

# Mobile application for electrophoretic gel image processing

Anna Mojžišová

Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication  
Brno University of Technology  
Brno, Czech Republic  
xmojzi09@vutbr.cz

**Abstract**— This paper describes a mobile application that assists in the analysis of 1D gel electrophoresis images. It provides functions for image processing, band detection, lane segmentation, and molecular weight approximation. The designed methods were extensively tested using a dataset of diverse gel images. The application provides a convenient and portable tool for the analysis of electrophoresis images on-the-go.

**Keywords**—1D gel electrophoresis, image analysis, quantitative analysis, mobile application

## I. INTRODUCTION

One-dimensional gel electrophoresis is a widely used technique for separating proteins or nucleic acids in a sample based on their size and charge [1]. The sample is loaded onto a polyacrylamide gel, which acts as a molecular sieve [2, 3]. The gel is then placed in a buffer solution and an electric field is applied to the gel [4]. The molecules in the sample migrate through the gel at different rates based on their charge and size, resulting in banding patterns that can be visualized using staining techniques [5]. By comparing the banding pattern of the sample to that of known molecular weight markers loaded onto the gel, it is possible to perform a quantitative analysis of the molecular weights of separated molecules. The distance traveled by each band is proportional to the logarithm of the molecule's molecular weight, allowing the estimation of the molecular weight of the separated molecules [6]. This technique can be used for the identification and quantification of specific proteins or nucleic acids in a sample and can be used to monitor changes in protein expression or DNA mutations over time.

## II. METHODS

### A. Grayscale conversion and image quality enhancement

First, the input image is converted to grayscale, and the mean pixel intensity is evaluated. This information is used to create a new grayscale image with dark bands and a lighter background. Numerous methods including histogram equalization [7, 8], gamma correction [7], piece-wise linear contrast adjustment [7, 9], and background thresholding [8][10] have been tested and evaluated. Out of these techniques, only gamma correction provided sufficient improvement in gel image quality.

Gamma correction is a technique used to adjust the brightness and contrast of an image by altering the gamma value of the image [11]. Gamma value was estimated by extracting pixel intensities from a representative region defined by a user

containing all the sample lanes and minimum of additional background, computing the mean intensity of the pixels in the representative region, computing the logarithm of the pixel intensities and the logarithm of the mean intensity, fitting a line to the logarithmic data using linear regression, and extracting the slope of the fitted line, which is the estimated value of gamma [12].

### B. Lane segmentation

The algorithm of lane detection was slightly modified from the method proposed by Škutková et al. and requires an input value of the number of lanes of the analyzed electrophoretic image [7]. The method involves two main steps. Firstly, the algorithm calculates the intensity mean value of each column in the upper third of the gel image and identifies peaks using findpeaks method to mark the position of the first pixel of each border. The algorithm automatically fills in the expected lane boundary if none is found. Secondly, the algorithm tracks the other pixels of each border line consecutively, comparing the intensity values of 3x3 pixels below the current pixel and selecting the column containing the pixel with the highest intensity. The tracking algorithm stops at the bottom of the gel. An example gel image with segmented lanes is shown in Fig.1.

### C. Median filtering

In order to obtain a 1D signal representation of the intensity of pixels from 2D grayscale profiles representing individual sample lanes in horizontal orientation, a grayscale profile is firstly converted to a 1D signal by flattening it. Median filtering

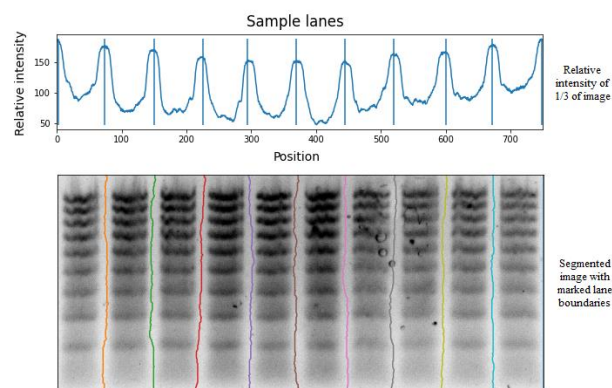


Fig. 1. Principle of gel image segmentation using tracking of lane boundaries

is then applied to the 1D signal using an appropriate window size, which is twice the width of the largest detected band [7].

