Charles University in Prague

Faculty of Mathematics and Physics

MASTER THESIS

Aleš Zita

Automatic analysis of videokymographic images by means of higher-level features

Department of Software and Computer Science Education

Supervisor of the master thesis: RNDr. Barbara Zitová PhD.

Study programme: Computer Science

Specialization: Software Systems

Prague 2013

I would like to express my greatest gratitude to Barbara Zitová PhD. for exceptional supervision, selfless support and guidance during the writing of my master thesis. Additionally I would like to thank Jan G. Švec, PhD. et PhD. and Christian T. Herbst, PhD. for the provided VKG image set and for valuable consultations on videokymography and laryngology; members of the Department of Image Processing, namely Jiří Sedlář PhD. and Adam Novozámský for the segmented VKG image dataset and Filip Šroubek PhD. for consultation on energy minimisation methods in image processing. Finally I thank all my teachers at MFF UK for their care and guidance during my studies, chairman and members of my Thesis defence committee and opponents of my Thesis for their time, my family for their support and my friends for their encouragement.

I declare that I carried out this master 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…..... date..... signature

Název práce: Automatická analýza videokymografických obrazů použitím vlastností vyšších řádů

Autor: Aleš Zita

Katedra / Ústav: Kabinet software a výuky informatiky / Matematicko-fyzikální fakulta Univerzity Karlovy

Vedoucí diplomové práce: RNDr. Barbara Zitová, Ph.D., Oddělení zpracování obrazové informace, Ústav teorie informace a automatizace, Akademie věd České republiky

Abstrakt: Diagnostika lidských hlasivek je i v dnešní době komplikovaným problémem. Důvodem jsou špatný přístup k orgánu samotnému a vysoké frekvence vibrací hasivek. Jednou z klinicky dostupných zobrazovacích metod řešících tyto problémy je videokymografie – technika založená na snímání lidských hlasivek pomocí speciální řádkové CCD kamery. Jednotlivé řádky poskládány za sebe podle času pak tvoří videokymografický záznam. Videokymografické snímky jsou vhodné pro počítačovou extrakci základních charakteristik hlasivek napomáhající snížení pracovní zátěže vyšetřujícího laryngologa. Za tímto účelem jsou v oddělení Zpracování obrazové informace v UTIA AV ČR vyvíjeny metody automatické detekce charakteristik hlasivkových vibrací, jež lze extrahovat z videokymografického záznamu. Jednou z důležitých, ale obtížně detekovatelných charakteristik je průběh ventrikulární řasy. Cílem této práce je navržení metody automatické detekce ventrikulární řasy na videokymografickém záznamu za pomocí technik digitálního zpracování obrazové informace.

Klíčová slova: videokymografie; hlasivky; ventrikulární řasa; digitální zpracování obrazu

Title: Automatic analysis of videokymographic images by means of higher-level features

Author: Aleš Zita

Department / Institute: Department of Software and Computer Science Education

Supervisor of the master thesis: RNDr. Barbara Zitová, Ph.D., Department of Image Processing, Institute of Information Theory and Automation of the Academy of Science, Czech Republic

Abstract: Human voice diagnosis is a complicated problem, even nowadays. The reason is poor access to the body itself and the high frequencies of vocal fold vibrations. One of the clinically available imaging methods to address these problems is Videokymography – a technology for capturing the vocal fold vibrations using a special line CCD camera. Individual lines stacked on top of each other form videokymographic recording. Videokymographic images are suitable for automatic characteristics extraction, therefore helping to reduce the laryngologist workload. For this purpose, the set of such methods is being developed in the Department of Image Processing in the Institute of Information Theory and Automation of the Academy of Science of Czech Republic. The ventricular band position and shape determination is one of the important, but difficult, tasks. The aim of this thesis is to propose new method of automatic detection of ventricular band on videokymographic recording using digital image processing techniques.

Keywords: Videokymography; vocal folds; ventricular band; image processing

Contents

Preface / Foreword

Diagnosis of human vocal cords function is very hard task by its nature. When dealing with frequencies as high as 80-240Hz doctors and medical practitioners have to face the enormous speed of the cord vibrations. One way to deal with the high frequencies is to record the vibration and analyse the recording afterwards. But this approach creates another problem: a huge amount of data per recording. As an example we can easily calculate, that if we want to replay a 10 second recording of the vocal cords vibration captured by a device with the temporal resolution of 10.000 frames per second, we would need about 2 hours 46 minutes of time, provided the replay speed would be 10 frames per second. (1) This makes it impossible for one-session data recording to be analysed by any man and therefore there is an ongoing search for the ways of processing the data automatically.

This thesis is a free follow-up of my previous work, the bachelor thesis (2) and the diploma thesis of David Hauzar (3). The assignment originated from the work of Jan G. Švec, the co-author of the vocal cords vibration capturing method called 'Videokymography' abbreviated as VKG. This method is based on recording of the glottal area not with a whole-frame camera, but rather with an one-line scanner. This type of capturing device allows achieving several orders of magnitude higher frequency while maintaining the cost of a standard camera. The rows captured in such manner are aligned together, resulting in a two-dimensional image with one spatial and one temporal dimension (Figure 4).

In chapter *Introduction* the insight to the problem and the Videokymography method description can be found. *Objective* contains a depiction of the assignment goals. *Problem analysis* is dedicated to deeper assessment of the situation, together with definitions of methods used in the thesis. In *Solution*, thorough explanation of the proposed method and the algorithm description can be found. Chapter *Results* contains description of the dataset used in experiments and results of processing the data by the proposed method, including the result analysis. Finally, *Conclusion* contains the analysis of proposed solution and suggestions for future work.

1. Introduction

To be able to explain the proposed method, I need to introduce the basics of human voice production and vocal cords anatomy. I will also explain the method of the vocal fold vibration acquisition to help the reader understand the problem and proposed solution.

1.1. Human vocal cords

1.1.1. Description and physiology

It is a well-known fact that the vocal cord vibrations are responsible for the production of human voice. When the lungs contract, the pressure underneath the vocal folds builds up, and the pressure levels below and above the closed vocal folds become unequal. After the pressure difference rises above a certain level, the vocal folds could not hold the pressure, and a small amount of the air is released into the laryngeal cavity. This quantum of air released creates a pressure wave. The process repeats and the sequence of released pressure waves produces a sound. By the use of vocal fold muscle tension, humans are able to adjust the vibration frequency, therefore the pitch of the tone. Typical base frequencies of produced tone are approximately 70 – 500 Hz and 150 – 1000Hz for man and woman respectively. But the range of the frequencies can vary greatly with each individual (4).

1.1.2. Glottal cycle

One cycle of the cord vibration is called the glottal cycle and consists of following phases:

- Closed glottis (Figure 1, bottom)
- Pressure build-up underneath the cords
- Glottis opening and release of one quantum of air
- Closing of the glottis

For the purpose of this thesis we will recognise the closed glottis (approx. 3ms), opened glottis, opening phase (approx. 2ms), closing phase and the opening amplitude (the state when the glottis is opened to its maximum).

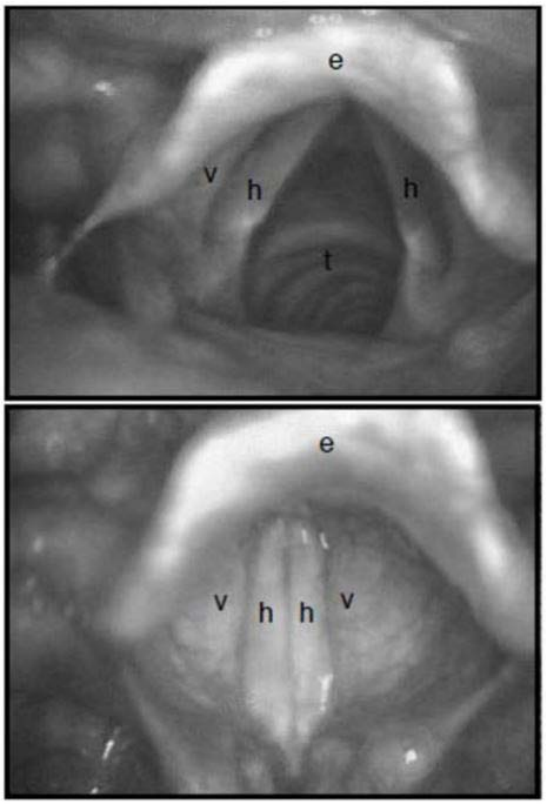

Figure 1. Human vocal cords: h-vocal folds; v-ventricular band; e-epiglottis; t-trachea. The image was taken from (4).

1.1.3. Ventricular band

Ventricular bands are pair of folds of membrane stretching across the laryngeal cavity, also called the vestibular fold or false vocal cord. It is a soft tissue overhanging the vocal fold (Figure 1 and Figure 2). In the case of healthy larynx, the ventricular band displays considerably less apparent vibration then the vocal folds. Abnormal vibration of the ventricular folds can be used as a sign of vocal cord disorder. It is often the case that with one vocal fold paralysed, the impaired

individual is able to utilise the ventricular band making it vibrate instead of the paralysed vocal fold.

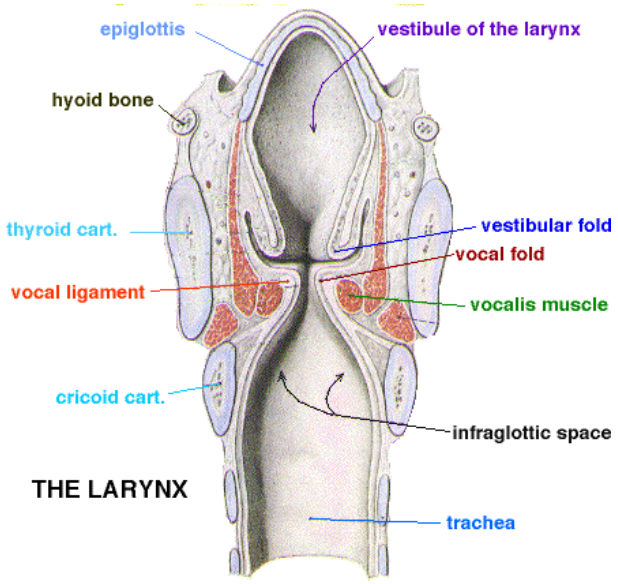

Figure 2. Human larynx. Please note the position of the ventricular bands, also known as vestibular folds. The image taken from (5).

1.2. Data acquisition and processing

1.2.1. Laryngoscopy

Laryngoscopy is a method used for larynx examination. By using the optical system doctors are able to take a look inside the larynx and its parts including the glottis.

The very first known attempt to explore the human voice organ dates over 150 years ago. In 1854 a vocal music teacher Manuel Garcia attempted to explore his own vocal cords using a dentist mirror and a small hand mirror. He published his observations in 1855 (6).

Two years later, it was independently demonstrated by Viennese neurologist Truck and Prague physiologist Johann Nepomuk Czermak (amongst other things the inventor of medical reflector), that this technique can be successfully used in

medicine (6). The imaging techniques suitable for the investigation of human vocal organ made an enormous progress since then.

Nowadays, the procedures used for larynx disease diagnosis in clinical practice are very complicated. They are usually based on the description of the patient's symptoms and empirical evaluations of the data obtained from instrumental and histological examinations. During the last two decades, many sophisticated imaging techniques have been developed allowing the accurate measurement of voice quality (7) (8).

For example, a huge leap forward was made by the use of computed tomography (CT) and magnetic resonance imaging (MRI) techniques for the purpose of vocal cords examination. These techniques allow displaying areas of the vocal cords otherwise hidden from examination by a conventional endoscope, or visualising the depth of the potential tumour (9). Today, these methods are used primarily for the malignant tumour treatment and preoperative examination for an appropriate planning of surgery (10).

1.2.2. Acquisition methods

There are several techniques of capturing human vocal cords vibrations for assessment of their functionality, diseases or disorders. The most commonly used are the *Videostroboscopy*, *High-speed videoendoscopy* and the latest, *Videokymography*.

