scene categorization require little attention? exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, 12(5):893–924, 2005.

[44] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[45] R. Fernandez-Gonzalez, T. Deschamps, A. Idica, R. Malladi, and C. Ortiz de Solorzano. Automatic segmentation of histological structures in mammary gland tissue sections. *Journal of Biomedical Optics*, 9(3):444–453, 2004.

[46] R. Finkel and J.L. Bentley. Quad trees: A data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, 1974.

[47] P. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, 1954.

[48] P. Fontelo, E. DiNino, K. Johansen, A. Khan, and M. Ackerman. Virtual microscopy: Potential applications in medical education and telemedicine in countries with developing economies. In *Proceedings of the 38th Hawaii International Conference on System Sciences*, 2005.

[49] A. Georgopoulos, J. Kalaska, and J. Massey. Spatial trajectories and reaction time of aimed movements: effect of practice, uncertainty, and change in target location. *J Neurophysiol*, 46:725–743, 1981.

[50] C. Gielen, K. Van den Oosten, and F. Pull ter Gunne. Relation between emg activation patterns and kinematic properties of aimed arm movements. *Journal of Motor Behaviour*, 17(4):421–442, 1985.

[51] J. Gilbertson, J. Ho, L. Anthony, Y. Yagi D. Jukic, and A. Parwani. Primary histologic diagnosis using automated whole slide imaging: a validation study. *BMC Clin Pathol*, 27(6):4–, 2006.

[52] J. Gilbertson and Y. Yagi. Histology, imaging and new diagnostic work-flows in pathology. *Diagnostic Pathology*, 3(Suppl 1):S14, 2008.

[53] K. Glatz-Krieger, D. Glatz, and M. Mihatsch. Virtual slides: high-quality demand, physical limitations, and affordability. *Human Pathology*, 34(10):968–974, 2003.

[54] F. Gómez, M. Iregui, and E. Romero. Virtual microscopy using JPEG2000. In *The 12th International Conference on Computer Analysis of Images and Patterns CAIP, Viena*, 2007.

[55] F. Gómez, M. Iregui, and E. Romero. Prediction of pathologist navigation patterns in virtual microscopy based on a soft-computing approach. In *HCI '08: Proceedings of the Third IASTED International Conference on Human Computer Interaction*, pages 150–155, Anaheim, CA, USA, 2008. ACTA Press.

[56] F. Gómez, F. Martínez, and E. Romero. *Predicting Complex Patterns of Human Movements Using Bayesian Online Learning in Medical Imaging Applications*, chapter 13, pages 283–306. Medical Information Science Reference, 2009.

[57] F. Gómez and E. Romero. A model for predicting pathologist's velocity profiles when navigating virtual slides. *To appear in Microscopy Research and Technique*, 2009.

[58] F. Gómez and E. Romero. Texture characterization using a curvelet based descriptor. In *CIARP '09: Proceedings of the 14th Iberoamerican Conference on Pattern Recognition*, pages 113–120, Berlin, Heidelberg, 2009. Springer-Verlag.

[59] F. Gómez, J. Villalón, R. Gutierrez, and E. Romero. Finding regions of interest in pathological images: an attentional model approach. In Nico Karssemeijer and Maryellen L. Giger, editors, *Proc. SPIE Medical Imaging 2009*, volume 7260, page 72603G. SPIE, 2009.

[60] F. González and E. Romero. *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques*. Medical Information Science Reference, 2009.

[61] J. Gortler, M. Berghoff, G. Kayser, and K. Kayser. Grid technology in tissue-based diagnosis: fundamentals and potential developments. *Diagnostic Pathology*, 1(1):23, 2006.

[62] G. Gottlieb, D. Corcos, G. Agarwal, and M. Latash. Organizing principles for single joint movements. iii. speed-insensitive strategy as a default. *Journal of Neurophysiology*, 63(3):625–636, 1990.

[63] P. Guerfali. *Research in Computer and Robot Vision*. River Edge: World scientific., 1995.

[64] A. Sussman E. Borovikov H. Andrade, T. Kurc and J. Saltz. On cache replacement policies for servicing mixed data intensive query workloads. In *Second Workshop on Caching, Coherence, and Consistency (WC3 '02)*, 2002.

[65] C. Harris and M. A Stephens. Combined corner and edge detector. In *Proceedings of the Fourth Alvey Vision Conference*, pages 147–151, Alvey Vision Club, Manchester 1998.

[66] J. Hui and C. Liang. Multimedia application using flashpix file format (Patent style). *U.S. Patent 6237010*, page June, 1997.

[67] M. Iregui. *Efficient Strategies for Navigation Through Very Large JPEG2000 Images*. PhD thesis, Université catholique de Louvain, 2008.

[68] M. Iregui, F. Gómez, and E. Romero. Strategies for efficient virtual microscopy in pathological samples using JPEG2000. *Micron*, 38:700–713, 2007.

[69] L. Itti and C. Koch. Computational modelling of visual attention. *Nat Rev Neurosci*, 2(3):194–203, 2001.

[70] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on patern analysis and machine intelligence*, 20:1254–1259, 1998.

[71] K. Jun, O. Sertel, K. Boyer, J. Saltz, N. Gurcan, H. Shimada, and S. Qualman. Computer-assisted grading of neuroblastic differentiation. authors' reply. *Archives of pathology & laboratory medicine*, 132(6):903–904, 2008.

[72] L. Junqueira and J. Carneiro. *Basic Histology, Tenth Edition*. MacGraw Hill, 2003.

[73] J. Kangasharju, Y. Kwon, and A. Ortega. Design and implementation of a soft caching proxy. *IEEE Signal Processing Society 1998 Workshop on Multimedia Signal Processing.*, 1998.

[74] B. Karacali and A. Tozeren. Automated detection of regions of interest for tissue microarray experiments: an image texture analysis. *BMC Medical Imaging*, 7(1):2, 2007.

[75] D. Karras, S. Karkanis, and D. Maroulis. Efficient image compression of medical images using the wavelet transform and fuzzy c-means clustering on regions of interest. *euromicro*, 02:469–473, 2000.

[76] K. Kayser, J. Gortler, T. Goldmann, E. Vollmer, P. Hufnagl, and G. Kayser. Image standards in tissue-based diagnosis (diagnostic surgical pathology). *Diagnostic Pathology*, 3(1):17, 2008.

[77] K. Kayser, H. Schultz, T. Goldmann, J. Gortler, G. Kayser, and E. Vollmer. Theory of sampling and its application in tissue based diagnosis. *Diagnostic Pathology*, 4(1):6, 2009.

[78] T. Keahey. The generalized detail-in-context problem. In *Proceedings Vis'98, IEEE Visualization*, Research Triangle Park NC, 1998.

[79] E. Krupinski, A. Tillack, L. Richter, J. Henderson, A. Bhattacharyya, K. Scott, A. Graham, M. Descour, J. Davis, and R. Weinstein. Eye-movement study and human performance using telepathology virtual slides implications for medical education and differences with experience. *Human Pathology*, 37(12):1543–1556, 2006.

[80] C. Kuglin and D. Hines. The phase correlation image alignment method. *Conference*, pages 163–165, September 1975.

[81] R. Kumar, G. Velan, S. Korell, M. Kandara, F. Dee, and D. Wakefield. Virtual microscopy for learning and assessment in pathology. *Journal Pathololgy*, 204(5):613–618, 2004.

[82] B. Lessmann, T. Nattkemper, V. Hans, and A. Degenhard. A method for linking computed image features to histological semantics in neuropathology. *Journal of Biomedical Informatics*, 40(6):631–641, 2007.

[83] R. Lewis and V. Torczon. Pattern search methods for linearly constrained minimization. *SIAM J. on Optimization*, 10(3):917–941, 1999.

[84] J. Li and H. Sun. On interactive browsing of large images. *IEEE Transactions on Multimedia*, 5(4):581–590, 2003.

[85] C. Lin and Y. Zheng. Fast browsing of large scale images using server prefetching and client caching techniques. In *Applications of Digital Image Processing XXII*, SPIE, pages 376–387, July 1999.

[86] H. Liu, X. Xie, W. Ma, and H. Zhang. Automatic browsing of large pictures on mobile devises. In *Eleventh ACM International Conference on Multimedia*, pages 148–155, 2003.

[87] J. Maintz and M. Viergever. An overview of medical image registration methods. *Symposium of the Belgian hospital physicists association*, 12:1–22, 1997.

[88] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[89] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41 (12):3397–3415, 1993.

[90] H. Mamata, W. Scott, and S. E. Maier. Efficient construction of histology slide mosaics via phase correlation registration of high resolution tiles. In *ICIP (1), Barcelona,*, pages 1117–1120, 2003.

[91] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal Applied Mathematics*, 11:431–441, 1963.

[92] S. Martello and P. Toth. *Knapsack problems: algorithms and computer implementations*. John Wiley & Sons, Inc., New York, NY, USA, 1990.

[93] J. McGee, P. Isaacson, and N. Wright. *Oxford textbook of pathology*, volume 1 Principles of pathology. Oxford University Press, 1992.

[94] V. Mezaris, I. Kompatsiaris, and M.G. Strintzis. Still image segmentation tools for object-based multimedia applications. *International Journal of Pattern Recognition and Artificial Intelligence*, 18:701–725, 2004.

[95] S. Mikula, I. Trotts, J. Stone, and E.G. Jones. Internet-enabled high-resolution brain mapping and virtual microscopy. *NeuroImage*, 35 (1):9–15, 2007.

[96] P. Morasso. Spatial control of arm movements. *Exp Brain Res*, 42:223–227, 1981.

[97] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *Int J. Med Inform*, 73(1):1–23, 2004.

[98] H. Nagasaki. Asymmetric velocity and acceleration profiles of human arm movements. *Exp Brain Res*, 74:319–326, 1989.

[99] M. Oger, P. Belhomme, J. Klossa, J.J. Michels, and A. Elmoataz. Automated region of interest retrieval and classification using spectral analysis. *Diagnostic Pathology*, 3(Suppl 1):S17, 2008.

[100] D. Okada, S. Binder, C. Felten, J. Strauss, and A. Marchevsky. Virtual microscopy and the internet as telepathology consultation tools: diagnostic accuracy in evaluating melanocytic skin lesions. *American Journal of Dermatopathology*, 21(5):525–531, 1999.

[101] J. Ortíz, V. Ruíz, and I. García. Improved jpip protocol with proxy caching. In *Actas de las XV Jornadas de Paralelismo. Computación de Altas Prestaciones*, ISBN 84-8240-714-7, pages 328–333. Computacion de altas prestaciones: actas de las XV Jornadas de Paralelismo, 2004.

[102] J. Ortíz, V. Ruíz, M. López, and I. Garcia. Interactive transmission of JPEG2000 images using web proxy caching. *Multimedia, IEEE Transactions on*, 10(4):629–636, June 2008.

[103] J. Pani, J. Chariker, and R. Fell. Toward a theory of qualitative visual reasoning in microanatomy. In *In J. de Kleer, & K. D. Forbus (Eds.), 18th International Workshop on Qualitative Reasoning*, pages 233–238, Evanston, USA, 2004. Northwestern University.

[104] J. Peking and R. Ogilvie. *Series: Advances in Pathology, Microscopy and Molecular Morphology*, volume 3. CRC Francis and Taylor Press, 2005.

[105] G. Pena and J. Andrade-Filho. How does a pathologist make a diagnosis? *Arch Pathol Lab Med*, 133(1):124–132, 2009.

[106] W. Philips, S. Van Assche, D. De Rycke, and K. Denecker. State-of-the-art techniques for lossless compression of 3d medical image sets. *Computerized Medical Imaging and Graphics*, 25(2):173–185, 2001.

[107] R. Plamondon. A kinematic theory of rapid human movements: Part i: movement representation and generation. *Biological Cybernetics.*, 72(4):295–307, 1995.

[108] R. Plamondon. A kinematic theory of rapid human movements: Part ii: movement time and control. *Biological Cybernetics.*, 72(4):309–320, 1995.

[109] R. Plamondon. Kinematic theory of rapid human movements: Part iii. kinetic outcomes. *Biological Cybernetics.*, 78:133–145, 1998.

[110] C. Preza and J.A. Conchello. Image estimation accounting for point-spread function depth variation in three-dimensional fluorescence microscopy. In *Proc. SPIE, volume 4964, No. 27*, page 1336, 2003.

[111] M. Rabbani and J. Rajan. An overview of the JPEG2000 still image compression standard. *Signal Processing: Image communication*, 17:3–48, 2002.

[112] V. Rankov, R.J. Locke, R.J. Edens, P.R. Barber, and B. Vojnovic. An algorithm for image stitching and blending. In *Proceedings of SPIE, volume 4964, No. 27*, volume 5701, pages 190–199, 2005.

[113] P. Rhodes and S. Ramakrishnan. Iteration aware prefetching for remote data access. In *In E-SCIENCE 05: Proceedings of the First International Conference one-Science and Grid Computing*, pages 279–286, 2005.

[114] A. Roche, G. Malandain, X. Pennec, and N. Ayache. Multimodal image registration by maximization of the correlation ratio. *INRIA Sophia Antipolis, Rapport de recherche 3378*, 16:48–62, 1998.

[115] O. Rohde, D. Healy, C. Berenstein, and A. Aldroubi. Measuring image similarity to sub-pixel accuracy. In *Biomedical Imaging: Macro to Nano, 2006. 3rd IEEE International Symposium*, pages 638– 641, 2006.

[116] M. Rojo, G. Garcia, C. Mateos, J. Garcia, and M. Vicente. Critical comparison of 31 commercially available digital slide systems in pathology. *International Journal of Surgical Pathology*, 14 (4):285–305, 2006.

[117] D. Romer and S. Suster. Use of virtual microscopy for didactic live audience presentation in anatomic pathology. *Ann. Diagn. Pathol.*, 7(1):67–72, 2003.

[118] E. Romero, F. Gómez, and M. Iregui. *Virtual Microscopy in Medical Images: a Survey*, pages 996–1006. Formatex, 2007.