#### D. Shading correction

In electrophoretic gels, the background can have a non-uniform shading due to factors such as uneven lighting or variations in the thickness of the gel [13]. By correcting the shading of the background, positions of bands can be more accurately measured and compared between different samples.

The steps involved in applying the local maxima method for envelope estimation include selecting an appropriate sliding window size, applying the window to the signal, computing the maximum value within the window, and repeating the process for each sample in the signal [14]. The result is then interpolated to obtain a curve covering the entire signal. The envelope is subsequently subtracted from the original signal, resulting in a corrected image with reduced background shading. The size of the sliding window is twice the size of the widest band in a signal. The window overlap is set as one quarter of the sliding window size and requires extension of both ends of the signal [7]. An example of input and output gel images is shown in Fig.2.

#### E. Band detection

The method involves analyzing a single-lane signal to detect bands. The first step is to invert the signal and then calculate the maximum amplitude. A prominence parameter is then calculated as a fraction of the maximum amplitude. This parameter is used to detect peaks in the signal using the find peaks function. The result is an array of peak indices representing the locations of the bands in the signal [15].

#### F. Molecular weights evaluation

This method involves creating a calibration curve to estimate the molecular weights of unknown protein or nucleic acid samples. The migration distance of each known molecular weight marker band and its corresponding molecular weight are obtained, and the calibration curve is created by interpolation of these data points. The molecular weights of the unknown bands in the sample lanes can be estimated by finding their corresponding distances in the 1D signal and then using the interpolation function to find their corresponding molecular weights [16].

### III. REALIZATION OF MOBILE APPLICATION

The structure of the application involves a client-side mobile application developed in Kotlin and a server-side application

developed using Python and FastAPI. Simplified workflow is shown in Fig.3. The client-side application allows users to capture or upload 1D electrophoresis gel images and select the area of interest for further analysis. The user also provides additional information about the gel image, such as the number of sample lanes, the lane containing the molecular weight marker, and the type of molecular weight marker. This information is sent to the server-side application via a FastAPI endpoint. The server-side application retrieves the image data and performs image processing and analysis using Python-based code. This includes grayscale conversion, gamma correction, lane segmentation, median filtering, shading correction, and band detection.

To ensure that the pairing of marker bands and their corresponding molecular weights is accurate, a manual step has been introduced. This step is triggered if the number of detected bands does not correspond to the number of molecular weights values and requires user input to resolve any discrepancies. The client-side receives an image of horizontally oriented marker lane along with its 1D intensity profile showing detected band peaks and corresponding molecular weights. The user can manually reposition the darts to ensure that all molecular marker bands are correctly annotated. The adjusted marker lane is sent back to the server-side.

A calibration curve is created and used to estimate the molecular weights of all the bands in the gel. The final result is an image that shows the molecular weights of individual bands. The new image is sent back to the client-side application via the FastAPI endpoint and displayed to the user. The user can save the image to their device or send it by email.

To add a new molecular weight marker, the user can request the addition through the client-side application. The request is sent to the server-side application via a new endpoint, and the server-side application updates the SQL database by creating a new record in the MarkerTypes table for the new marker type. The SQL database has three tables: Images, Bands, and MarkerTypes. The Images table stores information about the electrophoresis gel images uploaded by the users. The Bands table stores information about the molecular weight bands detected in the images, including their position and the molecular weight of the band if known. The MarkerTypes table stores information about the available molecular weight marker types, including the name of the marker and values of molecular weights of its consecutive bands.