Videostroboscopy

Probably the most popular method for diagnosis of human vocal cords is the videostroboscopy. It uses endoscope connected to camera for accessing and visualise patient’s larynx. Videostroboscopy is a method of capturing still images in precisely determined moments to simulate the lower frequency video of vocal fold vibration (Figure 3). A high intensity xenon flash is used to substitute fast camera shutter. This way the videostroboscopy recording shows false slow motion video. This can be viewed by the doctor in real time, during the patient examination. This approach has some flaws coming from the method principles. The most unpleasant is the inability to capture any irregularities in the vibrations or vocal cords with aperiodic vibration. (1)

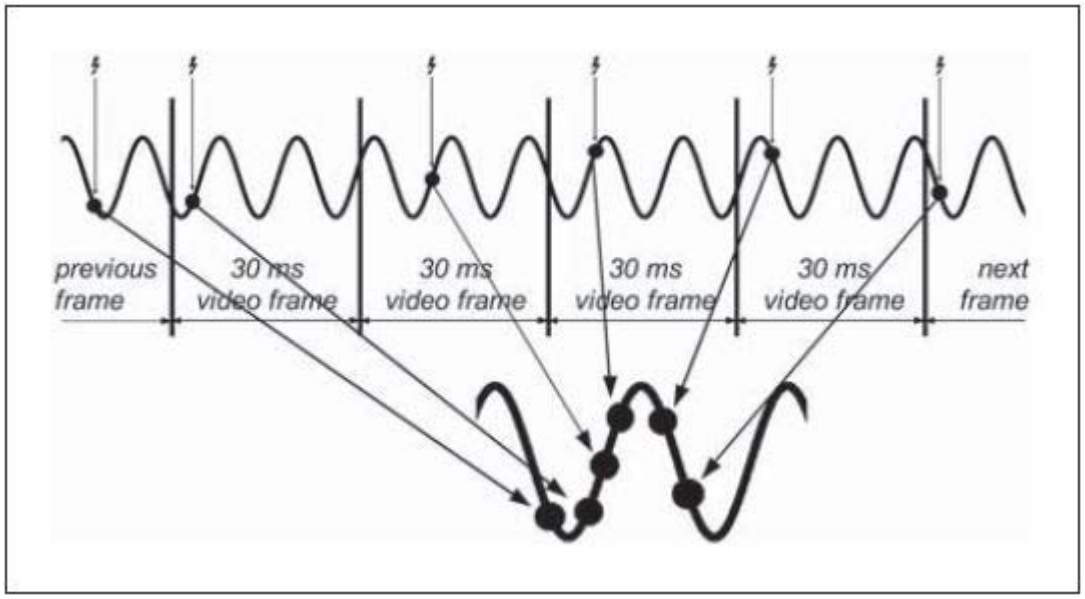

Figure 3. Principles of videostroboscopy: the strobe light flashes in precisely defined moments to capture the pseudo low frequency video. Image is taken from (1).

High-speed videoendoscopy

High-speed videoendoscopy is the ultimate method of recording the true whole-cycle video of human vocal cords vibrations. However, even this technique has its limitations and disadvantages. One of the limiting factors is the cost of the system and other is the amount of data it produces. The typical today's high-speed camera system has up to 10,000 fps temporal resolution with spatial resolution up to 800x600 pixels, true colour, or up to 30,000 monochromatic (1). It means that even very short Laryngologist session will necessarily result in tens or hundreds of Gigabytes of recorded data. As a result, the problems are not only the limits of storage capacity, but also the time for viewing and examination of the recordings. Although some algorithms for vibration characteristics extraction exist (11) (12), the doctors must always be able to confirm the algorithm findings. Because of these difficulties and because of the cost of the system, the high-speed videoendoscopy is not as widely spread. (1)

1.2.3. Videokymography

Videokymography (VKG) is the original Czech-Dutch method, developed in 1994 in Groningen (NL) as an inexpensive alternative to high-speed video recording.

VKG is designed to record mechanical vibrations in visible spectrum. The method has been developed mainly in order to improve diagnosis of vocal cord vibration in Laryngology and Phoniatry (13). The system consists of specially adapted CCD camera, which is able to operate in two different modes - standard (50 fps - interlaced) or high frequency (7 812.5 lines per second). In high frequency mode, the system records images of a single horizontal line of the selected camera row and stacks them one above the other (14) (Figure 4). The method allows a quick and easy recording of both regular and irregular oscillations of the vocal cords. The reason behind this is that the frequency 7 812.5 frames / sec is sufficiently greater than the base frequency of vocal cords (about 70 to 1000 Hz). A variety of important characteristics of oscillations can be extracted from the recording, potentially revealing clues of physiological or pathological condition of the vocal cords. Many of these characteristics cannot be assessed using current widely available methods such as laryngoscopy or laryngostroboscopy. The camera system can also be used for recording and evaluation of vibration of musical instrument string or any random industrial equipment. Additionally, the system does not require nonstandard recording medium, and recording can be made for example on a commercially available standard VCR. This is reflected in low price of the system, which makes it most economical. An interesting fact is that this method was first deployed into clinical practice in 1996 in Prague (14).

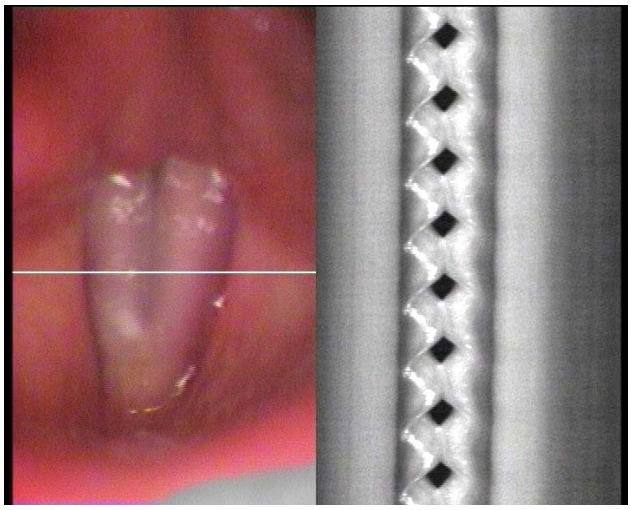

Figure 4. A VKG image example, displaing both the videokymographic camera modes of operation. The left part of the image shows the whole endoscope image with highlighted possition of the VKG scan line. The image on the right shows the actual VKG recordnig.

1.2.4. Digital image processing in larynges examination

There are several digital image processing methods addressing the problem of vocal cords vibration characteristics extraction. Good progress has been done in dealing with the high-speed video recordings. The tool for visualisation of high-speed video recordings was proposed by Deliyski (1) which helps to analyse the data. Another interesting method was proposed by Lohscheller (11) (12) for automatic high-speed video recordings evaluation by transforming the images to artificial coordinate system allowing better characteristics extraction.

Some work has also been done in the field of automatic processing of VKG images. Method of mucosal wave quantification was proposed by Jiang (15), automatic vibration properties extraction method based on the image segmentation done by Qui (16) and similar approach was shown by Manfredi (17). But until now, no complex solution addressing the problem of extraction of all the most common vibratory features exists. This is mainly due to the fact, that automatic processing of videokymographic images is harder, mainly because of high level of noise, low contrast and poor overall quality of VKG images.

2. Objective

One of the goals of laryngeal diagnostics is to find the relation between the healthiness of larynx, including the vocal folds, and the secondary vibrations of the ventricular bands, also known as the false vocal folds. For this purpose the need to automatically find and identify the ventricular band on a videokymographic recording have arisen. After the successful identification, the exact shape of the band needs to be determined for possible further processing and data extraction.

In most of the cases, the VKG images are recordings of unhealthy larynx. As a result, the images can contain a great variety of shapes and artefacts not seen on any other picture. Each patient is different, so each image is an original. It is therefore highly desirable, that the result of any automatic extraction method is further assessed by professional laryngologist to confirm the finding.

The main aim of this thesis is to propose a method of localising the ventricular band on the VKG recording image. The boundary of the ventricular band, which we are interested in, is usually recognised as a slightly darker contour adjacent to the middle region, where the vocal fold vibrates (Figure 5). The shape of this boundary can be used for further analysis of the vibratory features, such as base frequency, amplitude, moments, etc...

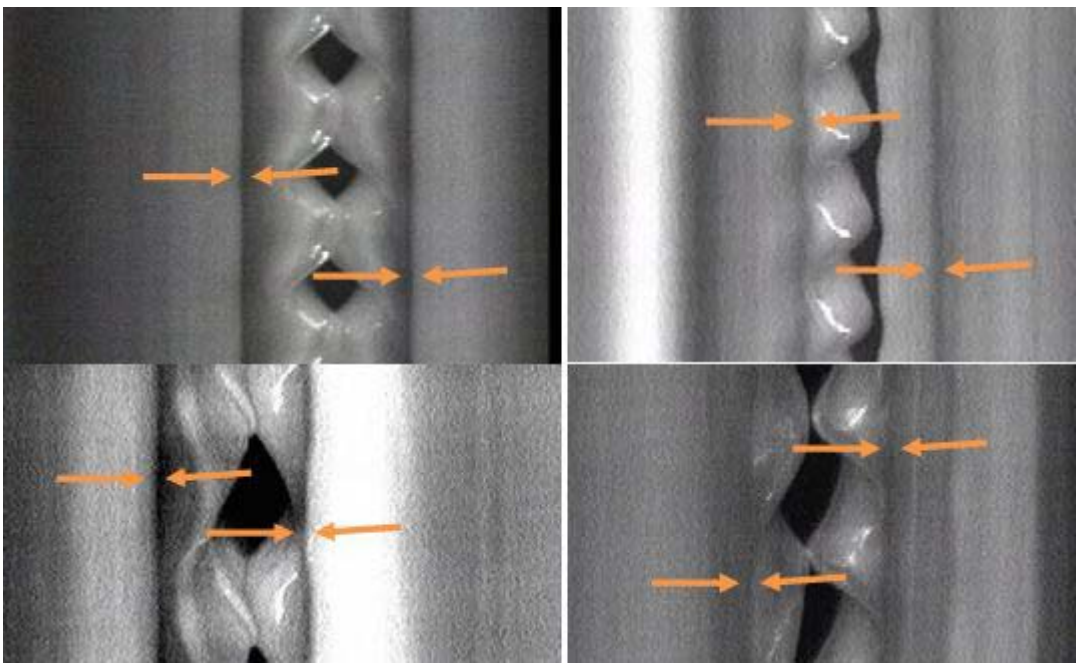

Figure 5. Left and right ventricular band edges as shown on the VKG images.

3. Problem analysis

3.1. Videokymographic recording

3.1.1. VKG image description

Each of the videokymographic recordings (VKG images) shows several openings and closings of the glottis. Typical VKG image also shows 3 distinguish horizontal regions. On the sides of the image, there are areas with little apparent vibration (in the case of non-vibrating ventricular band) which show recordings of ventricular bands. The area in the middle part of the image shows the vibrating vocal folds with dark regions representing the glottis openings. The areas of interest are the ventricular band edges, seen as borders between these regions. These are usually represented by the more or less pronounced dark contours running from the top of the image to the bottom with no or little vibration (Figure 5).

We can assume that in the middle of the image, where the vocal folds vibrate, the most of the information is contained, i.e. the scan line intensities vary in time due to the vocal folds motion. On the other hand, the sides of the image, representing the ventricular band are changing far less.

Unfortunately there are several randomly occurring issues with the VKG images. One of them is low quality of the VKG images, i.e. low resolution and contrast, and high level of noise. The other is the fact, that the images often contain many imperfections, such as the light reflections, caused by the mucosal secretions, or the non-uniformity of the tissue colour. The tissue itself can contain pronounced blood vessel, which manifest itself by the darker contour on the VKG image sometimes mimicking the ventricular band itself.

3.1.2. VKG image quality

Typical VKG image is a gray-scale image of poor resolution with high level of noise. Unfortunately there is nothing that could be done to improve the resolution of the image, but the same does not hold for the noise. In all of the cases the image contains the Gaussian additive noise caused by both the lack of proper light condition of the scene, and the heat of the CCD sensor. The reason for this is that powerful

light source would cause extensive heating of the laryngeal cavity and could effectively burn the patient (10). There could be solution for the heating problem of the sensor, but it would affect the cost of the system greatly. Instead, the de-noising techniques must be performed prior to further image processing to remove the unwanted noise. From the available de-noising techniques I chose the well-known convolution with Gaussian filter for noise reduction task. The experiments have shown that amount of blur, caused by Gaussian filter, does not affect the efficiency of proposed method. Furthermore, the convolution with averaging filter has been selected for the task of VKG image imperfection suppression. Each of the de-noising methods contains adjustable parameters, which must be well tuned for the particular purpose or situation. Specific parameter values usually depend on the resolution of the image, the noise amount and other factors.

3.1.3. VKG image issues

The regions of ventricular band as well as the edge of the ventricular band can show specular highlights caused by the endoscope point light source reflections, in most of the cases produced by the mucosal secretions on the tissue. These are represented on VKG image by small and very bright spots, often repeating itself with the vibration frequency.

