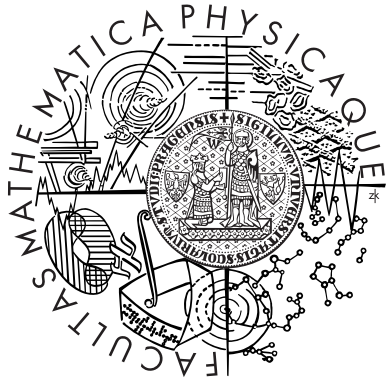
Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS



Miroslav Beneš

Digital Image Processing of Cross-section Samples

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic

Supervisor of the doctoral thesis: RNDr. Barbara Zitová, Ph.D.

Study programme: Computer Science
Specialization: Computer Graphics and Image Analysis

Prague 2014

I would like to express my thanks to Barbara Zitová for supervision of my thesis and colleagues at the Department of Image Processing for numerous advice, comments and ideas during my doctoral studies. I thank Janka Hradilová and David Hradil from ALMA laboratory for providing the cross-section images and invaluable insights into the field of material research and art restoration. Thanks also go to Jan Cendelín from Faculty of Medicine in Pilsen and to Martin Čapek from Institute of Physiology AS CR in Prague for provision of biological image.

Last but not least I thank my friends, my family and Terezka for support and never ending patience.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague Miroslav Beneš

Název práce: Digitální obrazové zpracování vzorků v příčném řezu

Autor: Miroslav Beneš

Katedra/Ústav: Ústav teorie informace a automatizace, Akademie věd České republiky, v.v.i.

Vedoucí doktorské práce: RNDr. Barbara Zitová, Ph.D., Ústav teorie informace a automatizace, Akademie věd České republiky, v.v.i.

Abstrakt: Tato disertační práce se věnuje digitální analýze a zpracování mikroskopických obrazových dat se zaměřením na vzorky v příčném řezu, které se odebírají z uměleckých děl, čímž práce tematicky částečně spadá do oblasti kulturního dědictví. Práce přispívá k řešení dvou problémů zpracování obrazu – segmentace obrazu a vyhledávání na základě obrazové podobnosti. Za účelem studia chování různých segmentačních metod jsou na sadě snímků příčných řezů porovnány a vyhodnoceny výsledky těchto metod. Na tomto základě jsou navržena doporučení pro volbu metod vhodných pro segmentaci mikroskopických snímků. Dále se práce věnuje přínosu kombinace/fúze segmentačních algoritmů a v textu je navrženo několik různých kombinačních postupů. Všechny výsledky jsou objektivně posouzeny a vyhodnoceny prostřednictvím několika indexů pro měření kvality segmentace v rámci mnoha experimentů. Použitelnost navržených řešení a zjištění je ověřena též na datech jiného než uměleckého původu. V druhé části se práce věnuje vyhledávání na základě obrazové podobnosti a představuje funkční řešení pro hledání podobných vzorků v databázi. Vyhledávání je implementováno v systému Nephele, který je určen pro zpracování a archivaci zpráv o materiálovém průzkumu. K tomu obsahuje implementované metody zpracování obrazu vyvinuté zvlášť pro aplikace v oblasti kulturního dědictví.

Klíčová slova: digitální zpracování obrazu, segmentace obrazu, obrazové vyhledávání, vzorky v příčném řezu, kulturní dědictví

Title: Digital Image Processing of Cross-section Samples

Author: Miroslav Beneš

Department/Institute: Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

Supervisor of the doctoral thesis: RNDr. Barbara Zitová, Ph.D., Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic

Abstract: The thesis is aimed on the digital analysis and processing of microscopic image data with a focus on cross-section samples from the artworks which fall into cultural heritage domain. It contributes to solution of two different problems of image processing – image segmentation and image retrieval. The performance evaluation of different image segmentation methods on a data set of cross-section images is carried out in order to study the behavior of individual approaches and to propose guidelines how to choose suitable method for segmentation of microscopic images. Moreover, the benefit of segmentation combination approach is studied and several distinct combination schemes are proposed. The evaluation is backed up by a large number of experiments where image segmentation algorithms are assessed by several segmentation quality measures. Applicability of achieved results is shown on image data of different origin. In the second part, content-based image retrieval of cross-section samples is addressed and functional solution is presented. Its implementation is included in Nephele system, an expert system for processing and archiving the material research reports with image processing features, designed and implemented for the cultural heritage application area.

Keywords: digital image processing, image segmentation, image retrieval, cross-section samples, cultural heritage

Contents

Chapter 1

Introduction

The improvement of digital imaging devices and technologies in recent years has led to dissemination of image acquisition to many different scientific fields. The result of this development is growing amount of data which is necessary to process, store and eventually retrieve again from dedicated database for further analyses. This would not be possible without comparable advances in information technologies. They provide means which allow to handle the continuous growth of images. One of the fields which can substantially help is digital image processing (DIP).

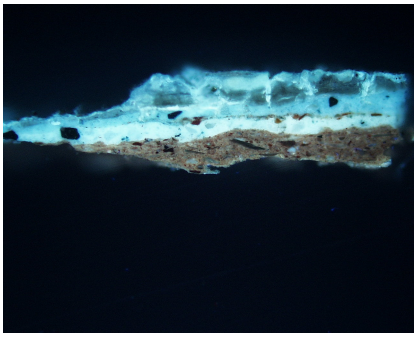
In this thesis we study how the DIP can contribute to the analysis and processing of microscopic image data. The progress of image acquisition technologies has made possible to capture the images of minute researched objects in higher resolution and different modalities using various devices. This applies to many scientific areas – biology, medicine, geology, mineralogy and cultural heritage to name a few. We concentrate mainly on the area of cultural heritage and study the cross-section samples taken from the artworks (see following section 1.1 for the description). However most of the achieved results are directly applicable also to other mentioned fields which is demonstrated in the text.

We focus on dealing with two different problems which DIP can help with. First, the image processing methods can facilitate the analysis of acquired microscopic samples or species. They are often visually processed and evaluated. Image processing algorithms can make this routine either completely automatic or at least they significantly simplify it. Moreover they can offer new solutions to existing problems and come up with new findings. We aim at a critical step of the sample image analysis – image segmentation. Image segmentation is a process which partitions an input image to meaningful non-overlapping regions – segments. Quality of the image segmentation algorithm affects the performance of the whole image analysis. Therefore it is always crucial to choose a right method for the very problem which needs to be solved. Unfortunately there is no universal method which would provide the best results possible. We evaluate performance of many segmentation algorithms and try to answer the question which of them are suitable for microscopic images and what can be done more to obtain even better results. We are not aware of any similar work in context of microscopic image data although there are papers working with different data, e.g. [42, 63, 38].

The retrieval stage is the second problem we focus on. DIP provides methods to improve searching in large databases using image information. Conventional databases with information systems usually contain only a text-based search which makes re-

(a)

(b)



(c)

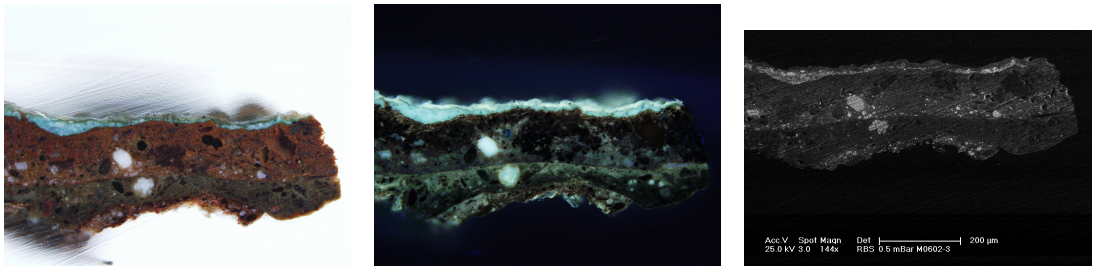
Figure 1.1: (a) The painting with marked location of the sample extraction; (b) image of sample 1 in visible spectrum; (c) image of sample 1 in ultraviolet spectrum. Image courtesy of Academy of Fine Arts Archives, Prague, Czech Republic, and ALMA.

trieval under certain circumstances a clumsy task. Retrieving functionality can be however extended by *content-based image retrieval* (CBIR). This approach enables fetching data by image similarity and thus help potential user to browse through a database in a different way.

The *Nephele* system implements the results achieved from solving both mentioned problems – image segmentation and image retrieval. Nephele is an expert system for processing and archiving the material research reports created during material research of the artwork. Incorporation of image-based search and image preprocessing methods for related image restoration simplifies the delivery of relevant reports and efficient work with the collection. Nephele reflects the best practices used in the art conservation and is intended to be used by art restorers, conservators, or art historians.

1.1 Cross-section samples of the artwork

Research of the thesis is based on a set of microscopic image data – images of cross-section samples taken from the artworks. They origin from painting restoration process, whose integral part is painting material research. Its aim is mainly to identify used painting substances which helps to evaluate painter's technique and style. During this stage the minute samples are taken away from representative places of the artwork. The locations should be well laid out across the artwork plane to provide competent view on different material layers composition (*stratigraphy*), morphology of the material grains and color. See the example in figure 1.1.



(a) VIS image (b) UV image (c) SEM image

Figure 1.2: The images of the cross-section samples are acquired in three modalities – visible spectrum (VIS), ultraviolet spectrum (UV) and scanning electron microscope (SEM). Image courtesy of ALMA.

Being an invasive method a scalpel or laser technique are used for sample withdrawal. Nowadays also the non-destructive methods of material research are exploited, e.g. inspection of the painting under different illumination (ultraviolet, infrared or roentgen spectra are used to name a few), and/or spectroscopic methods for measuring laser-induced atomic and molecular emission. However invasive sample extraction is still essential part of the research. After the withdrawal, the samples are embedded in a polyester resin, grounded at a right angle to a surface plane and ground to expose the material layers consisting of color pigments and binders (the procedure is well-described in [67][1]). The images in visible (VIS) and ultraviolet spectrum (UV) are taken, where UV image may reveal fluorescent property of certain materials. Study by scanning electron microscope (SEM) with chemical contrast is also performed. These constitute three modalities, which we work with throughout the thesis. See figure 1.2 for an example of images of cross-section sample.

The resulting data set consists of 156 cross-section samples captured in three modalities. They come from the Academic Materials Research Laboratory of Painted Artworks (ALMA)[2], where they help the art restorers to choose the proper materials and appropriate technique for the very restoration. The images do not always form a complete triplet for each sample. SEM modality is often missing, sometimes VIS or UV image are not present. Nonetheless there are 148 VIS images, 148 UV images and 89 SEM images in the data set. The SEM images are grayscale, the other two modalities are acquired in RGB colorspace.

1.2 Thesis outline

Thesis is divided to two different parts with the following contents. First part is dedicated to the problem of image segmentation and contains chapters 2, 3, 4, 5 and 6.

- **Chapter 2** introduces the reader to the problem of image segmentation and whole first part of the thesis.

[1] Alternatively see web page `http://chsopensource.org/2012/12/19/microscopy-for-art-examination-cross-sections-preparation/` with a video of cross-section sample preparation.

[2] `http://www.alma-lab.cz`

- **Chapter 3** deals with necessary preprocessing of the input images. They contain artifacts which have undesirable impact on the performance of the image segmentation algorithms. It is therefore feasible to remove or diminish them.
- **Chapter 4** is the key section of the first part. It presents set of studied image segmentation algorithms and indices for measuring their output quality (performance). The algorithms are compared there and findings of the evaluation are discussed. Chapter gives recommendations on using of the algorithms depending on the input image properties. Also the applicability to different image data is shown.
- **Chapter 5** shows that performance of even the best segmentation algorithm from chapter 4 can be improved by combination/fusion of multiple segmentation methods.
- **Chapter 6** concludes the first part of the thesis.

Part two of the thesis presents in chapter 7 the image processing system for art restoration – *Nephele* – with image retrieval facility. It also implements the results and findings of the first part. Last part summarizes main contributions of the thesis in chapter 8.

Part I

Image segmentation

Chapter 2

Introduction

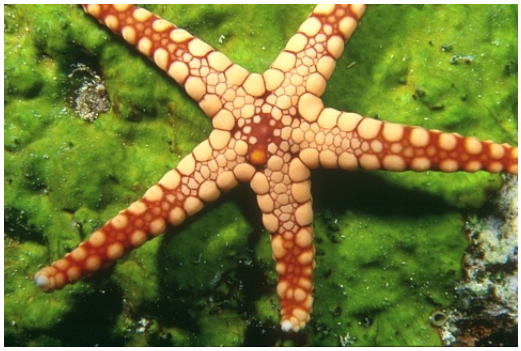
The fundamental objective of image segmentation is to partition the input image into meaningful non-overlapping regions – segments – for further analysis or visualization (see illustration in figure 2.1, subfigures (a) and (b)). There is a variety of approaches addressing this task, exploiting various image properties to achieve the given goal (see e.g. [69] for survey). They span from low-level techniques using intensity thresholds, edge tracing or region growing, over graph-based and statistical approaches, to model-based algorithms and other higher-level methods. Recently, the combination-based solution has been introduced, where the final partition is formed using a combination of results of several segmentation methods and thus inhibiting their shortcomings.

Despite the longtime effort to develop high quality segmentation algorithms, there has not been any universal segmentation method proposed. Under these circumstances, there is a dilemma which method to choose for given particular data set and whether the combination of segmentation results would be beneficial. This part of the thesis tries to answer these questions for defined category of image processing data – set of images of microscopic samples (described in section 1.1, see figure 1.2), moreover taken in different modalities. From the image processing point of view, the origin of the samples often does not play an important role. The factual meaning of particular intensity levels can be irrelevant for the segmentation algorithm.

We limit our study to the microscopic image data that contain the sample located in the inner part of the image, mostly not reaching to the top and bottom image borders. The data may come from an analysis of painting materials used in art restoration (figure 1.2), which is the case of the data set used in our evaluation. They can be samples of various biological materials, such as tissues, cells, or other biological structures. The task at hand can be seen as the two-target problem where an image has to be labeled with either foreground or background label and where the foreground is usually the inner part of the image and the background is separated and/or removed. The problem can be viewed as image binarization, too (illustrated in subfigure (c) of figure 2.1).

At first glance it might seem to be a simple task solvable by means of basic thresholding, however the situation is often more complex. Due to the setting of data collection process, acquired images are often unfit to the chosen segmentation method and following complications are usually inevitable – surroundings of analyzed samples can be semitransparent, with non-uniform cutting-plane and various debris, to name a few examples. High number of samples can negatively influence precision of sample scanning in terms of noise level and blurring.

The objective is to evaluate the non-interactive segmentation methods in terms of



(a) Input color image

(b) Image partitioned to salient non-overlapping regions

(c) Binarized image with only foreground and background

Figure 2.1: Illustration of partitioning the input image (a) to several salient non-overlapping regions (b). In (c) the input image is binarized only to foreground with one meaningful object and the rest in the background. Images come from The Berkeley Segmentation Database [63].

their accuracy, assessed by several indices used for measuring the output quality of image segmentation algorithms. First, necessary preprocessing of the input images from data set needs to be performed (chapter 3). In following chapter 4 the participating methods and indices are introduced, the very comparison of the algorithms is presented and we show the applicability of the achieved conclusions to different data set – the biological samples. Finally, efficiency of combination of segmentation results is addressed in chapter 5 as a way to further improve the performance of single segmentation algorithm.

Chapter 3

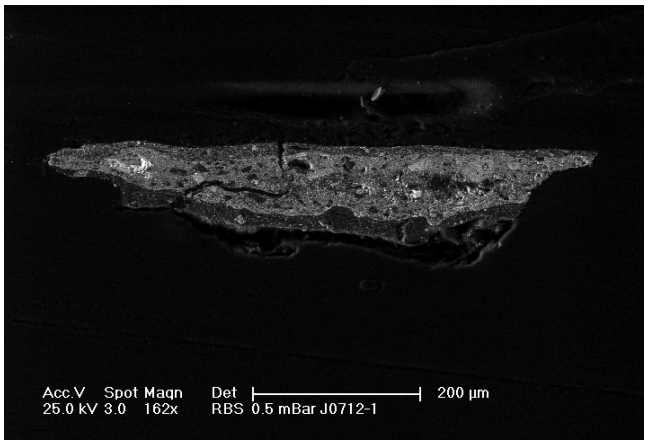
Preprocessing

This chapter deals with image preprocessing stage necessary to successful image analysis which follows in next chapters. Original images acquired during artwork restoration process are often unfit for direct image processing, i.e. image segmentation in chapter 4 or image retrieval in chapter 7. There are two main problems involved and covered by next two sections – legend in SEM images embedded by electron microscope and more importantly artifacts created by grinding of polyester resin during sample preparation process.

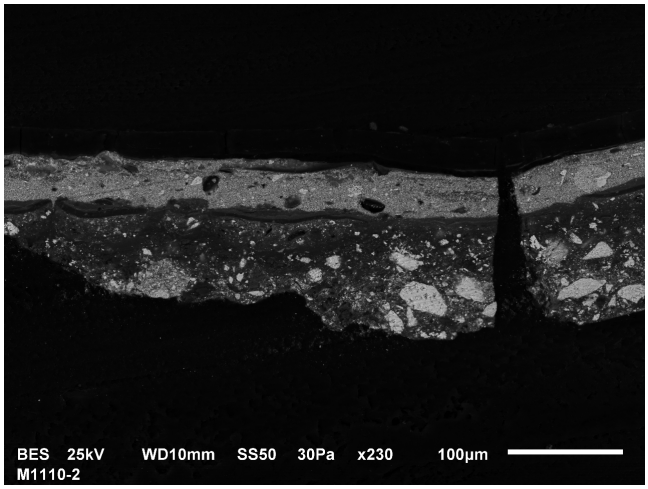
3.1 Removal of legend in SEM images

During SEM image acquisition process the scanning electron microscope embeds a legend into a final image it creates. The legend contains sample identification number, amount of magnification, image bar scale and other information possibly important for evaluation by the analyst. On the other hand with its white color it causes problems to image segmentation algorithms. They cannot produce desired output which would otherwise be in their capability. Under these circumstances it is better to remove the legend and to reconstruct the missing image parts appropriately. Since it is part of preprocessing stage the removal method should be simple and fast.

There are two kinds of SEM images in a data set coming from two different scanning microscopes. The first one (subfigure (a) of figure 3.1) has the legend only in white color, while the other one (subfigure (b)) has the letters anti-aliased (gray levels are present). The legend is vertically always in the same place in the image, only the horizontal length differs. This makes the localization of the legend easy and following steps are realized only on region with the legend inside. Replacement of white letters by one different color (e.g. black which is often close to the color of background) would not change a situation much. It would still influence the outcome of segmentation algorithms. Better approach. which would conceal the legend completely. is thus needed. Color of each pixel of the legend will be replaced with color close to the nearest neighboring non-legend pixels. First the letters are replaced by initial estimate of final color. The letters are identified by white color. However in case of the second type of SEM image this would be insufficient due to anti-aliasing. So few more neighboring pixels enclosing the whites are added (given by operator of mathematical morphology – dilation). Initial estimate of color is formed by masked local means of neighboring background pixel colors. Masked means that the legend pixels are not



(a)



(b)

Figure 3.1: Two SEM images with two different legends. In (a) the legend is only of white color, in (b) letters of the legend are anti-aliased. Partial closeups are in the right column. Image courtesy of ALMA.

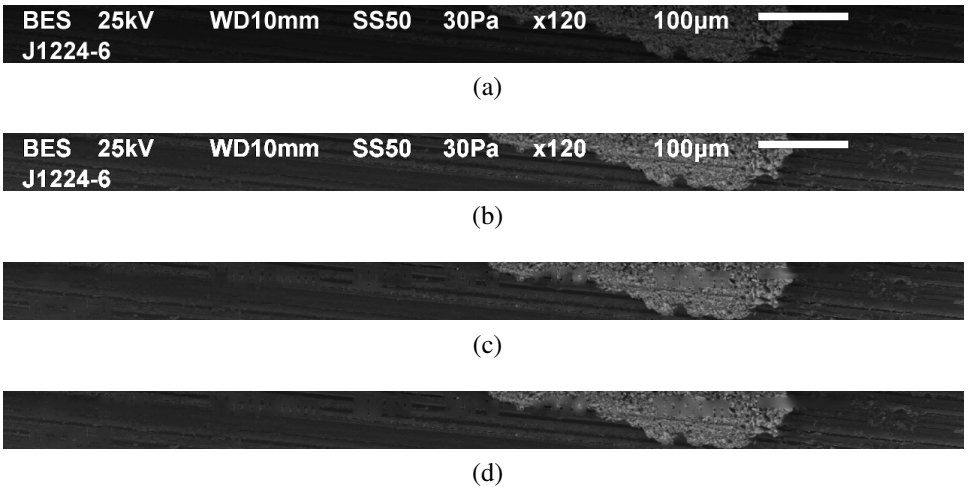


Figure 3.2: Process of legend removal. Subfigure (a) shows cutout with legend from original SEM image. Subfigure (b) shows the same with slightly enhanced contrast (this also holds for (c) and (d)). In subfigure (c) the letters are replaced with local means. In (d) there is a final output after mild averaging.

taken into account[1]. Second stage is to adapt the estimated pixel colors even more to the surrounding neighborhood of the former legend. This is achieved by two steps of mild average filtering. The legend is masked out in a resulting output in such a way that it does not further influence the outcome of segmentation methods. Figure 3.2 shows the whole process on the example where legend interferes with cross-section sample.

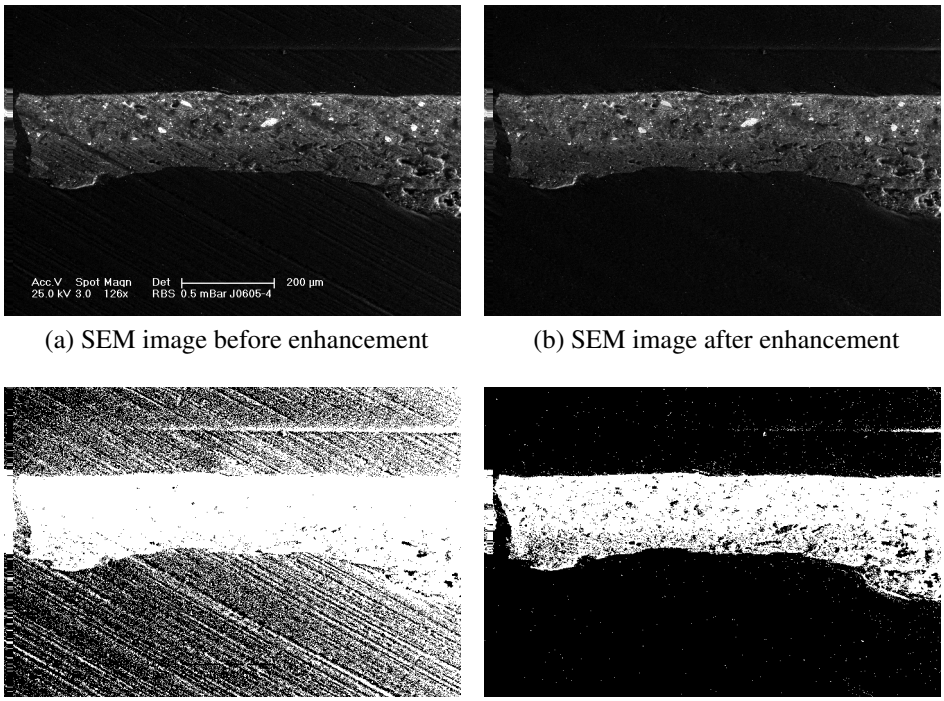
The UV and VIS images do not contain any such legend. However rarely the image bar scale is present also there. It is impossible to locate it and replace it automatically. The bar scale is thus removed manually in graphics editing program.

3.2 Removal of grinding artifacts

During acquisition process of the cross-section samples the polyester resin, which each sample is embedded in, is ground by fine sandpaper to reveal the painting layers for necessary analyses and measuring. Due to this grinding of the resin noisy artifacts form in the background. Unfortunately apart from being visually disturbing, the artifacts influence the results of image segmentation algorithms and also the comparison of their performance. It is therefore legitimate to remove these artifacts completely or at least to significantly diminish them. It is possible to grind the resin with much finer sandpaper and prevent creation of the artifacts in images. This would however complicate follow-up analyses with reflections and coarsening of the resin would be unavoidable. The image processing algorithm is thus needed. See figure 3.3 for example of segmentation method which is prone to do worse in presence of the artifacts. The proposed method was published in [9]. Several improvements have been made since then.

Figure 3.3 shows the properties of the artifacts. They are omnipresent and in form of parallel lines. The width and density of the lines depend on the grit size of the sandpaper. The visibility of the artifacts depends on the modality. In SEM and VIS they are clearly observable, while in UV they are not as evident due to their non-fluorescent

[1]It was confirmed by experiments that in the case of first type pure black color is sufficient as an initial estimate. The letters are very narrow and more complex approach is not necessary.



(a) SEM image before enhancement

(b) SEM image after enhancement

(c) Binarized SEM image

(d) Binarized enhanced SEM image

Figure 3.3: The background artifacts might influence the outcome of the segmentation algorithm. In (a) there is a SEM image with the artifacts, in (b) the image is after enhancement (artifacts are removed). Figures (c) and (d) illustrate the influence of artifacts (non)presence on segmentation method. The legend was removed before the actual segmentation via process described in section 3.1. Image in (a) courtesy of ALMA.

nature and dark background. The proposed method uses Fourier transform for detection and removal. In the following we assume that f is a matrix of real numbers of dimension $N \times M$ representing a grayscale input image with artifacts to be removed, j and k are spatial coordinates, u and v are frequencies.

First we compute discrete Fourier transform of the input image

$$F(u,v) = \sum_{k=0}^{M-1} \sum_{j=0}^{N-1} f(j,k) e^{-2\pi i (\frac{uj}{N} + \frac{vk}{M})}$$

and its amplitude

$$A(u,v) = |F(u,v)|$$

In amplitude spectrum we can observe considerable response in a direction perpendicular to the original artifacts lines (see figure 3.4(a), the spectrum is in logarithm scale and shifted using Matlab fftshift function for better visibility). It is useful to equalize the histogram of the input image to make the lines more distinctive and following processing easier. Otherwise the response would be too weak. Now the response and its orientation can be detected, masked out of the amplitude spectrum and after application of inverse Fourier transform the artifacts are removed in the image domain. It is achieved by following steps.

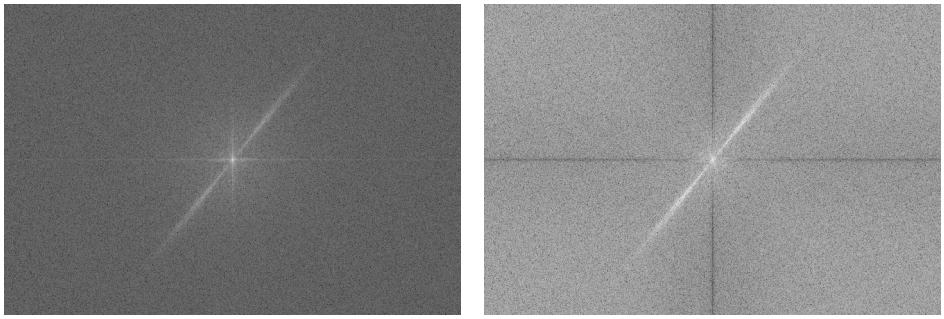
Amplitude spectrum $A(u,v)$ usually contains irregularities due to so-called spectral leakage (horizontal and/or vertical stripes induced by discontinuities of the image borders. See figure 3.4(a)). They could make the artifacts response detection difficult as they are very similar. Therefore it is important to remove these spurious flaws from the spectrum before the very detection is performed. It can be achieved by division of the spectrum by row and column sums respectively (the result is in figure 3.4(b)).

$$A'(u,v) = \frac{A(u,v)}{\sum_{j=0}^{N-1} A(j,v)}$$

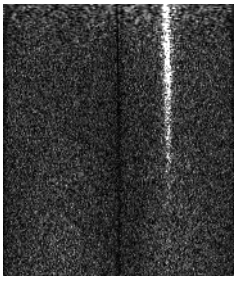
$$A''(u,v) = \frac{A'(u,v)}{\sum_{k=0}^{M-1} A(u,k)}$$

The alternative way is to suppress the spectral leakage via application of window function to the image domain [27]. Since the problem rises from image border discontinuity the solution is to assure that there are zero values on the image borders, preferably with smooth decrease from the image center. We can achieve this by multiplying the image with window function (or taper) such as Hamming or Hann (Hanning) functions. However in our case this method diminishes also the response of the grinding artifacts and final results are worse than with row/column sums normalization approach. Figure 3.5 exhibits comparison of these two approaches and it shows diminishing of response in case of window tapering.

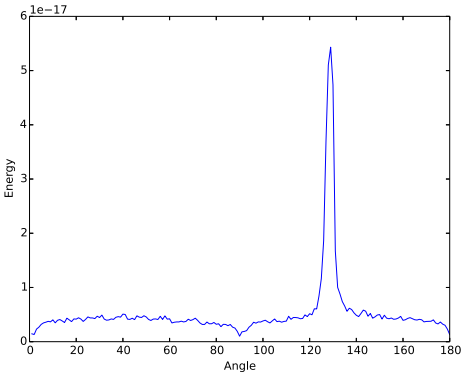
The response detection is based on following idea. We can rotate amplitude spectrum $A''(u,v)$ by one degree around the origin (assuming that the spectrum is shifted) for all angles between 1 and 180. In each step the values in columns are summed and the maximum of the column sums is taken (with regard to the character of Fourier transform it is always in the middle). Thus, for each angle we have maximum response. The highest peak of this function gives the orientation of the grinding artifacts in the image domain. Nevertheless the computation of the function using rotation approach

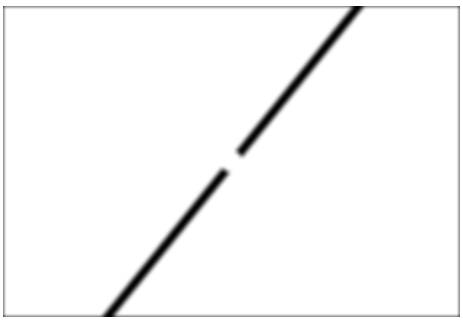


(a) (b)



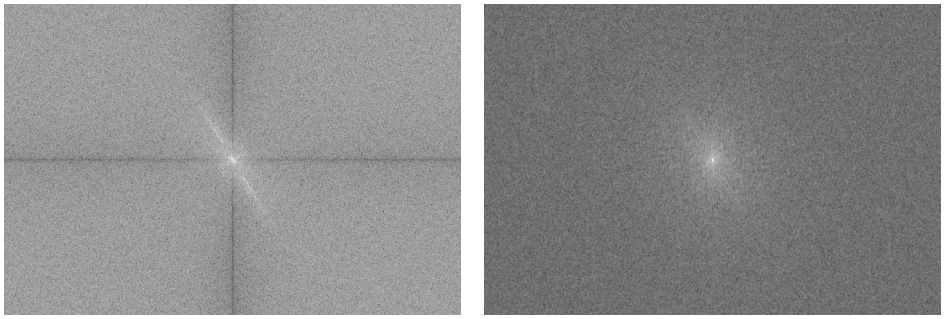
(c)





(d) (e)

Figure 3.4: Removal of grinding artifacts in stages. The image from figure 3.3(a) forms the input, figure 3.3(b) is the output. In (a) there is an amplitude spectrum with clearly visible response of the grinding artifacts, in (b) the spectrum is normalized using row/column sums to suppress the spectral leakage (highlighted cross in the middle of spectrum in (a)). The spectrum transformed to polar coordinates is in subfigure (c). (d) shows energy function $E(\theta)$ (directional energy for every angle). Finally, subfigure (e) exhibits mask constructed using detected orientation. All spectra are shifted, in logarithm scale and adjusted for better illustration.



(a) (b)

Figure 3.5: The comparison of two different approaches to suppress the spectral leakage. In (a) there is a result of normalization the original spectrum with row and column sums. In (b) there is a result of window tapering of the original image. Tapered spectrum is definitely smoother compared to normalized spectrum with dark cross. However the response of the grinding artifacts is significantly suppressed as well. The dark cross in (a) is not a problem because it does not create false high peaks as non-normalized spectrum. Both spectra are shifted, in logarithm scale and adjusted for better illustration.

would be inefficient and time demanding. The same can be achieved by transforming the amplitude spectrum $A''(u,v)$ to polar coordinates $A_p(r,\theta)$ and exploit that.

$$r = \sqrt{u^2 + v^2}$$

$$\theta = \text{atan2}(v,u)$$

The amplitude spectrum in polar coordinates $A_p(r,\theta)$ contains directional response for all angles in range of 1 and 180 degrees. Now a narrow horizontal band is removed from spectrum $A_p(r,\theta)$ for small values of r. The band corresponds to small circle of values surrounding the origin of spectrum $A''(u,v)$ in Cartesian coordinates. This part is heavily interpolated during the transformation to polar coordinates and as such it negatively influences following computation. Also the values of spectrum $A_p(r,\theta)$ are squared which emphasizes the artifacts response in spectrum even more. Formally we obtain a power spectrum with energy as values. Figure 3.4(c) exhibits amplitude spectrum transformed to polar coordinates.

$$A'_p(r,\theta) = \begin{cases} 0 & 0 \leq r \leq x \\ A_p^2(r,\theta) & \text{otherwise} \end{cases},$$

where x is parameter.