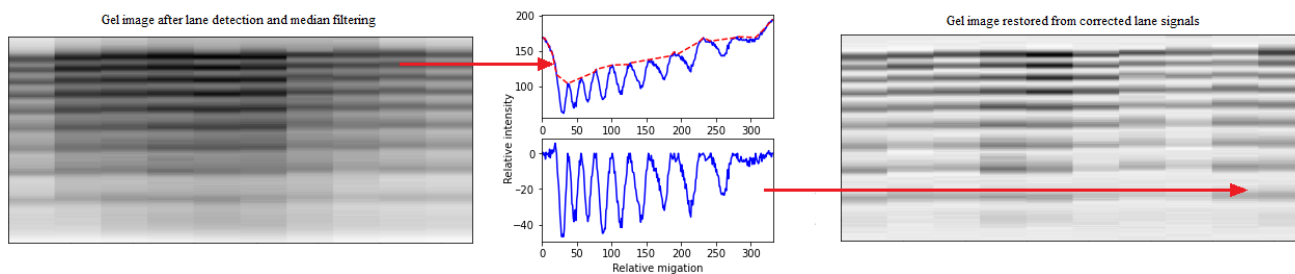


Fig. 2. The gel image after lane detection and median filtering (left), the signal of the lane with estimation of the signal envelope (middle), the reconstructed gel image with removed background shading (right).

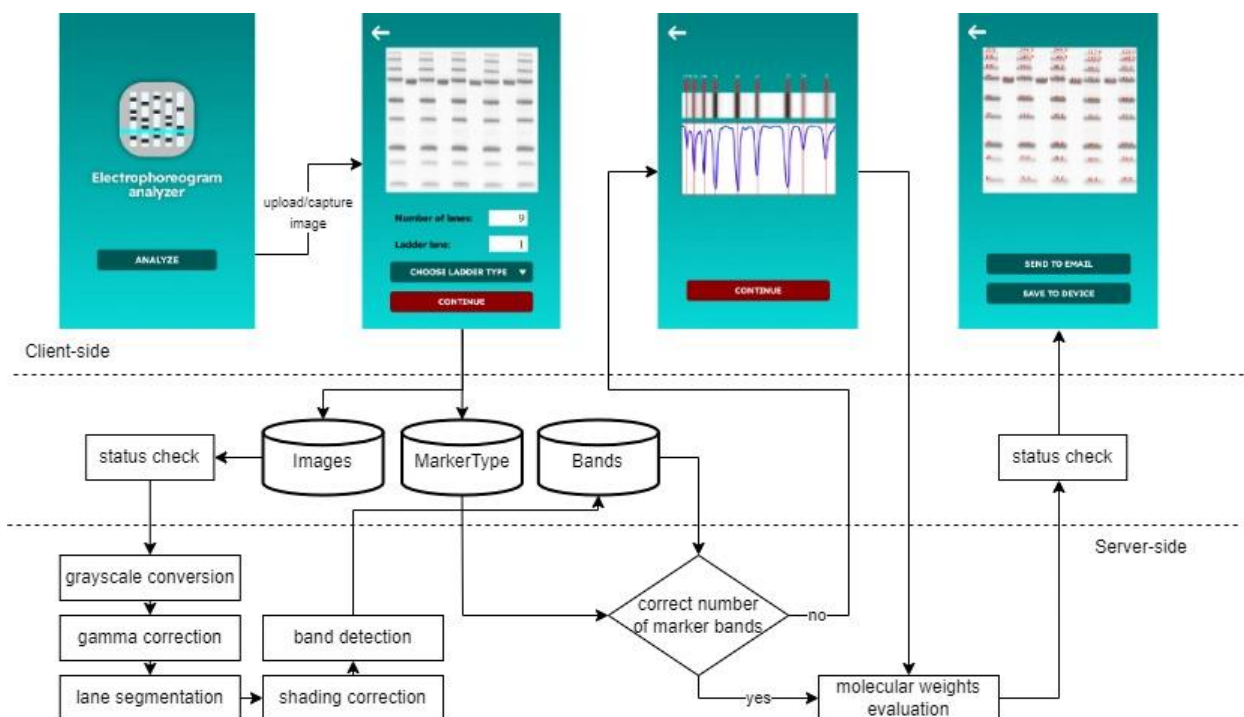


Fig. 3. Workflow of the mobile application.