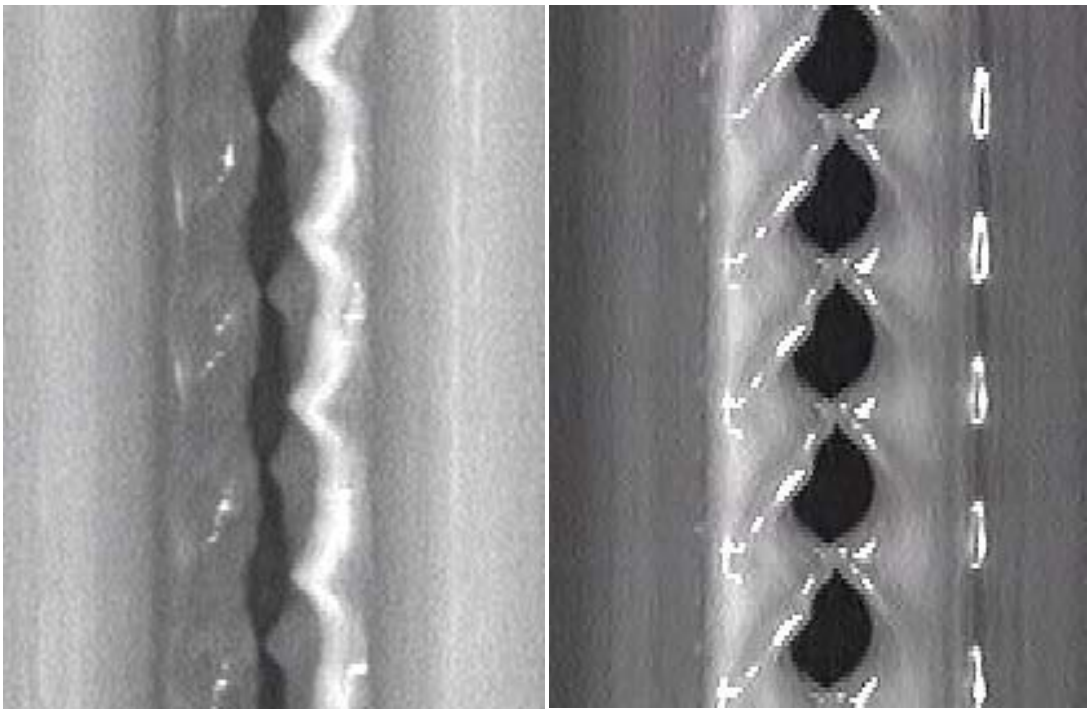

Figure 6. Examples of the unwanted mucus reflections. The reflections can be seen as very bright artefacts.

Another artefact making the finding of the ventricular band edge difficult is caused by the presence of distinct blood vessel on the tissue sample. Such an occurrence causes the dark line running along whole of the image from top to bottom vibrating, or not, with the tissue itself. This can mislead the algorithm or even non-trained human in the ventricular edge identification.

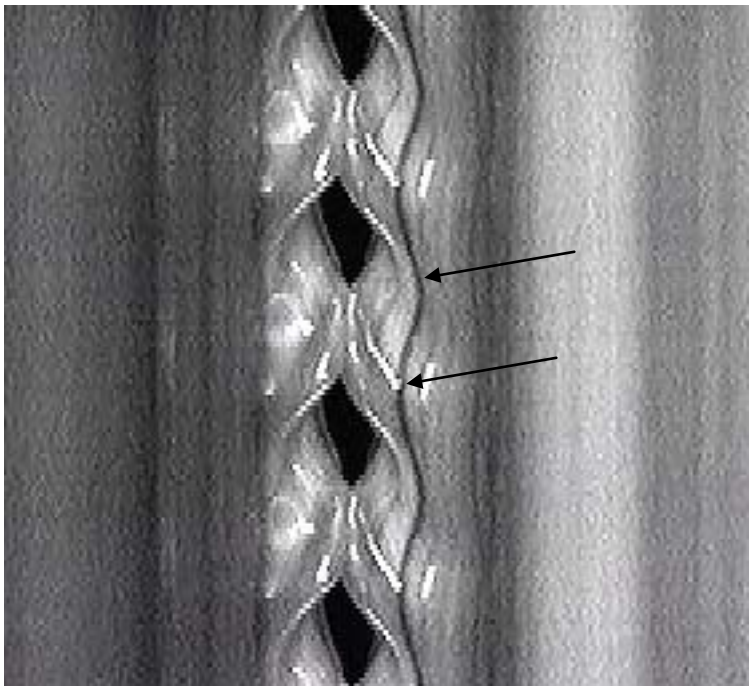

Figure 7. The blood vessel example. The dark line on the right from the glottis openings is created by vibrating vocal fold, which contains pronounced blood vessel.

3.2. Digital image processing techniques

In this chapter I will describe the methods I adapted for the use in the ventricular band detection algorithm.

3.2.1. Image pre-processing

Image pre-processing methods commonly serve as image enhancement techniques. The main role of such enhancements is to prepare the image for the main processing by operations designed to emphasize or suppress specific image feature, e.g. noise, contrast, edge information, colour, etc... For the purpose of this work the noise reduction and contrast enhancement methods need to be mentioned.

The noise

The noise that has to be dealt with in the VKG images pre-processing is the Gaussian additive noise. Mathematical model of additive noise is that the random value with the particular distribution is added to each and every pixel in the image.

For Gaussian noise the distribution is the Gauss normal distribution. This type of noise is typically caused by sensor heat. Because the noise is mostly pronounced in high frequencies, the noise reduction techniques are designed to diminish the high frequency information. Thus common attitude to this problem is to blur the image by the convolution with a specific kernel. There are several known and often used kernels such as the averaging kernel, Butterworth kernel or the Gaussian kernel (Equation 2). Convolution can also be looked upon as on the weighted sum with the weights determined by the kernel. The image is the 2D discrete function and the 2D discrete convolution takes for each pixel position its neighbourhood of size of the kernel and the pixel values are multiplied by the corresponding kernel value. Sum of these products are the resulting pixel values (Equation 1).

Equation 1. Discrete 2D convolution: g(u,v)-convolution kernel; f(x,y)-original image; t-size of the kernel; k(x,y)-resulting image.

$$k(x,y) = \sum_{u=-\frac{t}{2}}^{\frac{t}{2}} \sum_{v=-\frac{t}{2}}^{\frac{t}{2}} g(u,v) f(x-u, y-v)$$

Equation 2. Gaussian function: $g: \mathbb{R} \to \mathbb{R}$, where $\sigma > 0$ is the scale (also standard deviation in statistics) determining the width of the kernel; μ is the mean value.

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Image contrast enhancement

Because the VKG images are often of low contrast I opted to use the histogram equalisation technique to deal with the issue. The histogram equalisation is process of levelling all the intensity values in such manner, that all of the intensity values are evenly distributed in the image. It is a contrast adjustment technique done by first creating the histogram (which counts the pixels of each intensity level) and then transforming the image in such a way that the process makes the cumulative

histogram linear. I.e. darker regions of a dull image are darkened and lighter lightened.

The process is derived from the theory of probability. The histogram of any given image can be looked upon as a probability density function (PDF), estimating the probability of occurrence of a pixel of particular intensity in the image. As a result, the cumulative distribution function (CDF) can be calculated, matching the cumulative histogram. The linearization of the CDF, we want to achieve, can be done by applying the cumulative distribution function as a transform function for each pixel intensity value (Figure 8).

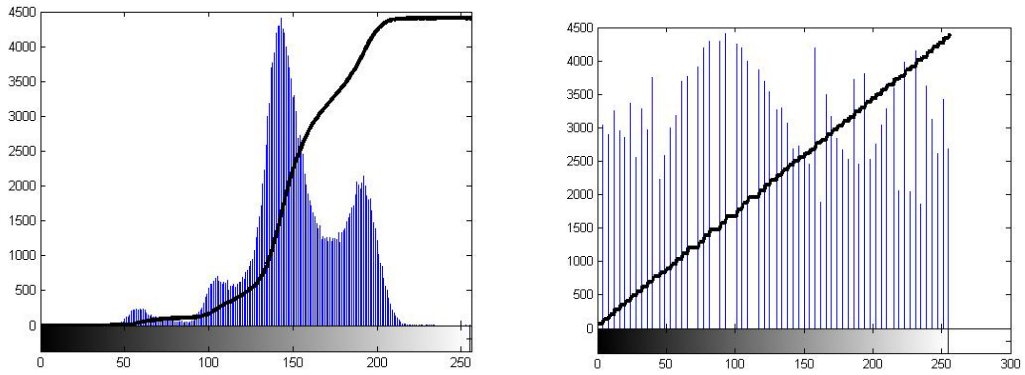

Figure 8. Image histogram equalisation: graphs show the histogram (blue) and the cumulative histogram (black) data of: left - an original VKG image; right - the same image after histogram equalisation.

In this thesis, the aim of image pre-processing is to enhance the contrast of the ventricular band edge locally, so the low contrast edge can be more pronounced. The global histogram equalisation does not perform well on VKG image, because the image often contains very dark regions of glottal openings and very bright regions of the light source reflections at the same time. To achieve local contrast enhancement, Contrast Limited Adaptive Histogram Equalisation (CLAHE) is utilised.

The adaptive histogram equalisation is similar to the regular histogram equalisation, but instead of histogram being computed globally for the whole image, it is calculated locally (Figure 9). For each pixel position histogram of pixel neighbourhood of given size is calculated and equalised. Every pixel is then transformed by the cumulative histogram computed from its neighbourhood. Adaptive histogram equalisation often leads to noise amplification. This can happen in the constant intensity regions with noise. In such a case, the noise contrast is

enhanced over the neighbourhood, resulting in the noise amplification. Contrast Limited Adaptive Histogram Equalisation (CLAHE) overcomes this by thresholding the histogram peaks, therefore limiting the local high contrast noise (18).

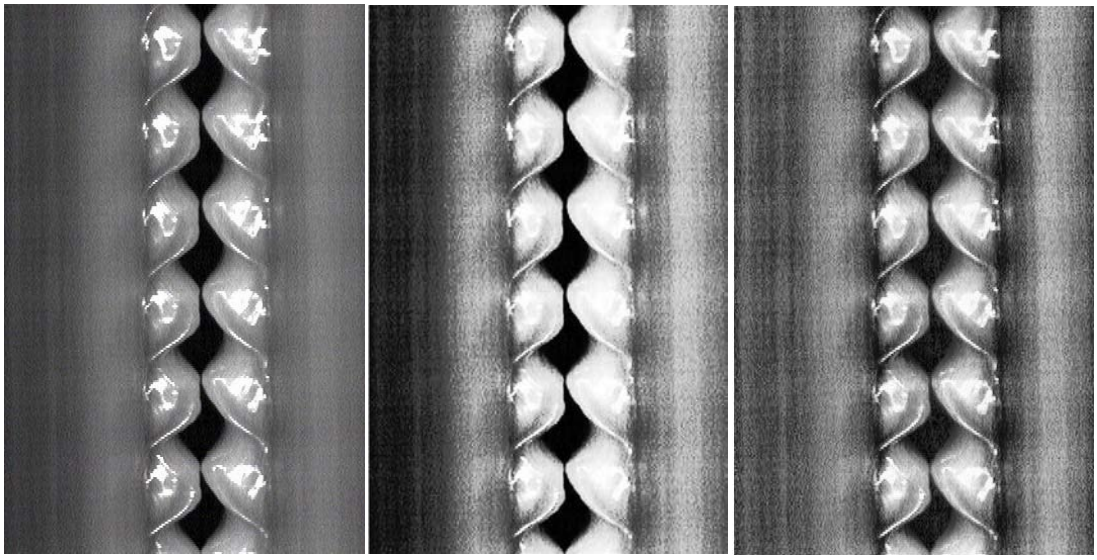

Figure 9. Histogram Equalisation vs. Adaptive Histogram Equalisation comparison: left-original image; middle-result of the histogram equalisation; right: result of the adaptive histogram equalisation. O the right image the local details are more pronounced.

3.2.2. Fourier transform

The Fourier transform is a process used to extract the frequency and phase spectra from any integrable function. Generally, the argument of the Fourier transform is a complex-valued function, but typically, in most of the cases, a real-valued function is used. The result of the transform is a complex-valued function, where a complex number describes the amplitude and phase of the corresponding frequency component (Equation 3).

Equation 3. Fourier transform in its 1D continuous form: For every real number k, and $f: \mathbb{R} \to \mathbb{C}$ integrable function, the Fourier transform $F: \mathbb{R} \to \mathbb{C}$ is defined as:

$$F(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x k} dx$$

Equation 4. Discrete 2D Fourier transform used in image processing: For resulting pixel coordinates k,l and f[m,n] being the original image of size M x N, the transform F is defined as:

$$F[k,l] = \frac{1}{\sqrt{MN}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f[m,n] e^{-2\pi i \left(\frac{mk}{M} + \frac{nl}{N}\right)}$$

In digital image processing, the original transform is adapted into a two-dimensional discrete version (Equation 4). Data contained in the Fourier frequency spectrum can be used to identify not only the frequency amplitudes of the two-dimensional signal, but also the direction of edges in the image. Because the image is in fact two-dimensional signal, the bases used in 2D Fourier transform correspond to 2D complex sinusoidal functions. The tilt angle of the particular basis corresponds to the tilt angle of the same basis response in the frequency domain. In that sense, the response in the Fourier frequency domain is orthogonal to the corresponding edge in the spatial domain (Figure 10).

