chromatoelectrophoresis, serum proteins,
watershed algorithms, image processing

Marcin WAŚKO[*], Patryk ORZECHOWSKI[**],
Norbert WNUK[**], Marcin KRZYCH[**]

A NOVEL APPROACH TO COMPUTER AIDED ANALYSIS OF HUMAN SERUM PROTEINS CHROMATOELECTROPHORESIS

Chromatoelectrophoresis is a laboratory technique which depicts changes in multiple protein fractions inclusively. The combined, simultaneous analysis based on two key protein features (charge and molecular weight) offers a unique opportunity for better understanding of serum protein composition as well as it indicates the fraction which needs more thorough investigation. Because of those promising features, the chromatoelectrophoresis with automated analysis is expected to be a valuable laboratory test, aiding the process of medical diagnosis. However, without information technology support, analyzing the results of chromatoelectrophoresis would be tedious and time-consuming. For better optimization, both an algorithm for output image analysis and application with user-friendly interface were developed. Planning, development and testing were conducted at AGH University of Science and Technology and Jagiellonian University, Medical College in Cracow, Poland. With this article, we present the results of the first year of cooperation, code-named ChromSee.org project. The algorithm for distinguishing, naming and analyzing the content of fractions, the application and their utility in real-life settings are described, as well as potential future developments.

1. CHROMATOELECTROPHORESIS

1.1. BACKGROUND

Chromatoelectrophoresis is a laboratory technique, combining thin layer chromatography [1] and gel electrophoresis [7]. It depicts multiple protein fractions simultaneously and provides the physician with the information on two key protein features - charge and molecular weight. Being more extensive than electrophoresis and chromatography performed independently, it conveys more clinically relevant data, as for example, which fraction needs to be investigated more thoroughly [10], for example in cases of acute inflammation [2], cancer [4,6] or hepatic disorders [3]. The technique of chromatoelectrophoresis is being developed constantly and new separation media for chromatography and electrophoresis are sought in order to improve the quality of the output image.

1.2. THE IMAGE

The chromatoelectrophoresis produces a two-dimensional image of serum protein fractions. The fractions are positioned according to increasing charge along the x axis and according to increasing molecular weight along the y axis. Usually, the output picture is composed of 11 to 15 greyscale spots, which represent the protein fractions, as schematically depicted in Fig.1.

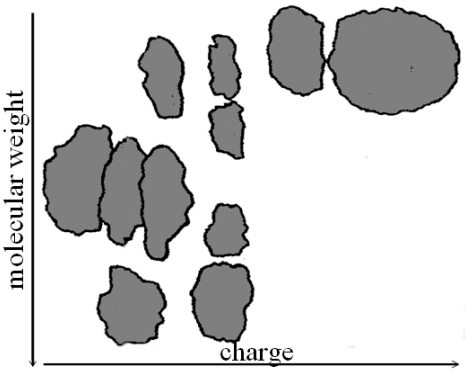

Fig.1 Schematic drawing of the chromatoelectrophoretic image

The intensity and dimensions of the spots depend on the amount of protein within them - indicating the concentration and thus, the quantity of protein. Although concentrations of all the fractions are known to lie within certain limits, the results

[*] Jagiellonian University, Medical College, Cracow, Poland
[**] AGH University of Science and Technology, Cracow, Poland

of chromatoelectrophoresis of healthy persons can vary in between, still within the normal range. That is why no referential image exists, nor the single normal result. [1].

1.3. PATHOLOGICAL ABNORMALITIES

Possible variations in chromatoelectrophoretic image include qualitative (increase or decrease in protein content of a given fraction, different fraction positioning or shape) and quantitative (e.g. absence of normal fraction or appearance of a new fraction) changes. Grossly, the pathological changes that could appear follow 5 patterns – increased or decreased intensity, horizontal or vertical shift and new fraction appearance.

2. ALGORITHM

2.1. PURPOSE AND RECOGNITION SCHEME

The proposed algorithm clusters visual images of human serum protein chromatoelectrophoresis in order to separate protein fractions. As the result, each fraction is represented as the set of neighbouring pixels. Different sets may contain the same pixels (so-called shared-pixels), as fractions may overlap. The fraction representation enables to easily apply different formats of data for recognition mechanism yet to be designed. For visualization purposes, an ecliptic scheme was applied – a structure contains information about position of centre of mass, its horizontal and vertical radius, and number of pixels covering and an average of their intensities. The recognition scheme aim is to automatically label the fractions, differentiate them between proper fractions and pathologies and finally – to count the amount of proteins covered (as sum of intensities of pixels belonging to the fraction).