Spectrum $A'_p(r,\theta)$ is then summed in columns to obtain directional energy for every angle (figure 3.4(d)).

$$E(\theta) = \sum_r A'_p(r,\theta)$$

This energy function $E(\theta)$ directly corresponds to function computed by rotation approach described above. Distinct peak of this function (and its maximum value at the same time) determines the orientation of the grinding artifacts.

$$\theta_{max} = \arg\max_{\theta} E(\theta)$$

Now we have enough information to construct binary mask $M(u,v)$, which will be used for removing the response from the Fourier spectrum. The aim is to mask out the artifacts response in the spectrum using the detected orientation. Mask must be smoothed, otherwise the ringing effect is involved [27]. See figure 3.4(e) for an example. Afterward, inverse Fourier transform is applied to modified spectrum and enhanced result is acquired (see figure 3.3(b)).

$$f(j,k) = \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} M(u,v) F(u,v) e^{2\pi i (\frac{uj}{N} + \frac{vk}{M})}$$

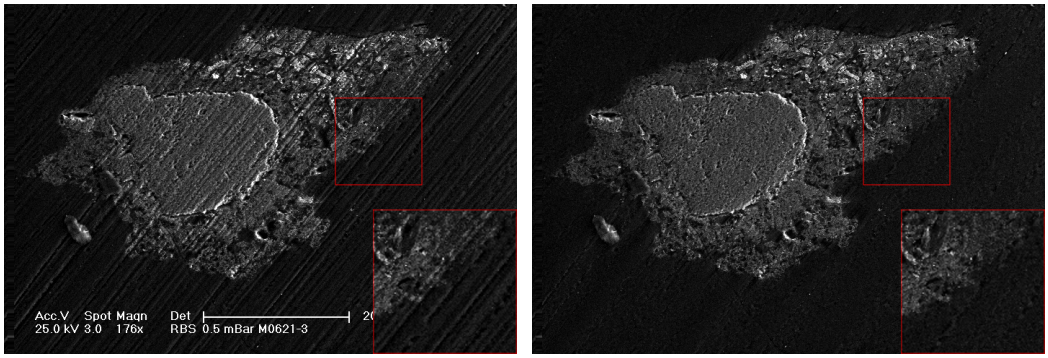
Finally, the parameters of the algorithm were tuned to have a minimal impact on the cross-section in terms of sharpness and level of detail.

Proposed method for artifacts removal is tested on the data set of cross-section images. Therefore, we have 148 VIS images, 148 UV images and 89 SEM images for testing purposes. Grayscale SEM images are processed directly, in case of color VIS and UV images artifacts orientation detection and mask construction is performed using intensity channel, mask is applied on each RGB channel separately. In VIS modality 114 images out of 148 are successfully enhanced (success rate of 77 percent), which means the artifacts are removed completely or significantly suppressed[2]. Concerning the remaining dissatisfactory results, either the response in Fourier spectrum is not distinctive enough or an appropriate peak is not correctly marked. The algorithm is much more successful in SEM modality – 83 out of 89 SEM images are correctly enhanced (success rate is thus 93 percent). The reason of better behavior is definitely in more contrastive nature of the grinding artifacts in SEM modality as oppose to VIS where the artifacts can blend with pale background. This naturally affects the Fourier spectrum and response detection. 126 UV images are successfully enhanced (85 percent). Originally the algorithm performed poorly in UV modality. The responses in spectra were almost negligible and detection of artifacts orientation was impossible. As a workaround the orientation detected on VIS image (where VIS image is present) is used for removal. This is not correct, since both images are not registered. However geometrical misalignment is small and final results are satisfactory. Nonetheless the practical impact of artifacts removal in UV modality is minute and it is done only for illustration purposes. The background of the UV images is dark and the artifacts are nearly unnoticeable, and thus they do not affect image segmentation and other image processing methods too much.

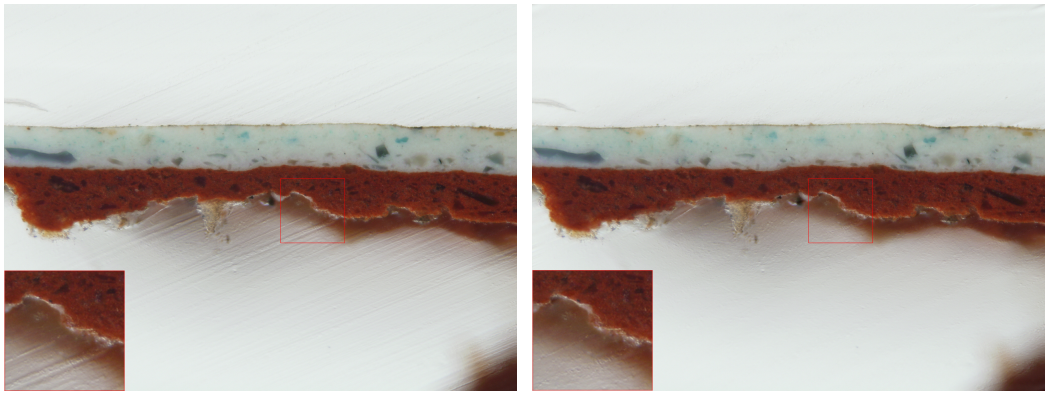
Figure 3.6 contains three examples of successful removal of grinding artifacts – one in SEM modality and two VIS images. The artifacts are satisfactorily suppressed in each case, while the texture of the cross-section is well preserved and ringing artifacts are kept at minimum. The positive effect of removal on image segmentation was already demonstrated in figure 3.3.

The proposed removal method can be also used in a different area of cultural heritage – processing of infrared (IR) images of old paintings. The IR light enables to see

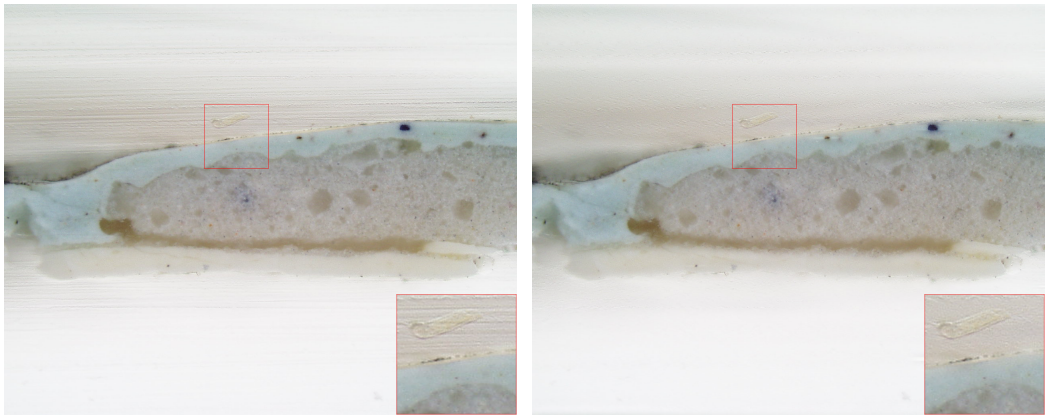
[2]There are few images where two different orientations of grinding artifacts are present. It originates from two phases of grinding by sandpaper. The proposed algorithm successfully removes one dominant orientation while the other stays in the image. However this second orientation is often not so disturbing. Nonetheless these cases are taken as successful.



(a)



(b)



(c)

Figure 3.6: Three examples of successful removal of grinding artifacts in cross-section images. In each pair original image is on the left side and enhanced image with artifacts removed on the right side. The close ups are in the red frame. In (a) there is result of algorithm in SEM modality. The legend was removed before processing via the method described in section 3.1. Two processed VIS images are in subfigures (b) and (c). The artifacts are satisfactorily suppressed, while the texture of the cross-section is well preserved and ringing artifacts are kept at minimum. Image courtesy of ALMA.

hidden underdrawing, which captures the painter's original intentions. Unfortunately, IR backlighting reveals also the canvas structure and its inhomogeneity. The algorithm with few minor tweaks helps to remove them which simplifies further analysis (e.g. it reveals several features around gemstone and eyes which were hard to notice before). See figure 3.7.

(a) Original image

(b) Image after enhancement using the proposed algorithm

Figure 3.7: Beside the underdrawings the IR backlighting shows also the canvas structure, which can be removed by the proposed method. The original underdrawing texture is well preserved and image details are more visible. Courtesy of Igor Fogaš, Moravian Gallery in Brno.

Chapter 4

Comparison of image segmentation methods

4.1 Segmentation algorithms and quality indices

A survey of the image segmentation algorithms analyzed in this thesis (i.e. studied set) and survey of indices used for measuring the output quality of the segmentation algorithms are presented in this section (subsections 4.1.1 and 4.1.2 respectively). It is not meant to be a comprehensive review, especially in the case of segmentation algorithms. The aim is rather to specify which methods are studied and by which means they are compared. If necessary, please consult given references. The abbreviations are assigned to each method and index for future references and their list is presented in table 4.1.

4.1.1 Segmentation algorithms

There is a variety of segmentation methods available to be used to solve the image segmentation problem which differ in many ways (see e.g. [69] for survey). The algorithms in our study are selected with respect to the following criteria. Methods with different fundamentals are considered to provide a diversity. The performance and computational (time) efficiency are taken into account with preference for short execution time. Finally, the public availability of the implementation and thus related popularity of the segmentation method are considered too. Last criterion is also important because it can be expected that potential users of image segmentation algorithms would choose exactly such popular methods. There exists a lot more segmentation algorithms (e.g. [61, 4, 36]) but inclusion of each of them is beyond the scope of this work.

The selected algorithms can be divided into groups according to their fundamental approach to solve the image segmentation problem. The following paragraphs briefly describe the groups and particular algorithms.

Thresholding

Thresholding is probably the most popular method for image segmentation. The aim is to find an optimum threshold which separates the input image to two distinct groups

Segmentation methods	
IMJ_*	Various thresholding methods from ImageJ [44, 74, 77, 58, 50, 53, 98, 68, 25, 85, 107, 106]
HT_*	Various thresholding methods from HistThresh [82, 34, 68, 77, 74, 23, 53, 98]
TNC	Tao's thresholding method [96]
RG	Region growing [73]
KM	K-means clustering [60]
MS	Mean Shift algorithm [17]
GC_FH	Felzenszwalb's method [28]
GC_R	GrabCut [83]
GC_CV	Daněk's optimization of Chan-Vese [21, 20, 13]
GC_RD	Daněk's optimization of Rousson-Deriche [21, 20, 84]
MNC	Multiscale normalized cut [19]
Quality indices	
HD	Hamming distance [39]
BHD	Boundary Hamming distance [54]
RI	Rand index [75]
ARI	Adjusted Rand index [46]
DC	Dice coefficient [24]
FMI	Fowlkes-Mallows index [30]
NMI	Normalized mutual information [94]
VI	Variation of information [64]
HAUSD	Hausdorff distance [88]
MASD	Mean absolute surface distance [88]

Table 4.1: List of image segmentation methods in studied set and of quality indices used for their comparison. The abbreviations widely used in text are in the first column.

of pixels by their intensity. Plenty of different methods for threshold detection exist and many of them are selected to participate in the evaluation.

The methods of the Auto Threshold plugin[1] for ImageJ software package[2] are included. Namely Huang method (IMJ_HUANG) [44] which minimizes the measures of background/foreground fuzziness, Intermodes (IMJ_IM) [74] with iterative histogram smoothing, IsoData (IMJ_ISO) [77] and its variation (IMJ_DEF) which iteratively update the threshold according to background and foreground intensity means, Li's method (IMJ_LI) [58] for cross entropy minimization, Kapur–Sahoo–Wong maximum entropy method (IMJ_MAXENT) [50], mean of the gray levels as threshold (IMJ_MEAN), iterative version of minimum error thresholding (IMJ_IME) [53], minimum method (IMJ_MIN) [74], moment-preserving method (IMJ_MOM) [98], Otsu's method (IMJ_OTSU) [68] for minimizing the intra-class variance, percentile method (IMJ_PER) [25], method using Renyi's entropy (IMJ_RENYI) [50], Shanbhag's extension (IMJ_SB) [85] to Kapur's maximum entropy method, geometric Triangle algorithm (IMJ_TRIANGLE) [107] and Yen's method (IMJ_YEN) [106] based on a maximum correlation criterion.

In addition to the plugin several other thresholding methods from MATLAB HistThresh toolbox[3] are studied[4] – concavity method by Rosenfeld (HT_CONCAV) [82], Glasbey's entropy method (HT_ENT) [34], maximum likelihood via EM algorithm (HT_MAXLIK) [23], Intermeans (HT_INTER) as equivalent to Otsu's method and its iterative version (HT_INTERI) which is equivalent to IsoData method mentioned above. Then there is median method (HT_MEDIAN) [34] which assumes that half of the pixels belong to the background and other half to the foreground, and non-iterative minimum error thresholding (HT_ME) [53].

Finally, a Tao's method for image thresholding (TNC) [96], which uses a normalized graph-cut to detect an optimum threshold, is included in the evaluation below.

Region growing

The region growing (RG) [73] is another common segmentation approach included in our selection. The algorithm partitions the input image to segmented regions by growing from the seed points (picked automatically or by the user) to the neighboring pixels depending on a membership criterion such as intensity or texture similarity.

Clustering methods

The goal of clustering methods is to group the input objects by their similarity or dissimilarity with respect to a given criterion such as color, spatial coordinates etc. K-means clustering and Mean Shift algorithm are selected representatives of this approach.

K-means clustering (KM) [60] assigns the input objects to the clusters with the nearest means which are iteratively updated. The method strongly depends on the

[1] `http://fiji.sc/Auto_Threshold`

[2] `http://rsbweb.nih.gov/ij/`

[3] `http://www.cs.tut.fi/~ant/histthresh/`

[4] There are more thresholding methods in the toolbox. Most of them are the same as in ImageJ plugin. However we found out that their implementation often slightly differed and so did the results of the segmentation. For this reason all methods are included in the studied set with corresponding suffices in their abbreviations (so there are e.g. both IMJ_MEAN and HT_MEAN in the studied set).

initialization and favors final clusters/segments of similar spatial extent. The Mean Shift algorithm (MS) represents more complex approach. Comaniciu and Meer [17] exploited the non-parametric mean shift procedure for detecting multiple modes in a feature space in order to delineate the final clusters in such space.

Graph-based algorithms

Graph-based image segmentation algorithms generally model the image as a graph in which the nodes represent the pixels and the edges of the graph correspond to some relation between pixels (usually their similarity or dissimilarity). A graph partitioning method is then used to obtain final partition and by doing so also the final segmentation of the input image.

In their paper [28] Felzenszwalb and Huttenlocher (GC_FH) introduced the efficient greedy algorithm for partitioning an image graph to obtain a final segmentation that is not too coarse or too fine given a dissimilarity predicate. GrabCut algorithm by Rother et al. (GC_R) [83] uses graph cut optimization technique (min-cut/max-flow algorithm) to minimize energy function derived from an input image using intensity values[5]. The OpenCV[6] implementation of this algorithm is examined. The graph cut minimization [21, 20] of both Chan-Vese active contour model for image segmentation (GC_CV) [13] and Rousson-Deriche Bayesian model (GC_RD) [84] is included[7]. A multiscale version of normalized cut graph partitioning framework (MNC) [19] is considered too. The multiscale adjustment added to the original algorithm by Shi and Malik [87] allows to segment large images thanks to its computational efficiency.

4.1.2 Quality indices

Quality indices form the second important part of the evaluation. To objectively evaluate the performance of the image segmentation methods and quality of their results, the quality indices (or measures) are necessary to adopt. The pursuit of objectivity is motivated by an effort to suppress the subjective (and still often empirical) evaluation of the segmentation algorithms in the original papers.

There exist two main approaches to design an objective measure – *unsupervised evaluation* and *supervised evaluation*. The unsupervised quality indices do not require comparison with any additional reference standard and their evaluation is solely based on a given segmented image. These indices usually exploit such criteria as intra-region homogeneity, inter-region difference etc. For a survey of unsupervised evaluation methods see [108]. Conversely the supervised performance evaluation approach requires the ground truth reference image (GT) which the actual segmented image is compared to. The ground truth image is often obtained manually by experts and reflects the optimum of the resulting segmentation. In our case the supervised evaluation is more appropriate because of the better ability to distinguish the slight disparities between the results of various segmentation algorithms thanks to the comparison with this ideal ground truth.

The following sections present quality indices used in this article. They are selected mainly to keep the diversity of the final set. On top of that they are widely used

[5] Although GrabCut is user interactive algorithm, its initialization can be done automatically with no effort (see section 4.2.1). Interactivity is thus no handicap.

[6] `http://www.opencv.org`

[7] `http://cbia.fi.muni.cz/projects/graph-cut-library.html`

in relevant papers. Each index usually favors certain properties of the segmentation results and penalizes others (they are biased in this sense). Therefore it is important to incorporate larger set of indices and handle their possibly different evaluation of given segmented image to keep the evaluation objective as much as possible. Only one or two indices would be insufficient and would probably distort the results.

It is worth mentioning that there exist more quality indices than there are described in this work. Nevertheless a lot of them are equivalent to the ones selected, like F-measure [78], Jaccard index [47] or Classification accuracy used e.g. in [55]. Some are inappropriate for the task, e.g. LCE and GCE [63], which try to deal with refinements in context of multilabel segmentation. We assume that the indices are correct, i.e. their values are meaningful and not random. The theoretical range of values is specified for each index[8]. Since we are not aware of any comprehensible survey on a supervised evaluation methods the indices are described in more detail and their formulas are included. In formulas I denotes segmented image for which the quality index is computed, GT is the corresponding ground truth, F and B subscripts denote foreground and background respectively.

Hamming distance

Hamming distance (HD) is well-known metric from the information theory [39]. Originally it counts differences between two strings. In image processing it can be used to count the number of misclassified or missegmented pixels. The distance is normalized with the total number of pixels and therefore the range is in the interval of 0 and 1, where 0 is for absolute mismatch and 1 for equality to the ground truth.

$$\mathrm{HD} = 1 - \frac{|I_B \cap GT_F| + |I_F \cap GT_B|}{|I|}$$

Huang and Dom introduced a variation called normalized Hamming distance [45], which can deal with multilabel and not only with binary segmentation. However in binary case Huang's normalized version is equivalent to plain Hamming distance[9].

Boundary Hamming distance

Boundary Hamming distance (BHD) introduced in [54] is the variation of Hamming distance that stresses the accuracy of the segmentation result on an object's boundary. Kohli et al. argue that the ordinary Hamming distance is not appropriate if the user is interested more in accurate object boundary (and so in the accurate segmentation as well), because a large qualitative improvement on the object border results in only a negligible increase of the performance measure. The quality in boundary version is then evaluated by counting the number of missegmented pixels in the region surrounding the object boundary with the specified width. As with the previous case, the distance is normalized and range is between 0 and 1.

$$\mathrm{BHD} = 1 - \frac{|I_B \cap GT_F|_{\mathrm{BOUNDARY}} + |I_F \cap GT_B|_{\mathrm{BOUNDARY}}}{|\mathrm{BOUNDARY}|}$$

In our case it makes sense to include both the Hamming distance and its boundary version, because even though we are interested in fine object boundary in the resulting

[8] Extremities of the range do not necessarily have to be reached in practice.

[9] Except for the matching problem between segmented regions. See the paper [45] for details.

image the complete missegmentation might happen and such case is better reflected (and penalized) by common Hamming distance.

Rand index and Adjusted Rand index

Rand index (RI) [75] and Adjusted Rand index (ARI) [46] are quality indices originally developed for comparing the clusterings. They are based on counting pairs of objects which two clusterings agree or disagree on (which leads to what is often called contingency table or confusion matrix). In the same manner they can compare segmentation results to the ground truth.

$$m_{ij} = |I_i \cap GT_j|,\ i,j \in \{F,B\}$$

$$m = \sum_{i,j\in\{F,B\}} m_{ij} \qquad m_{i+} = \sum_{j\in\{F,B\}} m_{ij} \qquad m_{+j} = \sum_{i\in\{F,B\}} m_{ij}$$

$$T = \frac{1}{2}\left[\sum_{i,j\in\{F,B\}} m_{ij}^2 - m\right]$$

$$P = \sum_{i\in\{F,B\}} \binom{m_{i+}}{2} \qquad Q = \sum_{j\in\{F,B\}} \binom{m_{+j}}{2} \qquad N = \binom{m}{2}$$

$$\mathrm{RI} = \frac{N + 2T - P - Q}{N}$$

The adjusted Rand index corrects the original RI for chance agreement between two clusterings by normalizing RI with its expected value. The range of RI (values between 0 and 1, where 0 is for absolute non-compliance with GT) is thus corrected to the interval of -1 and 1. It is questionable if this correction stays practical in the area of image segmentation where assumptions do not have to hold, but experimental results [102] show that it is worth considering.

$$\mathrm{ARI} = \frac{2(NT - PQ)}{N(P+Q) - 2PQ}$$

The RI and ARI are also in some sense equivalent to other well-known criteria like Cohen's Kappa statistic [15, 104] or Mirkin's metric [65], which is another adjusted form of RI [64].

Dice coefficient

Dice coefficient (DC) [24] is popular quality index for evaluating the results of image segmentation, especially in the medical imaging domain. Its range is again from 0 to 1 (1 for perfect match with ground truth).

$$\mathrm{DC} = \frac{2|I_F \cap GT_F|}{|I| + |GT|}$$

Other indices are equivalent to Dice coefficient, e.g. Jaccard index [47] and in binary case the popular F-measure [78].

Fowlkes-Mallows index

Fowlkes-Mallows index (FMI) [30] is another index based on the contingency table. It has different properties than both RI and ARI mentioned earlier. It handles the independent clusterings in a better way and behaves stably in the presence of noise (see the original paper). As with the RI the range of this index is between 0 and 1. The smaller the degree of missegmentation is the closer the index is to 1.

$$W_1 = \frac{T}{\sum_{i\in\{F,B\}} |I_i|(|I_i|-1)/2}$$

$$W_2 = \frac{T}{\sum_{j\in\{F,B\}} |GT_j|(|GT_j|-1)/2}$$

$$\mathrm{FMI} = \sqrt{W_1 W_2}$$

Normalized mutual information

Mutual information is information theoretic index which measures the amount of mutually shared information between two random variables (i.e. partitions or segmented images in our case). The more the segmented result resembles the ground truth the more information is shared. Since the mutual information has no argument-independent upper bound, Strehl and Ghosh [94] normalized it using the geometric mean of the entropies. The normalized version (NMI) thus ranges from 0 to 1 with 1 for equality to the ground truth.

$$\mathrm{NMI} = \frac{MI(I,GT)}{\sqrt{H(I)H(GT)}},$$

where $MI(I,GT)$ denotes the mutual information between I and GT, and $H(I)$ denotes the entropy of I.

Variation of information

The variation of information (VI) [64] is distance metric derived from the mutual information. Contrary to the mutual information it measures the amount of information (or entropy) which is not shared between two random variables. It would seem that VI is only a complement of NMI and their results would be equivalent. Comparison of the results however shows that they may differ, so both indices are used in evaluation. The non-normalized version of VI is used with values 0 for absolute match to the ground truth and positive values for the opposite.

$$\mathrm{VI} = H(I) + H(GT) - 2MI(I,GT)$$

Hausdorff distance and Mean absolute surface distance

Two last indices take the boundary of the segmented foreground into account. Hausdorff distance (HAUSD) measures the largest minimal distance between two boundaries. Mean absolute surface distance (MASD) measures the average minimal distance

between two boundaries (e.g. [88]). Both indices are symmetric and their values approach 0 with increasing resemblance between the segmented image and the ground truth. Both are directly connected to the distance distribution signature [45] which can be used for fast computation of these two indices.

$$d_{\min}(\mathbf{x}, B_j) = \min\left\{d_E(\mathbf{x}, \mathbf{y}) \middle| \mathbf{y} \in B_j\right\},$$

where $d_E(\mathbf{x}, \mathbf{y})$ denotes the Euclidean distance between points $\mathbf{x}$ and $\mathbf{y}$, B_j denotes set of boundary points of either I or GT. So $d_{\min}(\mathbf{x}, B_j)$ is the minimum distance of a point $\mathbf{x}$ (for example on boundary B_i) to boundary B_j.

$$h(B_I, B_{GT}) = \max\left\{d_{\min}(\mathbf{x}, B_{GT}) | \mathbf{x} \in B_I\right\}$$

$$\mathrm{HAUSD} = \max\left\{h(B_I, B_{GT}), h(B_{GT}, B_I)\right\}$$

$$\mathrm{MASD} = \frac{1}{2}\left[\bar{d}_{\min}(B_I, B_{GT}) + \bar{d}_{\min}(B_{GT}, B_I)\right],$$

where $\bar{d}_{\min}(B_I, B_{GT})$ denotes average (minimum) distance from all points $\mathbf{x}$ from B_I to B_{GT}.

4.2 Algorithms evaluation

The study of image segmentation algorithms performance is presented in this section. First, few remarks connected to the input data set and experimental setup are made. They are necessary to correctly interpret the results. Then the evaluation is carried out which mainly consists of answering two important questions – whether there is such segmentation method that would outperform the others in the studied set, and (if not) whether it is possible to choose method that is sufficiently good in the majority of cases. In final part (section 4.3) the results are analyzed in more detail and the generally applicable recommendations concerning the performance of the algorithms are proposed.

4.2.1 The input data set and evaluation setup

The algorithms for image segmentation in this thesis are evaluated on a data set of the cross-section images of the artworks described in the introduction (section 1.1). To briefly recall the images are captured in three modalities – VIS, UV and SEM – but they do not always form the triplet (SEM modality is often missing). There are 148 VIS images, 148 UV images and 89 SEM images in the data set. The SEM images are grayscale, but since the other two modalities are in RGB colorspace, we can evaluate the performance of the image segmentation algorithms in different colorspaces (or their subspaces) like LUV or LAB [73][10]. Before the very evaluation the images have to be preprocessed using procedures described in chapter 3. Especially handling the

[10]Naturally this applies only to UV and VIS images. SEM images are processed as grayscale. Also not every colorspace or its subspace is used for every segmentation method. Only those with meaningful results are included in the studied set. Finally, not every segmentation algorithm is able to work with full color data.

grinding artifacts is important as they influence the image segmentation results and their removal may improve the performance of specific methods (as was shown on example in figure 3.3). A study was conducted to find out which segmentation methods from the set are liable to grinding artifacts. The original or preprocessed image is used as an input for different segmentation methods in the following sections according to the study findings.

Next remark regarding the input data set concerns ground truth images as the reference standard for the evaluation of the image segmentation algorithms performance. They were obtained manually for each image in the input data set. The delineation of the sample boundary (i.e. the foreground) is a troublesome process even for the art restorer because of the difficulties mentioned earlier. The object boundary is not always clear. Sometimes the top or the bottom material layer is not even visible because the lack of contrast to the background or due to its properties (e.g. top varnish layer is not visible in SEM modality). This is the reason why the ground truth images of one cross-section sample may be slightly different for each modality. However the final binary masks produced in cooperation with ALMA represent suitable reference standard. There is an example of GT masks for one cross-section in all modalities in figure 4.1.

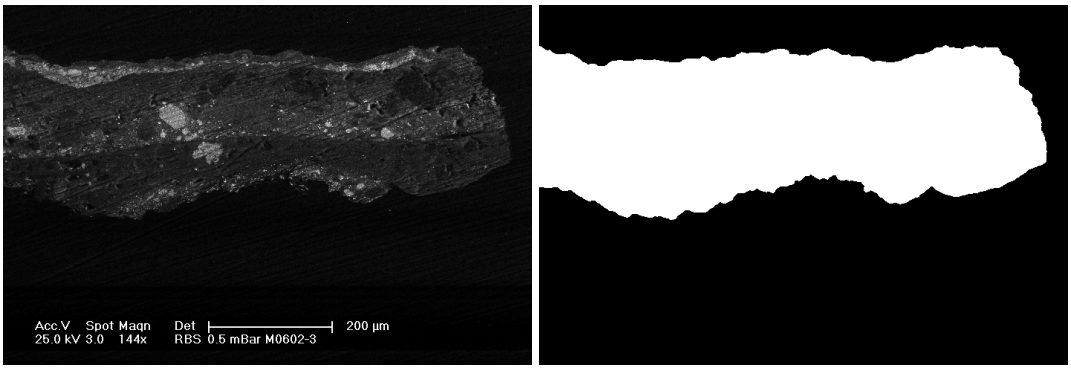
The second group of remarks is dedicated to the algorithms' parameters setting and their initialization. The behavior and so the output of the selected image segmentation algorithms can be considerably influenced by various setting of their input parameters. The parameters of some methods are plainly interpretable and as such they can be adjusted appropriately to obtain the best results. For the rest the experiments with different sets of parameters were performed and the parameter set with the best output was selected. The same goes for the parameter of BHD quality index, which is the only quality index with parameter.

The second issue is the initialization of some segmentation methods. For example the region growing demands the indication of the initial seed points. Considering the properties of the images the pixels with the most typical intensity on the border of the image (i.e. the mode) can be taken as the seed points. The algorithm then groups the pixels similar to the seeds by intensity with given tolerance (given as a parameter and added to the abbreviation, e.g. RG_25. There are 7 different parameters used in the studied set). The GrabCut algorithm requires user initialization in the form of a rectangle with a potential foreground inside. This task is done automatically in our case and the rectangle is set to cover the most of the image except for the narrow band of pixels around the image border.

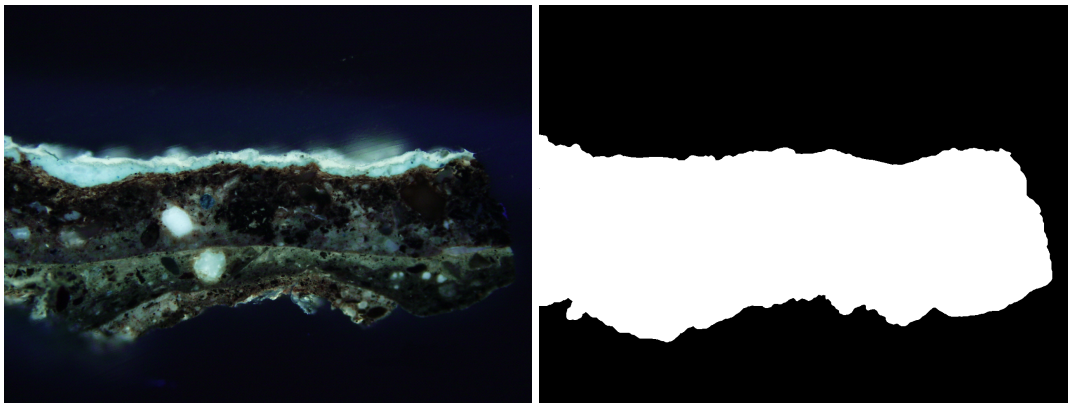
Finally, the aim is to obtain the final masks without small noisy regions in the background and with the smooth border of the foreground. Hence, the resulting binary masks after the segmentation are slightly post-processed using mathematical morphology.

4.2.2 Single best segmentation method

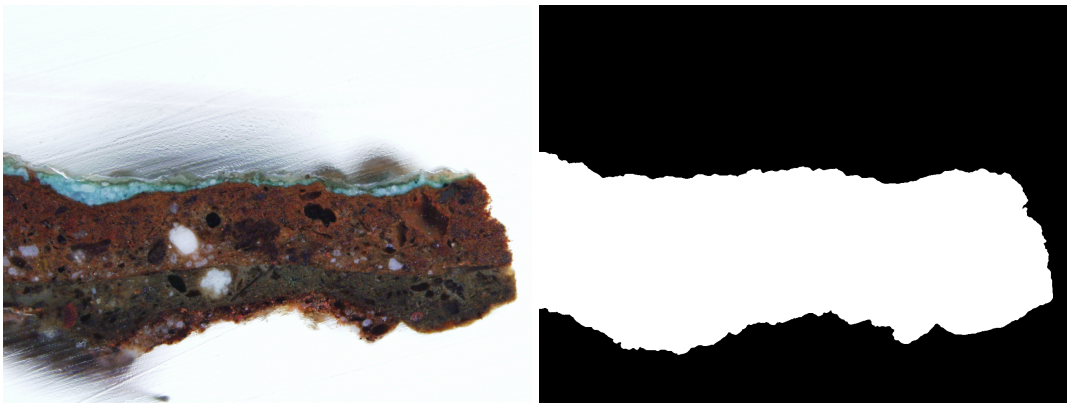
The goal of this subsection is to find out whether there is such image segmentation method in studied group of methods that solely outperforms the others in processing the input images in terms of quality. That means if there is method which gives better segmentation result for significant majority of images (or for each image in extreme case) in the data set than every other method in the group. If so, use of such method



(a) SEM



(b) UV



(c) VIS

Figure 4.1: Example of ground truth masks. In (a)-(c) in the left column there are images of one cross-section sample taken in all three modalities (SEM, UV and VIS). In the right column there are respective ground truth binary masks. White is for foreground, black is for background. Cross-section images courtesy of ALMA.