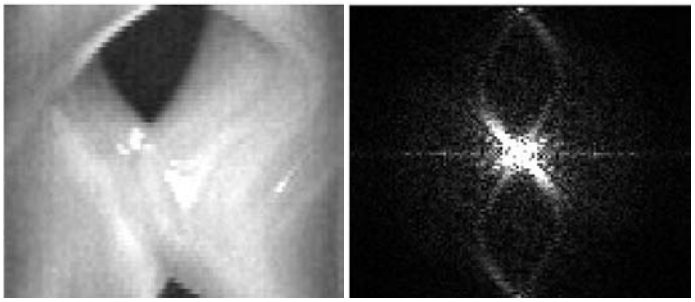

Figure 10. Fourier frequency spectrum: left image shows the original image; right image shows the Fourier frequency domain with centre of coordinate system in the middle. Note that the frequency response is orthogonal to the edges in the picture.

3.2.3. Graph shortest path algorithm

For the ventricular band shape determination I needed to find an algorithm, which would best find the contour with certain properties on given image. The properties of ventricular band edge are: the contour is slightly darker than its local surrounding; its intensity does not vary greatly along its path (in most of the cases); the knowledge of the approximate position can be extracted; the prevailing direction of the edge is horizontal; and the contour is an image long edge running from the top to the bottom of the image. After the extensive search, I decided to adapt the shortest path algorithm for this task.

Finding the shortest path in a graph is a problem belonging to the energy minimisation problems family (20). There are many algorithms used to find the optimal solution. The most popular algorithm is probably the Dijkstra's algorithm. Dijkstra's algorithm is based on the breadth-first search and is able to find the shortest path between the source and target for the given weighted graph in polynomial time.

In the image processing, Christina Gunkel et al. had shown the usefulness of this approach (21). In the publication they proposed a method of detection of micro cracks on a microscope image. First, they identify the crack clusters, and next, they extract the shapes of the cracks by computing the shortest path in the clusters by Dijkstra's algorithm.

4. Solution

4.1. Methods development

In order to find the solution to the problem, a set of methods was selected on the basis of extensive literature search and many experiments. I would like to mention few otherwise useful methods, which have proven to be non-effective for this kind of application.

4.1.1. Image pre-processing

Commonly used de-noising filter is the Gaussian filter. After some experiments I decided not to use it for the ventricular band position extraction. Instead, I recognised the averaging filter to be more suitable for this task. The reason was that the averaging filter has better results in reduction of small unwanted image features. I still use the Gaussian filter for noise reduction in the image designated for the computation of graph weights for shortest path search.

I have also considered altering the pixel intensity distribution prior to the processing by applying the logarithm or exponential operator. Those transformations have not shown any improvement in the later ventricular band detection.

4.1.2. The edge detectors

In my thesis I have tried several edge detecting techniques for the purpose of the ventricular band position extraction. I will mention the Canny edge detector, horizontal Sobel edge detector, convolution with Laplacian of Gaussian kernel (22) or the Active model contours algorithm (23). The main reason the edge detecting methods have failed lies in the diversity of the VKG recordings. Many of the images contain great number of false edges, caused by the non-uniformity of tissue colour, or reflections of endoscope light. Other reason is that many of the actual ventricular bands do not create good edges, or the edges they create are much less pronounced then other edges in the image (Figure 11).

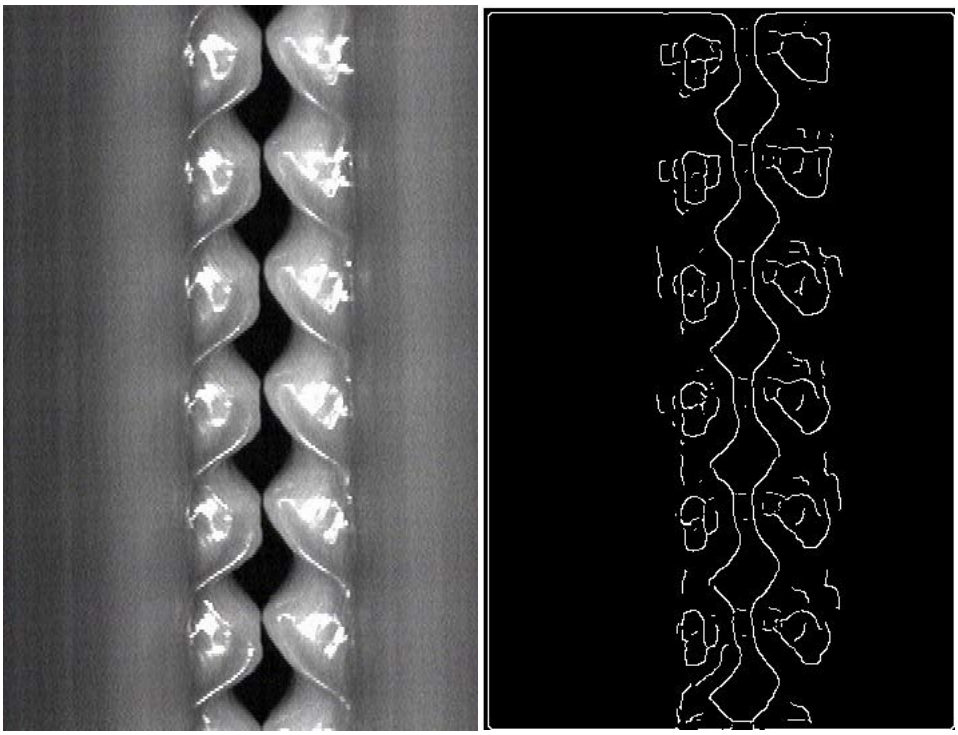

Figure 11. Result of Canny edge detector. The ventricular band is much less pronounced edge than the edges created by glottal waves and glottal openings.

4.1.3. Central region information assessment

In order to determine the position and size of the middle region of the VKG image with the vocal folds vibrations, I have tested several methods designed to estimate the amount of information in picture column.

First, I have tried the statistical approach in form of column variance. The values of resulting variances create the 1D function which domain is the spatial dimension (x-axis) of the VKG image. Such a function has the desired properties, i.e. in the region of vocal fold vibrations has usually its maximum, and in most of the cases, the variance values outside the vocal fold central region are below the function mean value. Unfortunately this approach has proven to be sensitive to noise and VKG image artefacts, such as reflections or tissue impurities. It is also sensitive to occurrence of columns with very low variance in region of the vibrating vocal fold (Figure 12). This can be partially solved by calculating variance of several columns together. But this approach fails to accurately differentiate the middle region from the rest of the image. The ventricular band is often represented by an edge, which cannot be distinguished from other edges in the image by using variance information

only. The reason is because the variance approach does not contain the information about the edges' directions.

I have also attempted to use the described variance as a weighted addition to the column sums or to the Fourier energy estimation method. Those procedures have shown partial improvements in most of the cases, but consequently caused unpredictable results in other.

Additionally I have considered using the column entropy calculation in similar manner as in case of the variance. This method showed the same issues as the variance extraction approach.

Finally the Fourier amplitude energy summation has been established as the most robust method. This technique does not operate with columns separately, but rather takes small region into consideration. Therefore it is not sensitive to local features mentioned above. As a result, the Fourier energy estimation has been used in the proposed method.

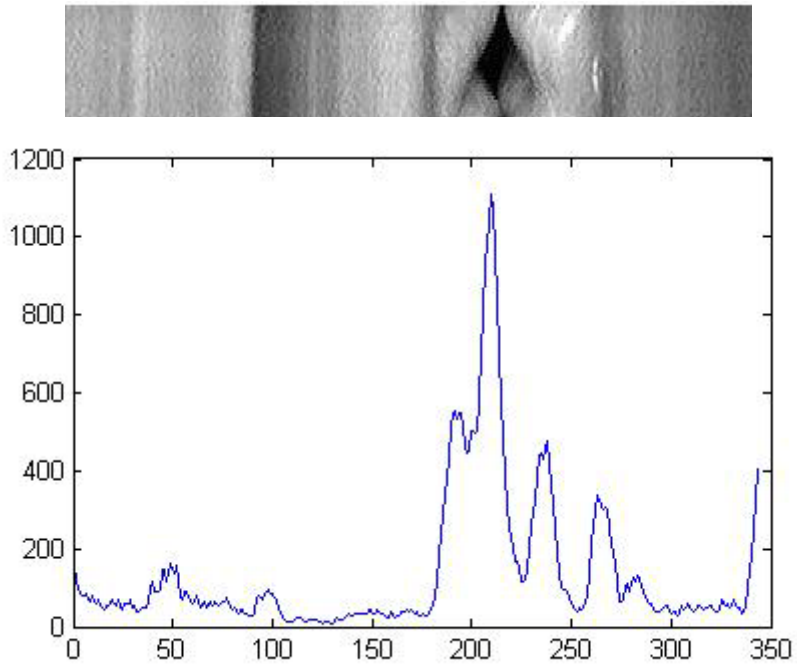

Figure 12. Image segment and corresponding graph of column variations. Note, the local minima in the region with vocal fold vibrations.

4.1.4. Global minimum

Another two approaches, which has been tested and failed, work with column sums of the image. Column sums form a function of overall column intensity (Figure 13). The tested methods for the column sums function were finding the global

minimum or the first minimum for the relevant part of the image, starting from the position of maximum glottal opening outwards. This approach has been surprisingly successful in most of the cases, but was not robust enough for the obvious reasons, i.e. mainly because the assumption of the searched edge being the darkest or first respectively, did not have to be true for all the images.

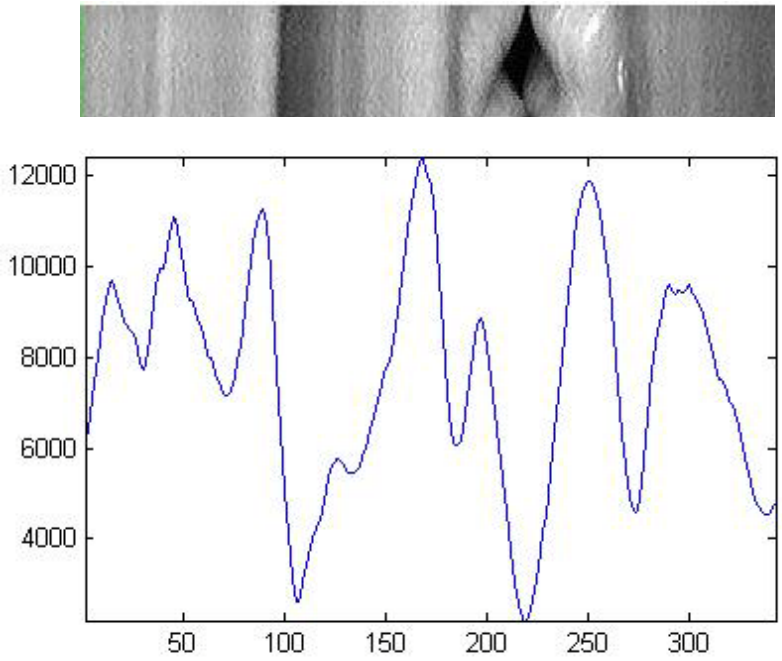

Figure 13. Column summation forms a graph, which is used to assess the overall column intensity.

4.2. Actual method description

4.2.1. Assumptions

For successful ventricular band position detection several assumptions must be made in order to create the unified problem model. Without the assumptions, the task itself would be a task of image understanding and interpretation belonging to the field of image semantics, rather than a field of digital image processing. The assumptions were stated on the basis of VKG image examinations as well as consultations with the experts. See Figure 5 for the ventricular band edge position examples.

First of the assumptions made is that the vocal folds vibrate, and are situated in the middle part of the VKG image. This part of the image lies in between the two ventricular bands, which are to be identified.

Second assumption is that the ventricular band is a tissue with much less apparent vibrations on the outer parts of the image.

Third, the ventricular band border, which is to be detected, is at least slightly darker edge, situated in between the vocal fold and the ventricular band.

For illustration please see Figure 14.