The client-side and server-side components are connected through the use of five API endpoints, which are responsible for handling requests and responses between the mobile application and the server. Upload Endpoint receives a POST request containing a cropped gel image and additional information about the image. This endpoint stores the image and data in a SQL database on the server. Marker Endpoint allows users to add new molecular weight marker types to the application. It accepts a POST request with marker information and returns a response with the newly created marker ID. Analysis Endpoint processes the image and additional data stored in the SQL database by the upload endpoint. This endpoint uses Python code to perform the image analysis described in the previous chapter and generate the analyzed image. Download Endpoint returns the analyzed image to the mobile application in response to a GET request. The URL of this endpoint is used by the mobile application to retrieve the analyzed image from the server. Status Endpoint returns real-time updates on the progress of the image analysis. This endpoint can be used by the mobile application to monitor the progress of the analysis and estimate how much longer the analysis is expected to take. The actual analysis time may vary depending on the complexity of the image being. Typically, the analysis time for a single image can range from 5 to 30 seconds, with more complex images taking longer to process.

#### IV. RESULTS

##### A. Dataset

A dataset of electrophoretic gel images with known ground truth information obtained from laboratory measurements and external sources [17, 18, 19] was created. The ground truth information includes locations of all bands in each image, with a smaller subset of images including corresponding values of molecular weight. The images have been divided into several

categories depending on their quality – standard, noise, non-uniform lane boundaries, blurred, and bubbles, to ensure that the program's algorithm can produce accurate results across different imaging conditions.

##### B. Lane segmentation

Accurate lane segmentation is crucial for the subsequent steps of band detection and molecular weight estimation. The lane segmentation method was applied to the dataset and the results were manually evaluated. The accuracy of the described method was calculated for different categories of image quality. The results are shown in TABLE I.

##### C. Band detection

Band detection functionality was evaluated by comparing the locations and sizes of the detected bands against the ground truth information for different image quality categories. Precision was calculated as the number of true positive detections divided by the total number of detected bands. Sensitivity was calculated as the number of true positive detections divided by the total number of ground truth bands. The results are shown in TABLE II.

##### D. Molecular weight approximation

To test the molecular weight approximation functionality of the program, a ground truth subset of images with known molecular weights was used. Approximated values of molecular weights were calculated and compared to the ground truth values. The accuracy of the method was evaluated using root mean squared error (RMSE), which measures the square root of the average squared difference between the predicted values and the true values and gives more weight to larger errors because of the squaring operation. The results are shown in TABLE III.

TABLE I. LANE SEGMENTATION RESULTS

Category	No. of lanes tested	Accuracy [%]
standard	112	100.0
noise	84	100.0
non-uniform	34	92.7
blurred	135	100.0
bubbles	47	100.0

TABLE II. BAND DETECTION RESULTS

Category	No. of bands tested	Precision [%]	Sensitivity [%]
standard	1038	99.8	99.6
noise	862	99.6	98.9
non-uniform	295	92.7	92.4
blurred	1264	95.0	94.8
bubbles	486	96.4	96.3

TABLE III. MOLECULAR WEIGHTS APPROXIMATION RESULTS

Category	No. of bands tested	Relative RMSE [%]
standard	148	3.4
blurred	184	5.4

## V. DISCUSSION AND CONCLUSION

The lane segmentation method was found to be 100% accurate for most image categories, with the exception of images with non-uniform lane boundaries, which is primarily a problem of gel preparation and not the algorithm itself. The band detection algorithm performed well in high-quality images as well as images with noise. However, there is room for improvement in images with non-uniform lane boundaries and blurred lanes. Due to the limited availability of ground truth data, molecular weight approximation was not extensively tested.

In addition to the improvements to the existing methods, the application has the potential to be expanded for sample classification. Moreover, the application could include more extensive image quality improvement techniques to handle images with non-uniform boundaries and blurred lanes. Another potential feature is the adjustment of the mutual positions of the bands to allow more accurate detection and analysis of molecular weights. These potential advancements could greatly enhance the usefulness of the mobile application.

In conclusion, the developed method for the analysis of molecular weights in gel electrophoresis images is accurate and efficient. It can analyze images on-the-go without requiring any specialized equipment or expertise, making it accessible to a wide range of researchers. While the method performs best on images without geometric distortions, it still provides highly accurate results for distorted images. Overall, this application has the potential to save researchers time and resources. Further development and refinement of the mobile application could lead to even greater accuracy and utility in the future.

## ACKNOWLEDGMENT

I would like to express my sincere gratitude to my advisor, Ing. Helena Škutková, Ph.D., for her invaluable help, advice, and guidance throughout this project.