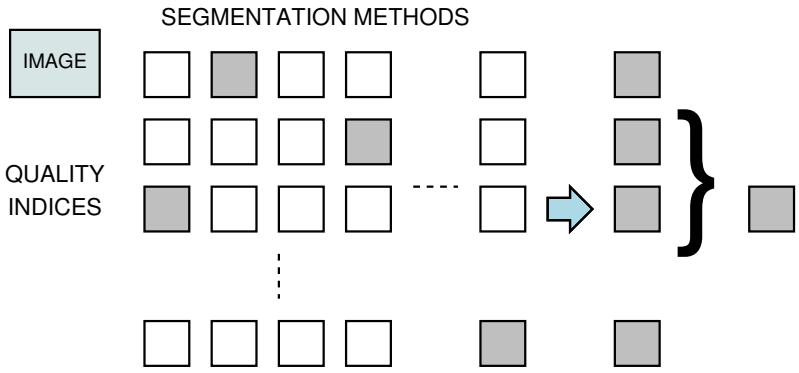


Figure 4.2: Workflow for denoting the best method for one image. Each quality index selects the best method (gray squares on the left side of the arrow). The majority rule is then applied to obtain the best segmentation method.

Algoritmus 4.1 Denotes the best segmentation algorithm for an image

Require: image I
  for all Q from the set of quality indices **do**
    $result \leftarrow$ empty vector
    for all M from the set of segmentation methods **do**
      compute Q on the result of M on I to obtain value val_Q
      $result(M) \leftarrow val_Q$
    end for
    $M_Q \leftarrow \arg\max_M \{result(M)\}$ {or min depending on the index}
  end for
  apply majority vote on all M_Q to obtain M_{BEST}
  return M_{BEST}

would be of general preference to solve background removal problem of similar data.

To study prevalence of any method first we need to denote the best segmentation algorithm for every image in input data set separately (see figure 4.2 for visualization of the procedure and algorithm 4.1 for pseudocode). Ten quality indices (described in section 4.1.2) have to be computed for every such image and every segmentation algorithm. Then the algorithm with the best result may be picked by each index for each image. It is the algorithm with the best correspondence to the respective ground truth, so the algorithm with maximum (or minimum) index value is picked. After this, there are ten possibly different segmentation methods selected by each quality index for every image. To obtain single decision for every image some combination rule has to be applied. Since the quality indices can be interpreted as ten different voters, voting rules can be successfully used in this situation. In our case the relative majority rule is considered. It means that for every image the segmentation method which is the most frequently selected as the best one by individual indices is the best segmentation method for the particular image overall. This gives us the best segmentation method for every image in input data set.

It would be useful to verify that the best segmentation method selected by quality indices according to the described procedure is also visually the best segmentation method from the set available for each image. Therefore visual comparison of all the segmentation results for every image was performed with extra focus on cases where

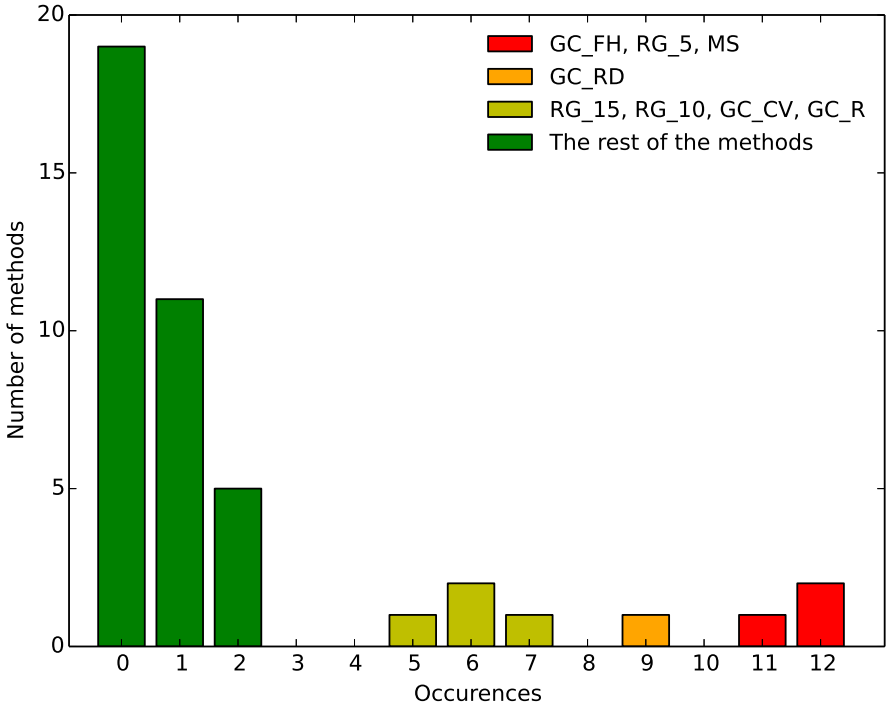
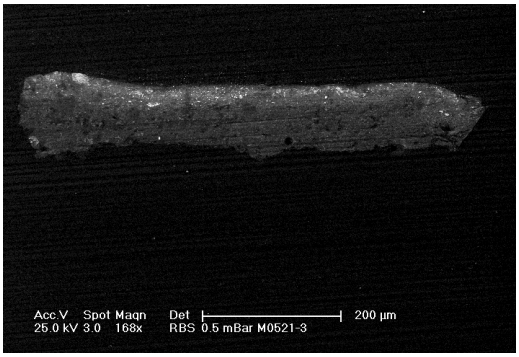


Figure 4.3: Graph of number of occurrences among the best segmentation methods for each method in SEM modality. Felzenszwalb's method, region growing (with parameter 5) and Mean Shift algorithm are the most successful methods. The majority of methods has however two occurrences at most.

the result of the selected best method was not too close to the ground truth (we need to verify that there is no better result available). The analysis leads to conclusion that the quality indices behave correctly in a vast majority of cases. The selected result is either one of the many proper ones or it is the only viable output. If there is no satisfactory result of any segmentation method, then the one visually most plausible is often selected. However there are some cases where the indices (or majority vote) do not decide entirely correctly. The selected result is not visually the best available though it is very similar to it. In such cases the decision of the indices is usually far from being unanimous. Each index may favor a different method and final decision using majority vote would be supported by small number of indices.

In any case, we have the best segmentation method denoted for every image in input data set. The key conclusion of this section is based on a distribution of segmentation methods among the best methods selected by quality indices and voting for each image. In this section we focus only on the most frequent segmentation methods which have potential to be the best. The results are presented separately for each modality. They naturally differ due to distinct character of those modalities and their input images. This gives us opportunity to study performance of the algorithms in different conditions.

The two most frequent segmentation methods in SEM modality are Felzenszwalb's method (GC_FH) and region growing (with parameter equal to 5 – RG_5) with 12 occurrences out of 89 possible (number of SEM images in total) each among the best methods. They are followed by Mean Shift algorithm (MS) and Rousson-Deriche ap-



(a) Original SEM image

(b) Ground truth image mask

(c) Final segmented mask of MS

(d) Final segmented mask of GC_RD

(e) Final segmented mask of IMJ_PER

(f) Final segmented mask of RG_5

Figure 4.4: Results of the four most successful image segmentation algorithms in SEM modality. In (a) there is an original SEM image. Ground truth mask is in subfigure (b). The image is segmented by GC_FH (c), MS (d), RG_5 (e) and GC_RD (f). Image in (a) courtesy of ALMA.

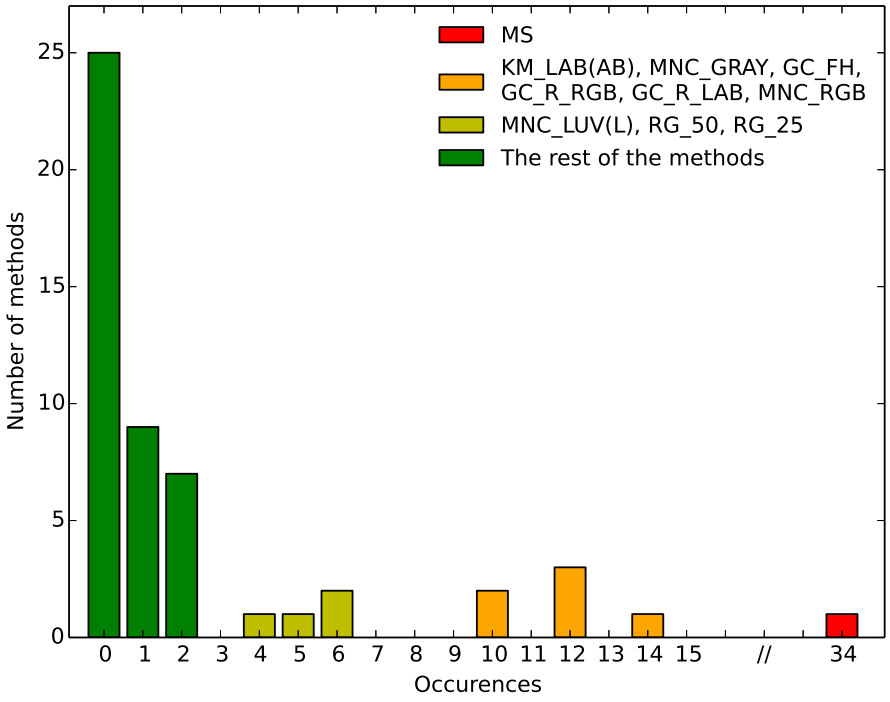


Figure 4.5: Graph of number of occurrences among the best segmentation methods for each method in UV modality. Mean Shift is by far the most successful method with colorspace versions of multiscale normalized cut, K-means, GrabCut and Felzenszwalb's method behind.

proach (GC_RD). The rest is featured in figure 4.3. 19 methods out of 43 have zero number of occurrences. Several important conclusions can be made based on this histogram. First and the most importantly, there is no segmentation method which clearly outperforms the others (12 occurrences for MS out of 89 are not sufficient enough). Second, region growing methods are quite successful, especially with smaller values of the parameter. Finally, thresholding algorithms do not perform well individually (though there are 16 occurrences in total for thresholding). Figure 4.4 shows an example of results of the four most successful methods.

The situation in UV modality is rather different. MS is clearly the most successful method. It is better than any other method in 34 cases out of 148 (the total number of UV images). K-means (KM, in AB subspace of LAB colorspace), GC_FH, GrabCut (GC_R, in RGB) and multiscale normalized cut (MNC, in grayscale) follow with 12–14 occurrences. Half of the methods (25 out of 52 precisely) are not among the best methods in at least one case. The rest is displayed in figure 4.5. As in SEM modality there is no clear winner which could be mechanically used for segmentation of UV images. MS is indeed very successful, but it outperforms the others only in quarter of cases which is not sufficient. Surprisingly, GC_RD and Chan-Vese approach (GC_CV) fail completely with one and zero occurrences respectively. Region growing does not perform that well as in SEM modality. Thresholding methods represent only a complement to more successful methods.

Finally, the results for VIS modality are presented. MS stays the most frequent among the best methods for each image with 40 occurrences out of 148 possible. Ver-

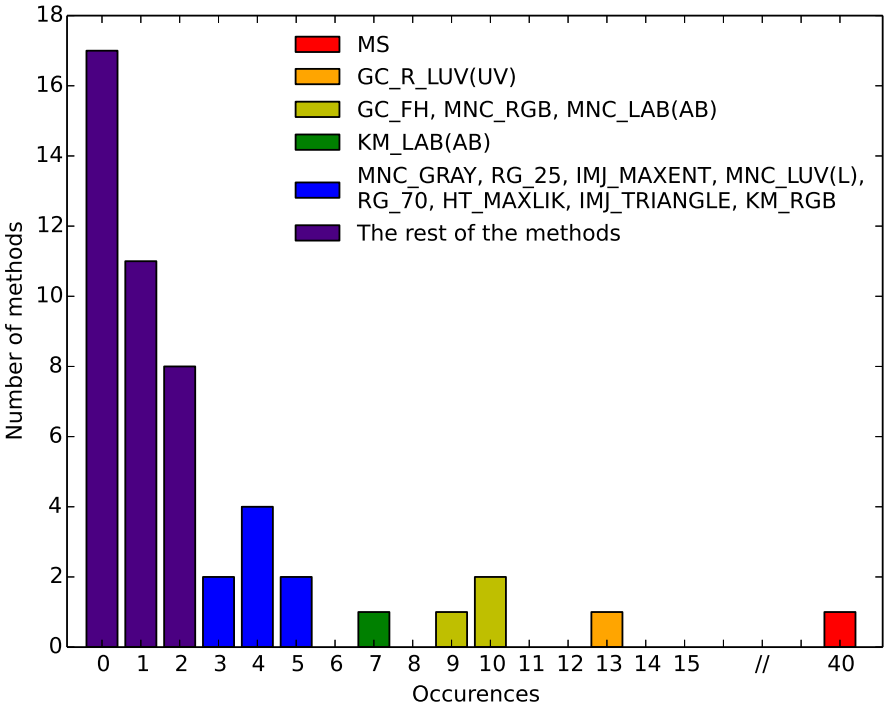
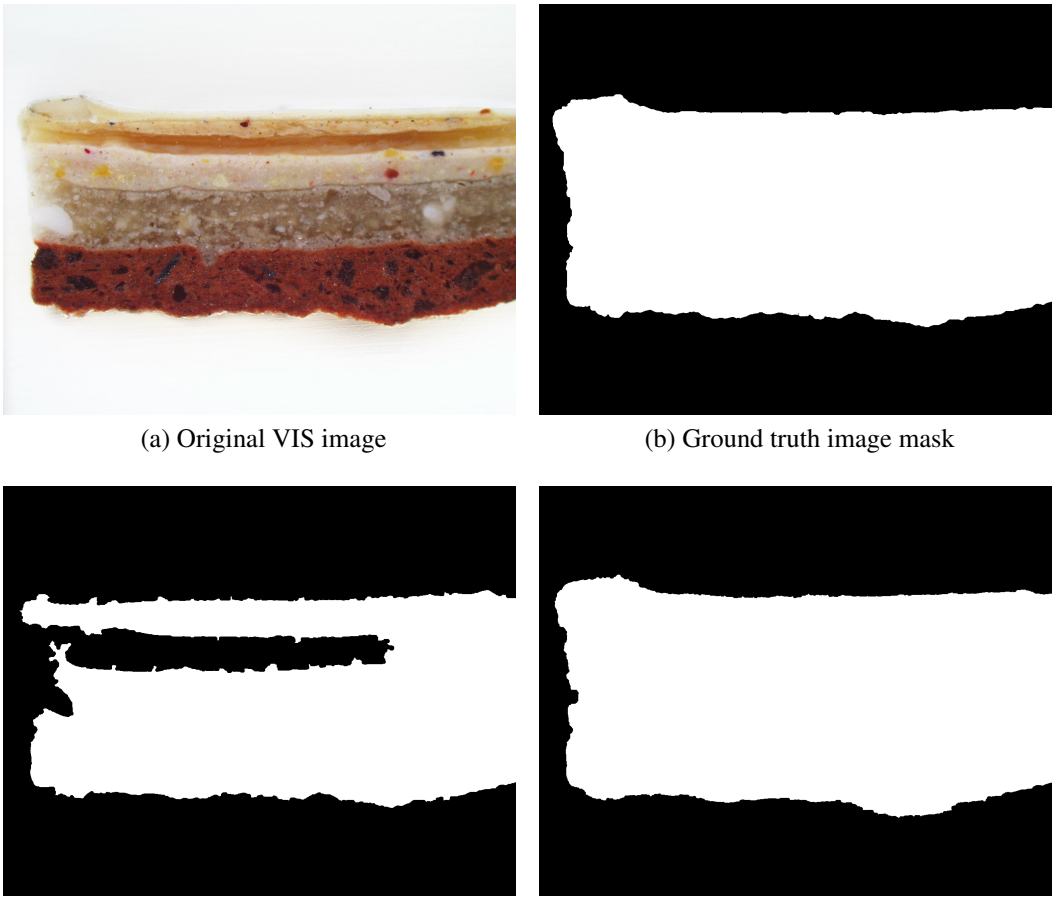


Figure 4.6: Graph of number of occurrences among the best segmentation methods for each method in VIS modality. Mean Shift is a method with the most occurrences. GrabCut follows with large gap and Felzenszwalb's method and colorspace variations of multiscale normalized cut are behind. Lots of methods have two occurrences at most.



(a) Original VIS image

(b) Ground truth image mask

(c) Final segmented mask of MS

(d) Final segmented mask of RG_25

Figure 4.7: Demonstration that the selected best method is not perfect for all images. The image in (a) is better segmented by region growing with parameter 25 (RG_25, in (d)) than by Mean Shift (MS, in (c)) which is the best method in VIS modality. RG does not perform nearly that well overall. In (b) there is a ground truth image for reference. Image in (a) courtesy of ALMA.

sions of GC_R and MNC in various colorspaces and GC_FH follow with roughly 10 occurrences. The rest can be seen in histogram in figure 4.6. 17 methods out of 50 are not selected as the best method at least once. The conclusions for UV modality hold also here. MS outperforms the other methods in lots of cases, nevertheless not in the significant majority. GC_RD and GC_CV approaches fail again. Region growing is not very successful and where it is, the bigger parameter values are used. In contrast to UV, thresholding methods represent alternative to more sophisticated methods. They are selected as the best ones for 31 images in total.

Based on these facts we can say that there is no segmentation method which significantly outperforms the other segmentation algorithms in the set. It happens only in fraction of cases (13–27 percentage depending on modality). The use of the most frequent method mentioned in previous paragraph (e.g. MS for UV modality) for background removal in images similar to those in our data set is not sufficient for achieving perfect results (see figure 4.7 for example of an image where the best method does not perform that well). It is important to keep in mind that potential user usually does not have the ground truth images, so he cannot select the individual best method for every

Algoritmus 4.2 Denotes the best average segmentation algorithm overall

for all Q from the set of quality indices **do**
  $result, medians \leftarrow$ empty vectors
  for all M from the set of segmentation methods **do**
    for all I from the set of input images **do**
      compute Q on the result of M on I to obtain value val_Q
      $result(M,I) \leftarrow val_Q$
    end for
    $medians(M) \leftarrow \underset{I}{\text{median}}\{result(M,I)\}$
  end for
  $M_Q \leftarrow \arg\max_M\{medians(M)\}$ {or min depending to the index}
end for
apply majority vote on all M_Q to obtain M_{BESTAVG}
return M_{BESTAVG}

sole image. Additional conclusions can be made from the results. MS, GC_FH, GC_R and MNC often perform well. But also more straightforward approaches such as RG or thresholding can be used to achieve good results.

4.2.3 Best average segmentation methods

The evaluation in the previous section is not entirely fair. The focus was on finding a segmentation method which was the best for significant majority of images. There was no such method in the studied set. However what if there is a method which is good enough (and not necessarily the best) for vast majority of the images. We look for method which is comparable to the best method in case of easy to segment images (majority methods can segment this image with satisfactory results) and does not completely fail in case of worse images (where most of the methods fail), i.e. the best average segmentation method. Such method (if found) could be used as number one choice to solve the image segmentation problem.

The starting point for the evaluation is the same as in the previous section. The values of ten quality indices are computed for each image and segmentation method. However following steps differ from the previous procedure (see algorithm 4.2). There are so many values as there are images for every pair of quality index and image segmentation method. Median of these values is the average performance of segmentation method according to the respective index. The best average method is thus the method with the highest median (or the lowest depending on the index). Finally the majority rule denotes the best average segmentation method as a consensus of all quality indices. The median is preferred over the mean because vectors of numbers often contain several outliers which would distort the results inappropriately[11]. Table 4.2 shows median values for each quality index and several selected segmentation methods in SEM modality.

Felzenszwalb's method (GC_FH) and Rousson-Deriche approach (GC_RD) are the two best average methods for SEM modality (they were selected equally by the

[11] Outlier means that segmentation method segments some image exceptionally well or poorly. Outlier is the value of the quality index for such image. We are interested in average performance which has to be stable despite the outliers. That is why the median is more suitable for the task.

Segmentation methods	Quality indices				
	BHD	HD	RI	ARI	VI
	$[0,1]$	$[0,1]$	$[0,1]$	$[-1,1]$	$[0,\dots)$
GC_RD	0.84 (0.12)	0.98 (0.03)	0.96 (0.06)	0.90 (0.15)	0.29 (0.31)
GC_FH	0.82 (0.13)	0.98 (0.03)	0.96 (0.06)	0.90 (0.20)	0.28 (0.24)
MS	0.82 (0.14)	0.97 (0.04)	0.95 (0.08)	0.88 (0.23)	0.33 (0.34)
GC_CV	0.84 (0.14)	0.97 (0.06)	0.95 (0.11)	0.88 (0.31)	0.32 (0.37)
IMJ_IME	0.81 (0.14)	0.97 (0.04)	0.94 (0.07)	0.89 (0.20)	0.32 (0.31)
RG_10	0.82 (0.15)	0.97 (0.05)	0.94 (0.10)	0.87 (0.23)	0.33 (0.38)
IMJ_TRIANGLE	0.77 (0.18)	0.97 (0.09)	0.93 (0.15)	0.86 (0.39)	0.39 (0.47)
GC_R	0.73 (0.22)	0.96 (0.10)	0.93 (0.17)	0.82 (0.41)	0.37 (0.45)
KM	0.63 (0.17)	0.87 (0.20)	0.78 (0.27)	0.40 (0.55)	0.77 (0.57)
IMJ_OTSU	0.61 (0.17)	0.84 (0.18)	0.74 (0.24)	0.38 (0.53)	0.82 (0.48)
TNC	0.49 (0.29)	0.81 (0.28)	0.70 (0.34)	0.01 (0.84)	0.81 (0.55)
RG_70	0.49 (0.09)	0.75 (0.19)	0.64 (0.17)	0.02 (0.19)	0.93 (0.28)
MNC	0.50 (0.05)	0.57 (0.17)	0.51 (0.05)	0.01 (0.09)	1.66 (0.35)
IMJ_SB	0.46 (0.04)	0.70 (0.19)	0.58 (0.12)	0.00 (0.00)	0.88 (0.24)

	FMI	DC	NMI	HAUSD	MASD
	$[0,1]$	$[0,1]$	$[0,1]$	$[0,\dots)$	$[0,\dots)$
GC_RD	0.96 (0.05)	0.97 (0.07)	0.82 (0.22)	40.31 (65.19)	4.57 (8.03)
GC_FH	0.96 (0.04)	0.96 (0.10)	0.83 (0.26)	32.60 (54.43)	4.43 (7.28)
MS	0.96 (0.07)	0.95 (0.11)	0.81 (0.23)	45.50 (68.99)	5.71 (10.63)
GC_CV	0.96 (0.08)	0.94 (0.16)	0.79 (0.33)	53.48 (71.93)	5.79 (13.93)
IMJ_IME	0.96 (0.06)	0.95 (0.09)	0.80 (0.25)	48.71 (94.05)	5.82 (11.98)
RG_10	0.95 (0.08)	0.95 (0.10)	0.78 (0.26)	57.24 (97.56)	6.40 (12.15)
IMJ_TRIANGLE	0.95 (0.10)	0.93 (0.26)	0.77 (0.38)	54.58 (127.65)	7.00 (22.40)
GC_R	0.94 (0.11)	0.92 (0.16)	0.74 (0.37)	56.04 (71.01)	7.98 (21.90)
KM	0.85 (0.17)	0.62 (0.46)	0.39 (0.43)	118.17 (177.56)	29.42 (47.81)
IMJ_OTSU	0.82 (0.16)	0.60 (0.48)	0.36 (0.40)	124.39 (192.89)	34.48 (58.20)
TNC	0.81 (0.19)	0.12 (0.89)	0.05 (0.71)	361.82 (449.15)	111.70 (129.26)
RG_70	0.78 (0.11)	0.11 (0.37)	0.08 (0.20)	370.18 (333.21)	98.91 (82.07)
MNC	0.59 (0.09)	0.39 (0.31)	0.02 (0.11)	197.50 (75.12)	63.94 (23.54)
IMJ_SB	0.76 (0.10)	0.00 (0.01)	0.01 (0.02)	523.64 (181.98)	144.72 (41.20)

Table 4.2: Table with median values and interquartile ranges in brackets (both rounded to two decimal places) of all ten quality indices for several selected segmentation methods in SEM modality. Median value is the average performance of a segmentation method on a set of images according to a quality index. There are the six most successful methods, several representative methods in the middle and the two worst methods according to evaluation in section 4.2.3 (in this order). SEM modality is chosen for demonstration due to bigger variance in indices values for different methods in different places of the ranked list than it is in other two modalities.

Quality indices

BHD	HD	RI	ARI	VI	FMI	DC	NMI	HAUSD	MASD
GC_RD	GC_RD	GC_RD	GC_RD	GC_FH	GC_FH	GC_RD	GC_FH	GC_FH	GC_FH
GC_CV	GC_FH	GC_FH	GC_FH	GC_RD	GC_RD	GC_FH	GC_RD	GC_RD	GC_RD
RG_15	MS	MS	IMJ_IME	GC_CV	GC_CV	IMJ_IME	MS	MS	MS
RG_10	GC_CV	GC_CV	GC_CV	IMJ_ME	MS	MS	IMJ_IME	IMJ_IME	GC_CV
MS	IMJ_IME	IMJ_IME	MS	RG_10	IMJ_IME	RG_10	GC_CV	GC_CV	IMJ_IME

(a) SEM

Quality indices

BHD	HD	RI	ARI	VI	FMI	DC	NMI	HAUSD	MASD
MS	MS	MS	MS	MS	MS	MS	MS	GC_FH	MS
MNC_RGB	GC_FH	GC_FH	GC_FH	GC_FH	GC_FH	GC_FH	GC_FH	MS	GC_FH
MNC_GRAY	MNC_GRAY	MNC_GRAY	MNC_RGB	MNC_GRAY	MNC_GRAY	MNC_GRAY	MNC_RGB	MNC_RGB	MNC_RGB
MNC_LAB(AB)	MNC_RGB	MNC_RGB	MNC_GRAY	MNC_RGB	MNC_RGB	MNC_RGB	MNC_GRAY	GC_R_LAB(AB)	MNC_GRAY
IMJ_LI	RG_25	RG_25	RG_25	RG_25	RG_25	MNC_LAB(AB)	MNC_LAB(AB)	MNC_GRAY	RG_25

(b) UV

Quality indices

BHD	HD	RI	ARI	VI	FMI	DC	NMI	HAUSD	MASD
MS	MNC_RGB	MNC_RGB	MNC_RGB	MS	MNC_RGB	MS	MS	GC_FH	MS
MNC_RGB	MS	GC_FH	MS	MNC_RGB	MS	MNC_RGB	GC_FH	IMJ_ISO	MNC_RGB
IMJ_HUANG	KM_RGB	MS	GC_FH	KM_RGB	IMJ_MOM	IMJ_HUANG	MNC_RGB	IMJ_OTSU	GC_FH
RG_20	IMJ_OTSU	KM_RGB	KM_RGB	GC_FH	KM_RGB	IMJ_MEAN	IMJ_HUANG	KM_RGB	MNC_LUV(L)
RG_25	IMJ_ISO	IMJ_OTSU	IMJ_OTSU	IMJ_OTSU	TNC	TH_MEAN	IMJ_MEAN	TH_INTERI	IMJ_HUANG

(c) VIS

Table 4.3: Lists with first five segmentation methods (rows) according to every quality index (columns) in all modalities. Lists are sorted by median values, thus by average performance of segmentation methods.

indices). If we look on the problem of finding the best average segmentation method even in more detail and consider first five methods for each quality index (assuming that the lists for each index are sorted by median values, thus by performance), we can see that GC_FH and GC_RD occupy the first two positions of almost every list (there is only one exception) in SEM modality (see table 4.3(a)). Considering the median values there is a noticeable gap between these two and next methods in the list. This second cluster is formed by Chan-Vese approach (GC_CV), Mean Shift (MS) and minimum error thresholding (IMJ_IME). Apart from them there are several occurrences of region growing with parameters 10 and 15 on lower positions. MS holds its superiority in UV modality even as the best average method (table 4.3(b)). It is first for 9 out of 10 quality indices (only HAUSD votes for GC_FH) with substantial performance gap from the second position which is occupied almost only by GC_FH (except for HAUSD naturally). Two colorspace versions of multiscale normalized cut (MNC, RGB and grayscale) fill the third and the fourth position. The last one with another noticeable loss in performance is mainly region growing with parameter 25 (RG_25). There are sporadic occurrences of other methods from studied set on lower positions, but nothing of importance. The result in VIS modality is not so clear. Majority vote denotes MS to be the best average method, since five quality indices vote for it (table 4.3(c)). Nonetheless four indices are for MNC (in RGB) and one for GC_FH. The rest of the first five positions is shared by plenty of different methods including thresholding, RG, K-means etc. The conclusion is that there exist four very good methods which can be used as number one choice depending on the modality. It is GC_FH and GC_RD for SEM, MS for both UV and VIS modality, in the latter case supported by MNC (in RGB).

The evaluation of previous paragraph can be done more rigorously with the removal of the following shortcoming in addition. The choice of the best average method (and four runners up) was based on the position within ten sorted list coming from ten quality indices. Unfortunately the situation when one method was chosen as the best one by several indices and given a lower rank by others was not taken into account because only first five positions were considered. Therefore, the results could be little bit inaccurate. This drawback can be amend by exploiting the information about performance of all the methods from all the indices, i.e. by processing complete sorted lists of indices' values. The goal is to combine all ranked lists to the single ordering which would express input preferences in the best way. This is called a rank aggregation problem and is extensively studied in different fields (elections, web search etc.). See for example [26] in context of web searching. We use RankAggreg package [72] for R statistical software[12] for our evaluation. It implements optimization techniques necessary to produce final ranked list[13]. As a result there is one list of image segmentation methods sorted by their performance (according to quality indices) for each modality. This list represents consensus of ten input lists as individual voters with preferences.

It would be reasonable to deeply analyze positions of every segmentation method in the final lists. Hence we focus only on several prominent methods, interesting results and general position of different approaches (comprehensive analysis is given below

[12] `http://www.r-project.org`

[13] Optimization is unavoidable because due to amount of data (ten relatively long lists) the exact solution cannot be computed in feasible time. However exact solution can be computed for short input lists and they more or less match the corresponding part of presented optimization results. Unfortunately implemented optimization algorithms do not necessarily find a global optimum and can get stuck in a local one. The scripts were therefore executed many times to obtain the solution as good as possible.