[119] E. Romero and C. A. Vargas. Low cost and efficient prototype of a motorized microscope. In *Electronics, Robotics and Automotive Mechanics Conference (CERMA2006)*, volume 1, pages 83– 86, 2006.

[120] A. Rosenfeld and A. C. Kak. *Digital Picture Processing*. Academic Press, 1982.

[121] B. Girod C. Beaulieu S. B. Gokturk, C. Tomasi. Medical image compression based on region of interest, with application to colon ct images. In *Proc. 23rd Annual Int. Conf. IEEE Engineering in Medicine and Biology Society*, 2001.

[122] S. Saltz. Virtual microscope: Databases and system software for multiscale problems, johns hopkins university. In *E-poster presented in Advancing Pathology Informatics, Imaging, and the Internet, APIII, Pittsburgh, PA, USA.*, 1999.

[123] T. Schrader, S. Niepage, T. Leuthold, K. Saeger, K. Schluns, P. Hufnagl, K. Kayser, and M. Dietel. The diagnostic path, a useful visualisation tool in virtual microscopy. *Diagnostic Pathology*, 1(1):40, 2006.

[124] C. Sempoux, Y. Guiot, D. Dubois, P. Moulin, and J. Rahier. Human type 2 diabetes: morphological evidence for abnormal beta-cell function. *Diabetes*, 50(1):S172–S177, 2001.

[125] N. Shivang, D. Scott, A. Shannon, M. Anant, F. Michael, and T. John. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In *Proceedings of the 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 284–287, Paris, France, 2008. IEEE.

[126] G. Snedecor and W. Cochran. *Statistical Methods, Eighth Edition*. Iowa State University Press, 1989.

[127] J. Soechting and F. Laquantini. Invariant characteristics of a pointing movement in man. *J Neurosci*, 1:710–720, 1981.

[128] S. Soenksen. Automated microscopic inspection of tissue microarrays using virtual microscopy. *Genomics & Proteomics Technology*, 1(1):28–31, 2003.

[129] C. Sun, R. Beare, V. Hilsenstein, and P.T. Jackway. Mosaicing of microscope images with global geometric and radiometric corrections. *Journal of Microscopy*, 224(2):158–165, 2006.

[130] D. Taubman. High Performance scalable image compression with EBCOT. *IEEE Transactions on Image Processing*, 9(7):1151–1170, 2000.

[131] D. Taubman and M.W. Marcellin. *JPEG2000 Image Compression, Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2002.

[132] D. Taubman and M.W. Marcellin. JPEG2000: Standard for interactive imaging. In *Proceedings of the IEEE, vol. 90, no. 8*, page 1336, 2002.

[133] D. Taubman and R. Prandolini. Architecture, philosophy and performance of jpip: Internet protocol standard for JPEG2000. In *International symposium on visual communications and image processing (VCIP2003)*, 2003.

[134] D. Taubman and R. Rosenbaum. Rate distortion optimized interactive browsing of JPEG2000 images. In *IEEE International Conference on Image Processing (ICIP2003)*, volume 3, pages 765–768, 2003.

[135] A. Tecnologies. Aperio image server performance measurements. Technical report, A. Tecnologies, 200.

[136] P. Thévenaz and M. Unser. User-friendly semiautomated assembly of accurate image mosaics in microscopy. *Microscopy Research and Technique*, 70(2):135–146, 2007.

[137] E. Tiersma, A. Peters, A. Mooij, and G. Fleuren. Visualising scanning patterns of pathologists in the grading of cervical intraepithelial neoplasia. *Journal of clinical pathology*, 59(9), 2003.

[138] A. Tosun, M. Kandemir, S. Cenk, and C. Gunduz-Demir. Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection. *Pattern Recogn.*, 42(6):1104–1112, 2009.

[139] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, January 1980.

[140] Y. Tsuchihashi, T. Mazaki, K. Nakasato, M. Morishima, H. Nagata, I. Tofukuji, H. Shirakata, K. Naito, and Y. Akasaka. The basic diagnostic approaches used in robotic still-image telepathology. *Journal of Telemedicine and Telecare*, 5(S1):115–117, 1999.

[141] P. Underhill, G. Goring, D.L. DuQuesnay, J. Kangasharju, Y. Kwon, and A. Ortega. Design and implementation of a soft caching proxy. *Computer Networks and ISDN Systems*, 30(22):2113–2121, 1998.

[142] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Englewood Cliffs, NJ : Prentice-Hall, 1995.

[143] P. Villegas and X. Marichal. Perceptually-weighted evaluation criteria for segmentation masks in video sequences. *IEEE Transactions on Image Processing*, 13(8):1092–1103, 2004.

[144] P. Viola. Alignment by maximization of mutual information. Technical Report AITR-1548, IST, 1995.

[145] J. Wang, J. Nguyen, K. Lo, C. Law, and D. Regula. Multiresolution browsing of pathology images using wavelets. In *Proc of AMIA Symposium*, page 430, 1999.

[146] H. Watanabe and T. Ogita. Transcoding by Automatic ROI Extraction from JPEG2000 Bitstream. In *Proceedings of Picture Coding Symposium 2003*, volume 1, pages 4–7, 2003.

[147] R. Weinstein. Innovations in medical imaging and virtual microscopy. *Human Pathology.*, 36:317–319, 2005.

[148] R. Weinstein, M. Descour, C. Liang, G. Barker, K. Scott, L. Richter, E. Krupinski, A. Bhattacharyya, J. Davis, A. Graham, M. Rennels, W. Russum, J. Goodall, P. Zhou, A. Olszak, B. Williams, J. Wyant, and P. Bartels. An array microscope for ultrarapid virtual slide processing and telepathology. design, fabrication, and validation study. *Human Pathology*, 35(11):1303 – 1314, 2004.

[149] M. Wertheimer. *Laws of organization in perceptual forms (partial translation)*, pages 71–88. A Sourcebook of Gestalt Psychology. Harcourt, Brace and Company, 1938.

[150] JTC1/SC29 WG1. JPEG2000 part I final committee draft version 1.0. Technical report, ISO/IEC, 2003.

[151] S. Wienert, M. Beil, K. Saeger, P. Hufnagl, and T. Schrader. Integration and acceleration of virtual microscopy as the key to successful implementation into the routine diagnostic process. *Diagnostic Pathology*, 4:3, 2009.

[152] F. Willemse. *A colored view on quantitative pahotlogy, aspects of true color image analysis in routine pahotlogy.* PhD thesis, Rijkuniversiteit Gronigen, 1996.

[153] D. Williams, J. Porter, G. Yoon, A. Guirao, H. Hofer, L. Chen, I. Cox, and S.M. Macrae. *How Far Can We Extend the Limits of Human Vision?*, chapter 3, pages 19–38. Slack Incorporated, 2004.

[154] S. Zhai and I. MacKenzie. Teaching old mice new tricks: Innovations in computer mouse design. In *Ergon-Axia 98 - the First World Congress on Ergonomics for Global Quality and Productivity*, pages 80–83, 1998.

[155] B. Zhang, Y. Zheng, and Q. Zhang. Extracting regions of interest based on phase spectrum and morphological approach. *2009 ISECS International Colloquium on Computing, Communication, Control, and Management*, 2009.

[156] Y. Zhang and J. Wang. Progressive display of very high resolution images using wavelets. *Journal of the American Medical Informatics Association, Proc. of AMIA Annual Symposium*, pages 944–948, 2002.

[157] L. Zheng. *Automated feature extraction and content-base retrieval of pathology microscopic images using K-means clustering and code run-length probability distribution*. PhD thesis, Pittsburgh University, 2005.

[158] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, pages 977–1000, 2003.

A Kinematic Method for Computing the Motion of the Body Center-of-Mass (CoM) During Walking: A Bayesian Approach

Journal:	*Computer Methods in Biomechanics and Biomedical Engineering*
Manuscript ID:	GCMB-2009-0216.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	13-Jan-2010
Complete List of Authors:	Martínez, Fabio; National University of Colombia, Bioingenium Research Group - School of Medicine Gómez, Francisco; National University of Colombia, Bioingenium Research Group - School of Medicine Romero, Eduardo; National University of Colombia, Diagnostic Imaging
Keywords:	Human movement, Gait analysis, Center of mass, Bayesian Filter, Particle filtering

Computer Methods in Biomechanics and Biomedical Engineering
Vol. 00, No. 00, October 2009, 1–18

RESEARCH ARTICLE

A Kinematic Method for Computing the Motion of the Body Center-of-Mass (CoM) During Walking: A Bayesian Approach

Fabio Martínez[a], Francisco Gómez[a] and Eduardo Romero[a]*

[a] *Bioingenium Research Group - School of Medicine, National University of Colombia, Bogotá DC - Colombia*

(*October 20009*)

The gait pattern of a particular patient can be altered in a large set of pathologies. Tracking the body center-of-mass (CoM) during the gait, allows a quantitative evaluation of these diseases at comparing the gait with normal patterns. A correct estimation of this variable is still an open question because of its non-linearity and inaccurate location. This paper presents a novel strategy for tracking the CoM, using a biomechanical gait model whose parameters are determined by a Bayesian strategy. A particle filter is herein implemented for predicting the model parameters from a set of markers located at the sacral zone. The present approach is compared with other conventional tracking methods and decreases the calculated RMSE in about a 56% in the x-axis and 59% in the y-axis.

Keywords: Human movement; Gait analysis; Center of mass; Bayesian Filter; Particle Filter

Index to information contained in this guide

1. Introduction

The gait pattern can be altered in a large set of pathologies, such as diabetes, brain palsy, cerebral vascular accidents, and neuromuscular dystrophies or from any kind of accident. The study of the human body movement or visual gait analysis is a modern tool that allows to objectively assess any of these pathologies. Examination is based on the follow-up of dynamic variables, whereby the disease severity can be quantified, and the gait compared with normal patterns(Gace 2001, 2004; Manal and Buchanan 2004).

The musculoskeletal dynamics, obtained from a gait analysis, is evaluated from the kinetic and kinematic perspectives. The kinematic analysis describes patient

*Corresponding author. Email: edromero@unal.edu.co

ISSN: 1741-5977 print/ISSN 1741-5985 online

DOI: 10.1080/1741597YYxxxxxxxx
http://www.informaworld.com

displacements in terms of the system components and its fundamental relationships, using variables such as the trajectory of the CoM, some specific angle joint variations or the step length, among others (Deluzio et al. 1997). The kinetic analysis quantifies the needed energy for the human movement to be produced, measuring electro-physiological states during displacement (Carson et al. 2001). A correct extraction and quantification of these variables is an open question since they are highly non-linear.

The CoM constitutes a fundamental descriptor for the clinical gait analysis because through its movements it is possible to describe both central nervous system and musculoskeletal disorders (Detrembleur et al. 2000; Gutierrez et al. 2003; Tucker et al. 1998). The more accurate CoM is usually estimated using a force plate, where a double integral of the ground reaction force in the time define the CoM displacement(Eames et al. 1999; Gard et al. 2004). This relationship is a simple dynamic equilibrium equation:

$$d_{CoM} = \int\int \frac{F_{GR} - mg}{m} dt^2 + v_0 t + d_0 \quad (1)$$

with F_{GR}= ground reaction force, m= body mass, g= gravity, t= time, v_0 and d_0 the integration constants of the initial velocity and position. This method is the gold standard for the CoM calculation. However, this method requires a patient steps upon the force plate, a difficult task in many musculo-skeletal disorders. Moreover, conventional gait laboratories have only two force plates, which results insufficient when gait analysis demands data from the whole gait cycle. This method is only used for validation of others techniques and its accuracy depends on the integration constants so that its utility is still limited in real clinical scenarios (Gard et al. 2004; Eames et al. 1999).

In the clinical routine, it is common to use the optic kinematic methods for estimation of the CoM because of its versatility and control. However, accurate location of the CoM is impossible because of the high inter-patient anatomical variability and its within-the-body location (Detrembleur et al. 2000). Usually, the CoM is tracked the closest marker in the video (Duff-Raffaele et al. 1996). This CoM can also be estimated from the body segments (Eames et al. 1999), but its extraction is a complicated task that requires many markers and changes the natural motion expression. These methods do not accurately estimate the CoM trajectory(Eames et al. 1999; Gard et al. 2004), since this is the result of a complex interaction of forces, neuromotor commands and joint movements of the lower limbs, and therefore it shows a high non-linear dynamics (McGeer 1990).

This article presents a precise and efficient strategy for estimating the temporal CoM location using a non-linear gait biomechanical model whose parameters are recursively adjusted by a Bayesian strategy, herein implemented as a particle filtering. The Main contribution of this work is to use a very simple methodology to follow an actual non linear dynamic. The whole strategy allows a natural and accurate tracking of the non-linear gait patterns with a high degree of noise robustness. The rest of this paper is organized as follows: Section Materials and Methods introduces the biomechanical model and our prediction method, Section Results demonstrates the effectiveness of the method. The last section presents a discussion and possible future works.

2. Materials and Methods

The strategy for tracking the CoM requires that the set of video frames, within two alternating heel strikes (gait cycle), are segmented. Tracking is carried out under a Bayesian framework that defines the trajectory of the CoM as a sequence of hidden states. At each step of the gait, the more probable parameters are calculated from a set of observations and the previous step parameters.