## REFERENCES

- [1] B. Alberts. Molecular biology of the cell, 5th ed. New York: Garland Science, 2008. ISBN 9780815341055.
- [2] K. Wilson. Wilson and Walker's principles and techniques of biochemistry and molecular biology. Cambridge University Press, 2018. ISBN 9781316614761.
- [3] J.L. Viovy. "Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms," in Reviews of Modern Physics, vol. 3, pp.813-872. 2000. ISSN 0034-6861.
- [4] J. Sambrook, D.W. Russell. Molecular cloning: a laboratory manual, 3rd ed. New York: Cold spring harbor laboratory press, 2001. ISBN 9780879695767.
- [5] S. Tseng, J.Y. Lin, J.P. Hsu. "Theoretical study of temperature influence on the electrophoresis of a pH-regulated polyelectrolyte," in Analytica Chimica Acta, issue 847, pp.80-89. 2014. ISSN 00032670.
- [6] V.T.T. Lan, P.T.T. Loan, P.A.T. Duong, L.T. Thanh, N.T. Ha, T.B. Thuan. "Straightforward Procedure for Laboratory Production of DNA Ladder," in Journal of Nucleic Acids, pp.1-4. 2012. ISSN 2090-0201.
- [7] H. Škutková, M. Vitek, S. Křížková, R. Kizek, I. Provazník. "Preprocessing and classification of electrophoresis gel images using dynamic time warping," in International Journal of Electrochemical Science. 2013.
- [8] N. Kaabouch, R.R. Schultz, B. Milavetz, L. Balakrishnan. "An analysis system for DNA gel electrophoresis images based on automatic thresholding and enhancement," in 2007 IEEE International Conference on Electro/Information Technology, pp. 26-31. IEEE, 2007. ISBN 9781424409402.
- [9] Y. Labyed, N. Kaabouch, R.R. Schultz, B.B. Singh. "Automatic segmentation and band detection of protein images based on the standard deviation profile and its derivative," in 2007 IEEE International Conference on Electro/Information Technology, pp.577-582. IEEE, 2007. ISBN 9781424409402.
- [10] L.G. Shapiro, G.C. Stockman. Computer vision. New Jersey: Prentice-Hall, 2001. ISBN 9780130307965.
- [11] C.A. Poynton. Digital video and HDTV: Algorithms and interfaces. Morgan Kaufmann, pp.260. ISBN 1558607927.
- [12] S. Rahman, M.M. Rahman, M. Abdullah-Al-Wadud. "An adaptive gamma correction for image enhancement," in J Image Video Proc, issue 35. 2016.
- [13] R.B. Paranjape. "Fundamental Enhancement Techniques," in Biomedical Engineering, Handbook of Medical Imaging, Academic Press, pp.3-18. 2000. ISBN 9780120777907.
- [14] A. Azqueta, K.B. Gutzkov, G. Brunborg, A.R. Collins. "Towards a more reliable comet assay: Optimising agarose concentration, unwinding time and electrophoresis conditions," in Mutation Research/Genetic Toxicology and Environmental Mutagenesis, vol.1, pp.41-45. 2011. ISSN 13835718.
- [15] S.W. Smith. "Moving Average Filters," in Digital Signal Processing, pp.277-284. Newnes, 2003. ISBN 9780750674447.
- [16] D.C. Tseng, Y.C. Lee. "Automatic band detection on pulsed-field electrophoresis image," in Pattern Anal Applic, pp.145-155. 2015.
- [17] M.M. Muharram, M.S. Abdel-Kader. "Utilization of gel electrophoreses for the quantitative estimation of digestive enzyme papain," in Saudi Pharmaceutical Jpurnal, vol.25, pp.359-364. 2017. ISSN 13190164.
- [18] T. Dvořáček. "Kompenzace geometrického zkršení obrazu gelové elektroforézy." Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2015. 107s. Vedoucí diplomové práce Ing. Helena Škutková.
- [19] RGP: 332 PCR-based genetic markers on rice chromosomes, National Institute of Agrobiological Sciences licensed under CC Attribution-Share Alike 2.1 Japan .