SEM	GC_RD, GC_FH, MS, GC_CV, IMJ_IME, RG_10, RG_15, HT_ME, IMJ_TRIANGLE, IMJ_MEAN, HT_MEAN, HT_IME, GC_R, IMJ_HUANG, RG_20, IMJ_LI, RG_5, RG_25, KM, HT_INTER, HT_INTERI, IMJ_DEF, HT_CONCAV, IMJ_ISO, IMJ_OTSU, RG_50, HT_MOM, IMJ_MOM, IMJ_PER, HT_IM, TNC, IMJ_IM, HT_MEDIAN, IMJ_RENYI, RG_70, IMJ_YEN, HT_MIN, HT_MAXLIK, HT_ENT, IMJ_MAXENT, IMJ_MIN, MNC, IMJ_SB
UV	MS, GC_FH, MNC_GRAY, MNC_RGB, RG_20, GC_R_LAB(AB), RG_25, GC_CV, RG_15, IMJ_TRIANGLE, KM_LAB(AB), HT_MEAN, IMJ_HUANG, RG_50, TNC, GC_R_LAB, IMJ_LI, IMJ_MEAN, RG_10, KM_GRAY, KM_LAB, RG_70, HT_INTER, HT_ME, HT_INTERI, IMJ_DEF, KM_RGB, MNC_LAB(AB), IMJ_OTSU, GC_R_LAB(L), HT_CONCAV, IMJ_ISO, MNC_LUV(L), HT_MOM, GC_RD, IMJ_MOM, HT_IM, HT_MAXLIK, IMJ_IM, GC_R_RGB, HT_MIN, IMJ_YEN, IMJ_MIN, IMJ_RENYI, HT_ENT, IMJ_MAXENT, IMJ_IME, RG_5, HT_IME, IMJ_PER, HT_MEDIAN, IMJ_SB
VIS	MS, MNC_RGB, KM_RGB, IMJ_OTSU, IMJ_ISO, IMJ_DEF, IMJ_HUANG, HT_INTERI, TNC, MNC_LUV(L), GC_CV, HT_INTER, KM_LAB, KM_GRAY, IMJ_MEAN, IMJ_MOM, IMJ_IME, HT_MEAN, HT_MOM, IMJ_IM, RG_70, RG_50, IMJ_LI, MNC_GRAY, HT_IM, IMJ_RENYI, GC_FH, IMJ_MIN, HT_MAXLIK, HT_MIN, GC_R_LUV(UV), IMJ_YEN, KM_LAB(AB), RG_25, MNC_LAB(AB), RG_20, HT_ENT, GC_R_LUV(L), RG_15, IMJ_TRIANGLE, GC_RD, HT_CONCAV, HT_ME, IMJ_PER, IMJ_MAXENT, HT_MEDIAN, RG_10, HT_IME, RG_5, IMJ_SB

Table 4.4: Final lists of segmentation methods sorted according to their average performance (the best in the first place) in all three modalities.

in section 4.3). The complete lists are appended in table 4.4. Rousson-Deriche approach (GC_RD) and Felzenszwalb's method (GC_FH) stay the best average methods in SEM modality with that GC_RD is the best one. This result is little bit surprising, because GC_RD was not so successful as the best method overall (in previous section 4.2.2) and nothing indicated that it would outperform the others on average. Mean Shift algorithm (MS) and Chan-Vese approach (GC_CV) follow the two. Iterated and normal version of minimum error thresholding is very successful (both ImageJ and HistThresh, i.e. IMJ_IME, HT_IME and HT_ME), as well as Triangle and Mean approaches (IMJ_TRIANGLE and IMJ_MEAN). Region growing (RG) with parameters 10 and 15 occupies position 6 and 7 in the list, other parameters are scattered in the middle. From already mentioned methods K-means (KM) and GrabCut (GC_R) rather disappoint with its results and multiscale normalized cut (MNC) completely fails with the last but one position.

MS is the best average algorithm in UV modality, which only confirms its dominance. It is followed by GC_FH and grayscale and RGB versions of MNC, which is the opposite to SEM modality, where grayscale version fails. Parameters 15, 20 and 25 of RG are suitable for UV modality as they are placed in top 10 also with GC_CV method. IMJ_TRIANGLE, IMJ_MEAN, IMJ_HUANG and IMJ_LI are the most useful thresholding methods. Several colorspace alternatives of KM are ranked in the top half. Contrary to SEM modality GC_RD method is not very good as it is ranked in bottom half of the list. The least successful method is Shanbhag (IMJ_SB) approach to thresholding. It is interesting that this method was voted as the best one overall for one image (previous section 4.2.2) despite its uselessness on average.

MS is the best average algorithm also in VIS modality, but otherwise the situation differs a lot compared to previous two modalities. In the second and third place there are RGB version of MNC and RGB version of KM algorithm. Apart from them top 10 consists further from thresholding methods, IMJ_OTSU, IMJ_ISO, IMJ_HUANG and Tao's thresholding method (TNC) to name several. GC_CV algorithm produces satisfactory results. GC_FH, GC_R or GC_RD do not perform very well. Concerning RG approach its results are generally worse than in the previous two modalities. However higher values of parameter like 50 or 70 are definitely better than smaller ones. IMJ_SB thresholding is again the worst segmentation method on average.

The evaluation in this section delivers very interesting results. The construction of lists of segmentation methods sorted by algorithms' performance according to ten selected quality indices is the most important. The ordering allows the future user to pick the suitable segmentation method for his problem and character of data (which are represented by different modalities in this work). The lists also provides an insight to performance of different segmentation methods and their comparison. The conclusions about the performance depend on the specific modality, but generally some resume can be made. Mean Shift algorithm performs very well in all three modalities and can be declared the best average method overall. Felzenszwalb's method, Rousson-Deriche and Chan-Vese approaches, and multiscale normalized cut may deliver excellent results as well. Region growing is not a bad choice either, but its performance depend on the chosen parameter. Thresholding can be good alternative too, but the choice of specific algorithm has to respect the properties of data. Segmentation methods which take place at the end of the lists perform badly on average, however that does not necessarily mean that they perform badly on every image (for example see figure 4.7, where region growing outperforms the best method on average – Mean Shift. RG_25 is ranked in

the bottom half.). Furthermore they may provide important diversity for segmentation fusion/combination or other processing (chapter 5). More discussion and conclusions are presented in section 4.3.

One remark concerning correctness of the above evaluation has to be made before closing this section. The comparison does not take into account the absolute values of quality indices. So it is possible that the best average segmentation method is certainly better than the rest of the methods in the studied set, but absolutely its performance is poor with useless results. However it is not the case. The segmentation methods at the top of the lists obtained relatively high values from the quality indices (and vice versa for the methods at the bottom). See table 4.2 for reference in case of SEM modality. The further evaluation was performed to support this conclusion more precisely. The output of segmentation method on one image was marked good if its index value was above specified threshold (and bad if it was below another). Afterward all the methods were ranked according to the number of their occurrences in a set of good outputs and a set of bad outputs. The results of this evaluation did not differ much with the results of this section described above.

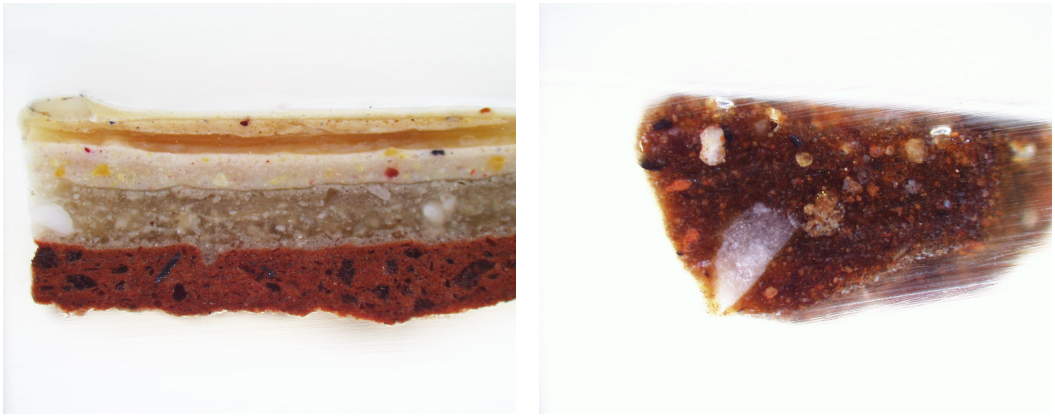
4.3 Discussion of the achieved results

In this section, deeper analysis of the evaluations and their results is presented. We will use it to give recommendations for the application of studied image segmentation methods in different situations, i.e. for different data. First the distinct features of each modality (SEM, UV and VIS as shown in figure 1.2) are examined in more detail. Then the performance of each segmentation approach and its connection to input images (or modality) is evaluated to make clear in which situations which image segmentation methods perform the best.

SEM modality images are products of scanning electron microscope. This technique enables to study chemical contrast of different materials. In the image it is expressed by varying texture of the cross-section in contrast to relatively homogeneous background. Thus the boundary edges between the cross-section as foreground object and the background are usually sharp and clear. The cross-section has generally different intensity values than the background. All this could make the segmentation quite easy. However in case of our data set the task is sometimes complicated with the artifacts induced by scanning microscope, and certain materials used in the paintings do not have sufficient contrast response so the boundary edge is not sharp enough.

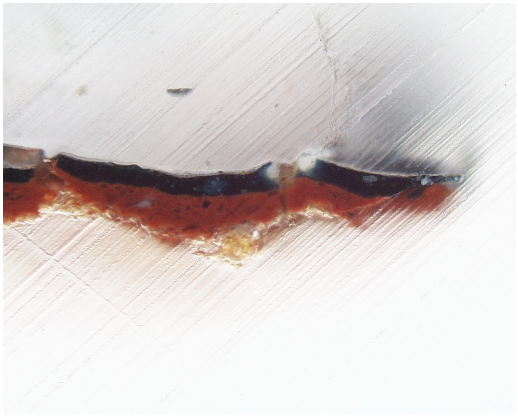
UV modality is similar to SEM in that the background is homogeneous. UV light reveals a possible fluorescent property of certain materials. Such materials have bright response (typically green, turquoise or blue) in the image. Non-fluorescent materials are on the other hand often dark and they blend with the background which is dark by definition due to absence of fluorescent property of polyester resin. Another problem is that the non-surface parts of the cross-section can shine through transparent resin and form blurred shadows on the borders of the cross-section. Satisfactory background removal can therefore be quite challenging.

VIS modality captures optical properties in visible spectrum. The sharpness of cross-section boundary varies from high contrast edge to fluent transition to background depending on the material color. The transparency of polyester also remains a problem in VIS modality. The difficulty of background removal is thus similar to UV modality in this aspect. In addition the background is not uniform. The lighting can

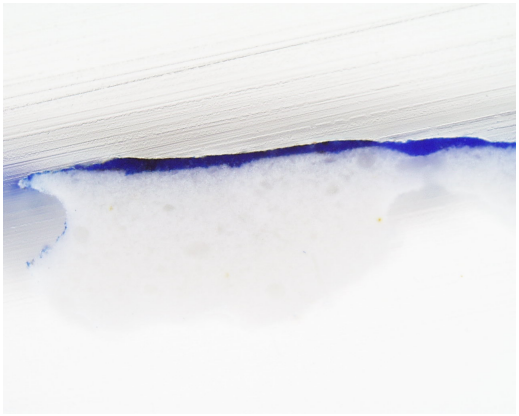


(a)

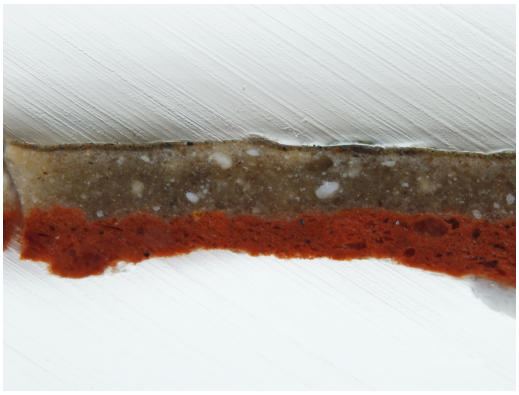
(b)



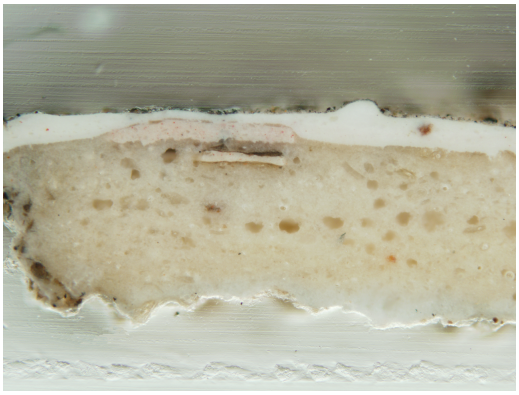
(c)



(d)



(e)



(f)

Figure 4.8: Set of six VIS images demonstrating different properties which cause problems for image segmentation. In figure (a) there is neat and relatively easy to segment image for comparison. Other images demonstrates non-uniform illumination of the background ((e), (f)), problematic transparency of the polyester resin ((b), (c)), grinding artifacts ((c)-(f)), air bubbles and defects in the background ((c), (f)) and finally unclear boundary edge between cross-section and background (d). Image courtesy of ALMA.

be reflected unevenly and there can be lot of different artifacts like air bubbles which are not visible in other modalities. Also grinding artifacts may be a problem as was mentioned before. Figure 4.8 gives examples of distinct properties of VIS images.

To summarize key properties of the modalities SEM modality generally represents images with sharp and contrast boundary edges, relatively homogeneous background and often clear separation of object and background intensity values. UV modality images have uniform background, but unclear boundary edges between background and certain (non-fluorescent in our case) parts of the foreground object, also transparency of the resin is the problem. VIS images are similar to UV in problems with unclear boundary edges and transparency of the resin. Difference is in more problematic background which is not uniform and contains artifacts.

Discussion about the usability of studied segmentation methods starts with simpler approaches, i.e. region growing, thresholding and K-means[14]. *Region growing* generally delivers satisfactory results when there is relatively homogeneous background and boundary between desired segmented object and background is apparent. In our case it is demonstrated on SEM and UV modalities where the background surrounding the cross-section is more or less uniform. Tolerance to non-uniformity is given by parameter. The smaller values of parameter are sufficient for images in SEM, while slightly higher values are required for UV to compensate the transparency mentioned above. Region growing is then placed in top 10 of the best average methods. VIS modality is different. The background there is more variable in such way that it almost prohibits compensation with high parameter values (region growing would easily cross the border between background and foreground object in that case). This being said high values of parameter are more suitable in VIS. Overall region growing approach can provide satisfactory results comparable to more complicated methods if the assumptions of relatively uniform background and clear border are met.

Thresholding methods (not only those in the studied set) differ in the way they find the threshold to divide pixels into two groups. Strictly bimodal histogram would be an optimum situation, however such case is not very common in our input data set (and in real images neither). Therefore some methods are more successful in handling non-optimum case than others. In SEM modality where the background pixels in histogram are easier to separate Triangle (IMJ_TRIANGLE), Mean (IMJ_MEAN) and minimum error method (IMJ_IME) are the most successful. On the other side of spectrum there are entropy-based methods (IMJ_MAXENT, IMJ_RENYI, IMJ_SB, HT_ENT) and several others (HT_MAXLIK, IMJ_YEN, IMJ_MIN). In UV modality the intensity values of the foreground often blend with those of the background, which is difficult condition for thresholding. Triangle, Huang (IMJ_HUANG), Mean and Li (IMJ_LI) methods handle it well on average. The spectrum of failing methods stays the same as in SEM modality. IMJ_IME produces disappointing results too. Though the image properties of VIS modality are similar to those of UV mostly different thresholding methods are satisfactory in VIS. Otsu (IMJ_OTSU and HT_INTER), IsoData (IMJ_ISO, IMJ_DEF and HT_INTERI) and Huang are among the most successful methods. Concerning Tao's thresholding approach (TNC) it succeeds in UV and VIS modalities, while it fails in SEM. Thus it deals better with visually hard cases with smooth transitions between background and foreground than in cases where the intensity values of the foreground object are clearly separated from those of the back-

[14]Concerning different colorspaces region growing and thresholding exploited only the grayscale information in all three modalities. K-means was evaluated in more colorspaces.

ground.

The results of *K-means* (KM) approach are highly dependent on colorspace (or subspace) which the input data are in and on overall color profile of the images in different modalities. Grayscale (the only one for SEM), LAB (plus AB only subspace version) and RGB variants are analyzed. KM in grayscale produces merely mediocre results on average in all three modalities. Same thing can be said on account of full LAB space variant (in case of UV and VIS) with slightly better results in VIS. However interesting results appear concerning KM in AB subspace of LAB and RGB. Both can perform well depending on color profile of the image. In UV modality where the images are mainly darker with dominant responses in blue or green the AB variant is placed in top positions of the ranked list. RGB variant performs much worse. The situation is opposite in VIS modality. RGB variant is the third best average method while AB variant takes place in two thirds of the ranked list. It is clear that successful use of K-means depends on the overall color dominance of input images. Generally its results can be quite satisfactory.

After more straightforward approaches were analyzed we will now focus on more complex segmentation methods in the studied set[15]. *Felzenszwalb's method* (GC_FH) performs very well being the second most successful average segmentation method in SEM and UV modalities. However it does not perform that well in the remaining VIS modality. The algorithm has apparent problems with converging to stable result when the border of the object is unclear and background is not homogeneous (and in that sense resembles the foreground object). In such cases the segmented result is often blank image. Apart from that GC_FH can be excellent method for segmentation which copes with other mentioned problematic image properties appropriately. *Daněk's optimization of Chan-Vese and Rousson-Deriche functionals* is very successful for the easy to segment images with clear and sharp border between object and surrounding background (GC_RD is the best average method in SEM, GC_CV being the fourth). Otherwise they struggle with unclear transitions and transparency. GC_RD fails in UV and VIS modality, GC_CV still manages to take position in top third of the average ranked list, but its results are often dissatisfactory. The results of *multiscale normalized cut* (MNC) differ with various colorspace configurations. MNC produces very good results when the original RGB colorspace is conserved (second place in VIS modality and fourth place in UV modality average ranked list). Also the exploitation of only the intensity channel (grayscale or lightness from LUV) can be profitable in case of UV and VIS. In all other cases MNC rather fails, especially in SEM modality. *GrabCut algorithm* (GC_R) provides perhaps the worst results from group of more advanced segmentation methods and cannot be recommended for unsupervised segmentation in similar setting. Originally it is based on user interaction and its power lies in additional adjustment of initial segmentation. *Mean Shift* is the last algorithm to discuss. According to the results of evaluation it is the best average segmentation method in the studied set. It can handle problematic image properties well and its outputs often outperforms the rest (see section 4.2.2).

With regard to the analyses above Mean Shift algorithm should be number one

[15]From those Felzenszwalb's method is applied to the images in original colorspaces. That means grayscale in case of SEM modality and RGB colorspace in case of UV and VIS. Processing in different colorspaces delivers comparable results. Mean Shift segmentation followed the original paper and LUV space is used. Daněk's version of Chan-Vese and Rousson-Deriche use the grayscale information. So only the performances of multiscale normalized cut and GrabCut algorithm are analyzed in different colorspaces.

Images in general	– Mean Shift algorithm would be number one choice
Image with relatively homogeneous background and apparent boundary edge between object and background	– Region growing with appropriate parameters – Felzenszwalb's method (even in the case of not so clear boundary edge and partial blending of the object and the background) – Chan-Vese and Rousson-Deriche approaches optimized by Daněk
Image with possibly unclear boundary edges between object and background, presence of shadows or halos around boundaries	– Multiscale normalized cut in RGB or applied to intensity/luminance channel
Image with easier to separate histogram	– Thresholding methods Triangle, Mean or minimum error thresholding
Image with more blended histogram	– Thresholding methods Triangle, Huang, Otsu or IsoData – Tao's thresholding approach
Image with color composition similar to UV modality	– K-means in AB subspace of LAB colorspace could deliver good results
Image with color composition similar to VIS modality	– K-means applied to whole RGB image could be good choice

Table 4.5: Table contains generalized findings of the evaluation. Mean Shift algorithm should be number one choice segmentation method. Use of other methods depends on the input image properties. Details and further results are described in the text.

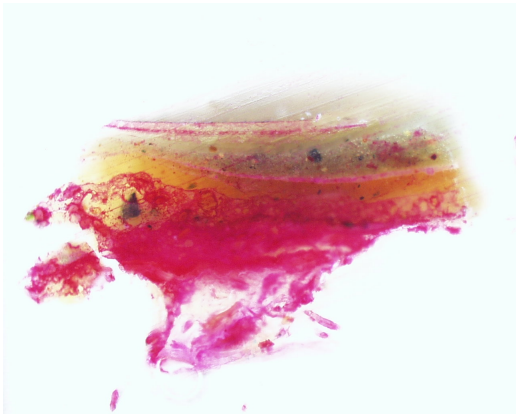
Segmentation methods	Quality indices								
	RI [0, 1]			DC [0, 1]			NMI [0, 1]		
	SEM	UV	VIS	SEM	UV	VIS	SEM	UV	VIS
All methods	0.75	0.75	0.85	0.64	0.75	0.84	0.39	0.42	0.55
GC_FH	0.96	0.90	0.88	0.96	0.91	0.84	0.83	0.71	0.65
GC_CV	0.95	0.82	0.87	0.94	0.83	0.86	0.79	0.51	0.59
GC_RD	0.96	0.74	0.71	0.97	0.76	0.74	0.82	0.44	0.44
MS	0.95	0.91	0.88	0.95	0.92	0.88	0.81	0.72	0.65
MNC	0.51	0.88	0.85	0.39	0.88	0.86	0.02	0.63	0.56
KM	0.78	0.77	0.87	0.62	0.71	0.86	0.39	0.42	0.58
IMJ_TRIANGLE	0.93	0.81	0.73	0.93	0.80	0.80	0.77	0.51	0.45
IMJ_OTSU	0.74	0.75	0.87	0.60	0.70	0.87	0.36	0.38	0.60
TNC	0.70	0.79	0.86	0.12	0.77	0.87	0.05	0.46	0.60
RG_10	0.94	0.76	0.67	0.95	0.80	0.76	0.78	0.47	0.38

Table 4.6: Table shows a change in performance of selected segmentation algorithms in three modalities for three different quality indices. In the first row there is an average performance (median) of all the methods in studied set in all three modalities according to the selected quality indices (RI, DC and NMI). In next rows there is performance of selected methods from the studied set. It is clear that more successful methods from the evaluation like MS perform the best in SEM modality, while in UV and VIS the average performance of all methods is the same or higher. There are some methods (e.g. MNC, KM) which perform better in VIS and UV than in SEM. All presented methods are the grayscale versions.

choice for image segmentation of related data. However several other methods could perform well while respecting above conditions, i.e. MNC, GC_CV, GC_RD or GC_FH. Should the execution time be an issue GC_FH especially would be an excellent choice. In that situation even plenty of thresholding methods or region growing could provide good results with some limitations. Table 4.5 sums up recommendations on the use of segmentation methods depending on the input image properties.

Concerning three modalities it is confirmed that SEM images are easier to segment for better methods thanks to clear boundaries between foreground object and relatively uniform background. The most successful segmentation methods like e.g. Mean Shift perform there generally much better than in UV and VIS where the segmentation is complicated by image properties. However the average performance of all algorithms is higher in VIS (and UV) than it is in SEM modality. In SEM there are lot of methods which underperform and lot of methods which deal with the problem excellently (interquartile range/variance of algorithms' performance is high). In VIS many methods deal with the problem comparably (interquartile range is much smaller). See table 4.6 for example of such methods and overall performance of the algorithms. Figures 4.9 and 4.10 show results of selected segmentation algorithms. More can be found in an appendix A.

One more evaluation was performed in addition to already described procedures. The idea was to find out what the various segmentation methods are sensitive to in the input images. For each method the images could be clustered to three groups – where

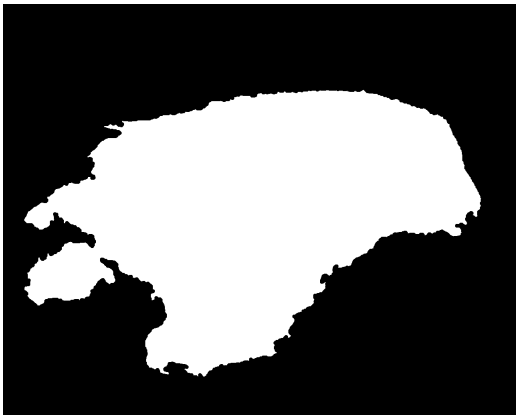


(a) Original VIS image

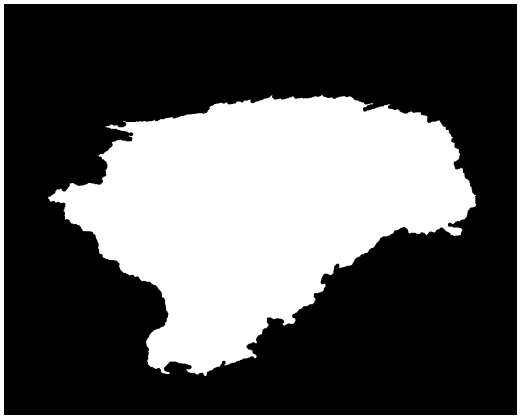
(b) Ground truth image mask

(c) Final segmented mask of MS

(d) Final segmented mask of KM_RGB

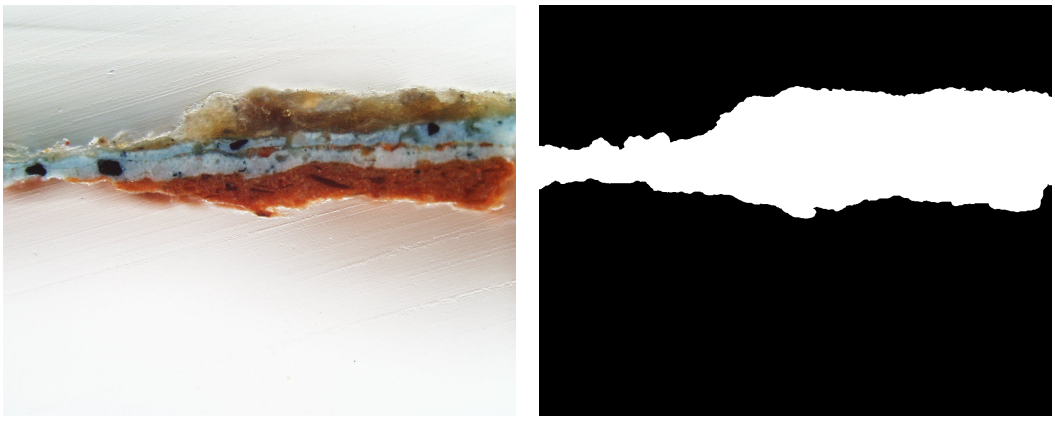


(e) Final segmented mask of IMJ_HUANG



(f) Final segmented mask of RG_25

Figure 4.9: Results of four image segmentation algorithms. In (a) there is an original VIS image. Ground truth mask is in subfigure (b). The image is segmented by MS (c), KM_RGB (d), IMJ_HUANG (e) and RG_25 (f). Image in (a) courtesy of ALMA.



(a) Original VIS image

(b) Ground truth image mask

(c) Final segmented mask of MS

(d) Final segmented mask of KM_RGB

(e) Final segmented mask of IMJ_HUANG

(f) Final segmented mask of RG_25

Figure 4.10: Results of four image segmentation algorithms. In (a) there is an original VIS image. Ground truth mask is in subfigure (b). The image is segmented by MS (c), KM_RGB (d), IMJ_HUANG (e) and RG_25 (f). Image in (a) courtesy of ALMA.

the output is good, bad and the rest. If some common features for the images in such groups could be found, it would provide a lead on which segmentation method should be used when such features happen to be present in an input image. Unfortunately no common features in addition to described properties could be found in defined groups.

Finally one remark to close the evaluation. It is important to keep in mind that behavior of some algorithms can be influenced with parameter setting. In our evaluation parameters are tuned to specific input data and we assume that same thing has to be done for different data set.

4.4 Demonstration of evaluation results applicability

We now demonstrate how the findings in previous section can be exploited in at least two different ways. First we verify that achieved conclusions hold also for different data set of cross-section images. In the second part the applicability of evaluation results is shown on distinct image data, i.e. biological images.

4.4.1 Testing data set of cross-section images

This section shows that the findings of previous sections hold also for different data set of cross-section images. The data set presented in section 1.1 and used so far in the text can be perceived as a training set. Now the same procedures of evaluation described in sections 4.2.2 and 4.2.3 will be applied on a smaller testing data set.

Testing data set consists of 28 SEM images, 27 UV images and 27 images[16]. The cross-section samples come from different artworks and some of the images are captured with different imaging devices. This data set is thus independent of previous training one and can be used for applicability test.

Foremost the conclusion that there is no such image segmentation method that can solely outperform the others in processing the input images in terms of quality (as stated in section 4.2.2) needs to be confirmed. We follow the same procedure to denote the most frequent segmentation methods among the best ones for every image. In SEM modality Mean Shift (MS) and Felzenszwalb's methods (GC_FH), two variants of region growing (RG_5 and RG_10) are the most frequent methods. Also percentile thresholding approach (IMJ_PER) is successful. This corresponds to the results on the training set (see graph in figure 4.3). The situation is similar also in UV modality. RGB version of Grabcut algorithm (GC_R_RGB), MS, GC_FH and grayscale version of multiscale normalized cut (MNC_GRAY) are the most frequent methods there. This is very close to graph in figure 4.5. Only MS is not so dominant here. It is however the most successful method in VIS modality with GC_R_LUV(UV), version of K-means (KM_LAB(AB)), MNC_LUV(L) and others behind. Thus the results correspond also in case of VIS modality (figure 4.6 with graph for training data set). The conclusion is that also in case of testing data set the single segmentation method is not sufficient to perfectly process the input images.

Next we compare the best average methods. The procedure described in section 4.2.3 is followed up to rank aggregation of sorted lists of methods created by ten quality indices. It is reasonable to expect that the order of segmentation methods in the aggregated lists would be similar to that achieved with training data set (except

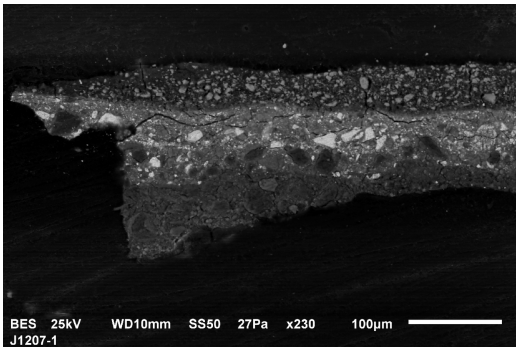
[16]In training data set they are 89 SEM images, 148 UV images and 148 VIS images.

for variations given by smaller number of images in testing set). In SEM modality this hypothesis is valid. The successful methods from training data sets are successful also here and conversely methods with dissatisfactory results are similar too. GC_FH and MS are the best average methods, Rousson-Deriche approach (GC_RD) with slightly worse results, IMJ_PER moves notably up the list. However the order of segmentation methods is more or less the same. Situation in UV modality is a little bit more confusing. MS algorithms takes second place after Chan-Vese approach (GC_FH). Also ranking of other segmentation methods is more shuffled in comparison to SEM modality. Nonetheless successful methods from training data set still perform very good and are placed in top positions of the list. VIS modality ranked list on the other hand does not deviate much from that of training set. MS is still by far the best average segmentation method and also the rest of the methods hold their positions.

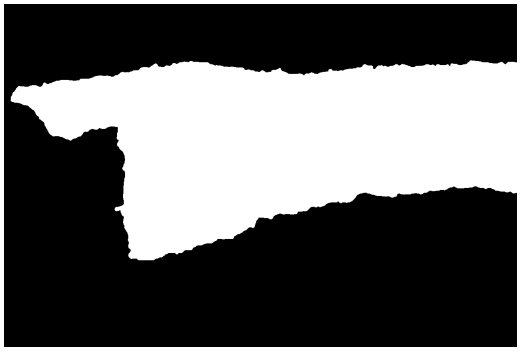
The conclusions of segmentation methods evaluation thus hold also for testing data set of cross-section images. There is no image segmentation method which would solely outperform the others on all images and the best average methods are the same for both training and testing data sets. Figures 4.11 and 4.12 show results of selected segmentation algorithms.

4.4.2 Biological images

In this section the applicability of evaluation results to different image data is shown. In figures 4.13, 4.14 and 4.15 there are segmentation results of biological images. The first figure shows the mouse retina. Specimen is colored with hematoxylin-eosin and captured with optical microscope in visible spectrum. It closely resembles VIS modality of cross-section images, because boundary edges are not clear enough and the background contains plenty of debris. The second figure shows transplant mouse cerebellum. Cells of the transplant generate enhanced green fluorescent protein (EGFP) so they are easily distinguishable from recipient tissue under a fluorescent microscope. The aim is to segment whole tissue (both original and transplant) from the background. Original tissue is however dark in contrast to green fluorescent transplant tissue and plenty of segmentation algorithms classify it as the background. The third figure shows 2D projection of 3D rendering of an early stage mouse heart, acquired by optical projection tomography. The image shows fluorescence excitation and emission. Last two figures resemble UV modality of cross-section images. The background is homogeneous and boundary edges are not so clear. The debris and other unwanted structures are also present in the background. Although it is not as visible as in the case of figure 4.13, it makes segmentation problematic. The best average segmentation method for UV and VIS modality is applied, i.e. Mean Shift algorithm. The results are depicted by red boundary line in respective figures.