The Bayesian approach for tracking aims to estimate the hidden states of a system from a set of observations, or at least to extract useful information under the fundamental hypothesis that the observation process do not completely destroy the link between the true and the observed variables, or that the observed and true data are somehow close together under a particular metric. A Bayesian strategy starts at defining the system state as a random variable x_t and its associated probability density function (*pdf*), i.e., the uncertainty level of the occurrence of the state x_t. Bayesian filters estimate such *pdfs* upon a state space following the sequence of observations. The belief $Bel(x_t)$ is defined as a posterior probability density function of x_t, conditioned to all the observed available data $z_1, z_2, \ldots, z_t$ at time t. This *pdf* addresses the question: *what is the probability that the system state is in* x_t *if the history provided by the measurements of the process are* $z_1, z_2, \ldots, z_t$*?* Since the number of observations increases through the time, the complexity of such *posterior density* grows exponentially. This estimation becomes computationally tractable by assuming a Markovian hypothesis: the current state of the system contains all the relevant information. Under this assumption, $Bel(x_t)$ may be computed efficiently without information losses. In practice, it is a required a system model $p(x_t|x_{t-1})$, that represents *how the system states change in a time step* and a likelihood model $p(z_t|x_t)$, which describes the *probability of making the observation* z_t *if the system state is in the* x_t state. Finally, the initial system state $Bel(x_0) = p(x_0)$ is also needed. Once this information is available, the belief $Bel(x_t)$ is calculated in two recursive steps:

- *Prediction*. The belief in the state x_t is computed by updating the previous belief $Bel(x_{t-1})$, according to the prediction given by the system model $p(x_t|x_{t-1})$, through the Chapman-Kolmogorov equation:

$$\widehat{Bel}(x_t) = \int p(x_t|x_{t-1}) Bel(x_{t-1}) dx_{t-1} \tag{2}$$

- *Update*. The predicted belief $Bel(x_t)$ is adjusted after the system observations:

$$Bel(x_t) = \frac{\widehat{Bel}(x_t) p(z_t|x_t)}{\int \widehat{Bel}(x_t^*) p(z_t|x_t^*) dx_t^*} \tag{3}$$

Bayesian filters provide a probabilistic framework for recursive and sequential estimations of the system state. This representation is important for obtaining good estimators in non-linear/non-Gaussian dynamics as it is the case of the CoM trajectory, herein tracked through a particle filtering which approached the belief function by discrete sampling.

2.1 Modeling the CoM movement

The human gait is a locomotion process which involves both lower limbs to help the body to keep the balance while it gains support and propulsion (Whittle 1996). This process comprises a cyclic set of movements, where one foot acts as a fixed point which supports the body swings, whereas the free foot moves forward until it reaches the floor and becomes the new fixed point.

* * * * * * * * * * * * * * * * *
INCLUDE FIGURE 1 ABOUT HERE
* * * * * * * * * * * * * * * * *

Schematically, the upper part of the body is represented by a mass which moves forwards with respect to each fixed point, describing a harmonic oscillating trajectory, similar to the inverted pendulum (Buczek et al. 2006). At the same time, the free foot swings with respect to this mass, such as a simple pendulum. Given that these processes are coupled together, the human gait is modeled by a double-articulated pendulum (see Figure 1). This model properly represents the gait trajectory and it has been widely used (Kuo 2007; Dong et al. 2008; Komura et al. 2005). Based on the double articulated pendulum model, MacGeer (McGeer 1990) formulated the theory of passive dynamic motion, which describes the movement without a complex control mechanism, i.e., it is more important the body's structure for understanding the gait rather than its control or muscular activity. Afterward, based on this theory, Garcia et. al. (Garcia et al. 1998) formulated a simplified gait model. Accordingly, the feet are relatively small with respect to the trunk and the heelstrike is subjected to a restriction rule. This model has been used for tracking other body's structures (J. et al. 2004; Goswami et al. 1998; Zajac et al. 2003). The model corresponds to two coupled non-linear differential equations:

$$\begin{aligned} &\ddot{\theta}(t) - \sin\theta(t) = 0 \\ &\ddot{\theta}(t) - \ddot{\phi}(t) + \dot{\theta}(t)^2 \sin\phi(t) - \cos\theta(t)\sin\phi(t) = 0 \end{aligned} \tag{4}$$

where θ is the angle of the stance leg at particular time t with respect to the slope and ϕ is the angle between the stance leg. The model also defines a transition rule that simulates the swing foot when it hits the ground at the heelstrike, this moment corresponds to $\phi(t) - 2\theta(t) = 0$.

2.2 Online Adaptation with a Bayesian Filter

The biomechanical model computes the trajectories for the CoM from temporal functions (θ_t, ϕ_t), which are hidden, and the set of markers, which are the only direct measurements. We shall focus on the markers located within the sacral zone, more specifically the $z_t = (M_t^1, M_t^2, M_t^3, M_t^4)$ located around the hip (two at the back and two at the front), as illustrated in Figure 2. Provided that locations of these markers are indirect measurements of the CoM movement, their location in

time t corresponds to the two hidden parameters $x_t = (\theta_t, \phi_t)$ which better match the observed marker locations $z_1, z_2, \ldots, z_t$ at time t. The most probable state x_t can be recursively found using a Bayesian filter. For doing so, we need to set an initial estimation of the CoM location, a prior of CoM evolution during the gait and the likelihood function which associates states and observations (marker locations).

* * * * * * * * * * * * * * * * *
INCLUDE FIGURE 2 ABOUT HERE
* * * * * * * * * * * * * * * * *

An estimation of the initial CoM location is herein calculated from a Gaussian distribution whose mean and covariance are computed from the set of four positions located closest to the actual CoM. These positions are marked by an expert in the first frame of the video gait recording.

The dynamic model of the CoM movements indicates how the angular functions (θ_t, ϕ_t) change during a gait cycle. Let us suppose that the gait satisfies Garcia's model (equation 4) including the heel strike rule. This model can be modified for including the gait variability so that gait dynamic reads as

$$P\left(x_t | x_{t-1}\right) = G(x_{t-1}, \Sigma_1^2) \tag{5}$$

where G corresponds to a multivariate Gaussian distribution, whose mean was experimentally set (Garcia et al. 1998) and σ describes an inherent gait variability. The covariance matrix Σ_1^2 was calculated using a maximum likelihood estimation (Kay 1993) from a data set given by the difference between the ideal CoM signal (drawn from (Eames et al. 1999)) and the signal obtained from the prior model (Garcia et al. 1998). On the other hand, the likelihood function should yield a maximal probability when the states closely follow the observations, i.e:

$$P\left(z_t | x_t\right) = G\left(\frac{1}{4}\left[\bar{z}_{x,t}, \bar{z}_{y,t}\right] - Le[\sin(x_{\theta_t,t}), \cos(x_{\theta_t,t})], \Sigma_2^2\right) \tag{6}$$

where $\bar{z}_{x,t} = \frac{1}{4}\sum_{i=1}^{4} M_{x,t}^i$, $\bar{z}_{y,t} = \frac{1}{4}\sum_{i=1}^{4} M_{y,t}^i$, $M_{x,t}^i$ and $M_{y,t}^i$ are the marker coordinates in the x and y directions. The CoM coordinates defined by $(Le[\sin(x_{\theta_t,t}), \cos(x_{\theta_t,t})])$ are directly computed from the geometric representation of Garcia's model, where Le is the leg length and $x_{\theta_t,t}$ is the angle of the stance leg with respect to the slope, at a particular time t. Adittionally, Σ_2^2 is a predefined covariance matrix calculated as a maximum likelihood estimation from a data set, given by the difference between the ideal CoM signal and observations. The mean of the Gaussian distribution stands for a measure of how well the hidden angles meet the observed marked positions. The use of this Gaussian is fully justified since the noise associated to the captured positions is independent of the anatomical details.

3. Particle Filter Implementation

A different approach to represent the Belief is a discrete approximation of the *pdf* using Monte Carlo techniques. These methods approximate the probability density function $p(x)$ using a large number of samples,

$$p(x) \approx \sum_{i=1}^{n} w^{(i)} \delta(x - x^{(i)}) \tag{7}$$

where $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ are a set of n discrete independent and identically distributed variables (i.i.d), $w^{(i)}$ are the weights which stand for the probability of occurrence of the sample x_i, and $\delta(x)$ is the Dirac function. Clearly, the larger the number of samples, the closer the description is to the *pdf*. This approximation allows useful quantities, such as the discrete expected value of any function $f(x)$:

$$\int f(x)p(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} w^{(i)} f\left(x^{(i)}\right) \tag{8}$$

and the samples are generated using methods such as the rejection or importance sampling (Arulampalam et al. 2002), better known as Particle Filters.

A particle filter is the discrete version of the Bayesian filter obtained when the belief is approximated using a Monte Carlo (MC) method. The belief is estimated with the point mass distribution defined in equation 7, when replacing the Chapman-Kolmogorov equation by the approximation defined in equation 8, and discretizing the update equation with the belief previously calculated. Since each particle corresponds to an independent state of the system, with a number of particles which is function of the desired precision, the method allows to simulate the evolution of the complete system with no restrictions regarding linearity or noise. Additionally, the belief can be modified for each of the simulated system states as to adapt to real observations. Finally, classic estimators such as the expected value of the state or the maximum a posteriori can be calculated from the simulation while the system is evolving for predicting future states.

The Particle filter is constructed as follows. First, the Belief is approximated by an empirical point-mass function:

$$Bel(x_t) \approx \sum_{i=1}^{n} w_t^{(i)} \delta(x_t - x_t^{(i)}) \tag{9}$$

where the weights $w_t^{(i)}$ are chosen using the principle of sequential importance sampling (SIS) (Arulampalam et al. 2002), which states that samples can not be directly sampled but rather found through a density importance function. Hence a weighted aproximation is given by:

$$w_t^{(i)} \propto \frac{P(x_0 \ldots x_t | z_1 \ldots z_t)}{q(x_0 \ldots x_t | z_1 \ldots z_t)} \tag{10}$$

where $P(x_0 \dots x_t|z_1 \dots z_t)$ is exactly the very same belief function, as defined in section 2 and $q(x_0 \dots x_t|z_1 \dots z_t)$ is an importance density function, chosen to meet:

$$q(x_0 \dots x_t|z_1 \dots z_t) = q(x_t|x_{t-1})q(x_0 \dots x_{t-1}|z_1 \dots z_{t-1}) \tag{11}$$

Now, particles evolve by dependences which are always of first order, so we can suppose that $\left\{x_{t-1}^{(i)}\right\} \sim q(x_t|x_{t-1})$. We can then assume that $q(x_t|x_{t-1})$ represents the system CoM dynamics as $P(x_t|x_{t-1})$, so:

$$q(x_t|x_{t-1}) = G(x_{t-1}^{(i)}, \Sigma_1^2) \tag{12}$$

Particles need then to be updated after the Bayes rule defined in equation (3), which can then be approximated as follows:

$$P(x_0 \dots x_t|z_1 \dots z_t) \propto P(z_t|x_t)P(x_t|x_{t-1})P(x_0 \dots x_{t-1}|z_1 \dots z_{t-1}) \tag{13}$$

By substituting (11) and (13) into (10), the weight updating equation reads as:

$$w_t^{(i)} \propto \frac{P(z_t|x_t)P(x_t|x_{t-1})P(x_0 \dots x_{t-1}|z_1 \dots z_{t-1})}{q(x_t|x_{t-1})q(x_0 \dots x_{t-1}|z_1 \dots z_{t-1})} \tag{14}$$

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{P(z_t|x_t)P(x_t|x_{t-1})}{q(x_t|x_{t-1})} \tag{15}$$

note that the left term is just the weight in the precedent time and the rest of the expression corresponds to the system dynamics, previously defined. Therefore, this entire expression can be re-written as:

$$w_t^{(i)} = w_{t-1}^{(i)} G\left(\frac{1}{4}\left[\bar{z}_{x,t}^{(i)}, \bar{z}_{y,t}^{(i)}\right] - Le[\sin(x_{\theta_t,t}^{(i)}, \cos(x_{\theta_t,t}^{(i)})], \Sigma_2^2\right) \tag{16}$$

After some iterations, all but one particle will have negligible weight, a recurrent drawback broadly documented in the literature and known as the SIS degeneracy phenomenon (Arulampalam et al. 2002). To cope with, we introduce a step of resampling that generates a new set of particles at each iteraction. Finally, The algorithm is explicitely described hereafter:

Algorithm 1 Particle Filter implementation

$\left[\{x_t^i, w_t^i\}_{i=1}^N\right] \sim Bel(\{x_t^i, w_t^i\}_{i=1}^N, z_k)$

FOR $i = 1 : N$

set $\left\{x_{t-1}^{(i)}\right\} \sim G(x_{t-1}^{(i)}, \Sigma_1^2)$

update weights according to $w_t^{(i)} = w_{t-1}^{(i)} G\left(\frac{1}{4}\left[\bar{z}_{x,t}^{(i)}, \bar{z}_{y,t}^{(i)}\right] - Le[\sin(x_{\theta_t,t}^{(i)}, \cos(x_{\theta_t,t}^{(i)})], \Sigma_2^2\right)$

END FOR

normalize weights $\sum_{i=1}^{n} w_t^{(i)} = 1$

resample $\{x_t^i, w_t^i\}_{i=1}^N$

3.1 *Gait data*

Validation was carried out on a set of gait cycles segmented from the humanEva dataset (Sigal and Black. 2006). The humanEva data consists of a set of videos, captured from four subjects in different activities, using a calibrated marker-based-motion-capture-system and multiple high-speed video capture systems. Every video is provided with an associated motion data in C3D format that describe the accurate 3D marker position. For evaluation purposes, we assume this capture has an associated Gaussian noise, an statement fully justified since the marker position is independent of the anatomical location and the capture process, as described before in subsection 2.2. The proposed strategy was assessed on twenty gait cycles from three different subjects ,i.e., 8400 frames which corresponded to a total of sixty cycles. Each cycle corresponds to the set of frames within two alternating heelstrikes. The initial and final points of each cycle were selected by an expert in the domain, who also verified that the extracted sequences corresponded to normal gait patterns. The locations of four hip markers were extracted for each frame, as well as the CoM location. For doing so, the evaluated subjects have a set of attached makers, following the VCM protocol (Vicon Clinical Manager), from which 4 were extracted and defined as:

- $LPSI$: Placed directly over the left posterior superior iliac spine
- $RPSI$: Placed directly over the right posterior superior iliac spine
- $LASI$: Placed directly over the left anterior superior iliac spine
- $RASI$: Placed directly over the right anterior superior iliac spine

Finally, a marker placed in the sacral region was set and defined as the CoM.