2.2. ANALYSIS DIFFICULTIES

Problems encountered during analysis of chromatoelectrophoretic image are – to some extent - similar to those encountered with 2-dimensional electrophoresis and include overlapping spots and varying background [5]. Another major difficulty with automatic clustering of protein fractions is very noisy input data, therefore some preprocessing needs to be used. Satisfactory results were received so far with use of combination of median filtering and opening operations for elimination of small background noise and edge smoothing, respectively.

2.3. PIXEL-GROUPING METHOD

The algorithm being used in our application bases on the idea of [9]. Parallel to his solution, the contextual information of specific pixel is used to form clusters and indexes of specific groups are hold in so-called “glue array”. The main difference between algorithms is the order of inspection – pixels are processed according to their greyscale level, not to their localization, as originally. This adds a watershed behaviour to the algorithm (with water pouring out from the strings on the hills, forming a basins until the threshold (minimal CUT-OFF) level is reached) and enables fraction differentiation/forming, but also implies having to check for each pixel all his neighbourhood, instead of analyzing neighbourhood of pixels already processed [9], as presented on Fig.2).

P(i-1, j-1)	**P(i-1, j)**	**P(i-1, j+1)**
P(i, j-1)	**P(i, j)**	P(i, j+1)
P(i+1, j-1)	P(i+1, j)	P(i+1, j+1)

Fig.2 Neighborhoods of processed pixels P (i, j) used in [9] (in bold) and in proposed algorithm.

2.4. THE WATERSHED ALGORITHM

For the reasons of clarity, a group is a set of pixels neighbouring each other, which doesn’t fulfil adequate conditions to be called a fraction (minimal condition being a minimal number of pixels), and a region – is a set of pixels shared by different fractions. Image negatives are considered for conceptual purposes.

The main part of the algorithm executes the following steps:

1. For each pixel from the current greyscale level, check its adjacent pixels in order to verify existence of fractions or regions in the neighbourhood.

a. If one or more fraction or region is discovered in neighbourhood, check their indexes. Add a new region to the region list if necessary (i.e. no region containing included fractions existed). Assign the pixel to fraction or unified region.
b. If one or more groups in the neighbourhood are found, then reassign (updating the index in a glue array) all the group pixels to the fraction or the unified region (if it existed). Otherwise – assign to the lowest index of the groups involved (assign the pixel appropriately and check if the merged group may be called a fraction, if it consists of more than MINIMAL_NUMBER of pixels, which can be set by the user)
c. If no group, fraction or region is met, consider the pixel as a new group.

2. Decrease the current greyscale level and go to step 1 until the threshold is reached.
3. Iterating through the glue array, reevaluate all the group indexes (i.e. take into account all the merges made between the groups during the greyscale level inspection)
4. Finally, re-index all the image pixels to their final values, basing on the glue array.

2.5. IMPLEMENTATION DETAILS

For implementation purposes, indexing table is used for storing pixels adhesion (see Fig.3), indexed_px vector – for easier access to pixel identifiers belonging to the same groups (or fractions), shared_px vector – for storing pixels belonging to barriers between different fractions. A separate glue table is used for merging groups in indexed_px, so that re-indexing all pixels during preprocessing wasn't necessary. Fraction table is used for covering indices of groups in indexed_px considered to be fractions and, nei_vect to store which fractions are neighbouring with each other.

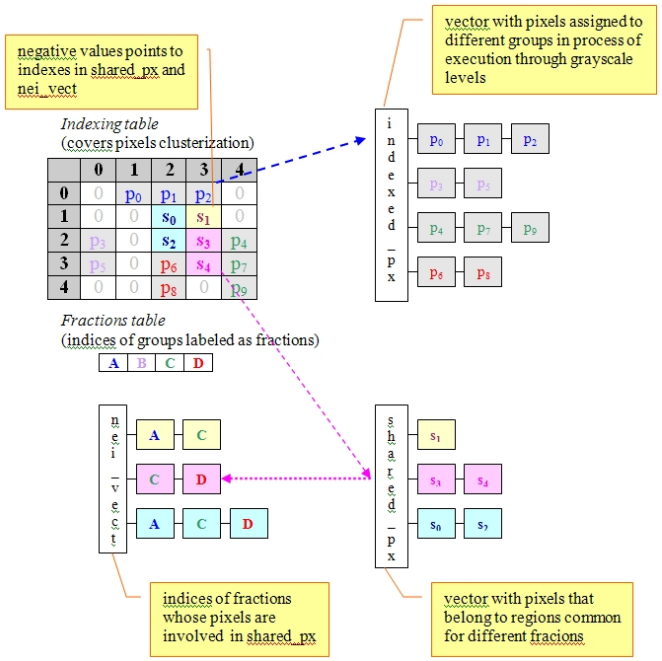

Fig.3 Algorithm indexing table. Notice that positive values in indexing table point to indexed_px, negative – to shared_px. The absolute value determines the position in a vector. Zero value is used for defining the background.