(a) Original SEM image



(b) Ground truth image mask

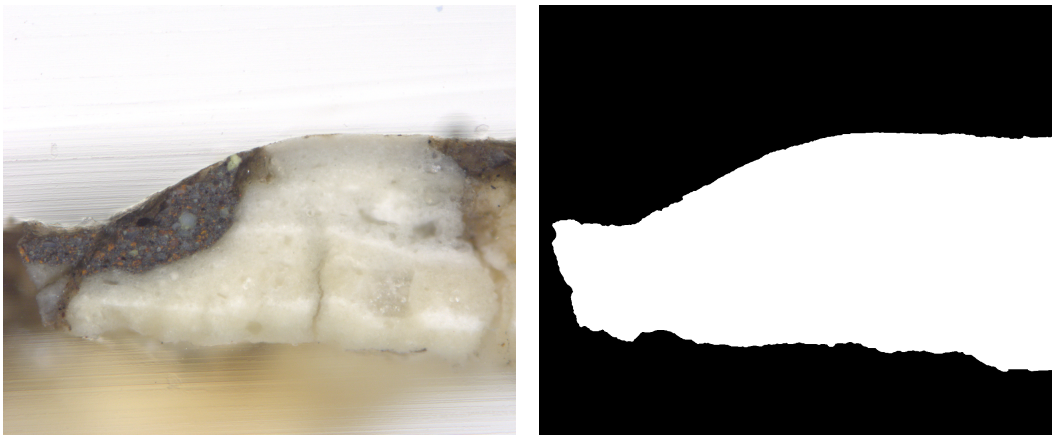
(c) Final segmented mask of MS

(d) Final segmented mask of GC_RD

(e) Final segmented mask of IMJ_PER

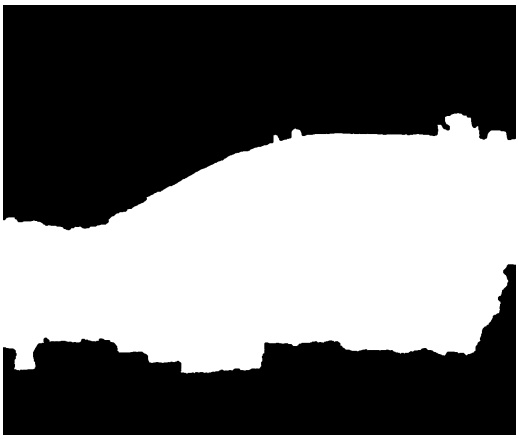
(f) Final segmented mask of RG_5

Figure 4.11: Results of four image segmentation algorithms. In (a) there is an original SEM image. Ground truth mask is in subfigure (b). The image is segmented by MS (c), GC_RD (d), IMJ_PER (e) and RG_5 (f). Image in (a) courtesy of ALMA.



(a) Original VIS image

(b) Ground truth image mask



(c) Final segmented mask of MS

(d) Final segmented mask of KM_RGB

(e) Final segmented mask of IMJ_HUANG

(f) Final segmented mask of RG_25

Figure 4.12: Results of four image segmentation algorithms. In (a) there is an original VIS image. Ground truth mask is in subfigure (b). The image is segmented by MS (c), KM_RGB (d), IMJ_HUANG (e) and RG_25 (f). Image in (a) courtesy of ALMA.

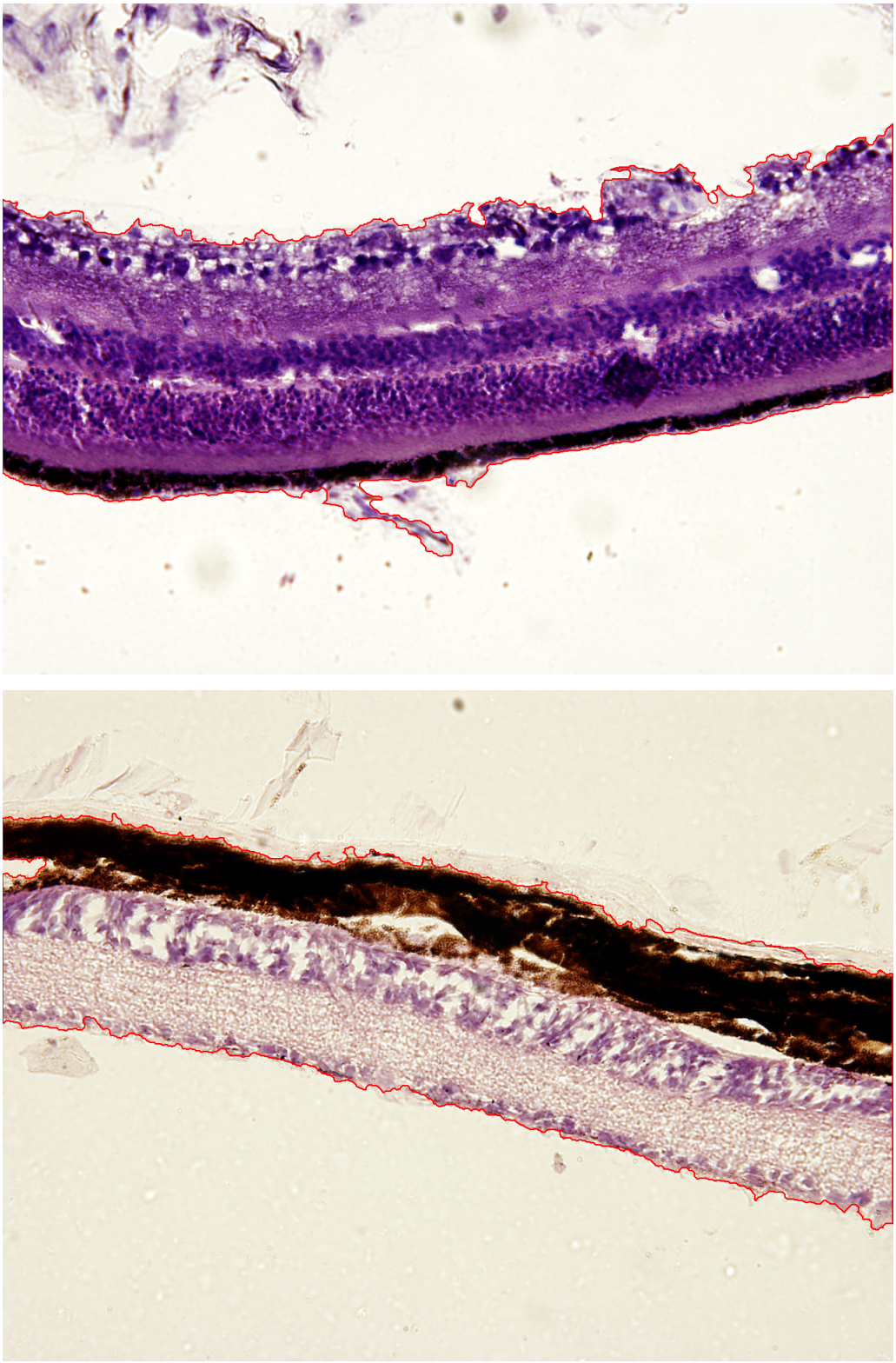


Figure 4.13: Mouse retina colored with hematoxylin-eosin. Boundary of segmented result by Mean Shift algorithm is depicted by red line. Courtesy of Jan Cendelín, Faculty of Medicine in Pilsen.

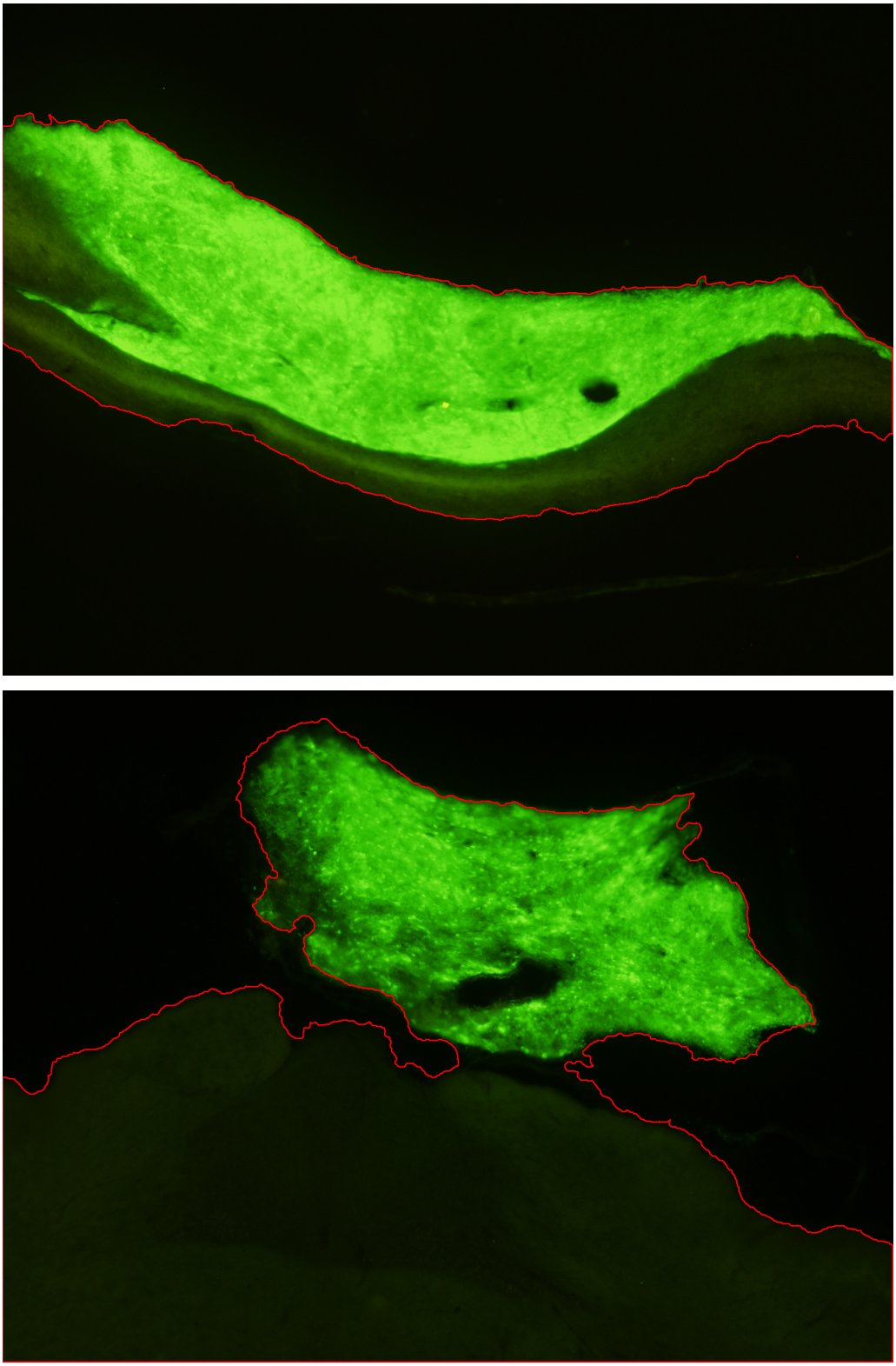
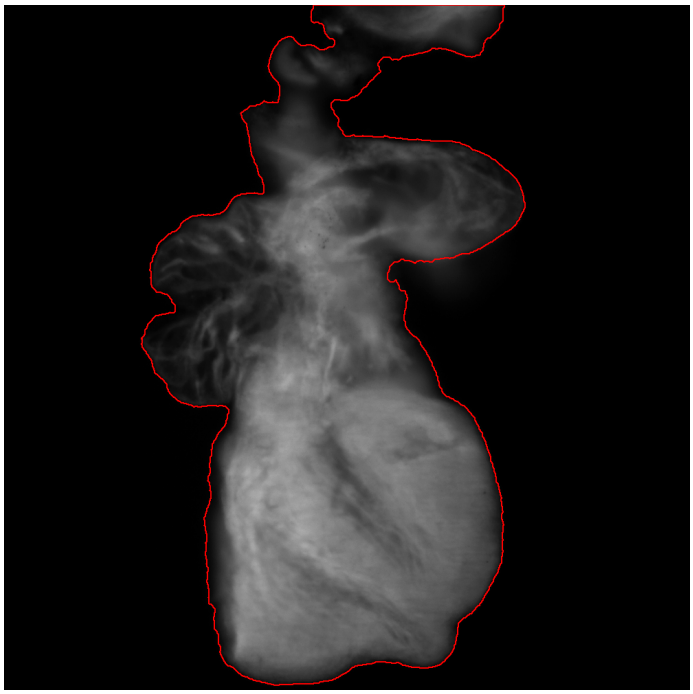
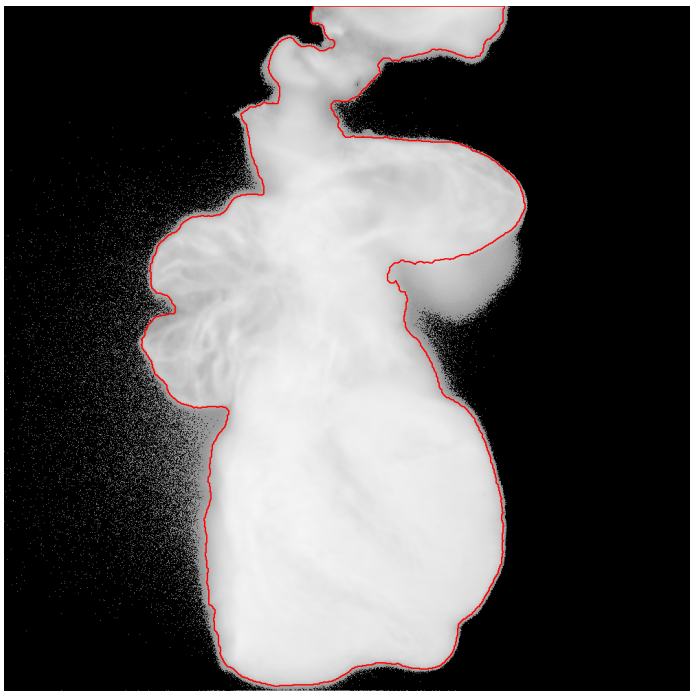


Figure 4.14: Transplant mouse cerebellum. Original tissue is dark, transplant tissue is light green. Mean Shift algorithm correctly segments both tissues. Boundary of segmented result is depicted by red line. Courtesy of Jan Cendelín, Faculty of Medicine in Pilsen.



(a)



(b)

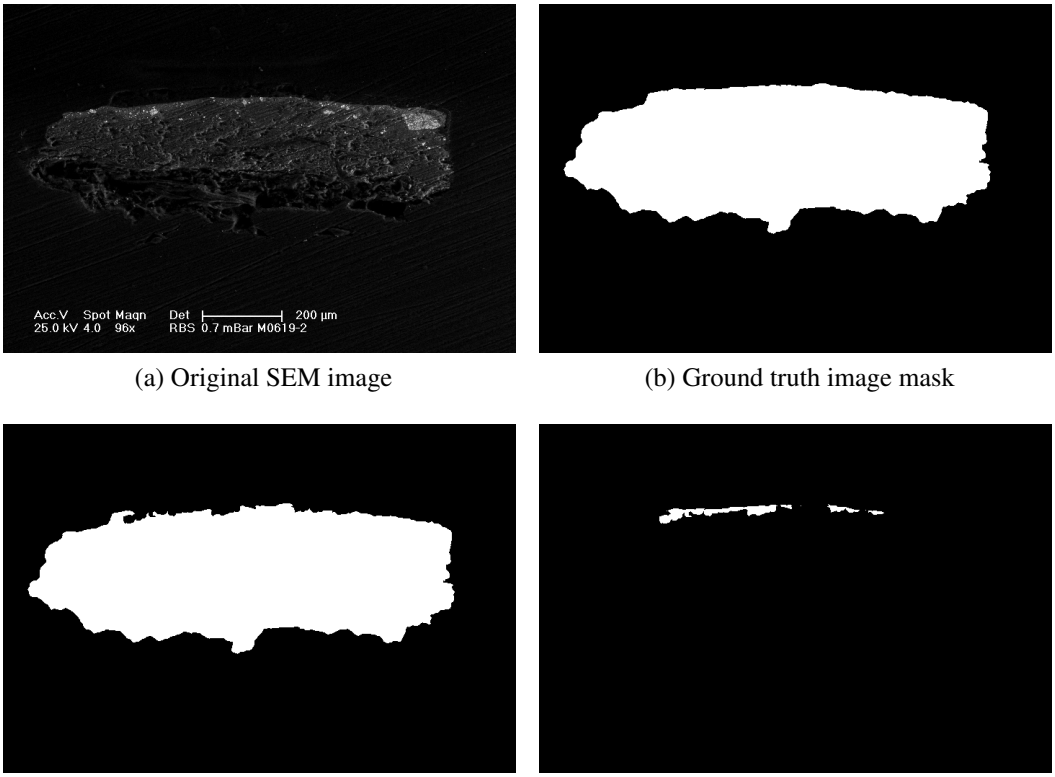
Figure 4.15: 2D projection of 3D rendering of an early stage mouse heart. Boundary of segmented result by Mean Shift algorithm is depicted by red line. In (b) the image is adjusted to reveal structures in the background which can influence image segmentation algorithms. Mean Shift however segments the foreground correctly. Courtesy of Martin Čapek, Institute of Physiology AS CR, Prague.

Chapter 5

Fusion of segmentation results

In section 4.2.3 we found (for each data modality) the image segmentation method which performed the best on average on input data set. The average means that this segmentation method often offers satisfactory results but sometimes it can fail (but not in such scale as other methods in the studied set). Next methods in the ranked list (the second, the third, ...) can behave differently (and due to their different fundamentals they often do) with failing on other images than the best method. Therefore it would be useful to somehow combine the results of several segmentation methods to remove unfavorable results and by doing so improve the overall performance of the segmentation process. The idea of combination comes from the classifiers domain. Kittler et al. in their paper [52] provided theoretical framework for combining classifiers. Key idea is to exploit advantages of different classifiers and eliminate their misclassification (sets of misclassified patterns do not necessarily overlap). Similar concept exists in clustering domain, i.e. cluster ensemble. Different clusterings of the same data set are combined to obtain final clustering of improved quality. Topchy et al. prove in [97] that cluster ensemble leads to better solution than individual clustering components. See [33] or [100] for an extensive survey of various combination methods and techniques. The idea of combination can be straightforwardly extended from classification and clustering also to the problem of image segmentation, because the segmentation method can be considered as a special kind of classifier or clustering method. See e.g. [31, 99] for application of cluster ensembles to image segmentation. Medical imaging represents another interesting domain where the concept of combination of segmentation results finds an application in so-called multi-atlas segmentation. Input volume (brain for example) is first transformed to match (it is registered) multiple known volumes in a collection (set of atlases) which are pre-segmented (usually manually by expert). The inverse transforms are then applied to known atlas segmentations and transformed label maps are combined to final segmentation result of the input volume (see [5, 80, 103] for examples).

In our case many of the published methods either cannot be used or they are not appropriate for the task. We deal with combination of binary results which makes the problem somewhat easier. Some of the methods (especially from multi-atlas segmentation domain) use also the intensity values of the original images next to results of the segmentation (labels) to improve the performance. In our problem domain this would however mean to substitute a behavior of image segmentation algorithm which would cause a vicious circle. Therefore we mainly exploit findings of previous chapter to better the outcome of image segmentation methods. To successfully resolve this problem



(a) Original SEM image

(b) Ground truth image mask

(c) Foreground pixel coverage

(d) "Non-segmented" pixels

Figure 5.1: Example of combination limits. In subfigure (c) there is foreground pixel coverage of SEM image in (a). It is obtained as an intersection of GT mask (b) and union of results of all segmentation algorithms. In (d) there is a difference of (b) and (c) – these pixels are not labeled as foreground by any segmentation method. Image in (a) courtesy of ALMA.

we have to decide which segmentation methods to combine and what method of combination to use. In the following sections we present couple of different approaches to these two questions. They all combine computed binary masks (segmentation results) on a pixel level. First the majority vote (MV) rule applied on limited subset of the segmentation methods is studied. Then the various methods of weighted voting with whole set of segmentation methods are addressed.

Performance of all combination methods is compared to that of the best average segmentation method and possible improvement is evaluated. However it is important to keep in mind that the combination could not overcome limitations given by input segmentation results. In some images there are foreground pixels (denoted by GT) that are not correctly labeled as foreground by any of the segmentation methods in the studied set. In such cases combination cannot overcome this fact and cannot approximate to the ground truth segmentation. Fortunately such cases are quite rare in the data set and "non-segmented" pixels are usually on the foreground boundary. See figure 5.1 for an example.

5.1 Majority vote rule and limited subset of methods

Generally it holds that the input set of methods (results, clustering or classifiers) has to be sufficiently diverse to achieve the best possible result of combination but at the same time if there are frequently failing methods included the final combination is spoiled (see e.g. [86] in context of neural networks classification). In terms of image segmentation we need to combine such segmentation methods which perform very well generally, do not fail too often and their results differ in important details (boundaries). We use evaluation results from previous chapter to achieve this. The best three average methods form the input set to combination in each modality. They perform the best from the studied set of methods, do not fail to often and their results are sufficiently diverse thanks to different fundamentals of each segmentation method. The combination of more than three methods was found dissatisfactory because the input results were more frequently bad which negatively influenced the output of combination (it is discussed in more detail below). Concerning combination method the majority vote is used. Therefore the pixel of an input image is labeled as foreground if at least two of the three methods label it as foreground. Otherwise it is background. We show that even such uncomplicated combination method can achieve considerable improvement of the image segmentation.

Results of segmentation combination are thus generated for every image in each modality using the three best average methods. It is Rousson-Deriche approach, Felzenszwalb's method and Mean Shift for SEM modality, Mean Shift, Felzenszwalb's method and multiscale normalized cut in grayscale for UV modality, and finally Mean Shift, multiscale normalized cut in RGB and K-means in RGB for VIS modality (see table 4.4). The aim now is to compare the results of the combination to the best average method. Again quality indices are necessary to ensure objective evaluation. We compute ten indices already used in evaluations of previous chapter for every image and compare them to those of the best average segmentation methods (Rousson-Deriche approach for SEM and Mean Shift or UV and VIS modalities). We use statistical evaluation with hypothesis testing to determine which of the two is better. The Wilcoxon signed-rank test [105] is used as good trade-off between plain sign test (which does not consider the magnitude of differences at all) and t-test (which considers the magnitude in much stronger way and also the stronger assumptions have to be met). Level of significance is set to 0.05.

Combination is statistically significantly better than the best average method in SEM and UV modality. In VIS modality the situation is little bit more complicated. Only four out of ten indices claim that the combination is significantly better. Conversely two indices claim that the best average method is significantly better. The rest stays rather undecided. Thus it cannot be decided which of the two approaches is better in VIS modality. If we compare combination to the second best average method (which is multiscale normalized cut in RGB) situation gets much clearer. Combination is significantly better in this case. For these reasons the choice of combination approach is appropriate even for VIS modality thanks to its robustness.

Visual evaluation was done as well to support the findings from statistical testing. Combination pays off also from this point of view. It is usually better than the best average methods in SEM and UV modality. In UV the difference is even more prominent and it is easy to see how combination of several segmentation methods amend inaccuracies of Mean Shift algorithm as the best average method (see figure 5.2 for

(a) SEM

(b) UV

(c) VIS

(d) VIS

Figure 5.2: Demonstration of improvement using combination of segmentation methods compared to the best average method. In each triplet in rows there is ground truth mask (left column), result of the best average method (middle column, GC_RD in SEM and MS in UV and VIS) and result of combination (right column). Last triplet corresponds to the images in figure 4.7. Combination there is certainly better than Mean Shift's result. However even better result can be achieved with pure region growing in this case as is shown in figure 4.7.

examples). Perhaps surprisingly the same holds for VIS modality. The results of combination are often more plausible. In those cases, where Mean Shift is better than combination, the difference is often minute. In the opposite cases difference between combination and Mean Shift is much larger and combination resembles ground truth more accurately[1].

Conclusion is that using majority vote rule applied on the limited subset of methods as a combination method for image segmentation can significantly outperform use of single (even the best average) segmentation method. This clearly holds for SEM and UV modality but also in case of VIS it is safe to use combination approach. Combination there is almost identical or only slightly worse than the best average method in vast majority of cases and occasionally it gives much better results. See figure 5.2 for examples of the results of segmentation combination using MV.

It remains to verify whether the limitation to only three best average segmentation methods is appropriate. Therefore we also compute the combination of five and ten best average methods and also the combination of all segmentation methods in the studied set. The results of the combination are worse with increasing number of methods to combine. This holds especially for VIS and UV modalities where the average performance of the methods at the top of the ordered lists is not as high as in SEM modality. Worse methods spoil the combination very quickly. In SEM modality the results of combination using of more than three methods are still quite good, because generally the image segmentation is more successful in this modality (see section 4.3). It is thus crucial for majority vote to use only good enough methods as an input for combination procedure. See [56] for a theoretical analysis of this aspect in classification domain.

We can also analyze benefit of combination approach on testing data set and biological images from section 4.4. Concerning testing set the combination is definitely better than the best average segmentation methods in all three modalities using visual evaluation. This holds especially for SEM modality. The statistical evaluation via Wilcoxon test confirms the significant improvement in SEM modality. In case of UV and VIS the improvement is not significant. However number of samples (images) is relatively small to consider statistical evaluation to be valid. In case of biological images the combination delivers similar results to those of Mean Shift as the best average method with negligible differences. MS results are already very good so there is not much space left for improvement (thus the benefit of combination will not be further analyzed in following section).

5.2 Weighted voting

To achieve good combination results using majority vote only the most successful methods could be used in previous section. The rest of the segmentation methods were set aside because otherwise they worsened the results. However these methods could still contain information useful for combination. It must be exploit in slightly different way though. Majority voting assigns the same influence on the result to every participating method. This influence can be adjusted with respect to performance of

[1] Wilcoxon test is incapable of capturing such subtleties. However t-test with greater sensitivity to magnitude of differences confirms this finding. According to this test combination is significantly better than the best average method.

the methods in image segmentation which is explored in chapter 4. The successful segmentation methods would have relatively big impact on the resulting combination while influence of less successful methods would be small. In this way we can exploit information provided by all segmentation methods and maintain performance level of the combination at the same time. Such approach is called weighted voting as the influence of impact is in fact weight of the method's vote.

In our binary case the final label L (F for foreground, B for background) of pixel x can be obtained as

$$L(x) = \arg\max_{i \in \{F,B\}} E_i(x)$$

where $E_i(x)$ is accumulated evidence that pixel x belongs to foreground or background respectively. It is computed as a sum of weights (influences mentioned above) of methods which denote the pixel as either foreground or background[2].

$$E_i(x) = \sum_{j=1}^{M} mw_{i,j}(x)$$

$$mw_{i,j}(x) = \begin{cases} w_j & \text{if method } j \text{ labels pixel } x \text{ as } i \in \{F,B\} \\ 0 & \text{otherwise} \end{cases}$$

M is the number of segmentation methods in the studied set and w_j is precomputed weight of particular method j[3].

In the following text we will look into different ways how the weights can be determined and compare it to majority voting from section 5.1 and the best average method. Weighted voting is also widely used in the area of multi-atlas segmentation. However the weight computation there uses different approach as it often exploits registration errors or image similarities. See e.g. [5].

5.2.1 Quality index as weight for voting

Straightforward idea is to use one of the quality indices as a weight for voting. Quality index offers natural measure of performance of each segmentation method. To be more precise the average (median) performance of each segmentation method on the image data set as computed in section 4.2.3 is suitable candidate for the weight. Normalized mutual information (NMI) happens to be a good choice from ten quality indices. It normalizes values to range of 0 and 1 and performances of segmentation methods covers the whole interval (dissatisfactory methods are close to 0, average methods somewhere in the middle and good methods draw near to 1). Table 5.1 presents NMI values for selected segmentation methods in all three modalities.

Using NMI as weight and previously described formulas the results of combination can be generated. Following evaluation procedure from section 5.1 we can compare the performance of this combination scheme to the best average method and majority vote combination to see whether there is any benefit of the NMI weighted voting.

Concerning the best average method combination does not improve its results in any modality. Statistically weighted voting is significantly worse or comparable at

[2]This corresponds to sum rule as described in [52].

[3]Majority vote is just the special case of this, in which weights of all participating methods equal 1.

Segmentation methods	NMI		
	SEM	UV	VIS
GC_FH	0.8267	0.7057	0.6496
GC_CV	0.7883	0.5079	0.5864
GC_RD	0.8224	0.4376	0.4354
MS	0.8078	0.7245	0.6524
MNC	0.0241	0.6307	0.5639
KM	0.3862	0.4214	0.5815
IMJ_TRIANGLE	0.7671	0.5084	0.4532
IMJ_OTSU	0.3586	0.3791	0.6046
TNC	0.0482	0.4643	0.5971
RG_10	0.7848	0.4720	0.3809

Table 5.1: Table shows NMI values of selected segmentation algorithms in all three modalities. All presented methods are the grayscale versions.

most. It holds also for visual comparison and regardless of input data set (training or testing).

Difference between NMI weighted voting and majority vote is more prominent. Majority vote is significantly better in all modalities according to Wilcoxon signed-rank test. Only in UV and VIS in case of testing data set the results of the test are undecidable. However visually the superiority of MV is evident (it is important to keep in mind validity objection of statistical testing mentioned above).

In the end, weighted voting using NMI as weights does not bring any improvement over the best average methods or majority vote. There is a visual comparison of majority vote and NMI weighted voting in figure 5.3.

5.2.2 Linear weights from quality index lists

The way to improve the results of weighted voting scheme may lie in exploiting information from all ten quality indices used in comparison of segmentation algorithms in chapter 4. Each index measures average performance of given algorithm individually and ordered lists produced by each index can thus differ significantly. These lists can be combined into one list by means of rank aggregation as is described above in section 4.2.3. The ranked list contains combined data about the average performance of each algorithm and as such can be used to determine their weights for voting process. However we will use more subtle approach and try to exploit data from each individual list before aggregation.

The weight w_m of method m is computed as

$$w_m = \sum_{i=1}^{10} pos_i(m)$$

where $pos_i(m)$ is a position of method m in an ordered list i (as there are ten quality indices and thus ten lists). Each list is sorted according to average (median) performance. This means that the method ranked as the best one by each index has w_m equal

(a) SEM

(b) UV

(c) VIS

(d) VIS

Figure 5.3: Comparison of majority vote and NMI weighted voting combination. In each triplet in rows there is ground truth mask (left column), result of the majority vote (middle column) and result of weighted voting (right column).

Segmentation methods	Voting weights		
	SEM	UV	VIS
GC_FH	0.9786	0.9725	0.7122
GC_CV	0.9262	0.8392	0.7347
GC_RD	0.9881	0.3824	0.1408
MS	0.9310	0.9980	0.9449
MNC	0.1595	0.9510	0.5673
KM	0.5381	0.5510	0.7224
IMJ_TRIANGLE	0.7952	0.8118	0.2429
IMJ_OTSU	0.4333	0.4216	0.8816
TNC	0.2857	0.7157	0.8020
RG_10	0.8786	0.5667	0.1327

Table 5.2: Table shows linear voting weights of selected segmentation algorithms in all three modalities. All presented methods are the grayscale versions.

to 10. Finally it is useful to normalize the weights to range between 0 and 1 (1 for the best method).

$$w'_m = \frac{max - w_m}{max - min}$$

where *max* and *min* are possible maximum and minimum values of w_m (minimum is clearly equal to 10).

Evaluation procedure stays the same. The results of combination are generated for every image using all the segmentation methods and weights for each modality. See table 5.2 which presents weights for selected methods in all three modalities. Quality indices are computed and statistical evaluation using Wilcoxon signed-rank test is used. Level of significance stays the same – 0.05.

First we need to answer the question whether there is some improvement of using all quality indices over one index, i.e. NMI. Statistically NMI is better only in SEM modality in both training and testing data set. However visually the differences are really small and both approaches are comparable. This also applies to UV and VIS modalities. The reason is that despite the distributions of values are not the same and NMI handles the best and worst methods slightly differently (see figure 5.4) the overall difference is not that big. In total when the voting is evaluated the results happen to be very similar.

Results of statistical evaluation between linear weighted voting and the best average methods are straightforward in case of SEM and UV modality. Combination there does not bring any improvement over the best average methods which are significantly better. In VIS modality the combination is comparable to the best average method as neither is significantly better than the other one. Visual evaluation confirms these conclusions. If we compare results of weighted voting to majority vote on the limited subset of methods from section 5.1, majority rule is significantly better in all modalities. In case of testing data set visually the results of weighted voting are comparable to the best average method in UV and VIS modalities. In SEM the best average method achieves better results. Also Wilcoxon test confirms this (however small set size objec-

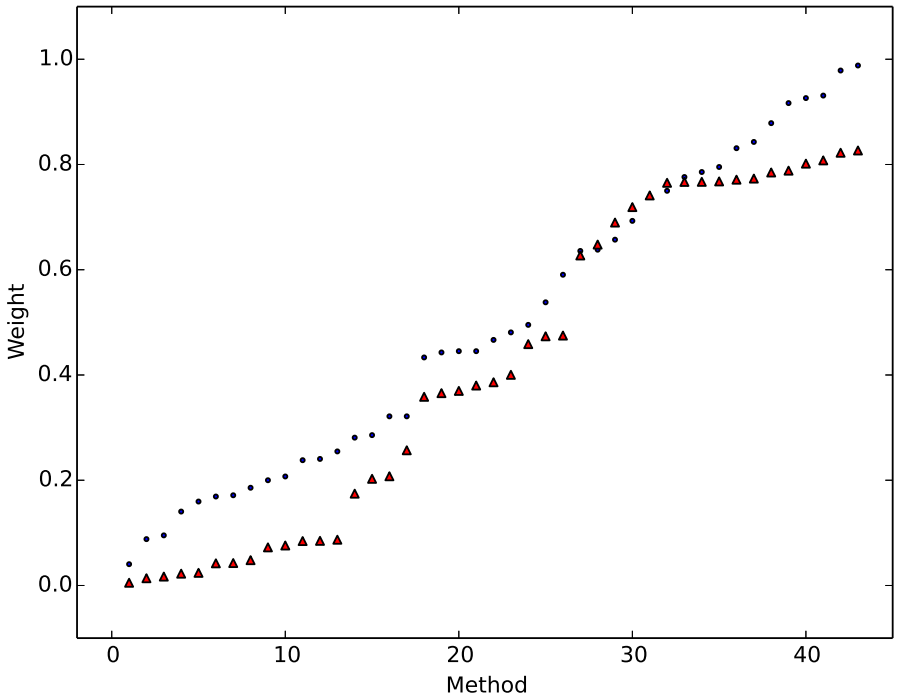


Figure 5.4: Comparison of linear (blue points) and NMI (red triangles) weights for segmentation methods in SEM modality. NMI handles the best and worst methods slightly differently but overall trend of the curves is more or less the same. Segmentation methods are sorted from the worst ones to the best ones (thus according to the weights).