4. Evaluation and Results

Evaluation was carried out by comparing the accuracy of the presented strategy with two standard tracking approaches: the location of the closest marker to the CoM and an exponential weighted moving average (EWMA) (Hunter 1986; Crowder 1987), a typical method for tracking time series. Control data were obtained from normal patterns, captured from force plates and reported in the literature (Eames et al. 1999) which thereby fully describe coordinates x and y of a normal cycle. The particle filtering algorithm was written in java (JDK 1.6.0_07) and run under an AMD turion 64 processor of 1.59 GHz and 3 GB in RAM. The Garcia's model was solved using a fourth order Runge-Kutta method, also implemented in Java.

* * * * * * * * * * * * * * * * *
INCLUDE FIGURE 3 ABOUT HERE
* * * * * * * * * * * * * * * * *

Figure 3 shows the gait trajectory decomposed into the y-coordinate at the left, and the x-coordinate at the right, during a complete gait cycle (control). The y-axis displays in both cases the CoM change, in percentage, weighted by the body height, while the x-axis is weighted by the entire cycle duration and also expressed in percentage. The two panels superimposed upon the same plot, the control gait path and the predicted trajectory of three different strategies: the closest marker to the CoM, a EWMA computed using the actual CoM observations along the recorded gait cycle and the approach herein presented (Bayesian tracking). The vertical movement or the y-axis spans the control pattern during 300 ms, along which the periodic movement is smoothly increasing within the first 40 % of the entire cycle. At this point, the height percentage decreases in a non-linear manner, because of the heelstreak, a task quite difficult to follow by any of the methods used. Importantly, the Bayesian tracking matches this non linearity much better than the other two, i.e., the Bayesian presents a little oscillation in this phase but rapidly decreases and closely follows the control pattern. On the contrary, the other two methods are highly oscillating after this phase and remain distant from the actual pattern until the cycle ends up. Regarding the x-axis displacements, there appear two different trajectories, a first one formed by the control and the Bayesian method and a second one by the other two methods, which follow nearly the same path but are shifted away from the control trajectory. These results demonstrate that the proposed method is more stable than the EWMA after the heel strike and follows much better a control pattern which is quite noisy. Interestingly, the particle filtering approximates better the non-linear plot of the control trajectory, a result which holds our original hypothesis about the necessity of using more robust tracking methods.

* * * * * * * * * * * * * * * * *
INCLUDE TABLE 1 ABOUT HERE
* * * * * * * * * * * * * * * * *

Table 1 presents the mean and the standard deviation of the Root Mean Squared Error (RMSE) between the tracked data and the control series for the three methods during twenty gait cycles, which correspond to about 30 s of gait, distributed among four different subjects. The higher tracking error is observed for the COM marker, a result that may be attributed to the fact that this is not actually placed close to the CoM as well as to the high noise levels in the acquisition process. Indeed, the EWMA decreases these error rates at smoothing these noise levels, but the error is still unacceptable for many applications, such as the prostheses fabrication, in which the prostheses mean life is absolutely dependent on the accuracy design (Jia et al. 2008; Pinzur et al. 1995). The particle filtering presents the lower errors, approximating better the control data since this model is not only

capable to adapt to the marker position variability but also to the non linearities in the trajectory. These results strongly suggest this method is steadier and more accurate for the tracking task.

Since the whole problem consists in following temporal series which are highly non linear and whose dynamics is therefore very difficult to determine, it is important to establish a metric on which it would be possible to measure the level of agreement between two trajectories. Herein we have measured this concordance level using the correlation coefficient of the temporal differences between two series. The correlation coefficient measures the degree to which two things vary together or to which two things draw apart together. In this case, the concordance level was evaluated as the correlation coefficient from the temporal differences between series.

* * * * * * * * * * * * * * * * *
INCLUDE TABLE 2 ABOUT HERE
* * * * * * * * * * * * * * * * *

Table 2 shows the correlation factor of the temporal differences between the kinematic data of the three methods and the control series during twenty gait cycles, in the x and y axes. Again, while the Marker CoM and the EWMA methods hardly follow the control data, the particle filtering tracking outperforms these two methods and definitely correlates with the control data (Correlation factor of about 0.9 in the two cases). A statement which is easily inferred from the correlation coefficients presented in table 2. Furthermore, figures in this table show a high correlation in the x and y axes, but importantly a better performance of the particle filtering method in the two dimensions, for instance, from 0.7 to 0.9 in the y-axis and from 0.5 to 0.9 in the x-axis. The results prove a natural tracking of the control signal obtained by the particle filtering method, i.e., the predicted data properly scale and shrink. In contrast, the CoM Marker or EWMA predicted data are inadequate for tracking the different phases of the gait and hence inappropriate to detect small changes, an issue which results crucial for identifying certain pathologies (Thirunarayan et al. 1996).

In despite of the so far better performance of the particle filtering method, provided that changes are in any case very small, it is very difficult to follow differences through the time and the correlation coefficient is a global measurement. In particular, it is really difficult to figure out the real gain of any method at any time, since results are always contaminated by the particular noise in measurements at any of the two dimensions and differences, as said before, are really small. In consequence, the quality of the prediction was weighted by the noise, using a logarithmic scale and measuring the difference between the expected (control data) and the predicted values. This SNR-like or quality performance measure reads as

$$q((x_c, x_m), (y_c, y_m))^{dB} = 10 \log \left[\frac{1}{(x_c - x_m)^2 + (y_c - y_m)^2} \right] \tag{17}$$

where (x_c, y_c) are the ground truth coordinates at the time t and the (x_m, y_m) are the signal coordinates at the same time. The great advantage with this mea-

sure is that it allows a temporal gain follow-up of the tracking. Overall, the most important component is the vertical (y-axis) since most pathologies alter mainly this vertical pattern (Gard et al. 2004; Thirunarayan et al. 1996). Therefore, this quality measurement was modified by weighting the horizontal and vertical components with a coefficient α, whereby we could tune the importance of the vertical direction. Then, this quality measure can be written as:

$$q((x_c, x_m), (y_c, y_m))^{dB} = 10 \log \left[\frac{1}{((1-\alpha)(x_c - x_m)^2 + (\alpha)(y_c - y_m)^2)} \right] \quad (18)$$

This evaluation was performed using the cycle for which the averaged difference between the control and each of the temporal series was the smaller, when comparing with the entire sequence of differences among gait cycles. Quality was then measured using α values of 0.9, 0.8 and 0.7 and results were plotted in figure 4. The four graphs show in the y-axis the gain in decibels for each of the methods and in the x-axis the cycle with smaller averaged difference. For comparing so, each cycle was expressed as percentage since overall each cycle spans a different time interval. The four panels are distributed as follows: Upper left panel is the first quality measure, that is to say, there is no α, upper right panel corresponds to the SNR-like measure with $\alpha = 0.7$ and left and right bottom panels correspond to this measure with α set to 0.8 and 0.9, respectively.

* * * * * * * * * * * * * * * * *
INCLUDE FIGURE 4 ABOUT HERE
* * * * * * * * * * * * * * * * *

In the image processing community, the PSNR measure is a term for the ratio between the maximum possible power of signal and the power of corrupting noise that affects the fidelity to the original data. Typical values for the PSNR in lossy image and video compression are between 30 and 50 dB, where higher is better. In the case of our evaluation a decibel entails a large difference and this accounts as a quality measurement of the fidelity to the control data. Overall, plots in the four panels show the same pattern, that is to say, the particle filtering always outperforms the other two methods. The pattern of the four panels results to be very alike at the beginning, but rapidly after the gait has reached a 30 % of the entire cycle, the particle filtering strategy shows a larger gain, between one and three decibels. The larger difference appears when the cycle has spanned about a 60%, as expected since at this time the gait pattern is much more non linear because of the heelstrike and the particle filtering is better suited to following this kind of discontinuities. Likewise, the introduction of a particular preference for the vertical direction in the SNR-like measurement has no influence on the results. Basically, the difference between plots with α set for instance to 0.7 (upper right panel) and the SNR-like measurement with no α (upper left panel) is about 0.5 decibels, and most punctual differences are slightly amplified, indicating that the x-axis or temporal evaluation contribute very little to the whole variability of the SNR-like measurement. This result agrees, with what is known in the literature, in the sense that most important evaluation must be performed upon the vertical direction (Gard et al. 2004; Thirunarayan et al. 1996). Finally, the analysis performed setting α to 0.8 and 0.9, shows the same trend, i.e., the vertical direction weights more and

differences are more important, demonstrating that the particle filtering strategy is able to closely follow these non linear patterns, and to more accurately predict this vertical movement.

The SNR-like assessment was also performed on the twenty cycles of the three patients and the mean with its standard deviation were calculated from this set of data.

* * * * * * * * * * * * * * * *
INCLUDE TABLE 3 ABOUT HERE
* * * * * * * * * * * * * * * *

Table 3 shows the mean and standard deviation of sixty cycles from three different patients. In general then, among the three methods, the higher quality gains are observed for the particle filtering, as was illustrated before in Figure 4 for the best cycle. These differences are larger when evaluation has no α value, as expected since the particle filtering outperformed the other two methods in the two directions. In this particular case, the trajectory generated by our model had an approximated quality of 2.17, compared to the averaged 1.34, obtained by the conventional methods.

When the vertical preference is introduced into the SNR-like measurement, differences decrease but the Bayesian tracking conserves the better performance at any of the set α values. The results with this modification show a considerable gain (mean of 3.17 for $\alpha = 0.9$), whereby the three methods track much better the vertical movement but the particle filtering presents still the best performance so that its utility in clinical application could be even better. Figures for the EWMA and Marker CoM methods look very similar, in despite the Marker CoM estimation is a very noisy.

The accuracy of the particle filtering method prediction is of course dependent on the goodness of the approximation to the *pdf*. This is essentially an issue which is function of the number of particles used to get this approximation. Accuracy was herein defined as the averaged RMSE for each of different number of particles. The Bayesian fidelity to the control data was thus assessed using a different number of particles so that we could establish the influence of this parameter. Since the whole performance picture includes also the computational time, the running time is also described in table 4.

* * * * * * * * * * * * * * * *
INCLUDE TABLE 4 ABOUT HERE
* * * * * * * * * * * * * * * *

Table 4 shows a high gain obtained with 100 particles when comparing with 10, but a smaller one when 1000 particles are used instead. On the contrary, the variance is comparable in the three cases, indicating a systematic error decreasing

with the number of particles, as expected. This statement holds true for the two dimensions (x and y) but this error reduction results very small when using 1000 particles. Computation times evidence an important increasing at using 1000 particles, when comparing with 100. It is very likely that a time of 7 s is incompatible with real time applications while the gain in error reduction is actually small, a trend which will be even bigger for larger numbers of particles since this exponential difference is a well known characteristic of this type of algorithm (Fox et al. 2003; Arulampalam et al. 2002).

Finally, it is worthy to point out an efficiency issue within this evaluation. Efficiency was herein conveniently defined as the precision per unit cost. For the present evaluation, the precision of a final estimate was expressed as the reciprocal of the data variance (σ^2) and the cost as the running time t. Hence, an index of efficiency E was computed as $\frac{1}{\sigma^2 t}$

* * * * * * * * * * * * * * * * *
INCLUDE TABLE 5 ABOUT HERE
* * * * * * * * * * * * * * * * *

Table 5 shows a larger efficiency when using 100 particles, a statement that was difficult to figure out from data in table 4 and which illustrates the needed balance between accuracy and the computational cost, an important demand in real time applications.

5. Discussion

This article has introduced a novel strategy for the CoM to be closely followed during actual body gait cycles, which is based upon some probabilistic considerations, namely, there exists an extensive domain knowledge about the gait physiology and the whole process can be modeled as a Markovian process so that the future system states are stochastic functions of the past system states. The particle filtering framework used in the present investigation nicely dealt with this highly non linear gait dynamics, based on the calculation of two complementary terms: a prior adapted from a well known mathematical gait model, whose aim is to predict the gait cycle (Garcia et al. 1998), and which is constantly tuned by a function that expresses the link between observations and system states (the likelihood). This approach is simple, easy to implement when the *pdf* is approximated by a particle filter and efficient in terms of computational running time and accuracy on the obtained measurements.

As extensively discussed before, the CoM has been a quantity difficult to establish due to its highly non linear dynamics. Previous works (Eames et al. 1999; Gard et al. 2004) have presented significant differences between optical methods and the CoM estimated from the double integration of the reaction force, measured in ordinary force plates. Many authors consider that the CoM trajectory, determined from the reaction force, is the gold standard in the CoM examination (Whittle 1996, 1998; Eames et al. 1999; Gard et al. 2004). Biomechanical literature is highly rich in methods whose aim is to estimate the CoM position. Nevertheless, their clinical use remain limited because most commercial laboratories are provided with one or

two force plates, a set up that hardly adapts to the step length variability since, at the best scenario, one entire gait cycle is assessed. Besides, repeatability of the gait analysis is poorer with these force plates and so the analysis becomes inconsistent. Overall, there have existed two main approaches: the first has been a very practical solution at associating the CoM temporal pattern to a marker located within the sacral zone, defined by an expert. Yet its accuracy is enough in many clinical studies, as it was shown in the present investigation, this estimation hardly follows the non linear CoM patterns, very likely because the marker dynamics is by itself non linear and not necessarily of the same type. This picture would be worst in those disorders in which the gait is so altered that it is really difficult to define even a cycle pattern. In practice, this approximation has consisted in setting a point where the CoM is supposed to be, namely, a 60 % of the body height, a point from which there exists considerable clinical evidence for being the candidate as the more probable CoM location (W.Zijlstra and A.L.Hof 1997; Gard et al. 2004). However, this method is particularly inaccurate in pathological situations in which the relative movements of the body segments are very distorted. On the other hand, a second strategy borrowed from optical methods consists in combining the CoM optical estimation for each of the body segments into a resultant CoM (Eames et al. 1999). The main drawback with this method is that the CoM location for each body segment is by itself a high non linear problem and therefore very inaccurate. Moreover, the error will be propagated systematically from each of these segment estimations because of the anatomical segment variability, as well as because the variable marker placement (Gard et al. 2004). Setting a convenient marker location may be particularly difficult in patients with high anatomical variability and unfeasible in patients using ortheses. In addition, movements of the body segments lead to relative displacements of the original marker locations, an intrinsic error which is almost impossible to avoid. Even worst, many pathologies are characterized by accentuated or attenuated movements and hence most marker placement protocols fail under these extreme conditions. This performance is improved by increasing the number of markers and so the CoM estimation, nonetheless a large number of markers alters the natural gait gesture. In despite that these optical methods are the more used in clinics because of their easy implementation and control, they are still very inaccurate in calculating with sufficient precision the CoM location (Gard et al. 2004) and therefore inadequate for many diagnosis and follow-up. The strategy herein proposed, although is also based on the marker measurements, it depends initially on the location of four markers, but above all upon the underlying gait dynamics which drives the movement. Moreover, the method introduces efficient mechanisms for tracking the gait trajectory involving non linear descriptions of the human movements. Among other advantages, this strategy requires a small number of markers (four for the present investigation) and can even be used in patients with ortheses since the sacral region is not invaded with any device, in these cases.