The algorithm covers very dark, intensive fractions (with high amount of proteins) as well as brighter ones with low intensity. Its main advantage is combining image indexing and clustering techniques, which reduces its complexity.

3. APPLICATION

3.1. COMPONENTS

The application, named ChromSee, was created with Open Source components, in order to minimize the costs and to allow easy widespread among academic users.

Graphic User Interface was designed in QT4 library [14], which makes the application very portable among Windows, Linux and MacOS platforms (with QT 4.4 this list has expanded even by portable devices controlled by Windows CE). Making the application and data accessible from various computers is an advantage that cannot be overestimated, especially in Life

Science and medical settings. Working on portable machines gives an opportunity to use application directly at the site of patient care. Schematic representation of application functionalities is depicted in Fig.4.

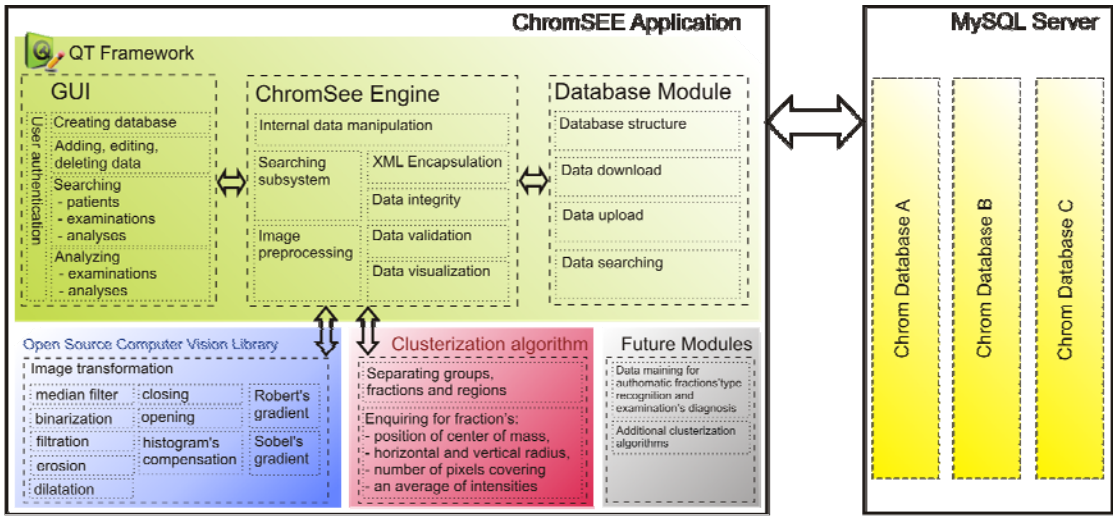

Fig.4 ChromSee application components scheme

3.2. DATABASE SPECIFICATION

Database subsystem was designed as an objective-relation solution run on MySQL server, version 5.1 [12]. MySQL was chosen for performance and scalability reasons, as well as for built-in tools important for data security [11]. A separated database engine gives also a possibility to access the data from different client applications or through a dedicated website. The database was designed to facilitate its future development by holding the part of the data (the most exposed for modification or changes in user's specification) inside XML structure. It enables quick application accommodation without changing databases structure which is much more flexible for development process and database security. Furthermore XML structures are supported by database engine which improves data operations like advance searching or integrity validation.

3.3. INITIAL IMAGE PREPROCESING

Image preprocessing with the use of the combination of predefined operations (like median filter, binarization, filtration, dilatation, erosion, opening, closing, histogram's compensation, Sobel's gradient, Robert's gradient, all of which can be repeated and linked together) is executed by OpenCV [13] library. This solution ensures a great performance even in complicated preprocessing combinations and could give a speedup of about an order of magnitude in comparison to common dedicated environments like Matlab.

3.4. MAIN FUNCTIONS AND DESTINY

The ChromSee enables a parallel access to patients' database, their medical history and analyses. Moreover, all of the data can be acquired, edited and the analyses run from distant locations on a central server. Thus, not only can several teams conduct their research simultaneously and use common database to increase the scientific yield, but also specialist from different centres can evaluate the results and comment on them (Fig.5), which gives a fair chance of prompt and effective diagnosis, particularly in ambiguous and pathological cases.