Segmentation methods	Voting weights		
	SEM	UV	VIS
GC_FH	0.9371	0.9199	0.3631
GC_CV	0.7945	0.5910	0.3966
GC_RD	0.9647	0.0559	0.0028
MS	0.8068	0.9941	0.8436
MNC	0.0041	0.8600	0.1826
KM	0.1558	0.1673	0.3771
IMJ_TRIANGLE	0.5029	0.5349	0.0143
IMJ_OTSU	0.0814	0.0749	0.6853
TNC	0.0233	0.3666	0.5159
RG_10	0.6782	0.1820	0.0023

Table 5.3: Table shows polynomial voting weights of selected segmentation algorithms in all three modalities. All presented methods are the grayscale versions.

tion holds). Similar findings apply also to comparison with majority vote. The results of both approaches are comparable in UV and VIS modalities while majority vote is better in SEM.

Overall conclusion is that weighted voting scheme with weights computed as above does not improve the results of the best average segmentation methods or majority vote. Majority vote on the limited subset of methods is still better choice. See figure 5.5 which exhibits comparison of the results of weighted voting combination to majority vote.

5.2.3 Nonlinear weights from quality index lists

The reason behind worse results of weighted voting in previous subsection is still relatively big influence of the worst segmentation methods in the set. They spoil the combination results too much in comparison to majority vote and the three best average methods. This handicap can be amended by different distribution of weights between segmentation methods. In previous section the weights are linear (see blue dots in figure 5.6). By applying some nonlinear function (e.g. polynomial) to the computed weights we can give more influence to more successful segmentation methods and at the same time penalize less successful methods in the studied set. Cubic function is a good trade-off since the weights decline fast enough for the best methods and yet it still give some influence to the methods in bottom half (see red pluses in figure 5.6 and comparison with linear weights (blue dots) in the same picture).

The polynomial weights w''_m are therefore computed as

$$w''_m = (w'_m)^3$$

The established evaluation procedure is followed. The results of polynomial weighted combination are generated for every image using the described formulas. Table 5.3 presents polynomial weights for the same selected methods as in case of linear weights for comparison. Quality indices are computed and statistical evaluation using Wilcoxon

(a) SEM

(b) UV

(c) VIS

(d) VIS

Figure 5.5: Comparison of majority vote and linearly weighted voting combination. In each triplet in rows there is ground truth mask (left column), result of the majority vote (middle column) and result of weighted voting (right column).

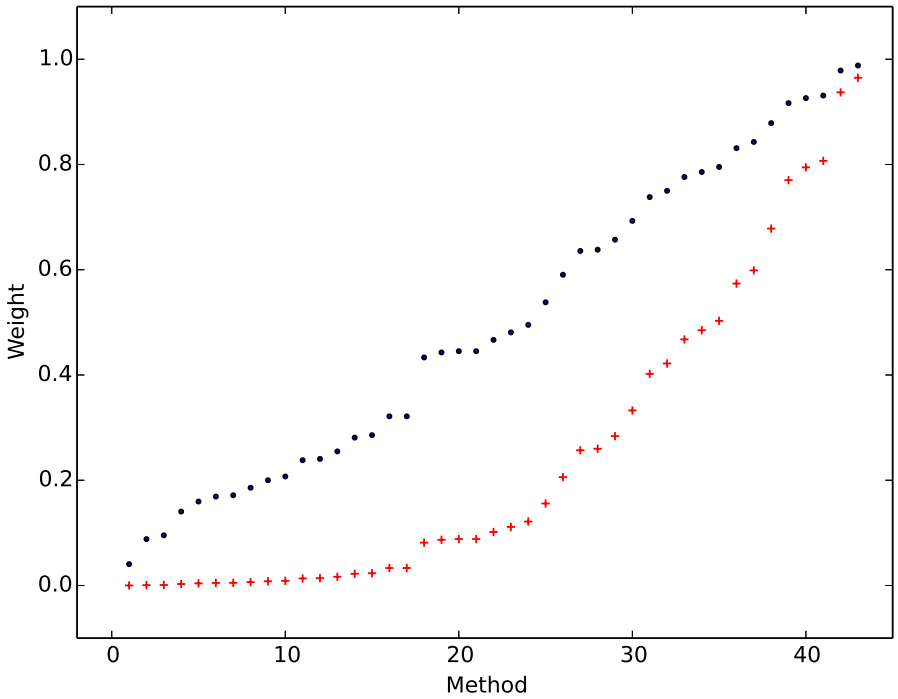


Figure 5.6: Comparison of linear (blue points) and polynomial (red pluses) weights for segmentation methods in SEM modality. Polynomial weights declines more rapidly towards zero. This eliminates the influence of worse segmentation methods. Segmentation methods are sorted from the worst ones to the best ones (thus according to the weights).

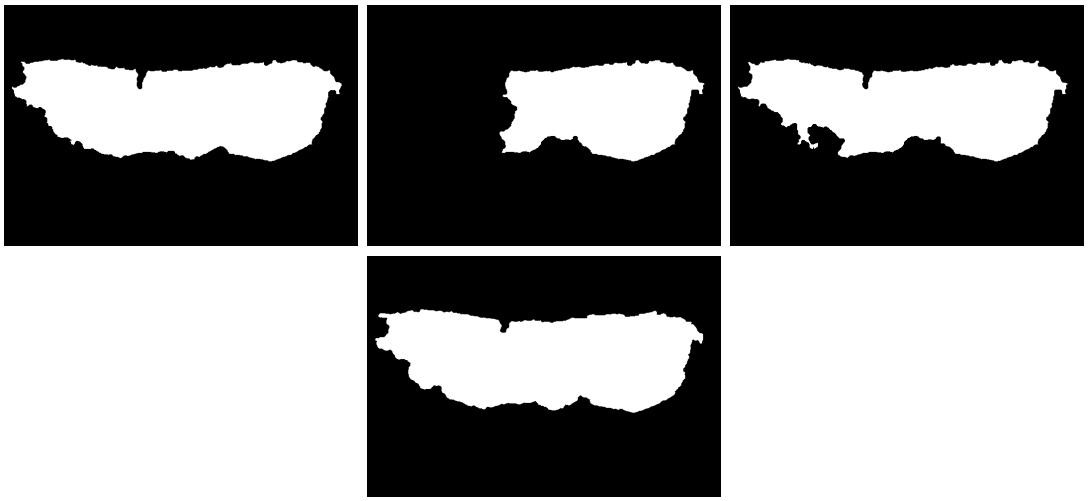
signed-rank test is used. Level of significance is still the same (0.05).

First it is necessary to analyze if there is even any benefit of polynomial over linear weights. The situation again differs in three modalities. In SEM and UV polynomial weights significantly improve the performance of weighted voting combination. This holds also for testing data set. In VIS modality linear weights are significantly better according to Wilcoxon statistical test. However when visually compared the difference between two approaches is minute. Linear weights are constantly slightly better which is the reason why statistical test states it is better, but polynomial weights offer almost the same results. In testing data set it is even more pronounced as Wilcoxon test fails to reject any null hypothesis (meaning neither weights are significantly better than the others). The conclusion is that polynomial weights are definitely better in SEM and UV modalities, and at least visually comparable in VIS modality (however statistically a bit worse in case of training data set). The same conclusion also applies to the comparison of polynomial weights and NMI weights from subsection 5.2.1.

Polynomial weighted voting behaves better than the linear also when the best average method is involved. The polynomial results are significantly better then those of the best average method in SEM modality (which is Rousson-Deriche approach). In UV and VIS the results are comparable which is still better conclusion than in previous case. In testing data set all the results are similar. It is safe to say that using polynomial weighted voting is better the best average methods as it is more robust and does not perform worse.

It remains to explore the relation of polynomial weighted voting to majority vote from section 5.1. Combination of three best methods using majority vote stays the best approach so far. In all modalities MV is statistically and visually better than polynomial voting. Only in case of UV and VIS in testing set the superiority of MV is not that dominant.

In conclusion, weighted voting combination scheme using polynomial weights definitely achieves better results then using linear weights. It is better or the same than the respective best average method. However majority vote combination is still superior. Figure 5.7 shows the comparison of three combination schemes to ground truth mask.



(a) SEM

(b) UV

Figure 5.7: Comparison of three combination methods. In each subfigure there is majority vote, linear weighted voting and polynomial weighted voting in a top row and ground truth mask in a bottom row. In subfigure (a) polynomial voting outperforms the linear one but is worse than majority vote. In subfigure (b) the result of polynomial voting is comparable to majority vote.

Chapter 6

Conclusion

The first part of the thesis addressed image segmentation problem and objective evaluation of image segmentation algorithms executed on microscopic image data. The input image data set was formed by images of cross-section samples extracted from the artwork during restoration process and captured in three different modalities – SEM, UV and VIS.

Chapter 3 dealt with preprocessing stage necessary to the following chapters of the thesis. Two problems were handled – legend removal from SEM images and more severe removal of grinding artifacts. The grinding artifacts origin from sample extraction process and they influence the image analysis. The automatic removal algorithm based on Fourier transform was proposed. Experiments proved the algorithm was successful in significant diminishment of present artifacts. Further the proposed method was used also in a different area of cultural heritage, processing of IR images. The algorithm removed the canvas structure and simplified following analysis of art restorers.

Chapter 4 analyzed the performance of segmentation methods on the input data set. The set of studied methods covered various approaches such as thresholding, region growing, clustering methods and graph-based algorithms. Ten quality indices were described to objectively evaluate the performance. The larger set of indices was incorporated to eliminate the bias of each individual index so the objectivity could be assured as much as possible. Concerning the very evaluation we first showed that there was no single segmentation method which significantly outperformed the others in the studied set. This means that there is no free lunch in image segmentation problem and it is necessary to compromise. Otherwise such segmentation method would be number one choice for solving background removal problem of similar data. This led to search for the best segmentation method on average which would be good enough. Such methods were found for every modality and also the lists of segmentation methods ranked according to their performances were produced through rank aggregation process. Mean Shift algorithm generally performed the best and can be considered the best segmentation method on average for related data. The results of the evaluation were thoroughly analyzed and circumstances, under which the segmentation methods performed well, were found. Recommendations on the usability of individual methods could be thus given. Finally, we verified the findings on separate testing data set and the applicability of the evaluation results was shown on different but related biological data.

Chapter 5 examined whether combination (or fusion) of segmentation methods could further improve the performance of even the best average method. Majority vote

with limited subset of segmentation methods and weighted voting with several methods of weight computation were considered. In both cases the findings obtained from evaluation of chapter 4 were exploited to achieve better results of combination. Majority vote applied on three best average methods and polynomial weights computed from quality indices lists proved to deliver significantly better or in some cases at least comparable results as the best average method itself. The combination approach is thus appropriate in all cases as it does not underperform and its advantage is robustness. From these two combination schemes majority vote method was superior. This result confirms the observations about good performance of majority vote from other domains – classifier combination [51] or multi-atlas segmentation [2, 41, 79, 81]. It remains to note that the superiority of majority vote applies to microscopic image data of similar properties as the cross-section images have. As in case of image segmentation there is no universal combination approach which would outperform the others in all cases [43, 5, 56]. All the conclusions of the chapter were supported with statistical testing.

Part II

The Nephele system

Chapter 7

The Nephele system

7.1 Introduction

Protection of cultural heritage remains an important issue. Next to the long-established methods new technologies are proposed to support this complex task. Advances in the digital image acquisition brought considerable momentum to the study of paintings – uncovering the author's intentions, verification of the authorship, or selection of the most proper restoration procedure for the artwork conservation. All this is possible thanks to the ability of multimodal imaging exploiting various spectral wavelengths (e.g. infrared, ultraviolet) to reflect material composition of the artwork, even in its invisible layers underneath the painting surface. Price lowering and wider accessibility of the hardware equipment for such multimodal acquisition have led to sharp increase in the volume of data acquired in variety of wavelengths and often in high resolution mode. Current methodologies, addressing the proper painting scanning, setting system parameters and lighting conditions, and/or solving geometric and radiometric degradations introduced during the data acquisition, have been no longer capable to handle these large quantities of information. Image processing methods able to extract higher level information have been introduced to the painting analyses. They can provide insight into brushstrokes, into pigment composition, or into canvas structure evaluation. These methods form an indispensable component of the current research in the field of painting conservation.

Our proposed system Nephele lies somewhere on the border between the pure data acquisition and a higher level analysis of the painting. Nephele is an expert information system for archiving the reports created during material research of the painting. The report contains all gathered information about the studied object – textual, numerical and also image-based. Typical database retrieving functionality is extended by *content-based image retrieval* (CBIR). This approach enables fetching reports by image similarity and thus leave out difficult and time consuming textual description of the required record. The cross-section images of minute samples described in section 1.1 are fundamental building blocks of the implemented image retrieval. In material research the cross-section images help to identify the used painting materials. In the Nephele system they are used to answer queries whether an artwork of similar material composition has been already analyzed. Thus, the Nephele system consists of database functions for data archiving, conventional search and of image processing functions which provide image preprocessing and analysis for the CBIR. The system has been developed in cooperation with Academic Mate-

rials Research Laboratory of Painted Artworks in Prague (ALMA) and it reflects all the best practices used in the material analyses of the paintings and their reporting. The system is continually evolved and extended and is published free of charge under GNU GPL (`http://zoi.utia.cas.cz/nephele.html`). The work was partially published in [8].

The chapter is organized as follows. Next section contains a brief survey of works related to use of image processing methods in art conservation and content-based image retrieval system. Section 7.3 describes the proposed Nephele system with material research context. It is followed by the section 7.4 dedicated to the image processing algorithms included in the Nephele system. Section 7.5 contains demonstration of the image retrieval together with screenshots of the system. The chapter concludes with section 7.6.

7.2 Related work

The Nephele system is an expert system for art restorers and conservators with additional image retrieval of cross-section images ability. Despite there is no other comparable system to the best of our knowledge, exploitation of the image processing methods in art conservation and restoration gains in importance. The image retrieval functionality can provide feasible means for exploration of the museum or gallery digitized collection. The next two sections present the related work concerning the use of image processing in art conservation and overview of content-based image retrieval systems.

7.2.1 Image processing in art conservation

The exploitation of the image processing methods in the art analysis and restoration grows and an application typology broadens too. Methods can be original or adaptations of algorithms used in existing applications, modified for the cultural heritage niche. Several main directions can be identified in this boom, which also follow the main steps of the art conservation process.

The first and the most important is the data acquisition category, which addresses the process of the object scanning and data gathering [70]. Proposed solutions offer capturing of the object (which is often the painting) in various image modalities (e.g. using different wavelengths) [16]. They are able to handle very high resolution data and solve eventual geometrical degradations and changes in lighting conditions of the scanning process. Acquisition in various wavelengths and modalities enables art restorers to see the underdrawings, the structure of the canvases and even individual pigments and other components of the paintings. The infrared (IR) and/or ultraviolet imaging belong to the most often used approaches. Less commonly used synchrotron radiation based X-ray fluorescence allowed to uncover the van Gogh's painting repainted later by its author [3].

Scanning process as described often produces very large data files. Methodologies how to handle, visualize and archive such objects are thus needed and form the next popular group of methods, which are in the center of interest. Algorithms for visualization of acquired multispectral information can properly complement the acquisition process [49]. Archiving and potential publication of data complete the processing pipeline. This last step calls for effective storage methods of huge number of images

with tools for their retrieval (see section 7.2.2). Nowadays, many museums and other institutions provide online access to their art collections via the Internet (e.g. Louvre, British Museum, Museum of Modern Art in New York, Google's Art project, ARTstor[1]).

The role of image processing methods in the mentioned data processing pipeline is essential but image processing can offer much more. Having the digitized version of the art piece, possibly in several modalities, image analysis provides an insight into the author's intentions by inspecting the underdrawings. The restoration interventions can be reviewed based on the pigment visualization in different modalities. Looking at the dynamics and shapes of the brushstrokes sheds light on the painting's authorship [48]. An estimation of the place and time of the painting's creation can be done based on the craquelure detection and analysis [10]. Even the author's method for the creation of painting can be explored [93].

Recently, complex software solutions offering the data acquisition and their further analysis have started to appear. ArtShop, an art-oriented image processing tool for cultural heritage applications, provides wide functionality from the multispectral scanning tools to the fuse-based visualization and symbolic description of the artworks [11]. Chip, an image processing software specifically designed for cultural heritage applications, offers, next to the basic image processing functionality, also higher level methods such as inpainting algorithms or feature extraction for further analysis [18]. Finally, all achieved data and analyses can lead to complex virtual restoration, which could demonstrate the original beauty of the art piece. Its colors can be rejuvenated [76] and craquelure removed [91].

7.2.2 Content-based image retrieval systems

Research in image retrieval has roots in the second half of the 70's of the 20th century. The necessity of image searching in huge collections became apparent and solutions based on a text annotation or tags emerged promptly (*text-based image retrieval*). As a result, the image retrieval was carried out by existing text retrieval system exploiting text tags which where tied to the image during insertion time. However a time-demanding image annotation process and perception subjectivity of human annotators showed the non-usability of such approach.

On the contrary, *content-based image retrieval* exploits visual features of the images in question and their visual similarity based on human visual perception. The features range from color or texture properties to objects shapes or object distribution across the image. The content-based image retrieval (CBIR) system implements finite set of descriptors of these features which were specifically chosen for given problem domain. Suitable similarity measure for comparing feature descriptor vectors of different images and eventual indexing data structure for fast retrieval response must be picked as well. Different CBIR system may also implement different query schemes or techniques. Query-by-example serves as a common type of query in which an image of interest is supplied by user. Another possibility is query-by-sketch technique. The user may draw approximate draft of desired results. For exhaustive survey on CBIR techniques and system properties see [89].

The true CBIR systems started to occur during the 90's of the 20th century. QBIC[2]

[1] `http://www.artstor.org`

[2] Abbreviation for Query By Image Content

developed by IBM [29] was perhaps then most well-known system. QBIC adopted both query-by-example and query-by-sketch techniques, and implemented a variety of image feature descriptors. Berkeley's Chabot system [66] made use of mentioned text tags in connection with content features. Image characteristics were accompanied by user's description (e.g. "search for all the sunsets above a blue lake"), which was translated to spatial relations between predefined objects. Photobook from MIT [71] linked CBIR system with relevance feedback, so the user could specify appropriateness of the retrieved images. The CBIR system called VisualSEEk [90] implemented different query technique similar to query-by-sketch. A diagram with approximate spatial arrangements of color regions served as the input to the retrieval process. An extensive list of contemporary CBIR systems can be found in [101].

An important change came with the expansion of the Internet. The CBIR systems could become online and had to deal with the growing amount of pictures. On the other hand, the development in indexing data structures and also in image processing research helped to handle novel challenges. CORTINA system [32] implements a web crawler to build a large-scale collection of images which serves as a base for CBIR. PIBE [6] focuses on browsing through a large image database and on user experience as well. LIRe is an open source CBIR library which among others includes a considerable range of global and local image feature descriptors [59].

In comparison to the ongoing research of general CBIR systems, image retrieval in art field stays rather behind in contrast to its potential merit for art galleries and museums. Artistic representation of ordinary objects, abstraction and creativity make the process of image analysis difficult. Nonetheless, CBIR systems specialized in this field exist and are used by galleries worldwide. ARTISTE [1], system deployed in several famous galleries in Europe (including the Louvre in Paris, the Victoria and Albert Museum in London, the Uffizi Gallery in Florence and the National Gallery in London), offers automatic annotating of picture collections and retrieval which respects specific needs of art historians. SCULPTEUR [35] extends the previous system with capability of retrieving 3D objects. A few other systems are also published (see [37, 12]).

7.3 The Nephele system

The Nephele system and database provide a solution for processing and archiving information about artwork and its material research in the course of art restoration. The material research of the painting helps art restorers to choose the proper materials for the actual restoration and determine the authorship of the artwork. Material analysis of the painting is described in the form of a report, which contains general information about artwork, its description, samples and chemical analyses which were hold. These reports can serve as a knowledge base for future restoration cases. The goal of Nephele system is to make such usage possible and efficient by using forms, which the report is generated from, storing them in the database and providing a user-friendly access. Up to now the reports have been created manually in the desktop publishing software, which certainly complicates further exploitation of gathered knowledge. To make this task easier the Nephele system implements the image retrieval functionality which is in detail discussed in section 7.4.

7.3.1 The material research report

The material research report captures that part of artwork restoration process which is aimed on the objective identification of the painting materials and on the evaluation of painting techniques and style. The precise identification of used substances helps the restorer to choose the proper materials and appropriate technique for the restoration. The optical and chemical analyses of minute surface samples taken away from artwork are the important parts of this process. The sample extraction and following manipulation is described in section 1.1. The images in VIS and UV spectra are captured as was mentioned, where UV image may reveal fluorescent property of certain materials (figure 1.2). Study by scanning electron microscope (SEM) with chemical contrast is performed together with a variety of chemical analyses. Energy-dispersive X-ray spectroscopy using SEM delivers qualitative information about elemental or chemical characterization (SEM/EDX). X-ray powder diffraction identifies the mineral components of the material layer. Fourier transform infrared spectroscopy (FTIR) is used for estimative determination of binders which is necessary for correct artwork classification. Matrix assisted laser desorption ionization time of flight mass spectroscopy (MALDI-TOF MS) serves as a means for protein binder differentiation. Finally, X-ray fluorescence (XRF) analysis exploits a fluorescent emission to determine an elemental composition of the sample.

Acquired findings are gathered and composed to the material research report. It includes general information concerning the artwork (e.g. author, owner, artwork's state, dating, painting technique), photodocumentation of the cross-section samples, results of the optical analyses and stratigraphy of the samples, the results of mentioned chemical analyses, interpretation of the results, and plenty of additional materials. In the end, art conservator or restorer draws the conclusion about material composition, which forms the final part of the report, and chooses the precise painting material for the very restoration. Figure 7.1 exhibits the parts of the material research report.

7.3.2 The Nephele system architecture

The Nephele system is the information system built on top of the database for material research reports. Besides a report archiving the primary goal of the system is to serve as a knowledge base for future usage. To do so the effective information retrieval must be implemented – text-based search, search in the results of chemical analyses and image retrieval facility (including the necessary image preprocessing modules). Text-based search implemented in the system permits the user to enter simple queries concerning artwork's author, report's author, report's number etc., and fulltext queries. Search based on the results of chemical analyses allows detection of the artworks with similar chemical composition as the researched one. The other option is looking for the artworks of given material composition. See section 7.4 for details about the image retrieval and image processing part.

The system is deployed in a real environment (Academic Materials Research Laboratory of Painted Artworks) which leads to certain requirements the system has to meet. The application should provide an easy access to all information, however such an access must be secured and sensitive data should be restricted only to precisely defined group of users (the information about detected materials could be misused for creation of falsifications). The system has to allow access of multiple users at the same time. Finally, the expandability can be crucial, since the material research reports are

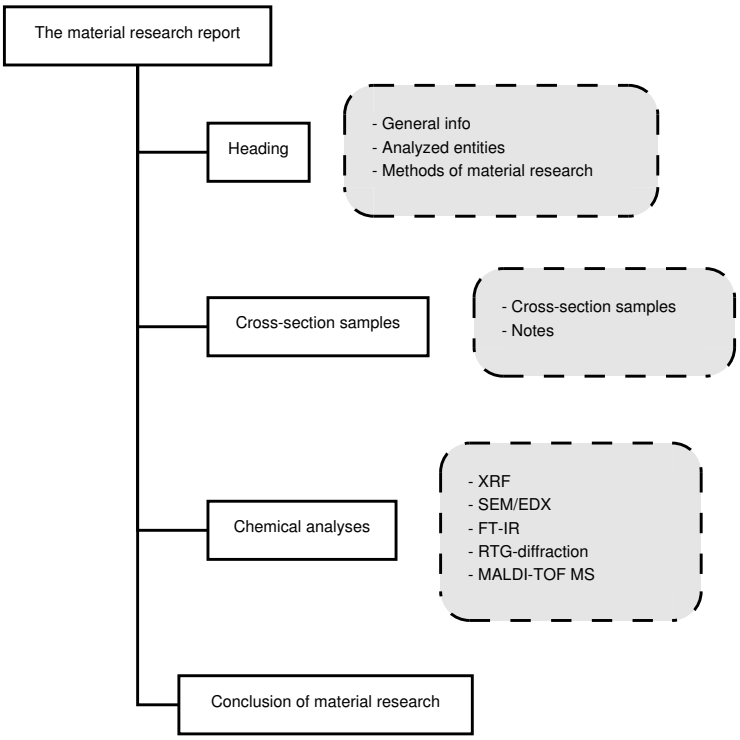


Figure 7.1: Contents of the material research report. Generally it has four parts: so called heading with general information concerning the artwork, its several parts if present (analyzed entities) and photodocumentation; cross-section samples part with photodocumentation, results of the optical analyses and stratigraphy, interpretation of the layers and notes; results of chemical analyses, and finally conclusion of material research.

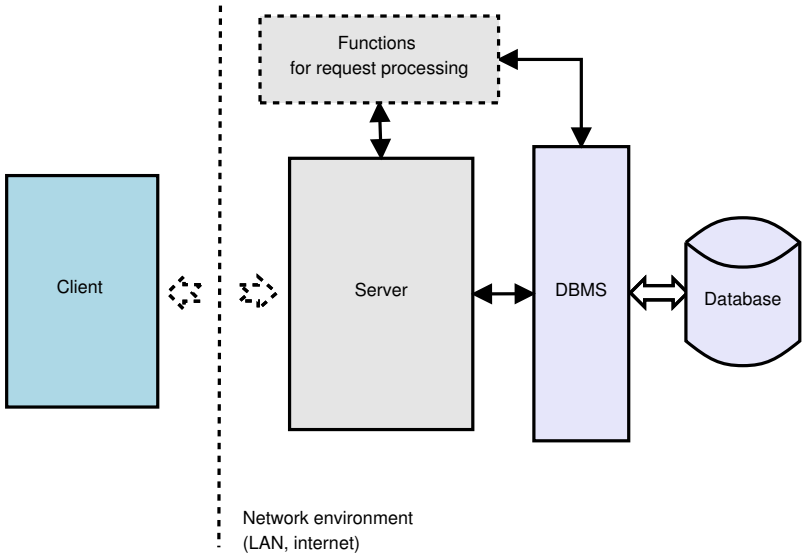


Figure 7.2: Client-server architecture model of the Nephele system. The server part acts as a service provider and communicates with database management system. The client part plays a service requester role.

from time to time enriched with a new analysis.

The Nephele system is based on the client-server architecture model (see figure 7.2). The server part acts as a service provider which implements most of the system functionality such as data management, information retrieval, users administration etc. The server communicates with database management system[3] through a series of SQL commands. The database was created using the entity-relationship model in cooperation with the experts. The programming part was implemented in C++.

The client part of the system on the other side plays a service requester role. It communicates with the user through graphical interface and dispatches user requests to the server by means of XML-based messaging model. Momentarily this part is implemented in .NET framework, but the client-server model and precisely defined communication protocol allows to design various clients (a web-based for example).

7.4 Image processing in the Nephele system

The image retrieval in the Nephele system, which provides extended functionality compared to the standard text-based searching, consists of chained image processing steps, which have to be performed. The preprocessing of acquired input data, feature extraction and the similarity retrieval are the key issues to be handled by the image processing blocks. Due to the preparation of the cross-section samples and the image acquisition process the multimodal input images can be degraded. Thus they cannot be used directly and certain preprocessing steps are necessary. A priori selected feature descriptors are then extracted from the input images and used as a feature vector to query a database and to retrieve a given number of similar cross-section images. Following sections describe mentioned steps in detail.

[3]MariaDB at present

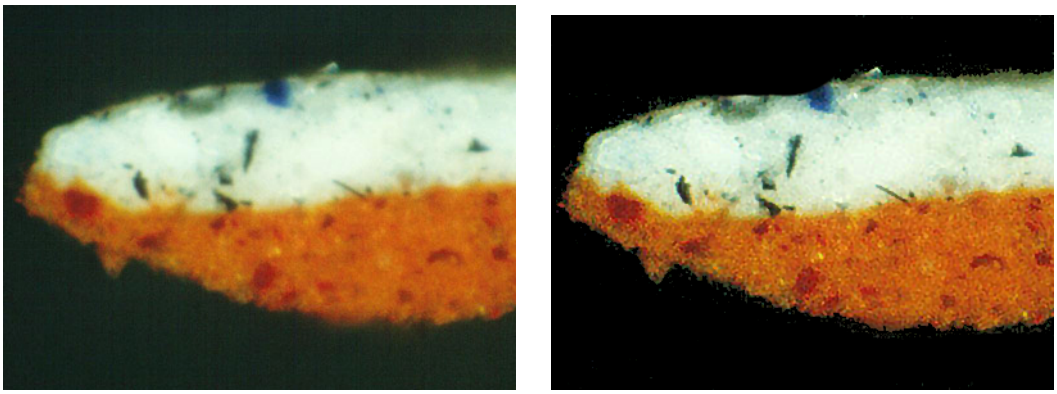


Figure 7.3: An example of blurred sample (left image) and the deblurred result (right image) achieved by method described in [92]. The structure of the seeds and rough texture is much more apparent now.

7.4.1 Image data preprocessing

The purpose of the image preprocessing phase is dealing with the degradations and the artifacts formed during the sample preparation and image acquisition, which can negatively influence the performance of the image retrieval. It includes the grinding artifacts but also during image acquisition in the microscope a foreign particles like dust might be present and create a noise off the very sample in the cross-section projection. All these unwanted fragments slightly change the result of the feature descriptors computation and therefore the outcome of the image retrieval, because the otherwise similar images would be more distant in the feature space.

The background removal from the cross-section image is a natural way to handle the problem[4]. The removal is accomplished by means of the image segmentation which leads to separation of the input image to two distinguished regions, noisy background and sample foreground in our case. We use the proposed method from the first part of the thesis to achieve this. Foremost the grinding artifacts have to be removed using the method presented in section 3.2 in order to improve the results of some image segmentation methods (and thus the image retrieval). Then the selected segmentation methods are applied and their outputs are combined using majority vote (see detailed description of the procedure in section 5.1).

Next to the described imperfections, the images of analyzed samples can be noisy and/or blurred. Blur can be introduced during the acquisition process. Its two most common sources are the out-of-focus blur due to the inaccurate setting of scanning parameters and the too shallow camera depth of field. It is not always possible to avoid these degradations, mainly due to high amounts of the analyzed data and due to operator's skills. The blind deconvolution approach can be utilized for removal of the introduced blurring, when the known a priori information is used to restore the original denoised and sharp image. Figure 7.3 demonstrates an example of blurred sample and the deblurred result achieved by method described in [92].

[4]The noisy fragments are indeed omnipresent and the background removal does not solve the problem entirely. Nevertheless, the sample foreground generally occupies only a smaller fraction of the whole image and thus the removal is always beneficial.

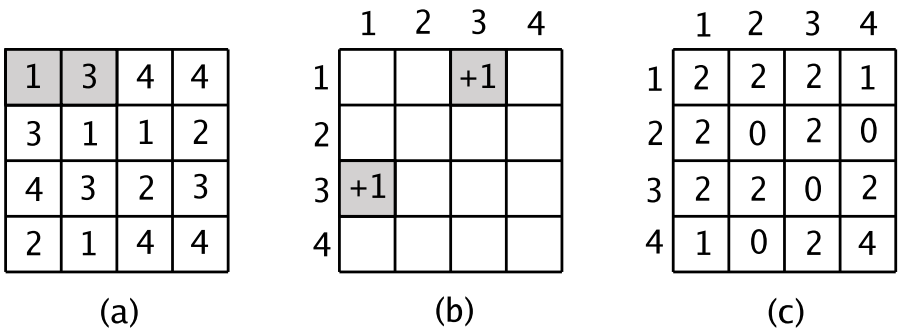


Figure 7.4: Computation of symmetric co-occurrence matrix (an example) with shape operator (1,0) and image of size 4×4 . In (a) two pixels of the original image defined by shape operator are highlighted, (b) demonstrates increment of one to appropriate places in co-occurrence matrix. In (c) there is final co-occurrence matrix of image in (a).