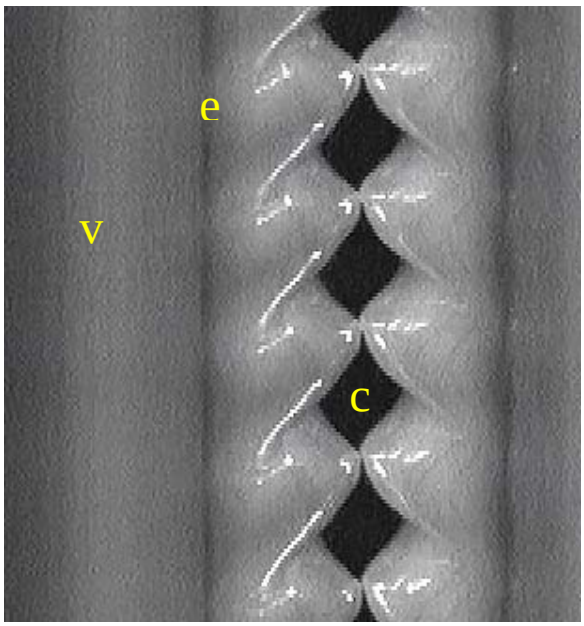

Figure 14. Problem assumptions: e-darker ventricular band edge; c-vibrating vocal folds in the middle; v-ventricular band with very little vibrations.

4.2.2. Phases

Based upon the assumptions stated above the proposed method can find the position of the ventricular band by identifying the borders of middle region with the vibration vocal folds. This is achieved by extracting the amount of edge information from each part of the image and estimating where it changes to the region with considerably less edge information. Then, the darker edge in vicinity is found, and finally the exact shape of the edge is extracted.

The method can be divided into following phases:

1. Divide the VKG image into parts, each approximately one glottal cycle long. Every part is processed separately, and the resulting position of the ventricular band is determined as a median value of the results.
2. Image part enhancement – de-noising and adaptive histogram equalisation.
3. For each of the VKG image part compute the most probable position of the ventricular band.
4. Use the modified shortest path algorithm to assess the shape of the whole ventricular band edge.

4.3. Image divisions

First, the image is automatically divided into several parts, each roughly one glottal period long. The reason for ventricular band per image part computation is that the image can be slightly tilted due to the endoscope movement. In smaller parts is the endoscope shift negligible. The parts should not be too small, so the edge information is preserved. For these reasons, the ideal size of one part has been determined as one glottal cycle long.

The size is calculated from the segmented image of glottal openings. First, the segmentation is inverted, so the zero values signify the glottal opening, and afterwards the row sums are calculated. In this way, the 1D graph of the opening row width is calculated. Next, the smoothing of the resulting curve is performed and the

minima of the graph are found by the use of the first and the second derivative of the glottal openings width graph. Some of these minima are then used as the starting points of each image part. In cases of some vocal disorders, more than one local opening minimum per cycle can be found. Therefore the suggested divisions are selected in such a way, that resulting image part height lies within pre-determined minimum and maximum values. But in general case the divisions correspond to the glottal cycle length (Figure 15).

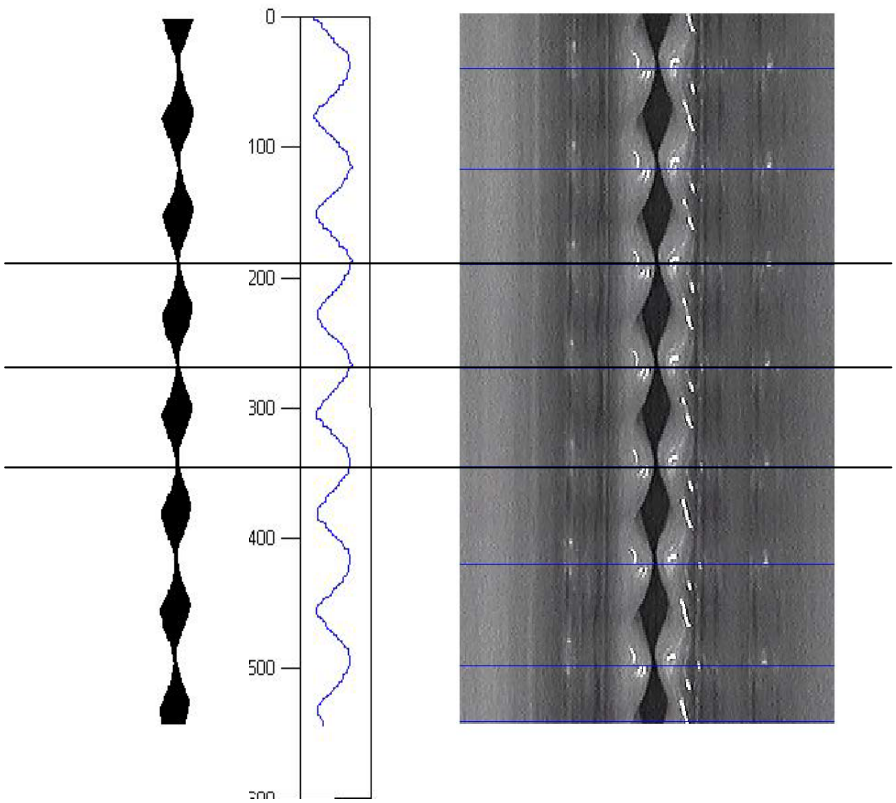

Figure 15. Example of image divisions for the processing (right) extracted from the row sums (middle) of the segmentation image (left).

4.4. Image pre-processing

To each part of the VKG image the adaptive histogram equalisation (CLAHE) is applied to enhance the local contrast of the ventricular band boundary (Figure 16). Next, the two-dimensional convolution with averaging filter kernel is performed. This makes the following computation techniques less prone to local imperfections and the image noise. After these transformations, the image is prepared for the ventricular band position extraction.

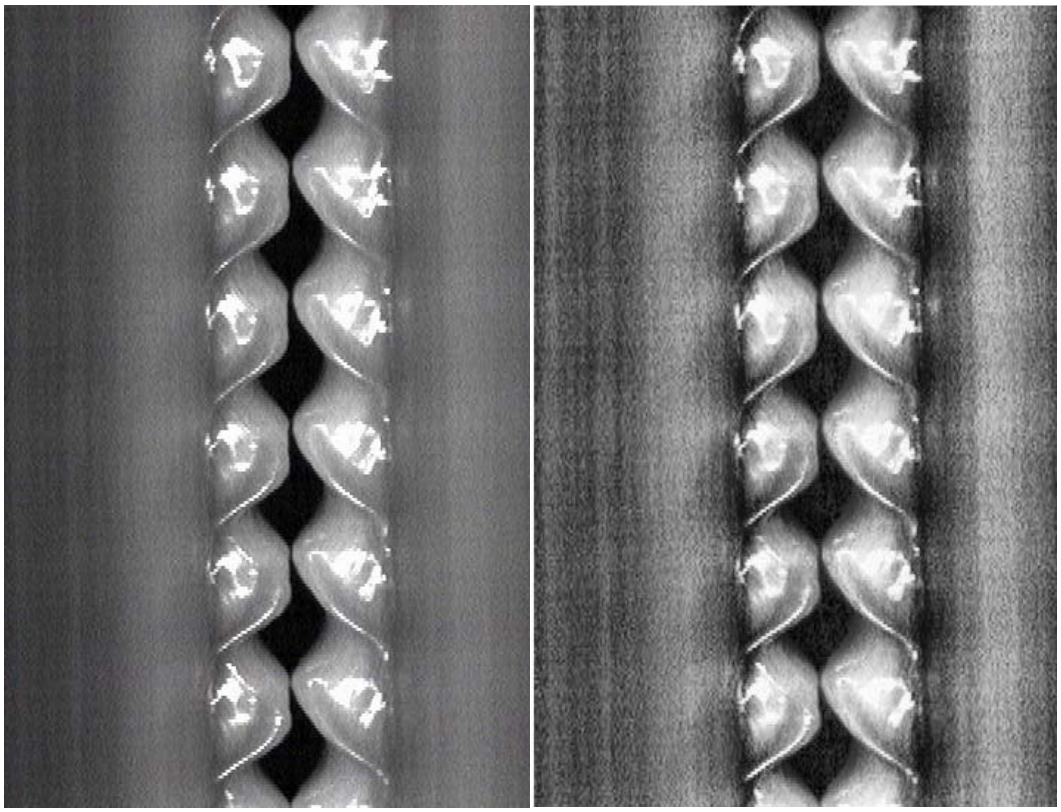

Figure 16.Adaptive histogram equalisation. On the left: original image; on the right: the enhanced image.

4.5. Ventricular band position determination

Based upon the assumptions, the pre-processed image is enhanced in such a way, that the distinctive middle region with the vibrating vocal folds should be enclosed between two moderately pronounced darker borders. These borders indicate the position of the ventricular band. It is not always so, but in most of the cases the assumption described above stands true. The middle region contains the information about the vocal fold vibration, within which majority of non-vertical edges can be found. The edges in this image region are in most of the cases related to the opening or closing of the vocal folds, e.g. the glottal wave, vocal folds, etc... To extract the non-vertical edge information from the image, the two-dimensional discrete Fourier transform is used. Furthermore, the areas of each the left and the right part of the image designated for the use by shortest path algorithm are established. Those are specified by inner and outer borders and serve as bases for the graphs creation.

Both the ventricular band position and the shortest path algorithm boundaries are found in following steps:

1) Fourier energy distribution is calculated for the given part of the image and the middle region containing vocal folds vibration is estimated.
2) Next, the position of the ventricular bands is determined.
3) The boundaries of the left and right image regions are established, to be used later in the shortest path algorithm.

4.5.1. Fourier transform

For each starting position of the whole width of the image part, 2D Fourier transform of the window of pre-determined size was calculated. So the Fourier frequency domain contains the information about the edges in given window, together with the information about the directions of the edges. The assumption is, that in the image region corresponding to the vocal fold vibrations, many non-vertical and non-horizontal edges can be found. This very information is contained in the Fourier frequency spectrum and can be extracted. From the Fourier frequency domain the particular regions indicating the non-vertical and non-horizontal edges are taken into consideration, and the energies in these parts are summed. The vertical and horizontal responses are caused by discontinuities in the image borders, and therefore are omitted. Please see the illustration in Figure 18.

Next, the array consisting of the resulting sums of energies for each VKG spatial position (x-axis) is created. This array can be considered a function of lateral position, returning the value proportional to the amount of oblique edges contained in the image region corresponding to the neighbourhood of given position (Figure 17).

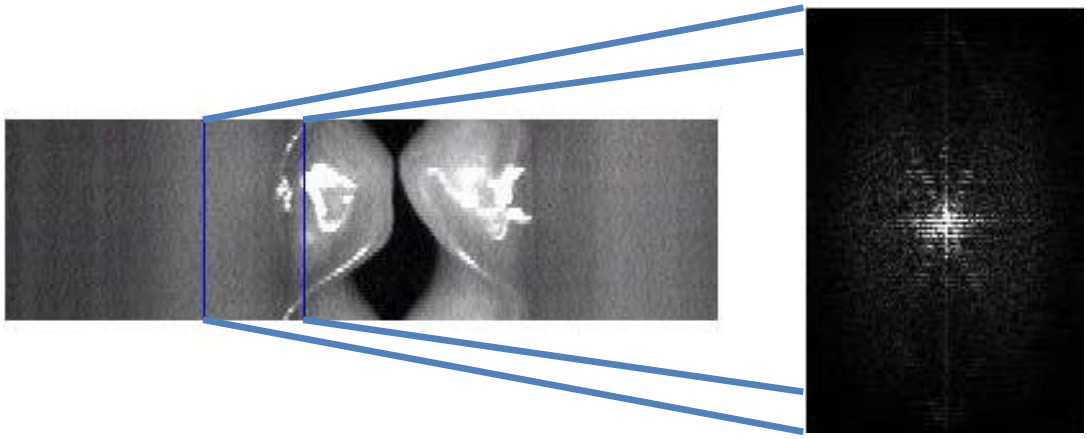

Figure 17. Fourier transform of the given image portion.

Figure 18. Illustration of the Fourier spectrum energy considered in energy summation.

4.5.2. Inner boundaries

After the successful Fourier energy sums computation, the rest of the extraction process occurs only for the relevant left or right side of the image segment. The left and right part's inner boundaries are defined by the corresponding glottal opening maxima.

To accomplish that, the space coordinate (x-coordinate) of the opening maximum, relevant for the particular part of the image, is extracted from the images with segmented glottal openings mentioned earlier. This is achieved by the column sum of the segmentation image part and searching for the first and last non-zero value. The extracted opening extremes are then used as the starting points for the ventricular band search, forming the inner boundaries of the left and the right part of the image at the same time.

4.5.3. Ventricular edge

Next, the mean values of the Fourier diagonal energies sums are calculated for each left and right parts of the image segment. For the given side, the first energy value, lower than the mean, is found. This value represents the loss of the Fourier diagonal energy in the particular part of the image, therefore the absence of oblique edges. It basically stands for the border of the region with pronounced vibration, in

this case the vocal fold vibration. Such a boundary is a potential candidate for the ventricular band edge.