The trajectory described by the CoM during a gait cycle is a global indicator that correlates with the gait efficiency but that also can be used as a dynamical variable which complements the standard gait analysis (Detrembleur et al. 2000). The CoM temporal path is distorted in different movement abnormalities and associated with a degree of illness (Detrembleur et al. 2000). The CoM is considered as an efficient indicator for assessment of pathologies such as hemiplegia, paraplegia or distonia. The optimal walking, in terms of the energy, can be defined as the movement of the CoM from a place to another with minimum energy expenditure. Therefore, a pathological gait can be analyzed in terms of energy using the CoM change as the transfer of potential to kinetic energy (recovery) ,i.e., normal gait patterns loss a

40 % of this energy in this transfer, a higher lost is pathological (Detrembleur et al. 2000). Indeed a proper gait analysis should be based on an accurate estimation of the CoM positions and for doing so, the analytical description of the CoM movement pattern is required. The inverted pendulum system is an appropriate mechanism to representing the CoM movement in energy terms, describing the exchange between kinetic and potential energies within the different gait cycles (Buczek et al. 2006; Detrembleur et al. 2000; Cavagna and Kaneko 1977). Garcia's model has succeeded about representing the whole system in these terms (Garcia et al. 1998) and is able to describe normal patterns. The strategy presented here takes this model as a prior, adapts it to the observations and demonstrates its effectivity for tracking such patterns with no major concern on the anatomical variabilities, i.e. this hypothesis allows to assume a fixed leg length. Importantly, the strategy herein described could be easily used to follow pathological patterns by simply replacing this prior (see (Komura et al. 2005) for models that describe pathological patterns), or changing the weight given to the likelihood function so that even with this prior, the mapping of the observations to the states could follow the actual pattern. We claim that this is possible since a mixture of Gaussians has been already used in other problems (Pennec and Joshi 2008) for tracking non linear dynamics with comparable accuracy rates. The prior here is needed because it gives physical meaning and more importantly, clinical meaning, to the possible altered patterns and so quantification makes sense.

An important part of the routine clinical examination is the CoM estimation per cycle, a basic descriptor of many neuromuscular and musculoskeletal disorders. So far the common point, among the different approximations, has been a heuristic detection of the CoM location. Yet these approaches have coped with many clinical needs, no endeavours have been dedicated, up to now to developing tracking methods (Thirunarayan et al. 1996; Eames et al. 1999). The point is that any strategy, devised to uniquely achieve detection, is inefficient because it is based on noisy position mesurements and highly dependent on their inital marker location (Duff-Raffaele et al. 1996; Detrembleur et al. 2000). In the present work, we propose an efficient strategy for tracking CoM locations, using a non-linear gait biomechanical model whose parameters are recursively adjusted by a Bayesian filter, implemented as a particle filtering. Implementation of any Probabilistic function distribution (*pdf*) can be achieved using several techniques, namely, extended Kalman filtering, multihypothesis tracking, grid-based or topology-based representations and particle filtering (Fox et al. 2003; Arulampalam et al. 2002). The Extended Kalman Filter approaches the system dynamic by a first order Taylor series expansion. This filter is useful if the state uncertainty is not too high, i.e., measures come from accurate sensors, which is obviously not the case for CoM tracking. The multihypothesis tracking represents the belief as a mixture of Gaussians and tracks each with a Kalman filter. This technique is computationally expensive and requires complicated heuristics to determine when to add or delete Gaussians. Additionally, the heelstrike rule, herein used for the dynamic model, introduces angular discontinuities that would require many Gaussians whose number would be impossible to establish beforehand. Grid-based approaches stand for a piecewise constant representations of the belief. This approach approximates discontinuities by refining the resolution grid cells, and therefore expensive computational methods are required when discontinuities are present. This grid complexity could be approached by topological representations, corresponding to a graph where each node is related to a state and each edge to the environments connectivity, but the computational cost is even higher. In contrast, each particle of a particle filtering can easily follow any discontinuity, with a low number of particles and minimum computational

cost. Moreover, the particle filter does not require accurate measurements because the particle weights are modified proportionally to their likelihood.

The whole strategy allows a natural tracking of the non-linear gait patterns with a high degree of noise robustness, under a non-linear minimal square criterion. Accuracy and efficiency were herein assessed under several different metrics. Firstly, under RMSE metric the particle filtering is a 56% smaller in the x-axis and 59% in the y-axis, when comparing with other two conventional tracking strategies. Also, correlation was assessed using the correlation coefficient and again the particle filtering outperformed the other two from 0.9 to 0.7 in the x-axis and from 0.9 to 0.6 in the y-axis. Finally, the SNR-like measurement permitted to measure the quality of the tracked signal among the different strategies under the Gaussian noise conditions, introduced from different independent sources. On average, measurements showed that the trajectory generated by our model was $0.86\,db$ higher, when comparing with conventional methods, indicating a large gain in accuracy. Importantly, the largest gain difference was observed in the heelstreak phase ($1.5\ db$), demonstrating the capacity of the proposed method to follow the non linear gait patterns. Finally, the method was computationally implemented through a particle filter because of the great advantages of this implementation when comparing with others (Fox et al. 2003), among others: accuracy, robustness, efficiency and easy implementation. The particle filter is herein used to estimate the gait Bayesian model. The Sequential Importance Resampling implementation of the particle filter approximates the *pdf* by a weighted set of particles whose importance is constantly evolving with the dynamics we introduced and then updated after system observations are available. This implementation, as shown by our results, is efficient in terms of the computational time and enough accurate as to follow the gait non linear patterns so that on-line data analysis is possible along with the routine capture of the other variables which compose an entire gait analysis.

6. Conclusions and Perspectives

This work presented a general framework for tracking complex human movements. The whole strategy consists in simulating the dynamics of the system, using some priori information of the particular problem. Simulation requires a discrete system model and an observation process. The model approximates the truth hidden system states, while the observation process adapts to non-linear dynamics using the Bayes rule, implemented as a particle filter. The method was successfully assessed. The presented procedure could be extended to other type of medical imaging problems, under the restriction that there exists a proper knowledge of the problem so that analytical or parametric expressions may be found.

References

Arulampalam S, Maskell S, Gordon N. 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing 50:174–188.

Buczek FL, Cooney KM, Walker MR, Rainbow MJ, Concha MC, Sanders JO. 2006. Performance of an inverted pendulum model directly applied to normal human gait. Clinical Biomechanics 21(3):288 – 296.

Carson M, Harrington M, Thompson N, O'Connor J, Theologis T. 2001. Kinematic analysis of a multi-segment foot model for research and clinical applications: a repeatability analysis. Journal of Biomechanics 34:12991307.

Cavagna GA, Kaneko M. 1977. Mechanical work and efficiency in level walking and running. Physiology 268:467–481.

Crowder SV. 1987. American Statistical Association and American Society for Quality; A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts. Technometrics 29(4):401–407.

Deluzio KJ, Urs P. Wyss BZ, Costigan PA, Serbie C. 1997. Principal component models of knee kinematics and kinetics: Normal vs. pathological gait patterns. Human Movement Science 16:201–217.

Detrembleur C, van denHecke A, Dierick F. 2000. Motion of the body centre of gravity as a summary indicator of the mechanics of human pathological gait. Gait Posture 12:243–250.

Dong H, Pelah A, Cameron J, Lasenby J. 2008. New York, NY, USA: ACM; The perceptual influences on gait transition of step parameters and optic flow in virtual environment locomotion simulators pp. 143–146.

Duff-Raffaele, Kerrigan, Corcoran, Saini. 1996. The proportional work of lifting the center of mass during walking. Am J Phys Med Rehabilit 75:375–379.

Eames M, Cosgrove A, Baker R. 1999. Comparing methods of estimating the total body centre of mass in three-dimensions in normal and pathological gaits. Human Movement Science 18:637–646.

Fox D, Hightower J, Liao L, Schulz D, Borriello G. 2003. Los Alamitos, CA, USA: IEEE Computer Society; Bayesian Filtering for Location Estimation. IEEE Pervasive Computing 02(3):24–33.

Fox V, Hightower J, Liao L, Schulz D, Borriello G. 2003. Bayesian filtering for location estimation. Pervasive Computing, IEEE 2(3):24–33.

Gace JR. 2001. Gait Analysis in Cerebral Palsy. NY: MacKeith Press.

Gace JR. 2004. The treatment of gait problems in cerebral palsy. Mac Keith Press.

Garcia M, Chatterjee A, Ruina A, Coleman M. 1998. The simplest walking model: Stability, complexity, and scaling. ASME Journal of Biomechanical Engineering 120:281–288.

Gard SA, Miff SC, Kuo AD. 2004. Comparison of kinematic and kinetic methods for computing the vertical motion of the body center of mass during walking. Human Movement Science 22(6):597 – 610.

Goswami A, Thuilot B, Espiau B. 1998. A Study of the Passive Gait of a Compass-Like Biped Robot. Journal of Robotics Research 17:1282–1301.

Gutierrez EM, saBartonek, Haglund-kerlind Y, Saraste H. 2003. Centre of mass motion during gait in persons with myelomeningocele. Gait & Posture 18(2):37 – 46.

Hunter JS. 1986. The exponentially weighted moving average.. J. Quality Technol. 18:203 –210.

J. M, G. C, C.G. A, Zeglin G. 2004. A simple reinforcement learning algorithm for biped walking. Proceedings of the 2004 IEEE. International Conference on Robotics & Automation 3:3030 – 3035.

Jia X, Wang R, Zhang M, Li X. 2008. Influence of Prosthetic Sagittal Alignment on Trans-Tibial Amputee Gait and Compensating Pattern: A Case Study. Tsinghua Science & Technology 13(5):581 – 586.

Kay SM. 1993. Fundamentals of Statistical Signal Processing: Estimation Theory. Secaucus. NJ, USA.

Komura T, Nagano A, Leung H, Shinagawa Y. 2005. Simulating pathological gait using the enhanced linear inverted pendulum model. Biomedical Engineering, IEEE Transactions on 52(9):1502–1513.

Kuo A. 2007. The six determinants of gait and the inverted pendulum analogy : A dynamic walking perspective. Human movement science 26:617–656.

Manal KT, Buchanan TS. 2004. McGraw-Hill; Biomechanics of human movement.. In Standard Handbook of Biomedical Engineering & Design p. 26.

McGeer T. 1990. Passive Dynamic Walking. International Journal of Robotics Research 9:62–82.

Pennec X, Joshi S. 2008. New York, USA: Geometrical and Statistical Methods for Modelling Biological Shape Variability. Proceedings of the Second International Workshop on Mathematical Foundations of Computational Anatomy .

Pinzur MS, Cox W, Kaiser J, Morris T, Patwardhan A, Vrbos L. 1995. The Effect of Prosthetic Alignment on Relative Limb Loading in Persons with Trans-tibial Amputation : A Preliminary Rep. Journal of Rehabilitation Research and Development 32:373–378.

Sigal L, Black. MJ. 2006. Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion.. , Dept. of Computer Science, Brown University. Technical report.

Thirunarayan MA, Kerrigan DC, Rabuffetti M, Croce UD, Saini M. 1996. Comparison of three methods for estimating vertical displacement of center of mass during level walking in patients. Gait & Posture 4(4):306 – 314.

Tucker CA, Ramirez J, Krebs DE, Riley PO. 1998. Centre of gravity dynamic stability in normal and vestibulopathic gait. Gait and Posture 8:117–123.

Whittle M. 1998. Three-dimensional motion of the center of gravity of the body during walking. Human Movement Science 16:347–355.

Whittle MW. 1996. Clinical gait analysis: A review. Human Movement Science 15:369–387.

W.Zijlstra, A.L.Hof. 1997. Displacement of the pelvis during human walking: empirical data and model predictions. Gait and Posture 6:249–262.

Zajac FE, Neptune RR, Kautz SA. 2003. Biomechanics and muscle coordination of human walking Part II: Lessons from dynamical simulations and clinical implications. Gait & Posture 17:1–17.

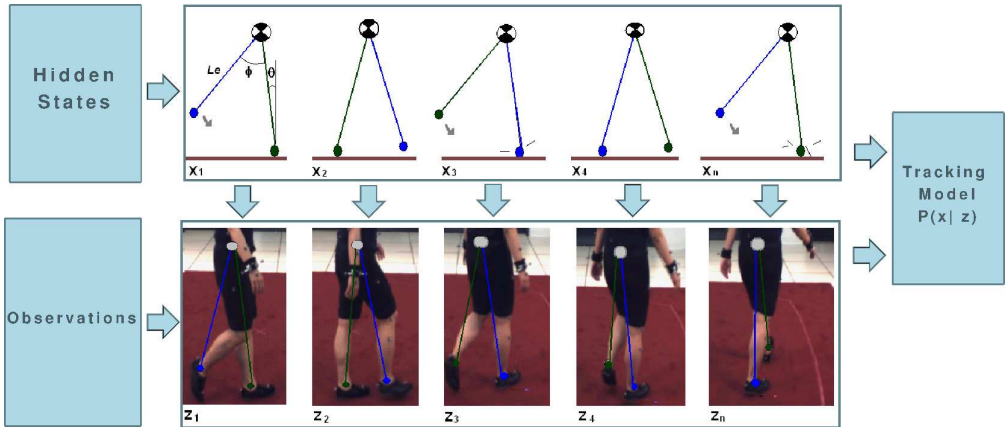
Figure Captions

Figure 1: Sequence of events of a complete gait cycle extracted from the humanEva database (Sigal and Black., 2006).