Contrast	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 M(i,j)$
Inverse Difference Moment	$\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{1+(i-j)^2}M(i,j)$
Entropy	$-\sum_{i=1}^{N}\sum_{j=1}^{N}M(i,j)\log M(i,j)$
Variance	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-\mu)^2 M(i,j)$

Table 7.1: Mathematical formulas of four selected Haralick's feature descriptors. The co-occurrence N-by-N matrix M is computed from the image, μ denotes the mean.

7.4.2 Image retrieval

Image retrieval of UV and VIS images

The image retrieval subsystem in Nephele follows the query-by-example concept. The user submits the pair of UV and VIS images from the artwork he is interested in and expects the set of similar pairs as a result. By doing so, the user receives the reports which are alike regarding visual similarity of respective cross-section images. The image feature descriptors are computed from the input pair and compared in feature space with the descriptors of the images already stored in the system database. The nearest neighbors in terms of the similarity measure are retrieved and returned to the user.

The descriptors must be carefully chosen with respect to the properties of the input images, since inconsiderate selection could negatively influence the result (as far as the number of false positives and negatives is concerned). The problem is complicated because of the difficult determination what the visual similarity means (for example the shape of cross-section samples is irrelevant due to its arbitrariness as a results of sample acquisition process). The proposed method is based on co-occurrence matrices and color descriptors. The co-occurrence matrix [40] reflects the joint occurrence of gray level pixel pairs with a defined spatial relationship formed by a shape operator (see ex-

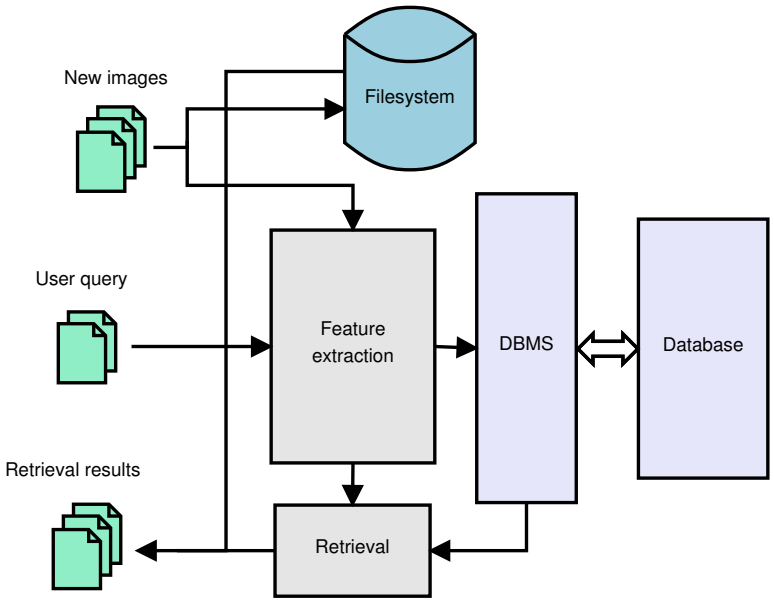


Figure 7.5: Workflow of the image retrieval in the Nephele system with image feature descriptors computation for each cross-section image

ample of computation small co-occurrence matrix in figure 7.4). In consequence they have a potential to capture the relationships between the material layers as well as different texture properties of a single layer. Shape operators of different lengths are used and all color channels are processed separately. Since the matrices by themselves are not very suitable for image retrieval, Haralick proposed a set of descriptors which can be directly computed from the matrices [40]. In our case Contrast, Inverse difference moment, Entropy and Variance showed the most promising results (see table 7.1 for mathematical formulas). Next to the Haralick descriptors two color descriptors, image mean color and standard deviation, are included to take the main color trends of the images into account.

The image retrieval process in the Nephele system consists of two parts (illustrated in figure 7.5). First part is the descriptors computation from existing images in the database, second is the very retrieval. The images are processed during the material research report insertion. The marked pair of VIS and UV images for every sample passes through the image processing step and the resulting vector of feature descriptors is stored in the database. The image retrieval step is alike. The query pair of VIS and UV images is processed and the query vector is compared via weighted Euclidean distance with all the vectors stored in the database. The specified number of the nearest (the most visually similar) neighbors is retrieved. The exhaustive search for neighbors would be time consuming and therefore M-tree indexing structure [14] is used to speed up the retrieval process. M-tree is a flexible data structure designed to combine advantages of metric tree and database access methods. It offers much better performance in a high-dimensional feature space than the R*-tree [7], which has been implemented in the Nephele system before.

Image retrieval of SEM images

Not only UV and VIS images can be used for image retrieval functionality to help the art restorer. As was mentioned before scanning electron microscope with energy-

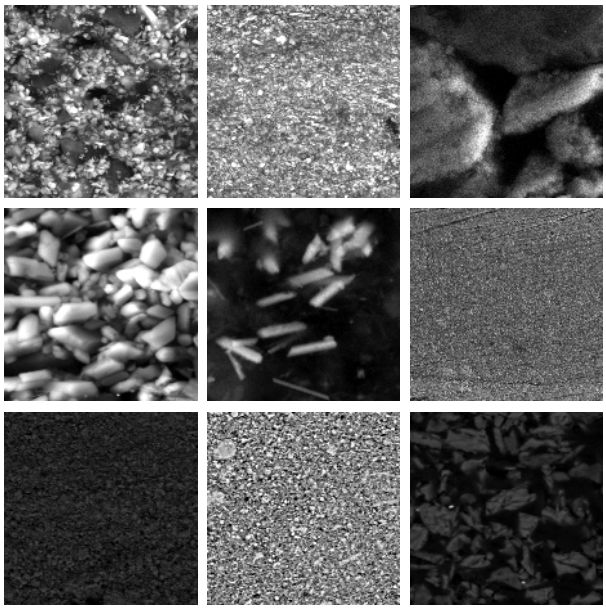
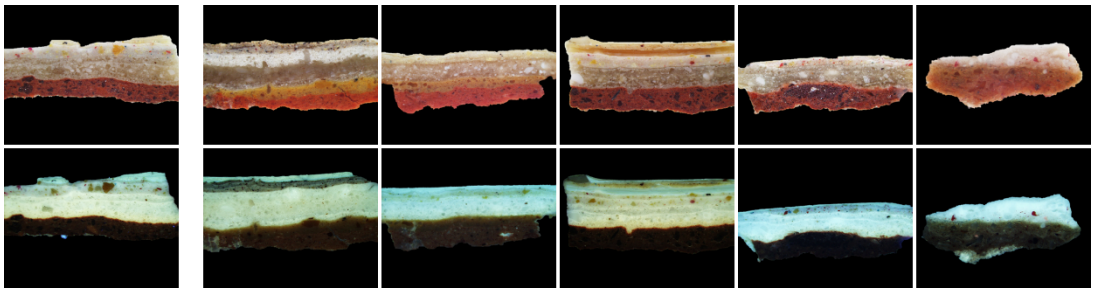


Figure 7.6: Several SEM cutouts which form a basis for image retrieval. Each exhibits texture of different properties in terms of coarseness, contrast, directionality etc. There are total of 296 cutouts in database. Image courtesy of ALMA.

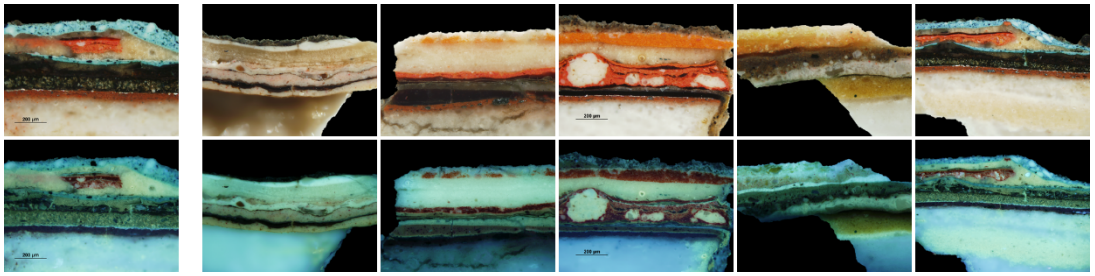
dispersive X-ray spectroscopy detector (SEM/EDX) has the ability to capture chemical contrast of different materials. Different materials thus have different texture in terms of coarseness, contrast or directionality [95] and this property of SEM images can be exploit to implement successful retrieval of similar painting materials. The user would submit the part of material texture image (in SEM) and would expect the set of similar materials to be retrieved. This would help him to find all the artwork reports in database where correspondent materials were used by the original author.

Cutouts of textures from SEM images form a basis for implemented image retrieval. Original SEM image generally contains several layers of different painting materials. It is feasible to cut out certain parts which contain texture of only one material and use this cutout as an input for image retrieval. Exhibit 7.6 shows several examples of different materials and textures. There are total of 296 cutouts in database for image retrieval.

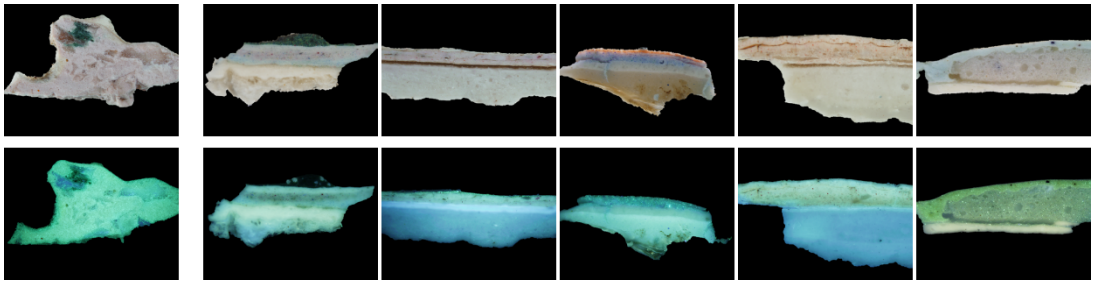
As in case of UV and VIS image retrieval of SEM cutouts is realized through the image feature descriptors and similarity measure for finding the nearest neighbors. The workflow corresponds to that in figure 7.5. However descriptors used here are different than in UV and VIS images case. Instead of co-occurrence matrices, which turned out to be good method to capture relationship between single material layers, the multiresolution wavelet transform is applied to SEM cutouts. Wavelet transform [62, 22] can reflect local details and changes in several scales levels and therefore it can be a very useful tool for characterization of textures. SEM cutout is decomposed to the depth of 2 using Daubechies orthogonal wavelet (db4), which proved suitable for the case. The decomposed texture patch is then represented using the nonnormalized Shannon entropy of the individual high frequency bands (HL, LH and HH using common wavelet notation) [57]. Entropy feature performed the best compared to other textural features like energy or variance (or other Haralick features mentioned above).



(a)



(b)



(c)

Figure 7.7: The favorable results of the image retrieval. Query pair of VIS and UV images is in the leftmost column, while the five nearest neighbor pairs are on the right side. The query pair is found as the nearest neighbor by retrieval system. It is omitted in the illustration. Image courtesy of ALMA.

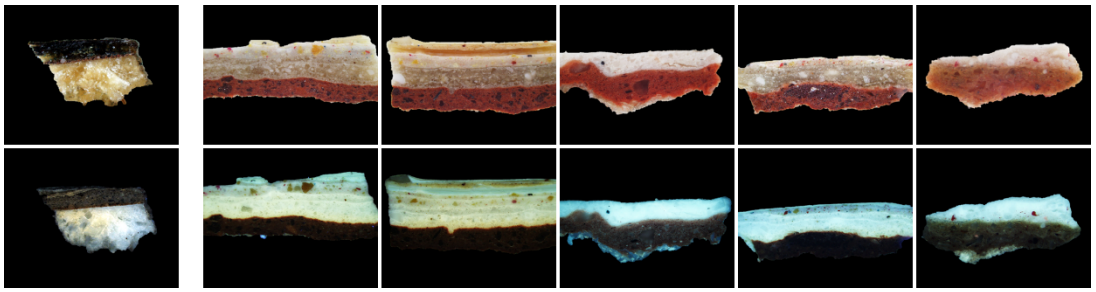
$$\text{Entropy} = -\sum_{i=1}^{N}\sum_{j=1}^{M} C^2(i,j)\log C^2(i,j)$$

where $C(i,j)$ is a wavelet coefficient.

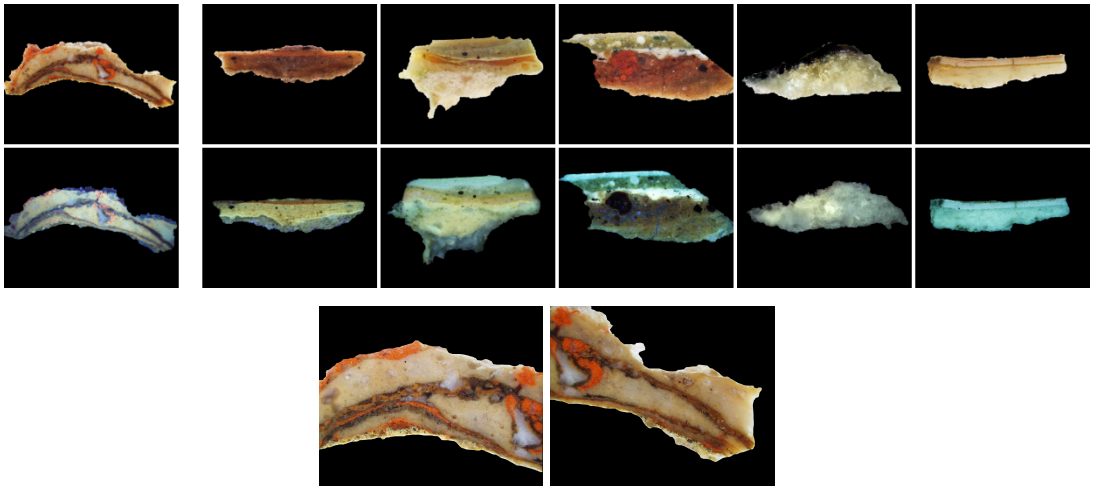
These entropy values of individual bands from wavelet decomposition form the feature vector. It is compared via Euclidean distance to all feature vectors stored in a database and specified number of the nearest neighbors is retrieved. The search can be sped up by incorporation of indexing structure as in previous case.

7.5 Demonstration of the Nephele system

This section presents several illustrative pictures of the Nephele system. First, results of the implemented image retrieval facility are introduced. The real life screenshots of the system form the second part of the section.



(a)

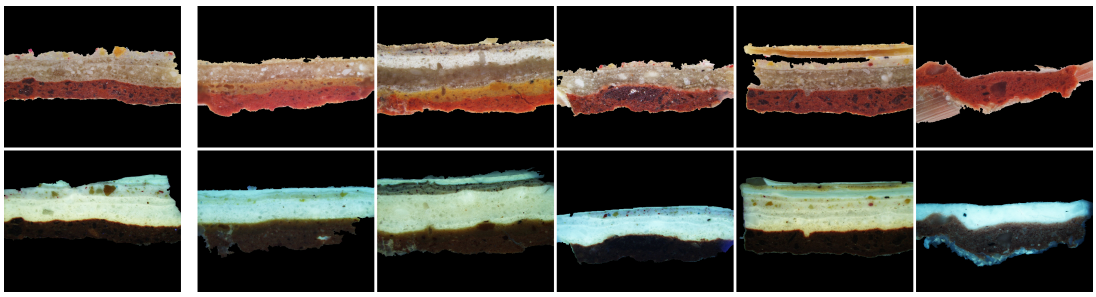


(b)

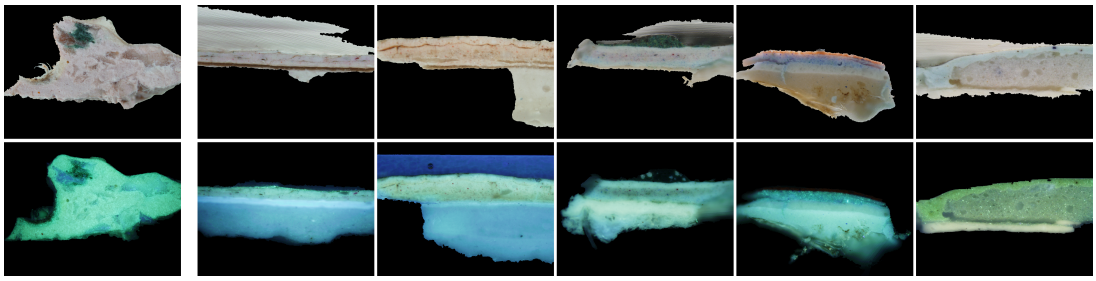
Figure 7.8: Demonstration of image retrieval failure. In subfigure (a) it is caused by the absence of resemblant image pairs in the database. In subfigure (b) the two visually similar VIS images are absent in the results. The query pair is always found as the nearest neighbor by retrieval system. It is omitted in the illustration. Image courtesy of ALMA.

The figures 7.7 and 7.8 show the results of the image retrieval of UV and VIS images. There were 146 pairs of VIS and UV images available in the database. In the leftmost column there is the query pair provided by user, in five columns on the right side there are retrieved nearest neighbors of the query pair (both VIS and UV images are displayed). In all cases the same image pair as the query is retrieved as the nearest neighbor (the image retrieval thus functions correctly). Nonetheless it is omitted in the figures for illustration purposes. To demonstrate the capabilities of the image retrieval system the cross-section images in the figures are processed by an "optimum" segmentation algorithm which delivers GT results. The figure 7.9 exhibits the results on the images which are preprocessed according to the workflow described in section 7.4. The results are comparable since only the order changes. It is caused by imprecise segmentation of cross-section samples in some cases. While the results of the image retrieval are often satisfactory in terms of visual similarity, the final decision of their usefulness and applicability is left to the art restorer.

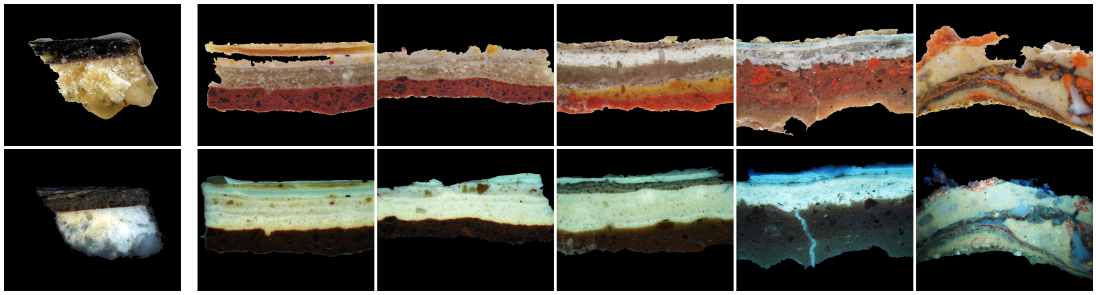
The figure 7.7 presents the favorable results of the retrieval. The five nearest neighbors are visually similar in terms of the texture (its smoothness, coarseness or periodicity) and color. Also the composition of material layers is often alike. Figure 7.8 represents example of rare cases when the CBIR does not return good results. The



(a)



(b)



(c)

Figure 7.9: The results of the image retrieval processed according to the workflow described in section 7.4. Query pair of VIS and UV images is in the leftmost column, while the five nearest neighbor pairs are on the right side. The query pair is found as the nearest neighbor by retrieval system. It is omitted in the illustration. Subfigures (a) and (b) correspond to subfigures (a) and (c) of figure 7.7. Subfigure (c) corresponds to (a) in figure 7.8. The retrieval results are approximately the same up to the order. Image courtesy of ALMA.

reason can be either absence of the resemblant image pairs (subfigure (a)) or simply the failure of the retrieval system (subfigure (b) with two VIS image which are visually similar yet missing in the results).

Figure 7.10 demonstrates the results of SEM cutouts image retrieval. There were 296 cutouts from SEM images of different materials or the same/similar material from different cross-section samples. In the leftmost column there is a query patch, in next columns to the right there are the three nearest neighbors to the query patch (the query image is again omitted as the nearest neighbor found by retrieval system).

The real life screenshots of the Nephele system are described next. The figure 7.11 shows the user interface of the image retrieval in the system. The left form provides the means to submit the query pair of VIS and UV images and to specify the number of the nearest neighbors. The right form contains the list of the retrieved results and the basic information for each of them. There are also buttons for consequent record administration (viewing, editing etc.). Figures 7.12 and 7.13 show the process of a new report insertion. On the left side of the form there is tree-based structure which serves as a navigation through the report. On the right side of the form there are addition of one of the samples with photo documentation and filling of the XRF results respectively.

7.6 Conclusion

The Nephele system for art restoration was proposed in this chapter. It can facilitate the work of material researcher and art restorer by providing the easy access to the material research reports they create and store in the database during artwork restoration process. Next to the traditional information system features Nephele is extended by image processing functionality – content-based image retrieval and related image preprocessing methods. The image retrieval simplifies the search among stored reports by exploiting the visual similarity of VIS and UV images of cross-section samples taken from the artwork. SEM images of the very same samples are used for retrieval of similar painting materials. The preprocessing stage implements findings of part I of this thesis. The removal of grinding artifacts is included, image segmentation necessary for successful image retrieval is implemented using several segmentation algorithms and their following combination via majority vote. The Nephele system reflects the best practices used in the art restoration, is open-source and distributed free of charge.

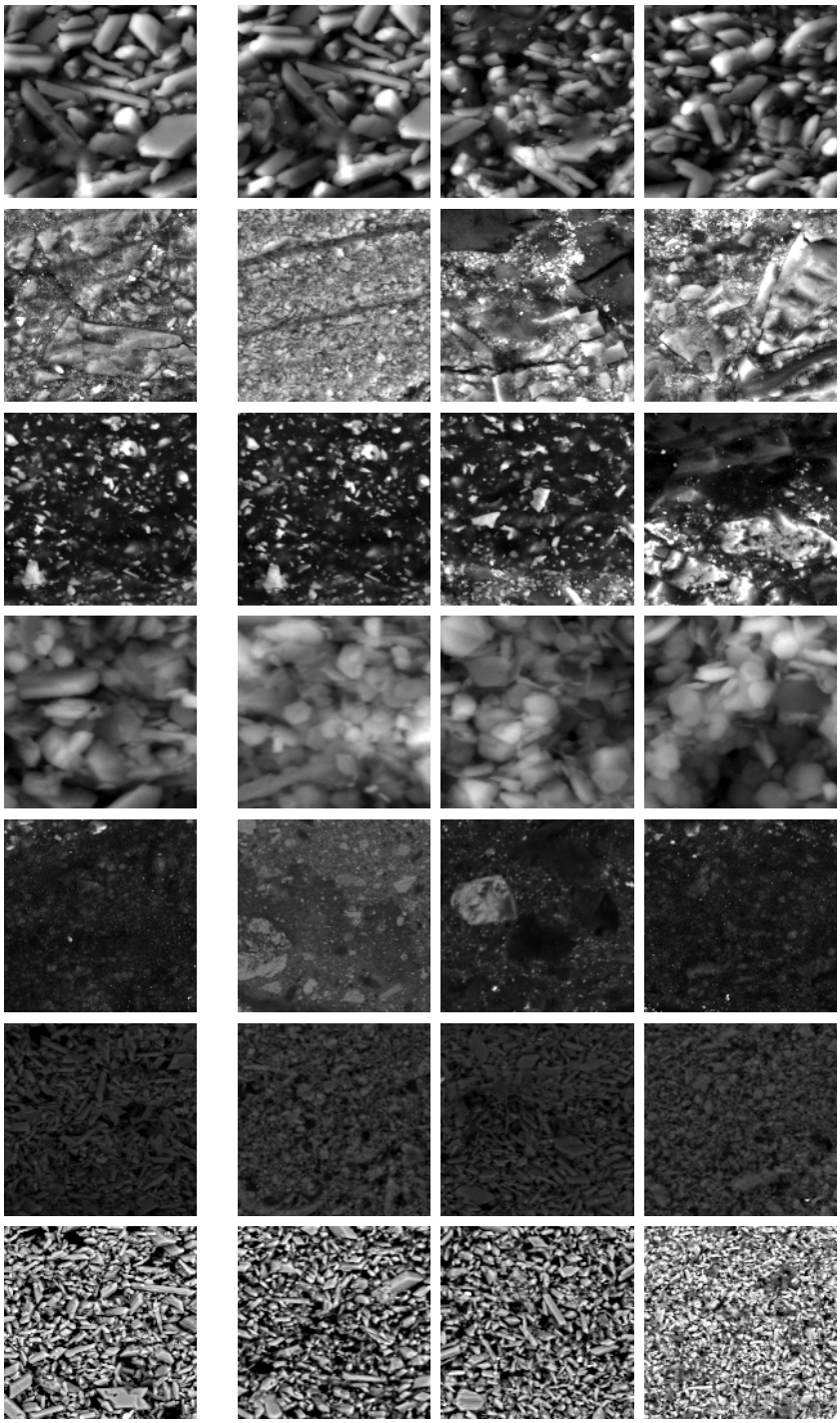


Figure 7.10: The results of the SEM cutouts image retrieval. Query patch is in the leftmost column, the three nearest neighbor patches are on the right side. The query image is always found as the nearest neighbor by retrieval system. It is omitted in the illustration. Image courtesy of ALMA.

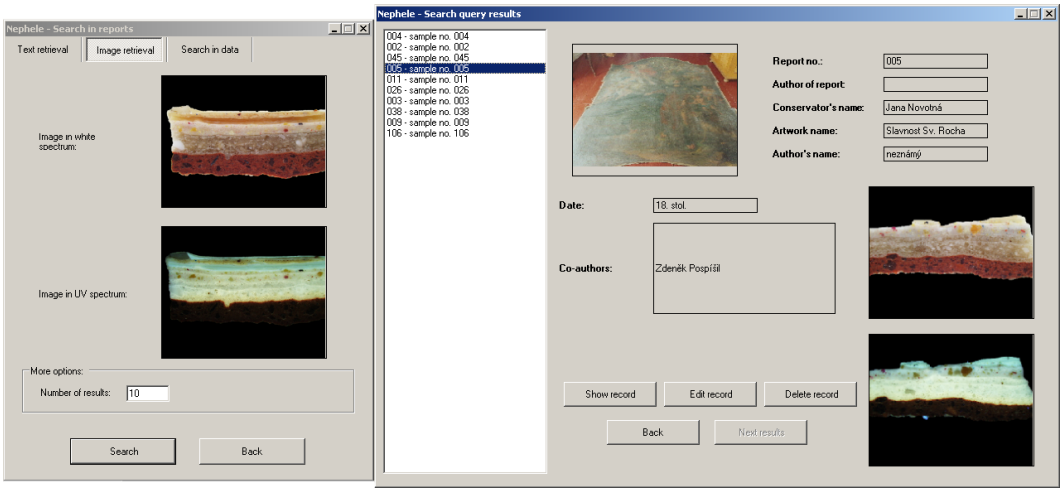


Figure 7.11: The graphical user interface for image retrieval functionality in the system. On the left side there is a form for submission of the query pair images. On the right side there is a form with list of results. Image courtesy of ALMA.

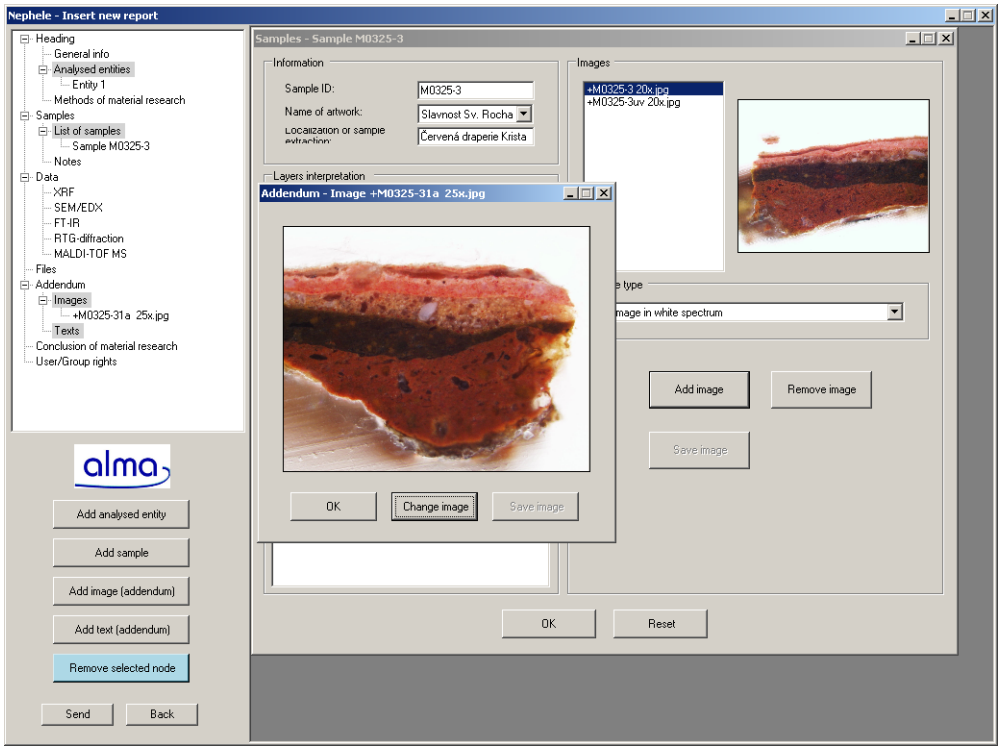


Figure 7.12: Demonstration of new report insertion with tree-based structure on the left side of the form for navigation through report. The rest of the form captures the addition of new cross-section sample with a photodocumentation. Image courtesy of ALMA.

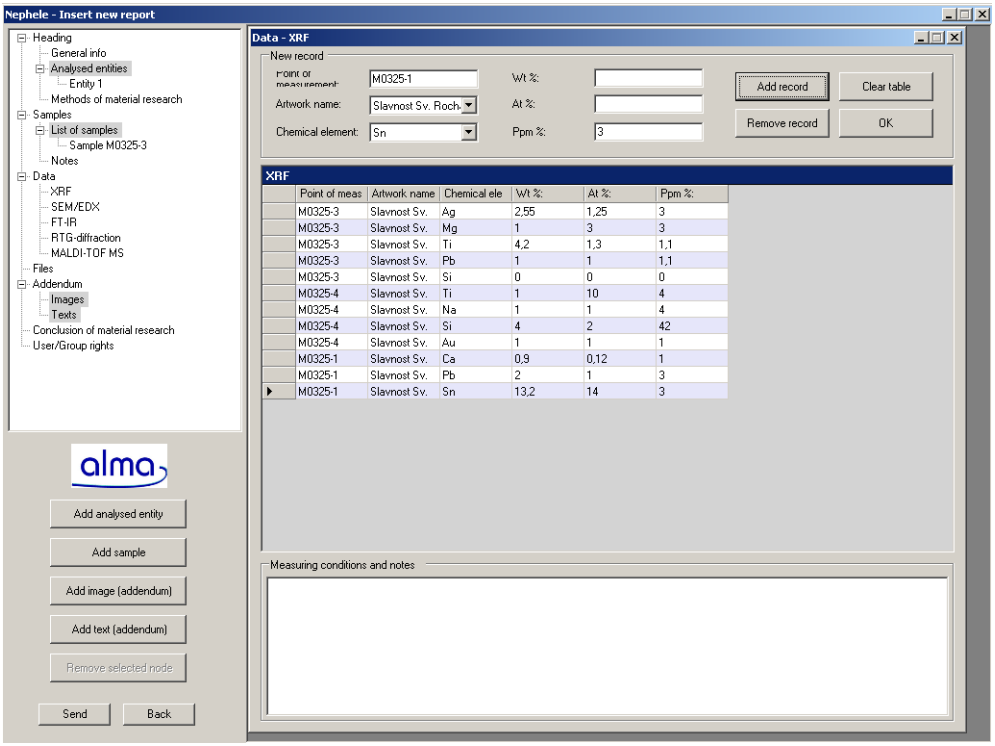


Figure 7.13: The figure demonstrates the filling of XRF analysis results for the new report

Part III

Chapter 8

Contribution of the thesis

The contribution of this doctoral thesis is related to two different problems of digital image processing – image segmentation and image retrieval. All the research was done on a data set of microscopic images, i.e. images of cross-section samples extracted from the artworks. The results are however applicable also to microscopic image data of different origin and similar properties.