Because in most of the cases (as assumed) the ventricular band edge is a bit darker than its surroundings, the search of the nearest minimum follows. First, the pixel intensities are summed along the spatial axis, forming one dimensional graph (Figure 13). The search for the local minimum is performed by the use of the first derivative of the graph. Next, the zero value of the first derivative is found and the second derivative of the same graph is used for the minimum determination. Such an established minimum of the column sums is used as the ventricular band approximate position. The position must be further refined by the use of the shortest path algorithm described below.

4.5.4. Outer boundaries

Further, the next local maximum of the column sum is calculated, starting from the ventricular band position outwards, to determine the outer border of the image crop, which will be used in the shortest path algorithm later. Finding this border in such a way is a result of the observation that the ventricular band tissue is generally of a round shape, forming a ‘hill’ brighter atop, due to the nature of the ventricular band illumination properties. The necessity of setting the image crop region for the further processing derives from the fact, that some of the images have dark edges, which could trick the shortest path algorithm.

4.6. Graph shortest path algorithm modification

In the previous section I described the algorithm for the ventricular edge position estimation for every segment of the image. Now, the exact shape of the ventricular band edge needs to be established. According to the assumption, the ventricular band edge is present in a form of darker edge along the whole picture. The problem of following the darker edge along the image running from the top row to the bottom is converted to the optimisation problem of finding the shortest path in the graph. The edge lengths or weights correspond to the pixel intensities and represent the energy to be minimised.

The graph is constructed as an 8-connected neighbourhood for every pixel position. The weights of the graph edges are derived from the corresponding pixel

intensity value. These weights are adjusted in such a way, that the prevailing direction of the path should be vertical, then diagonal and the least probable direction is horizontal.

The algorithm used for the path computation is the Dijkstra's shortest path algorithm. Prior to the processing, the pixel intensities are normalised. Next, the sparse 8 neighbourhood adjacency matrix is constructed using the established edge values. The size of the image used as a graph base is determined by the borders calculated in previous phase. The source and target for the shortest path is the top and bottom point of the image part respectively, where the space coordinate is calculated from all of the ventricular band positions extracted in the previous step of the process. To dismiss the outliers, the median of these positions is used (Figure 19).

The resulting vertices form a line, which corresponds to the connection of the source and target points having the minimum energy. For the path to best fit the shape of the ventricular band, the algorithm is executed two times. For the second time, the starting and ending point spatial coordinate is set as the mean value of coordinates of all the first computation resulting line pixels. This forces the path to settle best to the intensity 'valley' (Figure 19).

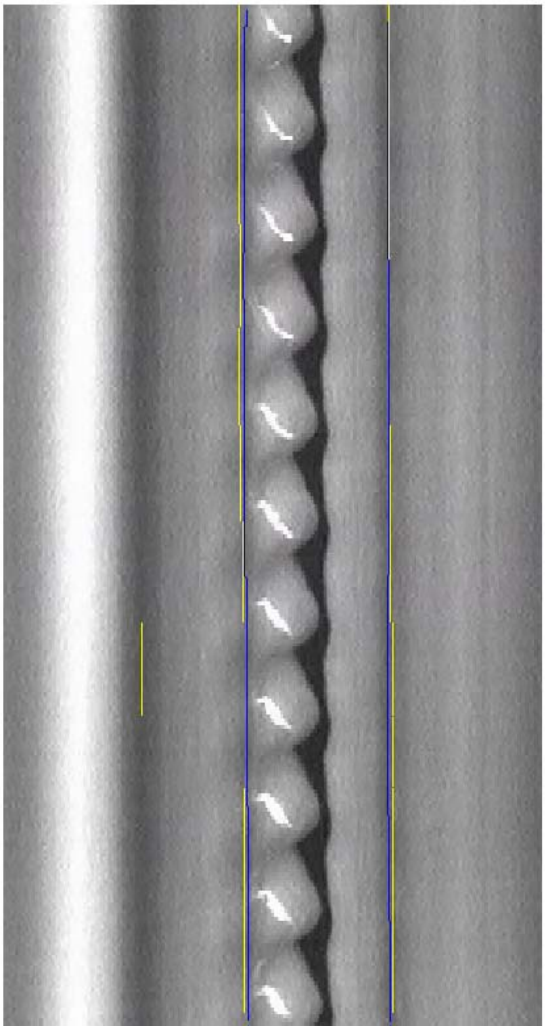

Figure 19. Results of ventricular band position estimation per image segment (yellow lines) together with the resulting ventricular band shape as calculated by the shortest path algorithm (blue lines).

4.7. Output

The data returned from the algorithm described above are the resulting ventricular band lines, both left and right, in the form of the line pixel coordinates. Such a representation is suitable for easy line visualisation by any common programming interface or language to create a visual result to be evaluated by an expert (Figure 20). The resulting data can also be easily used for further calculations and characteristics extraction.

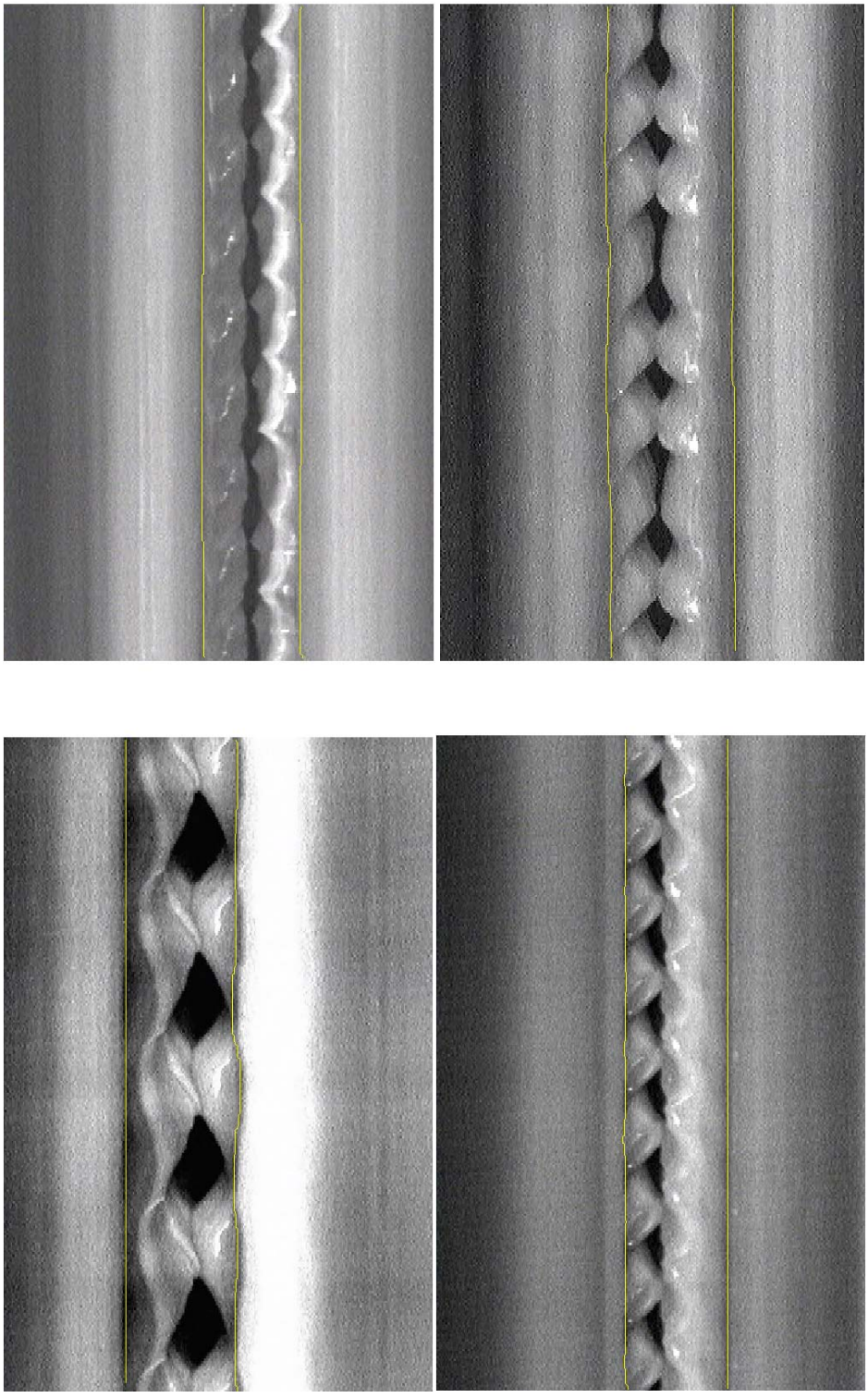

Figure 20. Example of the resulting ventricular band shapes (yellow lines).

5. Results

In this section the resulting performance of the algorithm can be found. The method was tested on a dataset, which has been also evaluated by an expert for comparison.

5.1. Data

The method performance was evaluated on set of 100 ventricular band instances contained in 50 original images of diseased vocal cords, together with the glottal openings segmentations of these images. The output of the method - the detected ventricular band – was outlined by yellow colour in the images for the visual evaluation of the algorithm performance. The localized ventricular band positions were compared to the manual detection done by an expert and the success rate of the method was established.

5.1.1. VKG images

The original VKG dataset has been provided by Jan G. Švec. It contains a set of 50 images of unhealthy vocal cord vibrations representing various vocal cord diseases.

5.1.2. Pre-segmented dataset

The dataset of VKG images has been segmented by means of the algorithm described in Sedlář's doctoral thesis (24), which is based on the Hauzar's algorithm for VKG segmentation (3). The program used for the images segmentation was implemented by Adam Novozamský (25). As a result of the segmentation algorithm black and white images were obtained, with intensity values of 0 for the glottis opening and 255 for the rest of the image (Figure 21). These images are used as a base for the glottal cycle lengths and amplitudes computation. The segmentation method is based on sophisticated thresholding by the use of modified graph-cut algorithm. The method itself is not a part of this thesis and therefore it will not be described in detail.

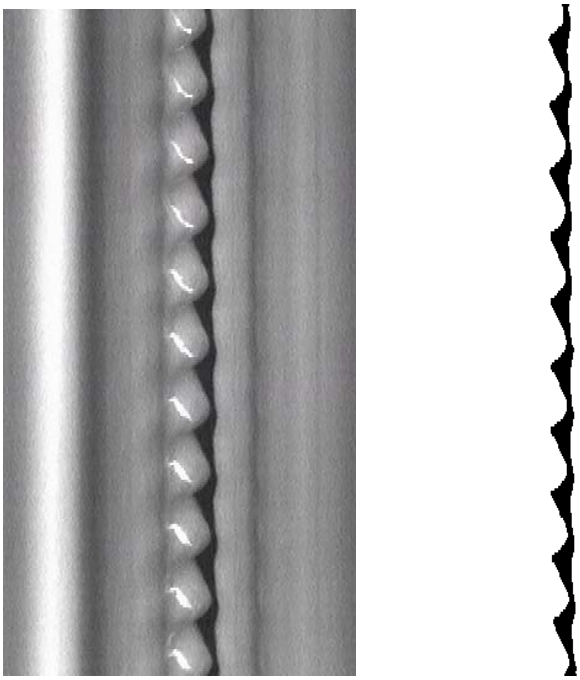

Figure 21. Segmented glottal openings example.

5.2. Outcome

In many of the cases the exact shape of the ventricular band cannot be determined with the pixel precision, even by experts. This prevents the accurate measurement of the detected shape variation. For this reason I decided to create six categories of the resulting algorithm performance and divide the results to those, with the help of an expert opinion. Christian T. Herbst, Austrian vocal scientist, Head of the Laboratory of Bioacoustics, has been very helpful in examining the same set of 50 images and determining the position of the ventricular band manually.

Table 1 shows the summary of the comparison of his results with the results of the proposed method.

- '*Precise results*' are the algorithm results, which cannot be distinguished from expert assessment.
- '*Slightly off*' are the results, where the ventricular band position was successfully determined in most of the cases, though the exact shape contains some errors. These are often caused by the fact, that the input image does not hold to one of the assumptions, e.g. the ventricular band is not darker than surroundings, or does not form an edge, etc...

- '*Algorithm failed*' category contains the cases, in which the algorithm was unable to determine the position of the ventricular band.
- '*Fundamentally hard*' category contains the cases, in which the ventricular band cannot be successfully recognised, even by an expert.
- '*Ventricular band missing*' counts the cases in which the ventricular band is not present on the image.
- '*Uncertain*' category contains the cases in which the expert is unable to pinpoint the position of the ventricular band.

Table 1. The results table contains number of images for each of the category of algorithm performance.

Category	Counts of cases
Precise results	81
Slightly off	12
Algorithm failed	1
Fundamentally hard	2
Ventricular band missing	3
Uncertain	1

This results show, that from all of the images, even the worst ones, the algorithm was not able to successfully determine the position of ventricular band in 3% of the cases and in 12% of the cases the exact shape of the ventricular band extraction failed. On the other hand the algorithm performed well on 81% of the cases and in 4% of the cases the positions of ventricular bands are undeterminable or not present in the image.