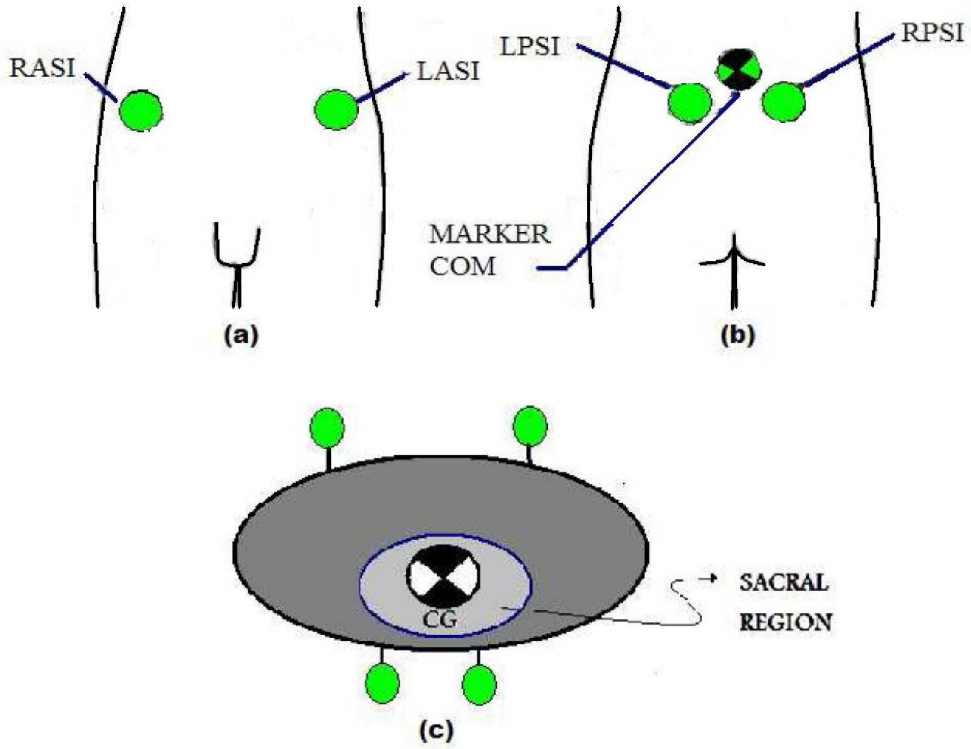
Figure 2: Marker locations in the sacral region.

Figure 3: The gait trajectory has been divided into its y (left) and x (right) components for better analysis. The y pattern is characterized by a non linear periodic path (Thicked line) which is closely tracked by the Bayesian strategy (squared dots) and hardly matched by the other two methods (dashed lines). The x plot complements the whole picture, the marker CoM and EWMA methods are shifted away from the control path, which is closely followed by the Bayesian strategy. The Bayesian method highly outperforms the two other methods, regarding non linear adaptation.

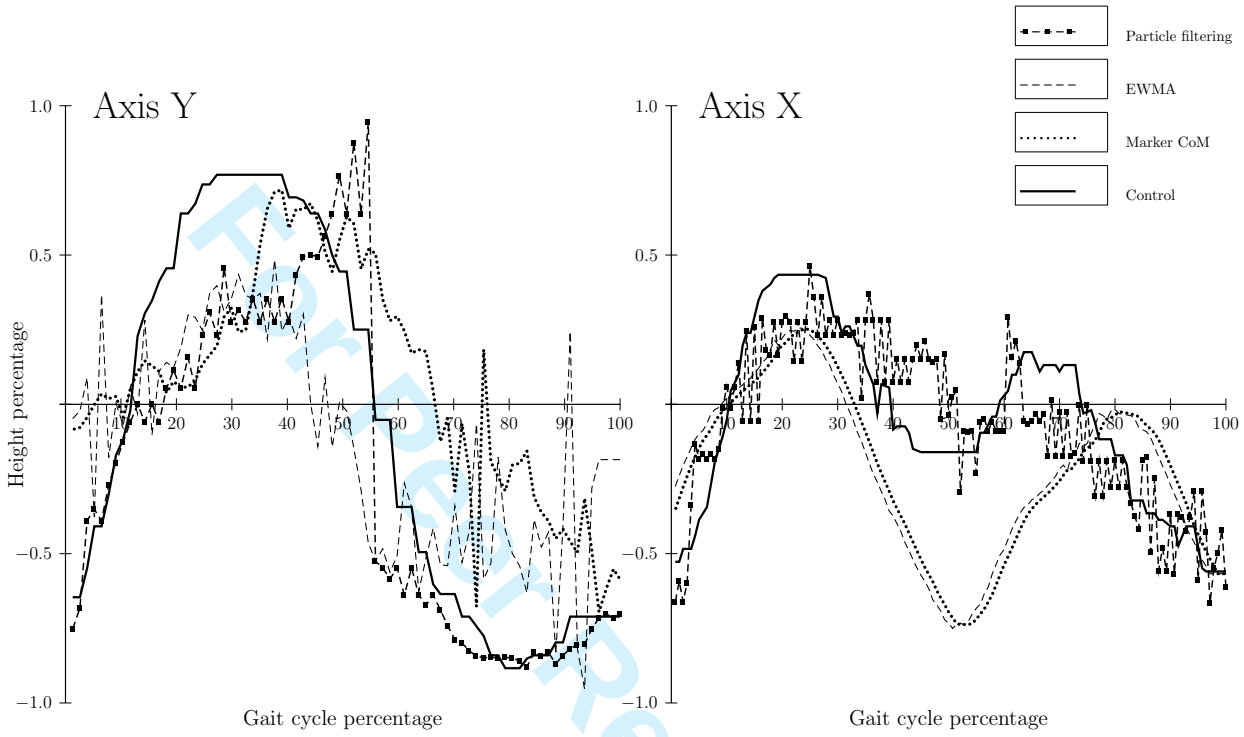
Figure 4: The SNR-like measurements assess the fidelity to the control data along the best cycle for each of the methods. The graph shows in four panels different evaluations: upper right panel displays the three methods when no specific weight is given to the vertical component i.e. there is no α while the other three panels show different α values, set to 0.7, 0.8 and 0.9 for upper-right, bottom-left and bottom-right panels, respectively. Notice that the Bayesian tracking outperforms the other two methods in about 1 to 3 decibels for the four different comparisons, indicating a higher quality of the prediction of the control data.



Sequence of events of a complete gait cycle extracted from the humanEva database (Sigal and Black., 2006)
633x267mm (600 x 600 DPI)



Marker locations in the sacral region.
341x266mm (600 x 600 DPI)



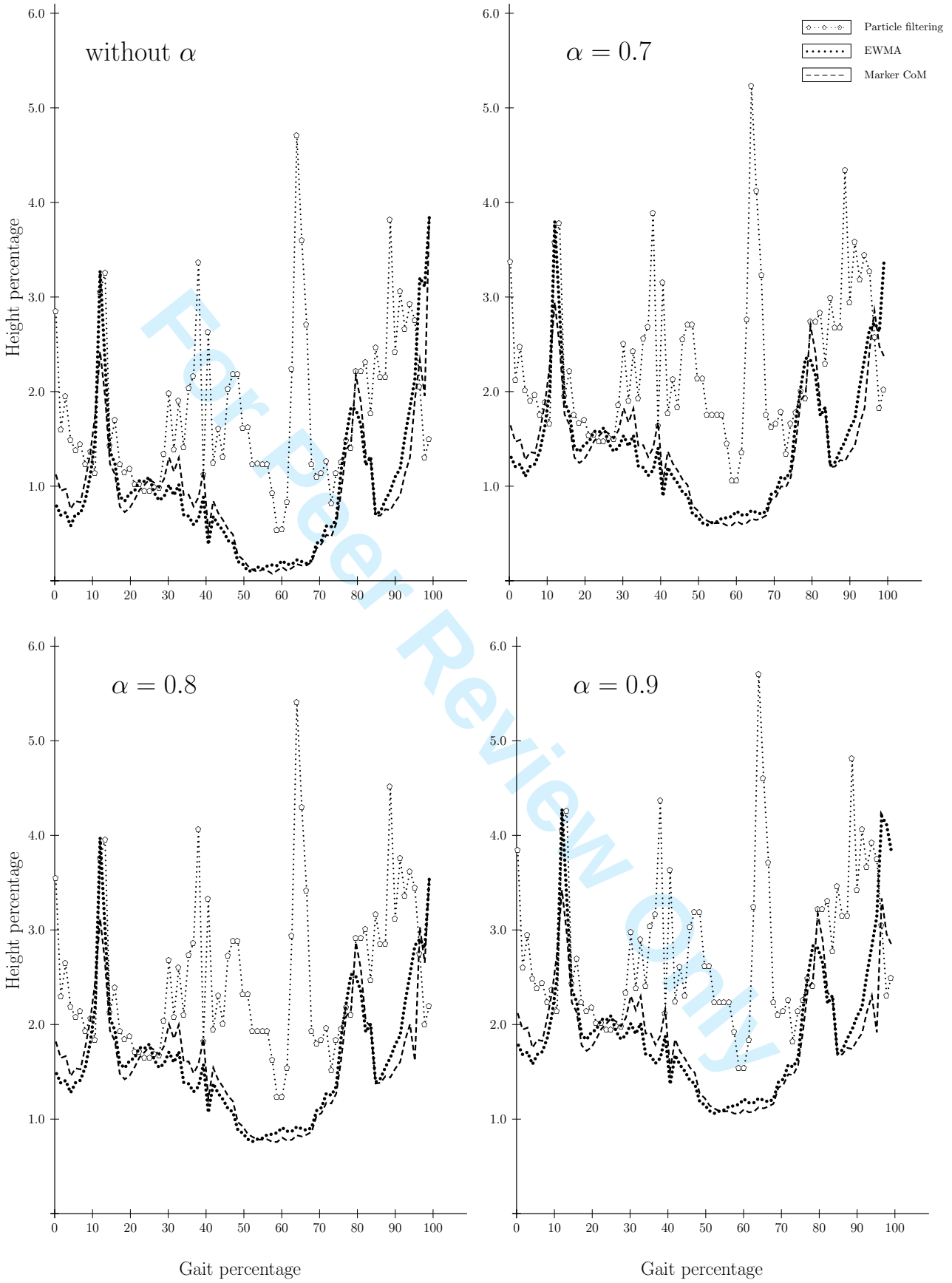


Table 1. Root Mean Squared Error and its tracking deviation for the three evaluated methods. These averages have been calculated from the differences between the control and any of the predicted data, expressed as percentage height.

Tracking Method	x-axis	y-axis
Marker CoM	0.414 ±0.05	0.4 ±0.081
EWMA	0.38 ± 0.05	0.41 ± 0.079
Bayesian tracking	0.232 ± 0.021	0.236 ± 0.028

Table 2. correlation factor of the differences between the control and the three assessed methods

Tracking Method	x-axis	y-axis
Marker CoM	0.54 $\pm$0.05	0.7849 $\pm$0.081
EWMA	0.56$\pm$ 0.05	0.7584 $\pm$ 0.079
Bayesian tracking	0.9 $\pm$ 0.021	0.91$\pm$ 0.028

Table 3. The mean and standard deviation of the SNR-like mesurement for a total of sixty cycles available in the present investigation

Tracking Method	without α	$\alpha = 0.9$	$\alpha = 0.8$	$\alpha = 0.7$
Marker CoM	1.35 ± 0.35	2.34 ± 0.39	2.04 ± 0.39	1.86± 0.36
EWMA	1.33 ± 0.46	2.32 ± 0.46	2.03 ± 0.46	1.85± 0.47
Tracking Bayes	2.17 ± 0.5	3.17 ± 0.5	2.86 ± 0.5	2.7± 0.5

Table 4. Accuracy index or the averaged RMSE with its standard deviation for different numbers of particles from the data of vertical and horizontal directions. The running time is included for illustrating the computational cost of any of them.

Particle	x-axis	y-axis	Duration ms
10	0.451 ± 0.022	0.423 ± 0.039	170
100	0.281 ± 0.023	0.29 ± 0.028	810
1000	0.253 ± 0.025	0.252 ± 0.031	7100

Table 5. Efficiency index for both dimensions and running time for a number different of particles.

Particle	x-axis	y-axis
10	0.02673	0.15082
100	0.05367	0.044091
1000	0.0056	0.00454

Rotation Invariant Texture Characterization using a Curvelet Based Descriptor

F.A. Gómez[a], E. Romero[a,*]

[a] *Bioingenium Research Group*
Medicine Faculty - University of Colombia, Bogotá DC - Colombia

Abstract

This paper introduces a highly discriminative, precise and simple descriptor of natural textures based on the curvelet transform. The proposed descriptor captures edge information from the statistical pattern of the curvelet coefficients. The image is mapped to the curvelet space and a statistical parametric model is calculated from each of the sub-bands, aiming to approach the subband marginal distribution. Once these parameters are estimated, they are subband energy sorted out, achieving so the invariance to planar rotations. Finally, the Kullback−Leibler divergence between the statistical parameters is used to estimate a distance between images. We demonstrated the effectiveness of the proposed descriptor for classification and retrieval tasks, obtaining significant improvements.