Contribution to image segmentation

Concerning the image segmentation many segmentation methods were analyzed in terms of their performance on microscopic image data. The set of studied segmentation methods covered various approaches such as thresholding, region growing, clustering and graph-based methods. The analysis provided an insight into their behavior which led to recommendations about their suitability for segmentation of such data (in table 4.5). In the process ten objective quality indices were used to evaluate the performance. More indices were incorporated to reduce the bias of each individual index so the objectivity of the evaluation could be assured as much as possible. The gathered findings thus have more value than simple visual evaluation. The set of studied methods and quality indices is summarized in table 4.1.

We showed that there was no free lunch in the presented segmentation problem as no studied method significantly outperformed the others. Mean Shift [17] algorithm performed generally the best but it outperformed the other methods only in fraction of cases (13–27 percent depending on modality). Among other successful methods were Felzenszwalb's algorithm [28], GrabCut [83], or multiscale normalized cut [19]. But also more straightforward approaches as region growing or thresholding managed to outperform the others in limited number of cases.

The best average methods were afterward determined and also the lists of methods ordered by their average performance established using the rank aggregation approach (table 4.4). Mean Shift algorithm was found to be the best average method for two out of three studied modalities (UV and VIS). In the third SEM modality Daněk's optimization of Rousson-Deriche [21, 20, 84] was the best average segmentation method. Already mentioned Felzenszwalb's method, multiscale normalized cut or Chan-Vese approach optimized by Daněk were also at the top of the lists. These findings are backed up by a large number of experiments and they were verified on testing data set. Their applicability was shown on a data of different origin (microscopic images of biological tissues).

The area of segmentation combination (fusion) was explored next. We showed that

there was a benefit in using a combination approach since it improved the segmentation results significantly (verified by statistical testing). Even in the cases where the results of the best average method and combination were comparable the combination approach was more suitable thanks to its robustness. Several combination schemes were proposed and compared to select the most appropriate. Majority vote using the three best average methods and weighted voting with differently computed weights were considered. Majority vote proved to be the best combination approach for given data set.

Finally, the method for automatic removal of grinding artifacts (necessary preprocessing for image segmentation stage) was found useful also for another cultural heritage application – removal of canvas structure from IR images of the artwork.

Contribution to image retrieval

Image retrieval was the second problem which the thesis addressed. Functional solution of content-based image retrieval concept was presented. It simplifies the work of art restorer with searching the database of cross-section samples by exploiting the visual similarity of UV and VIS images. Likewise SEM images were used as an input for finding similar painting materials. In both cases proper set of feature descriptors was selected specifically for given problem domain. In case of UV and VIS retrieval Haralick's features computed from co-occurrence matrices turned out to deliver satisfactory results. The SEM retrieval exploits entropy descriptor computed from wavelet coefficients. Daubechies orthogonal wavelet was found appropriate for the task. Retrieval system implements tree data structure (M-tree) to speed up the actual searching.

The implementation of image retrieval with supportive image processing methods is included in Nephele system, an expert system for processing and archiving the material research reports with image processing features, designed and implemented for the cultural heritage application area. The system is open-source and distributed free of charge.

The work contributed to number of projects which also supported my research, namely 72507 by the Grant Agency of Charles University; 1M0572 and MSM6046144603 by the Ministry of Education of the Czech Republic; GA203/07/1324, GA102/08/1593 and GAP103/12/2211 by the Grant Agency of the Czech Republic; M100750901 by the Academy of Sciences of the Czech Republic.

Bibliography

[1] ADDIS, Matthew, BONIFACE, Mike, GOODALL, S, GRIMWOOD, Paul, KIM, Sanghee, LEWIS, Paul, MARTINEZ, Kirk, and STEVENSON, Alison. *Integrated image content and metadata search and retrieval across multiple databases*. In Image and Video Retrieval, pp. 91–100. Springer, 2003.

[2] ALJABAR, Paul, HECKEMANN, Rolf A, HAMMERS, Alexander, HAJNAL, Joseph V, and RUECKERT, Daniel. *Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy*. Neuroimage, 46(3):726–738, 2009.

[3] ANITHA, Anila, BRASOVEANU, Andrei, DUARTE, Marco, HUGHES, Shannon, DAUBECHIES, Ingrid, DIK, Joris, JANSSENS, Koen, and ALFELD, Matthias. *Restoration of X-ray fluorescence images of hidden paintings*. Signal processing, 93(3):592–604, 2013.

[4] ARBELAEZ, Pablo, MAIRE, Michael, FOWLKES, Charless, and MALIK, Jitendra. *Contour detection and hierarchical image segmentation*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 33(5):898–916, 2011.

[5] ARTAECHEVARRIA, Xabier, MUNOZ-BARRUTIA, Arrate, and ORTIZ-DE SOLORZANO, Carlos. *Combination strategies in multi-atlas image segmentation: Application to brain MR data*. Medical Imaging, IEEE Transactions on, 28(8):1266–1277, 2009.

[6] BARTOLINI, Ilaria, CIACCIA, Paolo, and PATELLA, Marco. *Adaptively browsing image databases with PIBE*. Multimedia Tools and Applications, 31(3):269–286, 2006.

[7] BECKMANN, N., KRIEGEL, H.-P., SCHNEIDER, R., and SEEGER, B. *The R^*-tree: An efficient and robust access method for points and rectangles*. In 1990 ACM SIGMOD International Conference on Management of Data, vol. 19, pp. 322–331. 1990.

[8] BENEŠ, Miroslav, BARBARA, Zitová, JAN, Flusser, HRADILOVÁ, Janka, and HRADIL, David. *The image processing system for art specimens: Nephele*. In Proceedings of 14th European Signal Processing Conference. EUSIPCO, pp. 1–5. 2006.

[9] BENEŠ, Miroslav, ZITOVÁ, Barbara, BLAŽEK, Jan, HRADILOVÁ, Janka, and HRADIL, David. *Removing the artifacts from artwork cross-section images*. In 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 3537–3540. IEEE, 2011.

[10] BUCKLOW, Spike L. *A stylometric analysis of craquelure*. Computers and the Humanities, 31(6):503–521, 1997.

[11] CAPPELLINI, Vito, BARNI, Mauro, CORSINI, Massimiliano, DE ROSA, Alessia, and PIVA, Alessandro. *ArtShop: an art-oriented image-processing tool for cultural heritage applications*. The Journal of Visualization and Computer Animation, 14(3):149–158, 2003.

[12] CARNEIRO, Gustavo, DA SILVA, Nuno Pinho, DEL BUE, Alessio, and COSTEIRA, João Paulo. *Artistic image classification: an analysis on the PRINT-ART database*. In Computer Vision–ECCV 2012, pp. 143–157. Springer, 2012.

[13] CHAN, Tony F and VESE, Luminita A. *Active contours without edges*. Image Processing, IEEE Transactions on, 10(2):266–277, 2001.

[14] CIACCIA, Paolo, PATELLA, Marco, and ZEZULA, Pavel. *M-tree: An Efficient Access Method for Similarity Search in Metric Spaces*. In Proceedings of the 23rd International Conference on Very Large Data Bases, VLDB '97, pp. 426–435. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997. ISBN 1-55860-470-7.
URL `http://dl.acm.org/citation.cfm?id=645923.671005`

[15] COHEN, Jacob. *A coefficient of agreement for nominal scales*. Educational and Psychological Measurement, 20(1):37–46, 1960.

[16] COLANTONI, Philippe, PILLAY, Ruven, LAHANIER, Christian, and PITZALIS, Denis. *Analysis of multispectral images of paintings*. In 14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy. 2006.

[17] COMANICIU, Dorin and MEER, Peter. *Mean shift: A robust approach toward feature space analysis*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(5):603–619, 2002.

[18] COSTANTINI, L, MANGIATORDI, F, PALLOTTI, E, and SITA, P. *CHIP – Cultural heritage image processing tool*. In Communications Control and Signal Processing (ISCCSP), 2012 5th International Symposium on, pp. 1–6. IEEE, 2012.

[19] COUR, Timothee, BENEZIT, Florence, and SHI, Jianbo. *Spectral segmentation with multiscale graph decomposition*. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2, pp. 1124–1131. IEEE, 2005.

[20] DANĚK, Ondřej. Graph cut based image segmentation in fluorescence microscopy. Ph.D. thesis, Masarykova univerzita, Brno, 2012.

[21] DANĚK, Ondřej and MATULA, Pavel. *Graph Cuts and Approximation of the Euclidean Metric on Anisotropic Grids*. In VISAPP International Conference on Computer Vision Theory and Applications, pp. 68–73. Portugal: Institute for Systems and Technologies of Information, Control and Communication, 2010. ISBN 978-989-674-028-3.

[22] DAUBECHIES, Ingrid ET AL. Ten lectures on wavelets, vol. 61. SIAM, 1992.

[23] DEMPSTER, Arthur P, LAIRD, Nan M, and RUBIN, Donald B. *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society. Series B (Methodological), pp. 1–38, 1977.

[24] DICE, Lee R. *Measures of the amount of ecologic association between species*. Ecology, 26(3):297–302, 1945.

[25] DOYLE, W. *Operations useful for similarity-invariant pattern recognition*. Journal of the ACM (JACM), 9(2):259–267, 1962.

[26] DWORK, Cynthia, KUMAR, Ravi, NAOR, Moni, and SIVAKUMAR, Dandapani. *Rank aggregation methods for the web*. In Proceedings of the 10th international conference on World Wide Web, pp. 613–622. ACM, 2001.

[27] EASTON JR, Roger L. Fourier methods in imaging. John Wiley & Sons, 2010.

[28] FELZENSZWALB, Pedro F and HUTTENLOCHER, Daniel P. *Efficient graph-based image segmentation*. International Journal of Computer Vision, 59(2):167–181, 2004.

[29] FLICKNER, Myron, SAWHNEY, Harpreet, NIBLACK, Wayne, ASHLEY, Jonathan, HUANG, Qian, DOM, Byron, GORKANI, Monika, HAFNER, Jim, LEE, Denis, PETKOVIC, Dragutin, ET AL. *Query by image and video content: The QBIC system*. Computer, 28(9):23–32, 1995.

[30] FOWLKES, Edward B and MALLOWS, Colin L. *A method for comparing two hierarchical clusterings*. Journal of the American Statistical Association, 78(383):553–569, 1983.

[31] FRANEK, Lucas, ABDALA, Daniel Duarte, VEGA-PONS, Sandro, and JIANG, Xiaoyi. *Image segmentation fusion using general ensemble clustering methods*. In Computer Vision–ACCV 2010, pp. 373–384. Springer, 2011.

[32] GELASCA, Elisa Drelie, GUZMAN, Joriz De, GAUGLITZ, Steffen, GHOSH, Pratim, XU, JieJun, MOXLEY, Emily, RAHIMI, Amir M, BI, Zhiqiang, and MANJUNATH, BS. *CORTINA: Searching a 10 Million+ Images Database'*. In Proc. of VLDB, vol. 7. Citeseer, 2007.

[33] GHAEMI, Reza, SULAIMAN, Md, IBRAHIM, Hamidah, MUSTAPHA, Norwati, ET AL. *A Survey: Clustering Ensembles Techniques.* Proceedings of World Academy of Science: Engineering & Technology, 50, 2009.

[34] GLASBEY, Chris A. *An analysis of histogram-based thresholding algorithms*. CVGIP: Graphical Models and Image Processing, 55(6):532–537, 1993.

[35] GOODALL, Simon, LEWIS, Paul H, MARTINEZ, Kirk, SINCLAIR, Patrick AS, GIORGINI, Fabrizio, ADDIS, Matthew J, BONIFACE, Mike J, LAHANIER, Christian, and STEVENSON, James. *SCULPTEUR: multimedia retrieval for museums*. In Image and Video Retrieval, pp. 638–646. Springer, 2004.

[36] GRADY, Leo. *Random walks for image segmentation*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(11):1768–1783, 2006.

[37] GUNSEL, Bilge, SARIEL, Sanem, and ICOGLU, Oguz. *Content-based access to art paintings*. In Image Processing, 2005. ICIP 2005. IEEE International Conference on, vol. 2, pp. II–558. IEEE, 2005.

[38] HAINDL, M. and MIKEŠ, S. *Texture Segmentation Benchmark*. In Proceedings of the 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4. Los Alamitos: IEEE Computer Society, 2008. ISBN 978-1-4244-2174-9. ISSN 1051-4651. doi:http://dx.doi.org/10.1109/ICPR.2008.4761118.

[39] HAMMING, Richard W. *Error detecting and error correcting codes*. Bell System technical journal, 29(2):147–160, 1950.

[40] HARALICK, Robert M, SHANMUGAM, Karthikeyan, and DINSTEIN, Its' Hak. *Textural features for image classification*. Systems, Man and Cybernetics, IEEE Transactions on, (6):610–621, 1973.

[41] HECKEMANN, Rolf A, HAJNAL, Joseph V, ALJABAR, Paul, RUECKERT, Daniel, and HAMMERS, Alexander. *Automatic anatomical brain MRI segmentation combining label propagation and decision fusion*. NeuroImage, 33(1):115–126, 2006.

[42] HOOVER, Adam, JEAN-BAPTISTE, Gillian, JIANG, Xiaoyi, FLYNN, Patrick J, BUNKE, Horst, GOLDGOF, Dmitry B, BOWYER, Kevin, EGGERT, David W, FITZGIBBON, Andrew, and FISHER, Robert B. *An experimental comparison of range image segmentation algorithms*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 18(7):673–689, 1996.

[43] HU, Roland and DAMPER, Robert I. *A 'No Panacea Theorem' for classifier combination*. Pattern Recognition, 41(8):2665–2673, 2008.

[44] HUANG, Liang-Kai and WANG, Mao-Jiun J. *Image thresholding by minimizing the measures of fuzziness*. Pattern recognition, 28(1):41–51, 1995.

[45] HUANG, Qian and DOM, Byron. *Quantitative methods of evaluating image segmentation*. In International Conference on Image Processing, 1995. Proceedings., vol. 3, pp. 53–56. IEEE, 1995.

[46] HUBERT, Lawrence and ARABIE, Phipps. *Comparing partitions*. Journal of Classification, 2(1):193–218, 1985.

[47] JACCARD, Paul. *The distribution of the flora in the alpine zone*. New phytologist, 11(2):37–50, 1912.

[48] JOHNSON, C Richard, HENDRIKS, Ella, BEREZHNOY, Igor J, BREVDO, Eugene, HUGHES, Shannon M, DAUBECHIES, Ingrid, LI, Jia, POSTMA, Eric, and WANG, James Z. *Image processing for artist identification*. Signal Processing Magazine, IEEE, 25(4):37–48, 2008.

[49] KAMENICKÝ, Jan and ZITOVÁ, Barbara. *Contrast preserving color fusion*. In Proceedings of SPIE, vol. 7866, pp. 78660L–7. 2011.

[50] KAPUR, JN, SAHOO, Prasanna K, and WONG, AKC. *A new method for gray-level picture thresholding using the entropy of the histogram*. Computer Vision, Graphics, and Image Processing, 29(3):273–285, 1985.

[51] KITTLER, Josef and ALKOOT, Fuad M. *Sum versus vote fusion in multiple classifier systems*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 25(1):110–115, 2003.

[52] KITTLER, Josef, HATEF, Mohamad, DUIN, Robert PW, and MATAS, Jiri. *On combining classifiers*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20(3):226–239, 1998.

[53] KITTLER, Josef and ILLINGWORTH, John. *Minimum error thresholding*. Pattern Recognition, 19(1):41–47, 1986.

[54] KOHLI, Pushmeet, LADICKÝ, Lubor, and TORR, Philip HS. *Robust higher order potentials for enforcing label consistency*. International Journal of Computer Vision, 82(3):302–324, 2009.

[55] KUNCHEVA, LI, HADJITODOROV, ST, and TODOROVA, LP. *Experimental comparison of cluster ensemble methods*. In 2006 9th International Conference on Information Fusion, pp. 1–7. IEEE, 2006.

[56] KUNCHEVA, Ludmila I. Combining pattern classifiers: methods and algorithms. John Wiley & Sons, 2004.

[57] LAINE, Andrew and FAN, Jian. *Texture classification by wavelet packet signatures*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 15(11):1186–1191, 1993.

[58] LI, CH and TAM, Peter Kwong-Shun. *An iterative algorithm for minimum cross entropy thresholding*. Pattern Recognition Letters, 19(8):771–776, 1998.

[59] LUX, Mathias. *Content based image retrieval with lire*. In Proceedings of the 19th ACM international conference on Multimedia, pp. 735–738. ACM, 2011.

[60] MACQUEEN, James ET AL. *Some methods for classification and analysis of multivariate observations*. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, pp. 281–297. California, USA, 1967.

[61] MALCOLM, James, RATHI, Yogesh, and TANNENBAUM, Allen. *A graph cut approach to image segmentation in tensor space*. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pp. 1–8. IEEE, 2007.

[62] MALLAT, Stephane G. *A theory for multiresolution signal decomposition: the wavelet representation*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 11(7):674–693, 1989.

[63] MARTIN, David, FOWLKES, Charless, TAL, Doron, and MALIK, Jitendra. *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*. In Eighth IEEE International Conference on Computer Vision, 2001. Proceedings., vol. 2, pp. 416–423. IEEE, 2001.

[64] MEILĂ, Marina. *Comparing clusterings: an information based distance*. Journal of Multivariate Analysis, 98(5):873–895, 2007.

[65] MIRKIN, B.G. Mathematical Classification and Clustering. Mathematics and Its Applications. Kluwer Academic Publishers, 1996. ISBN 9780792341598. URL `http://books.google.cz/books?id=brzLe4X4ypEC`

[66] OGLE, Virginia E. and STONEBRAKER, Michael. *Chabot: Retrieval from a relational database of images*. Computer, 28(9):40–48, 1995.

[67] ORION ANALYTICAL. *Preparing layered samples for microscopic examination in cross-section*. `http://www.orionanalytical.com/Preparingcross-sections.pdf`. Accessed: 18.4.2014.

[68] OTSU, Nobuyuki. *A threshold selection method from gray-level histograms*. Automatica, 11(285-296):23–27, 1975.

[69] PAL, Nikhil R and PAL, Sankar K. *A review on image segmentation techniques*. Pattern Recognition, 26(9):1277–1294, 1993.

[70] PELAGOTTI, Anna, DEL MASTIO, Andrea, DE ROSA, Alessia, and PIVA, Alessandro. *Multispectral imaging of paintings*. Signal Processing Magazine, IEEE, 25(4):27–36, 2008.

[71] PENTLAND, Alex, PICARD, Rosalind W, and SCLAROFF, Stan. *Photobook: Content-based manipulation of image databases*. International Journal of Computer Vision, 18(3):233–254, 1996.

[72] PIHUR, Vasyl, DATTA, Susmita, and DATTA, Somnath. *RankAggreg, an R package for weighted rank aggregation*. BMC Bioinformatics, 10(1):62, 2009.

[73] PRATT, William K. Digital Image Processing: PIKS Scientific Inside. Wiley-Interscience, 2007, 4th ed. ISBN 0471767778.

[74] PREWITT, Judith and MENDELSOHN, Mortimer L. *The analysis of cell images*. Annals of the New York Academy of Sciences, 128(3):1035–1053, 1966.

[75] RAND, William M. *Objective criteria for the evaluation of clustering methods*. Journal of the American Statistical Association, 66(336):846–850, 1971.

[76] RIBES, Alejandro, PILLAY, Ruven, SCHMITT, Francis, and LAHANIER, Christian. *Studying that smile*. Signal Processing Magazine, IEEE, 25(4):14–26, 2008.

[77] RIDLER, TW and CALVARD, S. *Picture thresholding using an iterative selection method*. IEEE transactions on Systems, Man and Cybernetics, 8(8):630–632, 1978.

[78] RIJSBERGEN, C. J. Van. Information Retrieval. Newton, MA, USA: Butterworth-Heinemann, 1979, 2nd ed. ISBN 0408709294.

[79] ROHLFING, Torsten, BRANDT, Robert, MENZEL, Randolf, and MAURER JR, Calvin R. *Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains*. NeuroImage, 21(4):1428–1442, 2004.

[80] ROHLFING, Torsten, BRANDT, Robert, MENZEL, Randolf, RUSSAKOFF, Daniel B, and MAURER JR, Calvin R. *Quo vadis, atlas-based segmentation?* In Handbook of Biomedical Image Analysis, pp. 435–486. Springer, 2005.

[81] ROHLFING, Torsten and MAURER JR, Calvin R. *Shape-based averaging*. Image Processing, IEEE Transactions on, 16(1):153–161, 2007.

[82] ROSENFELD, Azriel and DE LA TORRE, Pilar. *Histogram concavity analysis as an aid in threshold selection*. Systems, Man and Cybernetics, IEEE Transactions on, (2):231–235, 1983.

[83] ROTHER, Carsten, KOLMOGOROV, Vladimir, and BLAKE, Andrew. *Grabcut: Interactive foreground extraction using iterated graph cuts*. ACM Transactions on Graphics, 23(3):309–314, 2004.

[84] ROUSSON, Mikaël and DERICHE, Rachid. *A variational framework for active and adaptative segmentation of vector valued images*. In Motion and Video Computing, 2002. Proceedings. Workshop on, pp. 56–61. IEEE, 2002.

[85] SHANBHAG, Abhijit G. *Utilization of information measure as a means of image thresholding*. CVGIP: Graphical Models and Image Processing, 56(5):414–419, 1994.

[86] SHARKEY, Amanda J C. *On combining artificial neural nets*. Connection Science, 8(3-4):299–314, 1996.

[87] SHI, Jianbo and MALIK, Jitendra. *Normalized cuts and image segmentation*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(8):888–905, 2000.

[88] SLUIMER, Ingrid, PROKOP, Mathias, and VAN GINNEKEN, Bram. *Toward automated segmentation of the pathological lung in CT*. Medical Imaging, IEEE Transactions on, 24(8):1025–1038, 2005.

[89] SMEULDERS, Arnold W. M., WORRING, Marcel, SANTINI, Simone, GUPTA, Amarnath, and JAIN, Ramesh. *Content-based image retrieval at the end of the early years*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(12):1349–1380, 2000.

[90] SMITH, John R and CHANG, Shih-Fu. *VisualSEEk: a fully automated content-based image query system*. In Proceedings of the fourth ACM international conference on Multimedia, pp. 87–98. ACM, 1997.

[91] SPAGNOLO, G Schirripa. *Virtual restoration: detection and removal of craquelure in digitized image of old paintings*. In SPIE Optical Metrology, pp. 80840B–80840B. International Society for Optics and Photonics, 2011.

[92] SROUBEK, Filip and FLUSSER, Jan. *Multichannel blind deconvolution of spatially misaligned images*. Image Processing, IEEE Transactions on, 14(7):874–883, 2005.

[93] STORK, DG. *Optics and realism in Renaissance art.* Scientific American, 291(6):76–84, 2004.

[94] STREHL, Alexander and GHOSH, Joydeep. *Cluster ensembles – A knowledge reuse framework for combining multiple partitions*. The Journal of Machine Learning Research, 3:583–617, 2003.

[95] TAMURA, Hideyuki, MORI, Shunji, and YAMAWAKI, Takashi. *Textural features corresponding to visual perception*. Systems, Man and Cybernetics, IEEE Transactions on, 8(6):460–473, 1978.

[96] TAO, Wenbing, JIN, Hai, ZHANG, Yimin, LIU, Liman, and WANG, Desheng. *Image thresholding using graph cuts*. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 38(5):1181–1195, 2008.

[97] TOPCHY, Alexander P, LAW, Martin HC, JAIN, Anil K, and FRED, Ana L. *Analysis of consensus partition in cluster ensemble*. In Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on, pp. 225–232. IEEE, 2004.

[98] TSAI, Wen-Hsiang. *Moment-preserving thresolding: A new approach*. Computer Vision, Graphics, and Image Processing, 29(3):377–393, 1985.

[99] VEGA-PONS, Sandro, JIANG, Xiaoyi, and RUIZ-SHULCLOPER, José. *Segmentation ensemble via kernels*. In 2011 First Asian Conference on Pattern Recognition (ACPR), pp. 686–690. IEEE, 2011.

[100] VEGA-PONS, Sandro and RUIZ-SHULCLOPER, José. *A survey of clustering ensemble algorithms*. International Journal of Pattern Recognition and Artificial Intelligence, 25(03):337–372, 2011.

[101] VELTKAMP, Remco C and TANASE, Mirela. *Content-based image retrieval systems: A survey*. 2000.

[102] VINH, Nguyen Xuan, EPPS, Julien, and BAILEY, James. *Information theoretic measures for clusterings comparison: is a correction for chance necessary?* In Proceedings of the 26th Annual International Conference on Machine Learning, pp. 1073–1080. ACM, 2009.

[103] WARFIELD, Simon K, ZOU, Kelly H, and WELLS, William M. *Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation*. Medical Imaging, IEEE Transactions on, 23(7):903–921, 2004.

[104] WARRENS, Matthijs J. *On the equivalence of Cohen's kappa and the Hubert-Arabie adjusted Rand index*. Journal of Classification, 25(2):177–183, 2008.

[105] WILCOXON, Frank. *Individual comparisons by ranking methods*. Biometrics Bulletin, 1(6):80–83, 1945.

[106] YEN, Jui-Cheng, CHANG, Fu-Juay, and CHANG, Shyang. *A new criterion for automatic multilevel thresholding*. Image Processing, IEEE Transactions on, 4(3):370–378, 1995.

[107] ZACK, GW, ROGERS, WE, and LATT, SA. *Automatic measurement of sister chromatid exchange frequency.* Journal of Histochemistry & Cytochemistry, 25(7):741–753, 1977.

[108] ZHANG, Hui, FRITTS, Jason E, and GOLDMAN, Sally A. *Image segmentation evaluation: A survey of unsupervised methods*. Computer Vision and Image Understanding, 110(2):260–280, 2008.

List of Tables

List of Abbreviations

ALMA Academic Materials Research Laboratory of Painted Artworks in Prague

ARI Adjusted Rand index

BHD Boundary Hamming distance

CBIR Content-based image retrieval

DC Dice coefficient

DIP Digital image processing

EGFP Enhanced green fluorescent protein

FMI Fowlkes-Mallows index

FTIR Fourier transform infrared spectroscopy

GC_CV Daněk's optimization of Chan-Vese approach

GC_FH Felzenszwalb's method

GC_R GrabCut

GC_RD Daněk's optimization of Rousson-Deriche approach

GT Ground truth (reference image)

HAUSD Hausdorff distance

HD Hamming distance

HT_* Various thresholding methods from HistThresh

IMJ_* Various thresholding methods from ImageJ

IR Infrared

KM K-means clustering

MALDI-TOF MS Matrix assisted laser desorption ionization time of flight mass spectroscopy

MASD Mean absolute surface distance

MNC Multiscale normalized cut

MS Mean Shift algorithm

MV Majority vote

NMI Normalized mutual information

RI Rand index

RG Region growing

SEM Scanning electron microscope, corresponding modality

SEM/EDX Scanning electron microscope with energy-dispersive X-ray spectroscopy detector

TNC Tao's thresholding method

UV Ultraviolet spectrum, ultraviolet spectrum modality

VI Variation of information

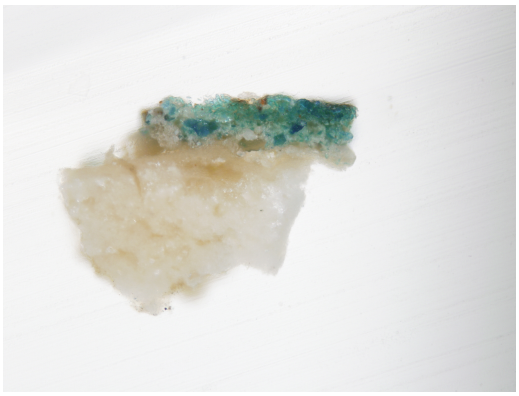
VIS Visible spectrum, visible spectrum modality

XRF X-ray fluorescence

Appendix A

Results of image segmentation algorithms

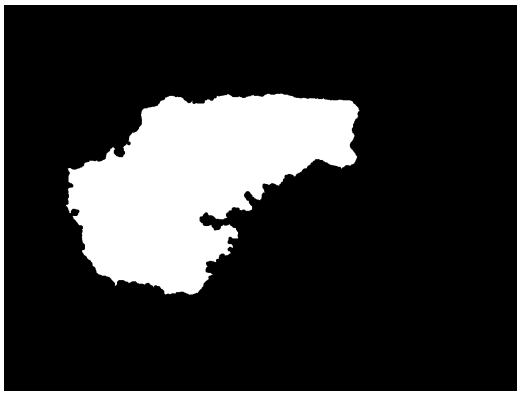
This appendix contains four figures which show results of selected image segmentation algorithms applied on cross-section images. In figure A.1 there is segmentation of VIS image. In figures A.2, A.3 and A.4 there are segmentation results of one cross-section sample in each of three modalities (SEM, UV and VIS).



(a) Original VIS image

(b) Ground truth image mask

(c) Final segmented mask of MS

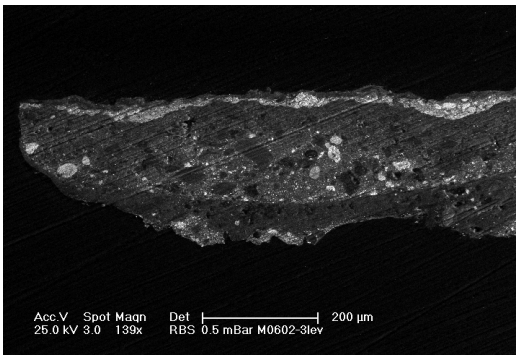


(d) Final segmented mask of KM_RGB

(e) Final segmented mask of IMJ_HUANG

(f) Final segmented mask of RG_25

Figure A.1: Results of four image segmentation algorithms. In (a) there is an original VIS image. Ground truth mask is in subfigure (b). The image is segmented by MS (c), KM_RGB (d), IMJ_HUANG (e) and RG_25 (f). Image in (a) courtesy of ALMA.



(a) Original SEM image

(b) Ground truth image mask

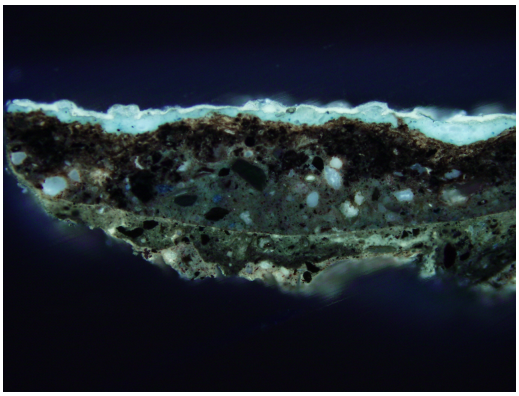
(c) Final segmented mask of MS

(d) Final segmented mask of GC_RD

(e) Final segmented mask of IMJ_PER

(f) Final segmented mask of RG_5

Figure A.2: Results of four image segmentation algorithms. In (a) there is an original SEM image. Ground truth mask is in subfigure (b). The image is segmented by MS (c), GC_RD (d), IMJ_PER (e) and RG_5 (f). Image in (a) courtesy of ALMA.



(a) Original UV image

(b) Ground truth image mask

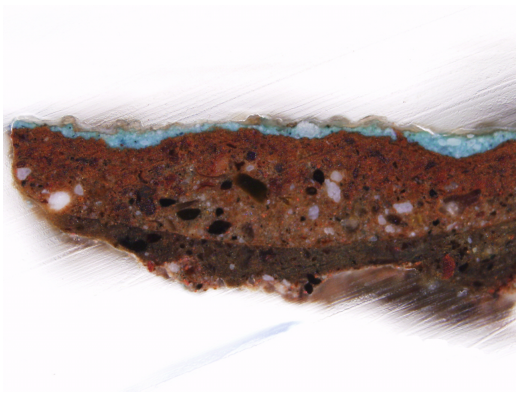
(c) Final segmented mask of MS

(d) Final segmented mask of GC_FH

(e) Final segmented mask of IMJ_TRIANGLE

(f) Final segmented mask of RG_15

Figure A.3: Results of four image segmentation algorithms. In (a) there is an original UV image. Ground truth mask is in subfigure (b). The image is segmented by MS (c), GC_FH (d), IMJ_TRIANGLE (e) and RG_15 (f). Image in (a) courtesy of ALMA.



(a) Original VIS image

(b) Ground truth image mask

(c) Final segmented mask of MS

(d) Final segmented mask of KM_RGB

(e) Final segmented mask of IMJ_HUANG

(f) Final segmented mask of RG_25

Figure A.4: Results of four image segmentation algorithms. In (a) there is an original VIS image. Ground truth mask is in subfigure (b). The image is segmented by MS (c), KM_RGB (d), IMJ_HUANG (e) and RG_25 (f). Image in (a) courtesy of ALMA.