5.3. Result examples

In this section a few result examples are shown to illustrate the functionality and weak spots of the proposed method.

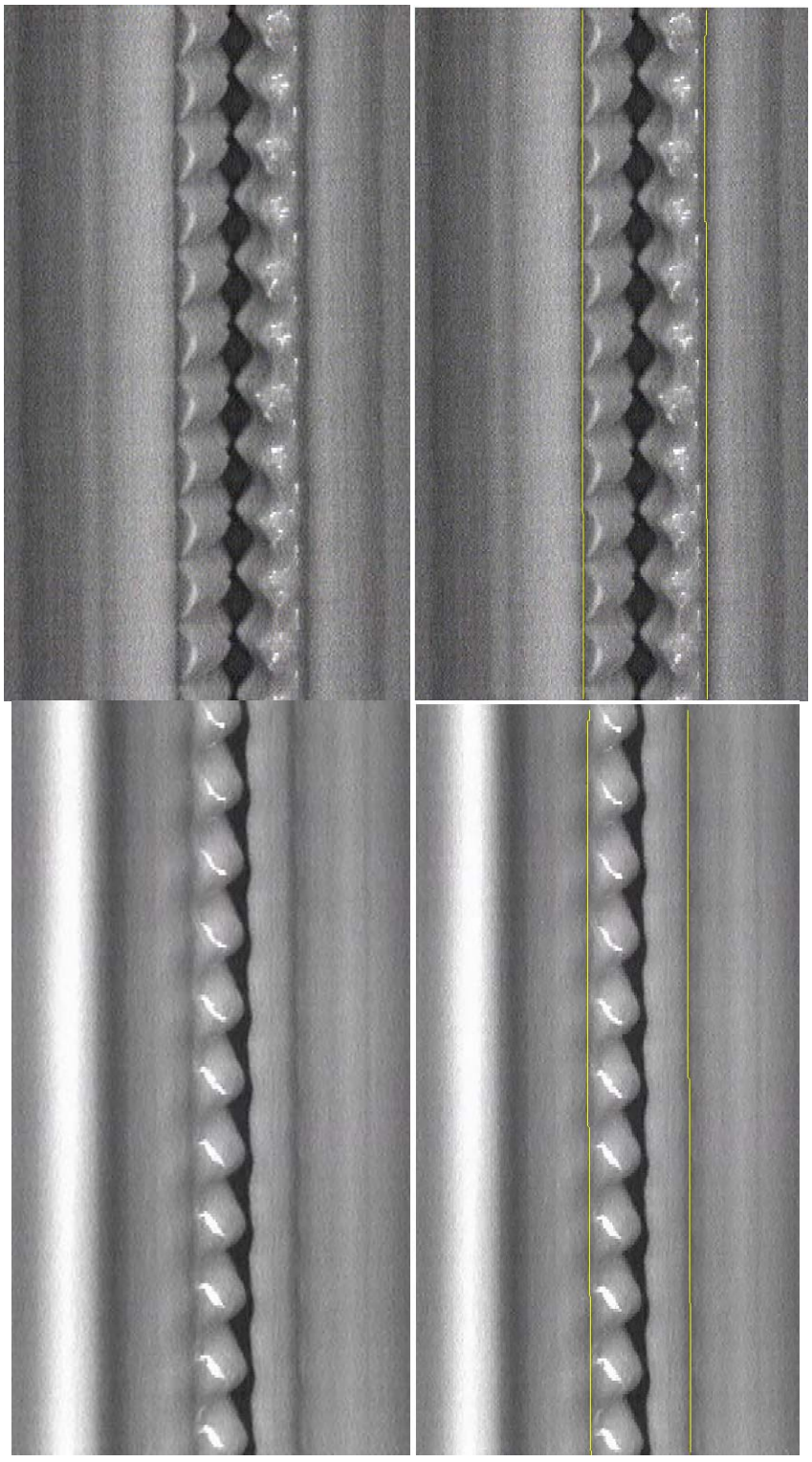

Figure 22. Examples of good results (yellow): Images on the left are the original VKG images. Right images show results of the proposed method.

Figure 22 shows examples of good results with both ventricular bands correctly detected, in conformity with the expert's findings. These are the images, which hold the assumptions, so the ventricular band position can be easily extracted and the shortest path algorithm has no problem following the ventricular band darker edge. The available dataset, used for testing, constituted of more than 80% of such ventricular band instances.

Another example (Figure 23) shows the case of one ventricular band completely missing from the image. The left side ventricular band was recognised successfully, but the algorithm was confused by the right side missing ventricular band.

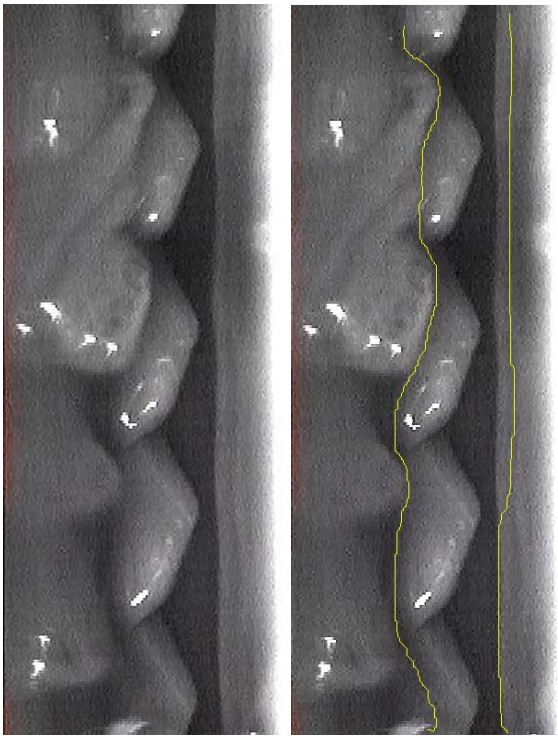

Figure 23. An example of image with missing ventricular band (on the right side of the image): The left image is original image; the right shows the results (yellow).

Next example (Figure 24) shows the case, in which the algorithm was mislead by complicated image. In this instance the left side ventricular band is vibrating in such a way, that it creates waves manifesting itself on a VKG image by pattern very similar to the vocal fold glottal wave. This led to false determination of the vocal fold region and caused the calculated position to be shifted outwards (right image). The correct location of the left side ventricular band is possible by good image understanding and interpretation. Consequently, it is a hard task for the digital image processing techniques.

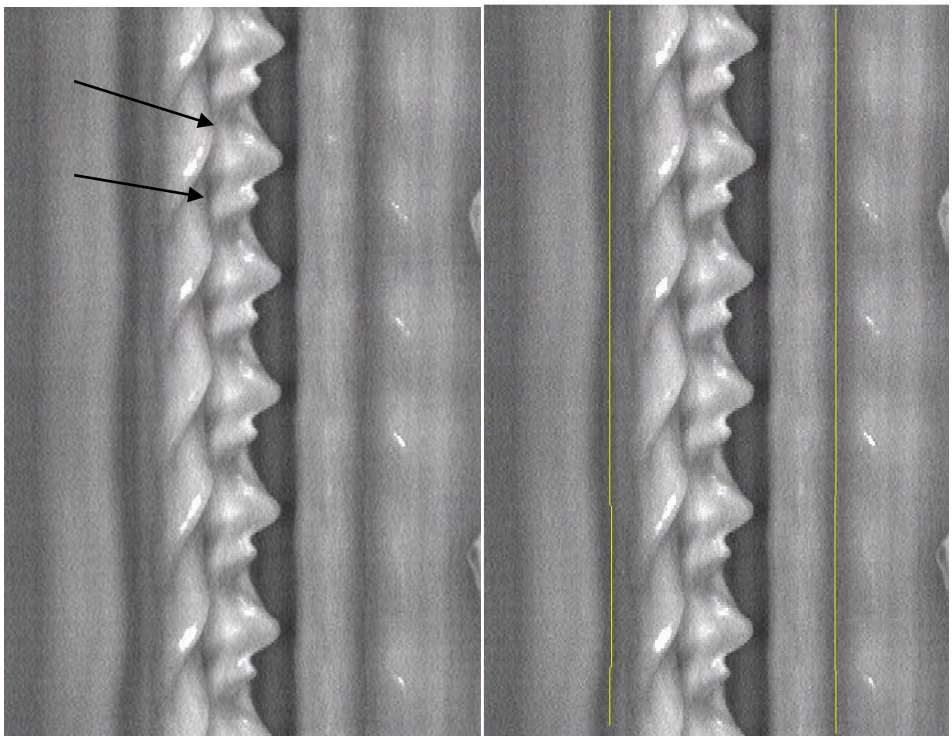

Figure 24. An example of image containing ventricular band, which position is difficult to determine (left). The method is tricked by edges on the ventricular band (left). The correct position is denoted by arrows. Image on the left is an original image, algorithm results (yellow) are illustrated on the right.

In the example, depicted by the Figure 25, the case of imprecise position of ventricular band is shown. In such a case the general position of the ventricular band is correctly calculated, but the shortest path algorithm failed to determine the shape of the band edge, because its exact location is unclear.

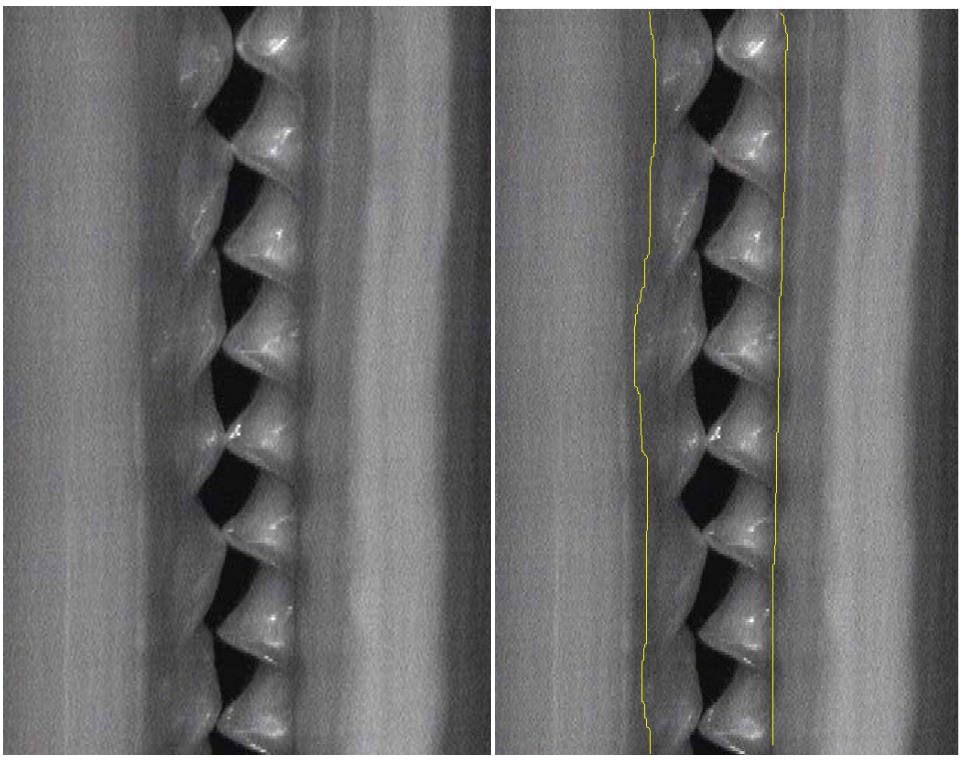

Figure 25. An example of ventricular band with imprecise position: left image shows original image; right show the result of the method (yellow).

The last example (Figure 26) illustrates the cases, in which the position of the ventricular band proximity is correctly determined, but the shortest path algorithm failed to ‘settle’ into the ventricular band edge. The problems here are caused by the endoscope light reflections, which makes the lowest energy solution, the ‘darkest path’, to be slightly off the correct position.