Keywords: texture characterization, rotation invariance, curvelet transform, generalized Gaussian distribution, Kullback-Leibler divergence

1. Introduction

The capacity of a mapping to generate features with highly discriminant textural characteristics, results of paramount importance for the problem of classification and/or retrieval. Typical applications include microscopical or satellite images [1]. Formally, the feature extraction process is thought of as a mapping of an image collection to a characteristic space, which provides a representation where similar images are close and different images are far, i.e., this property is known as the discriminating space power. Images projected onto this space are characterized by features which capture some image properties, typically statistical data properties. In the particular case of textures, the most popular characteristic spaces are currently the discrete cosine, wavelets, Gabor transforms [2, 3]. Unfortunately, these spaces are sub-optimal for this problem because textures are naturally entailed with geometrical, scale and directional properties which are poorly described with these transforms [4]. Some of the features, already used in this problem, capture information of the energy coefficient distribution and include the total energy, the mean and the variance [2]. These features result insufficient to capture the statistical properties of natural images [5]. In addition, the success of any comparison between images depends on the metrics one selects for the specific problem. The usual metrics is either Euclidian or an estimation of statistical dependence such as the Kullback−Leibler divergence [6]. In these terms, the problem of texture characterization consists in constructing a feature with high discriminative power that takes into account the statistical image contents.

The problem of texture characterization with curvelets was already addressed by Semler [7], who studied the performance of several features, namely: the energy, entropy, mean and standard deviation of the curvelet subbands. Results showed significant improvement when comparing with wavelets, but this characterization did not take into account the particular statistical patterns of the curvelet coefficients [8]. Sumana [9] also proposed the curvelet subband mean and variance as features and the Euclidian distance as similarity measurement. Results showed again improvement, when comparing with

*Corresponding author
Email address: edromero@unal.edu.co (E. Romero)

Gabor features. However, texture curvelet subbands are not described by simple Gaussians so that mean and variance result insufficient to describe the observed distribution [8].

In this paper we present a new global descriptor, entailed with the previously described properties. The curvelet space is used to capture information about edges, which are in fact one of the most discriminating features [10]. These features are the moments of a generalized Gaussian density (GGD), a good approximation to the marginal curvelet subband distribution [8], whilst the Kullback−Leibler divergence estimates differences between curvelet coefficient distributions. A main contribution of this paper is that an entire statistical characterization of the curvelet coefficient, results in a highly discriminative, precise and simple descriptor of natural textures. The rest of this paper is organized as follows: Section materials and methods introduces the new feature, Section Results demonstrates the effectiveness of this descriptor in classification and retrieval tasks. Finally, the last Section concludes with a discussion and future work.

2. Materials and methods

Two input images are curvelet-represented and their frequency subbands are statistically characterized, using the moments of a GGD. Invariance to planar rotation is obtained via the circular shifting process [11], based on the subband energies. Finally, the Kullback-Leibler divergence computes a distance estimation between the two representations. This strategy will be further explained hereafter:

2.1. The curvelet transform

The curvelet transform is a multiscale decomposition, developed to naturally represent objects in 2D, improving the wavelet limitations for representing geometrical information [12]. Curvelets are redundant bases which optimally represent 2D curves. Besides the usual information about scale and location, already available from a wavelet, each of these frame elements is able to capture orientation information.

A curvelet can be thought of as a radial and angular window in the frequency domain, defined in a polar coordinate system, upon which the different scales are represented as different rings with different level of frequential detail from the inner (low frequencies) to the outer (high frequencies) rings. This representation is constructed as the product of two windows: the angular and the radial dyadic frequential coronas. The angular window provides a directional analysis and the radial dyadic window is a bandpass filter, used to analyze image details at different scales (see Figure 1). Frequency cuts in both windows are selected, following the parabolic anisotropic scaling law $width \approx length^2$ (see Figure 1). The motivation behind this selection is to efficiently approximate a smooth discontinuity curve by "laying on" basis elements with elongated supports along the curve [12]. Curvelet bases were designed to fully cover the frequency domain, in contrast to other directional multiscale representations such the Gabor transform, case in which some information is always lost. Thanks to the anisotropic scale, curvelets adapt much better to scaled curves than Gabor transform, improving the representation at different scales and noise robustness [13]. All these statements have been experimentally demonstrated by comparing wavelets, curvelets and Gabor in retrieval tasks [9]. The Figure 1 shows a curvelet multiscale decomposition example.

2.2. Statistical characterization

Psychophysical research has demonstrated that two homogeneous textures are not discriminable if their marginal subband distributions are alike [10], i.e., the frequency subband distributions have a highly descriptive capacity, at least for the texture problem. This discriminative power was also experimentally verified for wavelet and Gabor representations [2]. In the curvelet case, each subband contains information about the degree of occurrence of similar curves within the image, that is to say, edge energy levels with similar direction and size. Figure 2 shows a typical example of the curvelet coefficient histogram of an image subband. The kurtosis in this case is about 7.4 so that a Gaussian density is not enough as to match the observed energies. Therefore, the mean and variance calculated from a Gaussian, used in a previous works [7, 9] will have a very poor descriptive capacity. In general, the curvelet coefficient distribution in natural images is characterized by a sharper peak centered at zero with symmetrical smooth tails. This shape is associated to the sparse property of this transformation, i.e., relatively few large coefficients capture most of the information. This leptokurtic pattern has been previously observed in curvelets [8, 14]

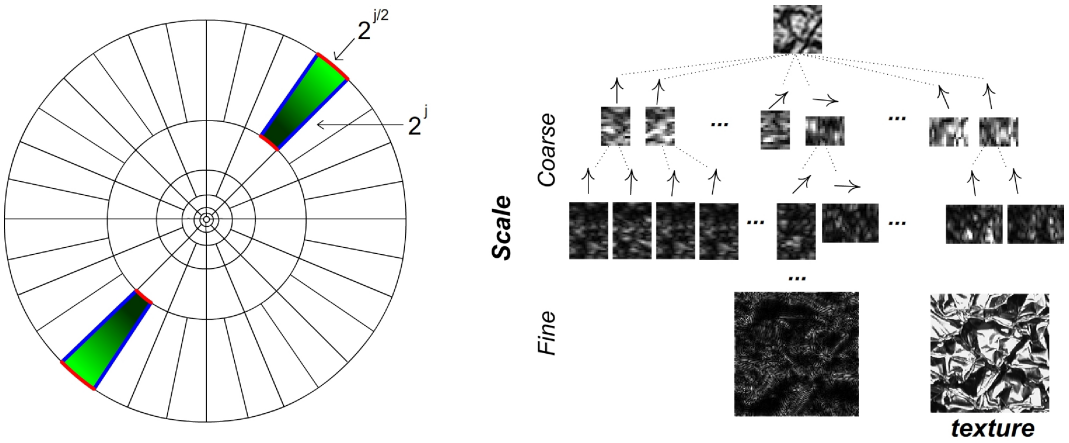
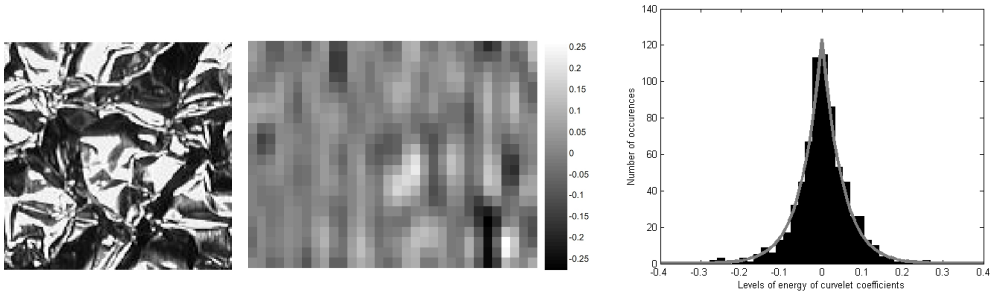


Figure 1: Curvelet transform. Left, continuous curvelet tilling generated by the product of angular and radial coronas, frequency cuts are selected to satisfy the anisotropic scaling law. Rigth, illustration of a curvelet texture decomposition: from top to bottom, increasing levels of detail, from left to right, different orientations.



(a) Texture. (b) Curvelet Subband (c) Curvelet histogram.

Figure 2: Curvelet histogram example (scale 3 and orientation 16).

as well as in wavelets [15]. This work proposes a texture characterization via the marginal distribution of the subband curvelet coefficients, specifically using the parameters of a generalized Gaussian density. Recent experimentation in natural images [8] shows that the generalized Gaussian density provides a good adjustment to the marginal density of the curvelet coefficient, within each subband. The GGD reads as $p(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^{\beta}}$, where $\Gamma(z) = \int_0^{\infty} e^{-t}t^{z-1}dt$, $z > 0$ is the Gamma function, α is the variance and β is related to the decreasing rate of the GGD. The parameters α and β are estimated from the subband data using Maximum Likelihood, as detailed in [15]. These parameters (α,β) are herein used as descriptor of the probability density function of the energy levels inside each curvelet subband.

2.3. Rotation Invariance

Previous texture characterizations have failed when the image is rotated, basically because similar textures with different orientations have very different statistical subband moments. Some works [11, 16] have tried to overcome this limitation by using the curvelet rotation shifting property, that establishes that the curvelet subbands of a rotated image are a shifted version of the original subbands. These approaches have independently performed a circular shifting on each scale level, assuming that the energy of the dominant orientation usually spreads between two neighboring subbands. Nevertheless, our experiments on the Brodatz database rapidly drive us to the conclusion that this statement was true only for some patterns. Indeed, orientation information is not homogenously distributed between the different scale levels and its calculation is not therefore independent. We decided then to com-

pute as the dominant orientation, the sum over all the subbands at the different scales, and the circular shifting is ordered using as reference the first curvelet level.

2.4. Similarity measure

The similarity between subband curvelets is estimated using the Kullback-Leibler divergence (KLD) of the corresponding GGDs:

$$D(p(.;\alpha_1,\beta_1)||p(.;\alpha_2;\beta_2)) = \log\left(\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)}\right) + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma((1/\beta_1)} - \frac{1}{\beta_1}$$

where (α_1,β_1) and (α_2,β_2) are the GGD parameters estimated for each subband. This estimation needs not additional normalization and shows good performance under other multiscale domains [15]. Finally, assuming independence of different subbands, the similarity between two images I_1 and I_2 is the sum of the divergences between corresponding subbands $D(I_1,I_2) = \sum_{\forall s}\sum_{\forall\theta} D(p(.;\alpha_1^{s,\theta};\beta_1^{s,\theta})||p(.;\alpha_2^{s,\theta};\beta_2^{s,\theta}))$, where $(\alpha_1^{s,\theta},\beta_1^{s,\theta})$ and $(\alpha_2^{s,\theta},\beta_2^{s,\theta})$, are the GGD parameters estimated for corresponding subbands, i.e., subbands in the same scale s and orientation θ.

3. Results

Provided that our main goal was to assess the discriminative power of the curvelet descriptor, the feature performance was assessed in both a multiclass classification and a retrieval problem. For doing so, we used two databases, the KTH-TIPS [1] image texture and the Brodatz databases. The former was used to evaluate the descriptor capacity in the multiclass classification problem. This database provides several variations of scale, pose and illumination, increasing the intra-class variability and reducing the inter-class separability and then augmenting the difficulty of the classification task [17]. In this database, there are ten texture categories: sandpaper (*sn*), aluminium foil (*af*), styrofoam (*sf*), sponge (*sp*), corduroy (*cd*), linen (*ln*), cotton (*ct*), brown bread (*bb*), orange peel (*op*), cracker (*cr*). These real world images come from different natural scenes and have different poses and scales. For our experiments, 45 images of each category in nine scales were converted to gray-scale levels (computed from the luminance component) and cropped to 128×128. Figure 3 displays examples of the original textures. A second image collection was used to test the planar rotation invariance property in a retrieval experiment. It consists of sixteen 512×512 Brodatz texture images that were rotated to various degrees before being digitized. From these images, we constructed the rotated image set by taking nine non-overlapping 128×128 from the original image at 0, 30, 60, 90 and 120 degrees. The database contains 720 (9×5) images that come from 16 texture classes: D3 (*reptile*), D9 (*grass*), D16 (*weave*), D19 (*wool*), D21 (*canvas*), D24 (*leather*), D29 (*sand*), D53 (*cloth*), D55 (*matting*), D57 (*paper*), D64 (*rattan*), D68 (*wood*), D77 (*cotton*), D78 (*straw*), D84 (*raffia*) and D92 (*pig skin*). Figure 4 displays examples of the original textures.

A real digital curvelet transform with 4 scales and 32 orientations was used, resulting in 66 subbands. The coarsest curvelet level was excluded in order to obtain robustness to changes in illumination. The algorithms were written in Matlab and run on a Intel Xeon X5460 Quad-Core 3.16 *GHz* with 8 *GB* in RAM. The discriminative capacity of the proposed descriptor was assessed in a multiclass problem using the most simple classifier, a nearest neighbor, and compared with other curvelet representation methods, namely: energy of the curvelet subband plus Euclidian metric [7, 9] (energy based feature), mean and variance plus Euclidian metric [9] (mean-variance feature) and the herein described proposal GGD plus KLD metric (GGD based feature). Sumana [9] has previously compared Gabor, wavelets and curvelets, obtaining a better performance for the latter so that our work was focused on characterizing the curvelets. The three classifiers were tested under a leave-one-out cross-validation and the corresponding confusion matrix was calculated. In the retrieval experiment, each image in the database was used as a simulated query. The relevant images for each query are defined as the set of images which belong to the same class of the tested image. The performance was evaluated in terms of the percentage of relevant images from 15 retrieved images [18] and also calculated for other curvelet features, rotation invariant, namely: circular shifting either with independent frequency energy [11] or mean-variance frequency [16] as descriptors, while the Euclidian metric was used as

[1] http://www.nada.kth.se/cvap/databases/kth-tips/index.html

similarity measurement.

		Assigned										Total	% Agree
		sn	*af*	*sf*	*sp*	*cd*	*ln*	*ct*	*bb*	*op*	*cr*		
True	*sn*	21	0	14	7	0	0	0	2	1	0	45	47
	af	0	42	1	0	0	0	0	1	1	0	45	93
	sf	5	1	35	3	0	0	0	0	0	1	45	78
	dp	2	0	1	38	2	0	0	1	1	0	45	84
	cd	0	0	0	2	31	0	0	6	6	0	45	69
	ln	0	0	0	0	1	42	0	0	1	1	45	93
	ct	0	0	0	1	0	0	40	3	1	0	45	89
	bb	2	0	0	1	1	0	0	36	4	1	45	80
	op	0	0	1	1	2	0	0	2	27	12	45	60
	cr	0	0	1	0	0	0	0	1	2	41	45	91
	Total	30	43	53	53	37	42	40	52	44	56	450	**78**

Table 1: Confusion matrix for the energy based feature.

		Assigned										Total	% Agree
		sn	*af*	*sf*	*sp*	*cd*	*ln*	*ct*	*bb*	*op*	*cr*		
True	*sn*	29	0	9	5	0	0	0	2	0	0	45	64
	af	0	41	1	0	0	0	0	1	0	2	45	91
	sf	4	0	38	2	0	0	0	0	0	1	45	84
	sp	3	0	0	39	2	0	0	0	1	0	45	87
	cd	0	0	0	2	31	0	0	7	5	0	45	69
	ln	0	0	0	0	1	44	0	0	0	0	45	98
	ct	1	0	0	0	2	0	41	1	0	0	45	91
	bb	2	0	0	2	1	0	0	35	5	0	45	78
	op	0	0	0	1	2	0	0	1	36	5	45	80
	cr	0	0	0	0	0	0	0	0	1	44	45	98
	Total	39	41	48	51	39	44	41	47	48	52	450	**84**

Table 2: Confusion matrix for the mean-variance based feature.

		Assigned										Total	% Agree
		sn	*af*	*sf*	*sp*	*cd*	*ln*	*ct*	*bb*	*op*	*cr*		
True	*sn*	31	0	4	5	0	0	0	5	0	0	45	69
	af	0	45	0	0	0	0	0	0	0	0	45	100
	sf	3	0	38	1	0	0	0	0	0	3	45	84
	sp	2	0	0	38	2	0	0	3	0	0	45	84
	cd	0	0	0	2	32	0	0	6	3	2	45	71
	ln	0	0	0	0	0	44	0	0	1	0	45	98
	ct	0	0	0	0	2	0	43	0	0	0	45	96
	bb	4	0	0	2	0	0	0	39	0	0	45	87
	op	1	0	1	1	0	0	0	0	42	0	45	93
	cr	0	1	0	0	1	0	0	1	1	41	45	91
	Total	41	46	43	49	37	44	43	54	47	46	450	**87**

Table 3: Confusion matrix for the GGD feature.

The confusion matrices for the classification experiment are shown in Tables 1, 2 and 3. The correct classification varies, overall, as 78%, 84% and 87%, showing a high discriminative capacity. The curvelet descriptor shows a better classification rate for both the average and the individual classification, as observed in table 3. Note that textures *linen (ln)* and *cotton (ct)* present a high density of lines and are correctly classified in a large number of cases. Likewise, texture *aluminium foil (af)*, which presents gross edges, is correctly classified using the curvelet descriptor. Finally, the confusion matrices show that misclassifications occur mainly in similar textures, for example, *sandpaper (sn)* and *styrofoam (sf)*, probably because of a similar edge distributions. Importantly, the curvelet descriptor shows much less classification errors, showing that in textures with higher levels of variability, the proposed method outperforms the previous approach. In the second series of experiments, we tested the rotational invariant property of the proposed curvelet descriptor. Figure 5 shows the performance comparison among the three methods previously described (see Section 2). The proposed rotational invariant method highly outperforms the other two methods. We see that textures as *grass*, *wool* and *sand* are not significantly affected by rotation, because they have no strong directional information, resulting in a similar performance. On the contrary, in textures with strong directional information (*wood*, *straw*), the proposed method results in a better representation. Finally, the average retrieval for the proposed method improved in a 17%, when compared with the previous rotational invariant methods. This improvement is consistent with the strong directional information in these textures. Regarding the computational complexity, the curvelet implementation runs in $O(n^2 \log(n))$ for $n \times n$ Cartesian arrays [12], less than 300 ms for each image, while the statistical characterization for the curvelet subbands runs in less that 1 second.

4. Conclusions and Discussion

We have introduced a new texture image descriptor, based on the curvelet transform and a statistical model of the frequency distribution of the curvelet coefficients in natural images. By applying the curvelet transform and adjusting the levels of energy for each subband to a generalized Gaussian model, we obtain a robust representation which captures the edge distribution at different orientation and scales. Experimental results indicate that the new feature improves the classification performance in a multiclass classification problem when compared with other features, also based on curvelets. Likewise, the descriptor has shown to improve retrieval tasks in actual texture applications. Future works includes the feature scale invariance and extensive experimentation in large texture databases.

References

[1] R. Valerie, Introduction to Texture Analysis: Macrotexture, Microtexture and Orientation Mapping, CRC Press, Amsterdam, 2000.

[2] T. Randen, J. Hå, IEEE Trans. Pattern Anal. Mach. Intell. 21 (1999) 291–310.

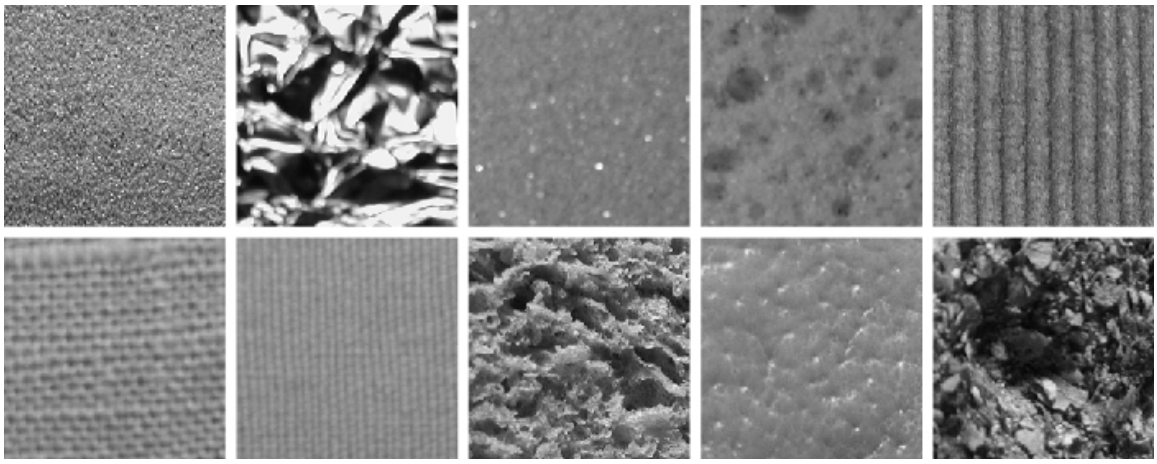


Figure 3: Texture images from the KTH-TIPS database colection. From left to rigth and top to bottom: *sandpaper, aluminium foil, styrofoam, sponge, corduroy, linen, cotton, brown bread, orange peel, cracker*

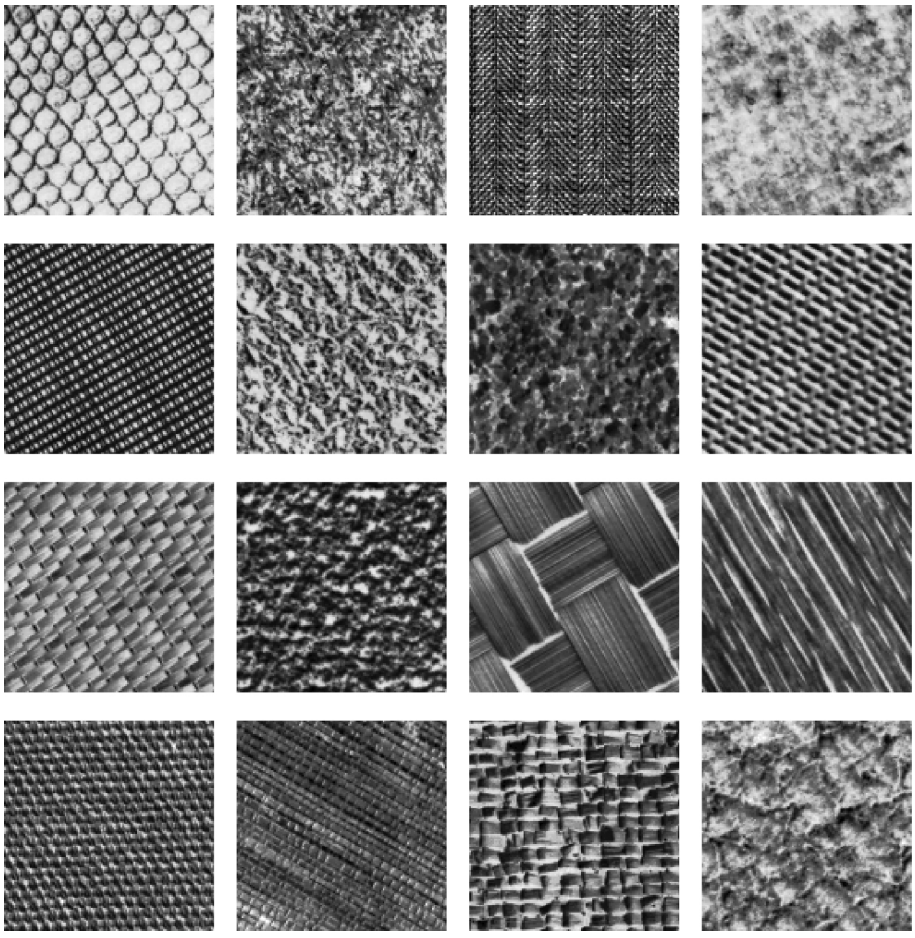


Figure 4: Texture images from the Brodatz rotated collection. From left to right and top to bottom: *reptile, grass, weave, wool, canvas, linen, leather, sand, cloth, matting, paper, rattan, wood, cotton, straw, pig skin*

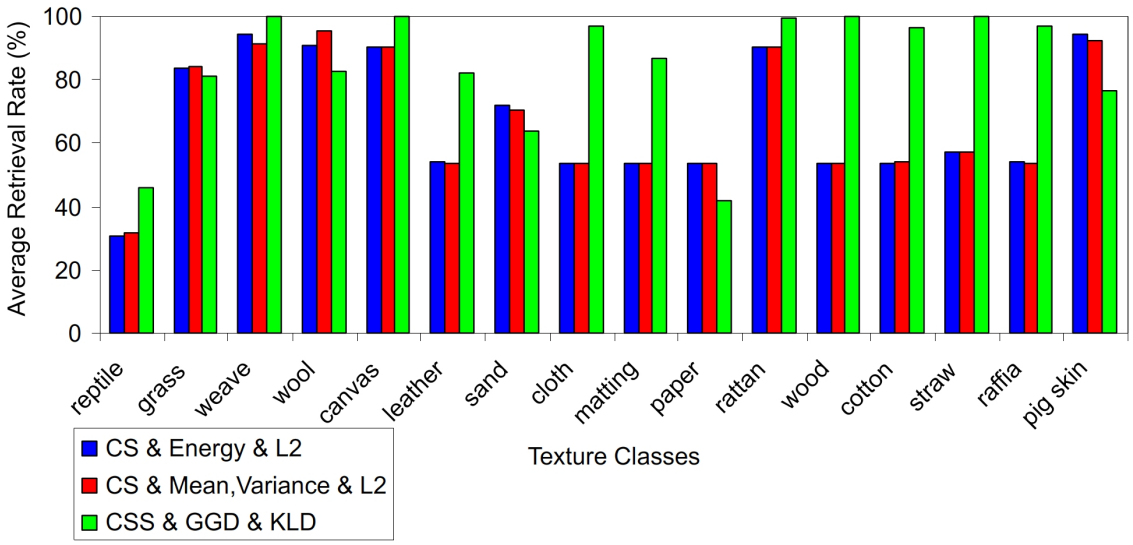


Figure 5: Average retrieval rate for the energy based feature (blue), the mean-variance based feature (red) and the GGD based feature (green)

[3] M. Petrou, P. García Sevilla, Image processing : dealing with texture, John Wiley & Sons Inc, Chichester, 2006.

[4] G. Welland, Beyond Wavelets, Academic Press, San Diego, CA, 2003.

[5] G. V. D. Wouwer, P. Scheunders, D. V. Dyck, IEEE Transactions on Image Processing 8 (1999) 592–598.

[6] S. Kullback, The American Statistician (1987) 340–341.

[7] L. Dettori, L. Semler, Comput. Biol. Med. 37 (2007) 486–498.

[8] A. Alecu, A. Munteanu, A. Pizurica, W. P. J. Cornelis, P. Schelkens, pp. 1617–1620.

[9] I. Sumana, M. Islam, D. Zhang, G. Lu, in: Proc. of IEEE International Workshop on Multimedia Signal ProcessingMMSP, pp. 11–16.

[10] D. J. Field, Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems, Elsevier-Health Sciences Division, pp. 151–193.

[11] L. C.-M. Shang Y, Diao Y-H, in: Proceedings of the 7th International Conference on Machine Learning and Cybernetics, pp. 3032–3036.

[12] E. Candes, L. Demanet, D. Donoho, L. Ying, Multiscale Modeling and Simulation 5 (2006) 861–899.

[13] E. Candes, Signal Processing 82 (2002) 1519–1543.

[14] L. Boubchir, F. M.J., in: Eighth International Conference on Signal Processing and Its Applications - IEEE ISSPA 2005, pp. 747–750.

[15] M. Do, M. Vetterli, IEEE Transactions on Image Processing 11 (2002) 146–158.

[16] M. M. Islam, D. Zhang, G. Lu, in: ICME'09: Proceedings of the 2009 IEEE international conference on Multimedia and Expo, IEEE Press, Piscataway, NJ, USA, 2009, pp. 562–565.

[17] S. Kondra, V. Torre, in: ICVGIP, pp. 429–434.

[18] B. S. Manjunath, W. Ma, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI - Special issue on Digital Libraries) 18 (1996) 837–42.