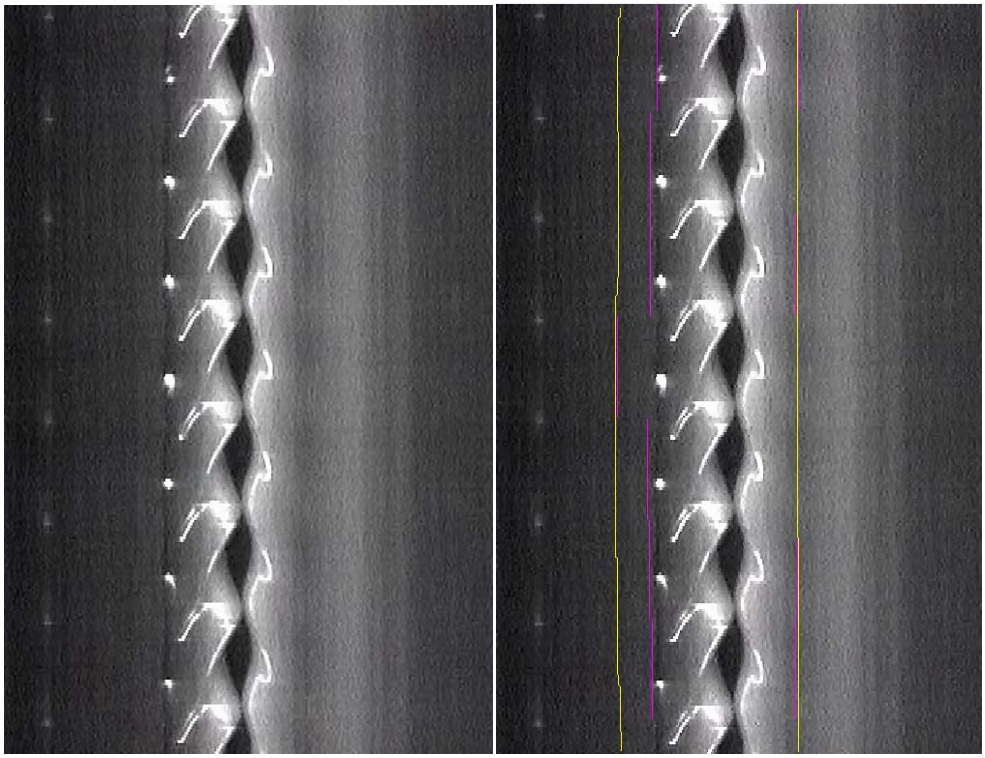

Figure 26. A typical case of ventricular band (left side) position being correctly estimated by Fourier energy computation (magenta), but the shortest path algorithm (yellow) being confused by the endoscope light reflections: left - original image; right - the result illustration

6. Conclusion and future development

The main goal of this thesis was to lessen the burden of the time-consuming image examination on the laryngologists by the task automation and the searched result suggestions. This goal was achieved by the presented method, which is able to determine the position and shape of the ventricular band on most of the videokymographic images. Although, by its nature, the algorithm cannot cope with the images fundamentally hard, because they do not meet the assumptions described above, such images are rare. They are often the recordings of very unhealthy vocal cords, where even manual recognition attempted by an expert is very difficult, confusing and sometimes nearly impossible. The task has proven to be on the edge of automatic image processing and it often crosses the border of image understanding and interpretation.

This method will become a part of an automatic VKG image processing algorithms family, which are now being developed in the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic. Together, they will form a standalone application, which is aimed to be deployed to the clinical practice in the near future.

As a next step in the development the extracted data can be further processed, e.g. for the ventricular band characteristics extraction. This could include the band oscillation amplitude, frequency, phase, if present. Furthermore, additional information can be added to the method by extracting information about the vocal folds structure from the whole endoscope image, as seen in Figure 4.

Summary:

- I have developed the first method addressing the problem of locating the ventricular band on the VKG image.
- This method has been chosen on the basis of extensive literature search and many experiments with other approaches, for example variance, global minima and several edge detecting techniques. As a result of this research I have decided to use the Fourier energy extraction for image middle region localisation, leading to the ventricular band position estimation, and the shortest path algorithm for its shape extraction.

- The applicability of the proposed method has been verified on 100 ventricular band images of unhealthy human vocal folds with variety of diseases. Given set of images was independently examined by an expert for comparison and algorithm performance assessment. As a result, more than 90% of the ventricular band instances were successfully localised and in more than 80% its shape was precisely extracted.

Epilogue

It is my utmost believe that my work will lead to better understanding of the human vocal cord problematic and diagnosis, mainly because this organ, along many others in the human body, is susceptible to malignant tumours, where the early diagnostic is crucial for successful treatment.

Bibliography

1. **Deliysky, DD, et al.** Clinical Implementation of Laryngeal High-Speed Videoendoskopy:Challenges and Evolution. *Folia Phoniatrica et Logopaedica.* 2007.

2. **Zita, Aleš.** Zpracování videokymografických záznamů [Analysis of videokymographic images]. *Bachelor Thesis.* Prague : Faculty of Mathematics and Physics, Charles Univesity, 2011.

3. **Hauzar, D.** Zpracování digitálních snímků videokymografických záznamů jako podpůrný nástroj pro diagnostiku hlasivek [Analysis of digital videokymographic images as a tool for vocal folds diagnostics]. *Diplomová práce.* Praha : MFF UK, 2010.

4. *Fyziologická akustika zpěvního hlasu: Nový pohled na starý problém.* **Švec, Jan G.** Kouty : s.n., 2000. 60. akustický seminář & 36. akustická konference.

5. Larynx part II. *Voice & speech source.* [Online] Eric Armstrong. http://www.yorku.ca/earmstro/journey/larynx2.html.

6. **Alberti, PW.** The history of laryngology: a centennial celebration. *Otolaryngol Head Neck Surg.* 1996, Vol. 114, pp. 345-354.

7. **Mafee, MF, Valvassori, GE and Becker, M.** *Imaging of the neck and head.* 2. Stuttgart : Thieme, 2005.

8. **Uloza, V, Saferis, V and Uloziene, I.** Perceptual and acoustic assessment of voice pathology and the efficacy of endolaryngeal phonomicrosurgery. *J Voice.* 2005, Vol. 19, pp. 138–145.

9. **Gallivan, KH and Gallivan, GJ.** Bilateral mixed laryngoceles: simultaneous strobovideolaryngoscopy and external video examination. *J Voice.* 2002, Vol. 16, pp. 258–266.

10. **Verikas, A, et al.** Advances in laryngeal imaging. *Eur Arch Otorhinolaryngol.* 2009, Vol. 266, pp. 1509-1520.

11. **Lohscheller, Jorg, et al.** Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing the underlying laryngeal dynamics. *IEEE Transactions on medical imaging,.* 3, 2008, Vol. 27.

12. **Lohscheller, Jorg, et al.** Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endosvopic digital high-speed videos. *Medical imge analysis.* 11, 2007, pp. 400-413.

13. **Švec, Jan G and Schutte, HK.** Videokymography : high-speed line scanning of vocal fold vibration. *J Voice.* Vol. 10, pp. 201-205.
14. *Videokymography in 2000: the Present State and Perspectives of the High-Speed Line-Imaging Technique.* **Švec, JG, Šram, F and Schutte, HK.** Jena : s.n., 2000. Advances in Quantitative Laryngoscopy, Voice and Speech Research. pp. 57-62. ISBN:3-00-005636-X.
15. **Jiang, Jack J., et al.** An automatic method to quantify mucosal waves via videokymography. *Laryngoscope.* 2008, 118, pp. 1504-1510.
16. **Qui, Quingjun, et al.** An automatic method to quantify the vibration properties of human vocal folds via videokymography. *Folia Phoniatr Logop.* 2003, 55, pp. 128-136.
17. **Manfredi, C., et al.** Objective vocal fold vibration assessment from videokymographic images. *Biomedical signal processing and control 1.* 2006, pp. 129-136.
18. **Pizer, Stephen M., et al.** Adaptive Histogram Equalization and Its Variations. *Computer Vision, Graphics, and Image Processing.* 1987, 39, pp. 355-368.
19. **Zuiderveld, Karel.** Contrast Limited Adaptive Histogram Equalzation. *Graphics Gems IV.* 1994, pp. 474-485.
20. **Cormen, Thomas H., et al.** *Introduction to algorithms (2nd Edition).* s.l. : Mc Graw Hill, 2001.
21. **Gunkel, Christina, et al.** Micro crack detection with Dijkstra'a shortest path algorithm. *Machine Vision and Applications.* 2012.
22. **Gonzalez, RC and Woods, RE.** *Digital image processing.* Upper Saddle River : Prentice-Hall, 2008. ISBN:978-0-13-168728-8.
23. **Kass, Michael, Witkin, Andrew and Terzopoulos, Demetri.** Snakes: Active Contour Models. *International Journal of Computer Vision.* 1988, pp. 321-331.
24. **Sedlář, Jiří.** Image Analysis in Microscopy and Videkymography. *Doctoral thesis.* Prague : MFF UK, 2013.
25. *Computer-assisted evaluation of videokymographic data.* **Novozámský, Adam, et al.** Prague : European Federation for Medical Informatics, 2013.
26. **The MathWorks, Inc.** Matlab: The Language of Technical Computing. [Online] 2009. http://www.matlab.com.

27. **Gleich, David.** gaimc:Graph Algorithms In Matlab Code. *Matlab File Exchange.* [Online] 2009. http://www.mathworks.com/matlabcentral/fileexchange/24134-gaimc-graph-algorithms-in-matlab-code.

List of Tables

List of Abbreviations

CCD	Charged Coupled Device is electronic component for capturing image.
CDF	Cumulative Distribution Function is a statistical function, which describes the probability, that the random variable will be found at a value less than input value.
CLAHE	The Contrast Limited Adaptive Histogram Equalisation is a modified adaptive histogram equalisation technique designed for local contras enhancement of images with the noise. (18)
CT	Computed Tomography is a radiological method which is using x-rays for visualisation of inner parts of human body.
MRI	Magnetic resonance imaging is a method that uses strong magnetic fields and electromagnetic waves with high frequency to examine the internal organs of human body.
VCR	Videocassette Recorder is a device for capturing the video and audio signals to an analogue tape.
VKG	Videokymogram/videokymographic – the method of capturing vocal fold vibrations by CCD camera converted to a line scanner.

Attachment 1

CD content

The CD attached to the thesis contains the following files and folders:

- Documentation/Master Thesis.pdf ... computer version of this document
- Program/ . ... the method implementation in Matlab
- Input/Images/ ... the source images dataset
- Input/Masks/ ... the pre-segmented set of glottal masks
- Results ... resulting set of images

Attachment 2

Program documentation

Program distribution

The program is distributed in form of source code files, a set of input images and a set of pre-segmented images of glottal openings of corresponding input image. These data are necessary for the program execution.

System requirements

The program is running under Matlab r2009b (26) environment, with the Image Processing toolbox.

Installation

The program can be installed by copying the source folders to any chosen folder on the target machine. The folders necessary for the program execution are: *Program* folder containing the source code; *Input* folder containing the input data sets.

Program execution

To execute the program successfully, Matlab environment must be running and the working directory must be set to the target folder containing the source files (*Program/*). The program can be executed from Matlab prompt by one of the following commands:

- `Extract()...` executes the processing of single VKG image
- `RunAll()...` executes the processing of all the VKG images in the Input folder

Syntax:

```
$result = Extract([image_name])
$result = Extract([image_name], dir)
$result = Extract([image_name], dir, debug)

$RunAll()
$RunAll(x)
$RunAll(x, dir)
$RunAll(x, dir, debug)
```

Where :

- `result` is a 4 vector consisting of 4 numbers : `[ yr, xr, yl, xl ]`. Those are coordinates of pixels representing the ventricular band.
    - `yr, xr`: y-coordinates of right ventricular band, x-coordinates respectively and
    - `yl, xl`: y-coordinates of left ventricular band, x-coordinates respectively
- `image_name` is a name of image to be processed, without extension (.jpg assumed); example: `VKG-005`
- `dir` is a optional subfolder, where the results will be stored
- `debug` is an optional logical variable, which, if set to true, forces the program to run in debug mode
- `x` is an optional sequence number of the 1st picture to be processed by `RunAll(),` example : `RunAll(4)` will process images in the *Input* folder, starting from the 4th file in alphabetical order onwards

Attachment 3

Program architecture

The program consists of several Matlab function files. All of the files are contained in Program folder. In *Program/shortestpath* folder the fast implementation of Dijkstra's shortest path by David Gleich (27) is included.

- `DisplayGraph.m` - Utility for displaying graph data – used in debug mode
- `Extract.m` - Divides the image to parts, on each part runs `ExtractPart`
- `ExtractPart.m` - Extracts the ventricular band position from image part
- `FindLocalMax.m` - Finds the local maximum of 1D function
- `FindLocalMin.m` - Finds the local minimum of 1D function
- `GetFEnergyDist.m` - Computes the distribution of Fourier energy of oblique edges
- `GetOpenings.m` - Computes the glottal opening left and right maxima
- `GetStarts.m` - Computes the image parts starting coordinates
- `GetVentricularBand.m` - Ventricular band shape approximation using Dijkstra's shortest path algorithm
- `Preprocess.m` - Image preprocessing: denoising and contrast enhancement
- `RunAll.m` - Runs all the images in *Input* folder
- `ShowPos.m` - Utility for debug displaying given values in graph

Program dataflow