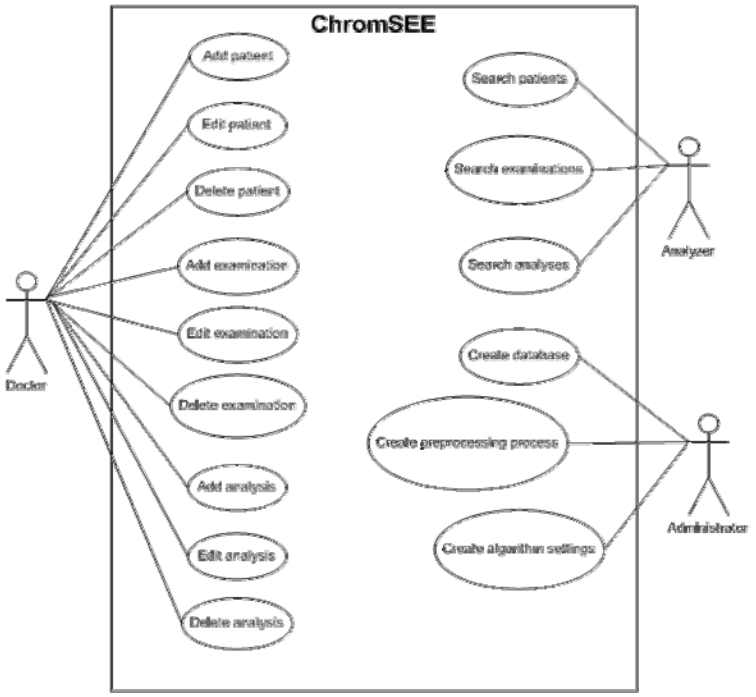

Fig.5 ChromSee use-case view

ChromSee supports a doctor with selective access to all significant information in the process of analyzing the results of an examination. These include among others:

1. Patients' medico-legal information. The application gives a flexible possibility to extend this collection in case of integration with external (e.g. hospital) database system.
2. Detailed information about each chromatoelectrophoresis process itself, such as time of sample run, sample's grade, electric field voltage, source of the serum, place where examination was executed and diagnosis (Fig.6 – left upper corner).
3. Algorithm processing results, such as number of recognized fractions, localization of each fraction (i.e. barycenter and radius), average fraction intensity (mean of all pixels intensities). Also relative information between fractions is provided to accelerate the searching. Each fraction may be separately named by a doctor/researcher by double clicking its name (Fig.6 – right bottom corner).
4. The visualization of results with elliptic markings around the detected fractions.
5. Comment window for sharing interpretation diagnosis between doctors. Every adding, editing, deleting operation leaves a mark with information about its author, time of edition and location, from which the change was made.

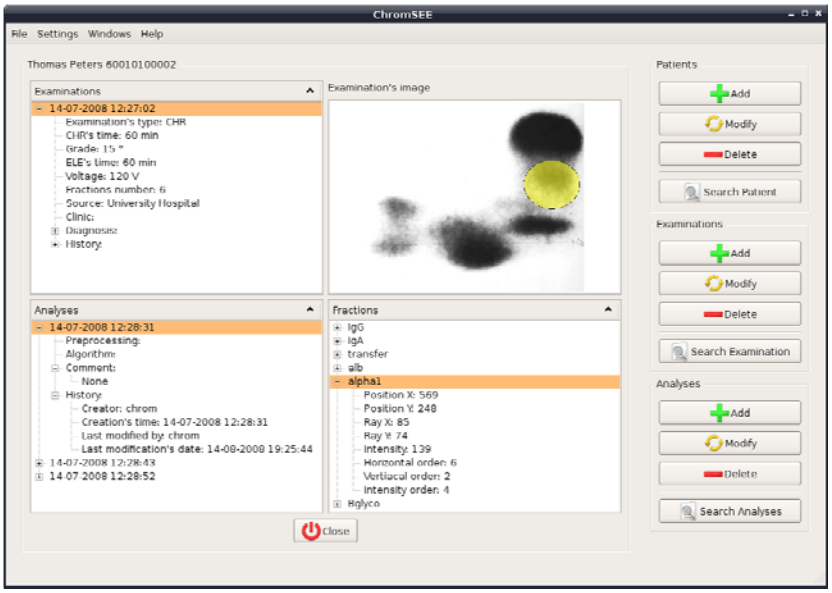

Fig.6 ChromSee analysis dialog window

ChromSee supports multiple examinations for one patient – each can be made at different time and with different parameters, at different cooperating locations (e.g. different hospital wards). Each may be analyzed with different initial image preprocessing and algorithm's settings customized by users to local requirements (like special characteristic of examination's image dependent on used reagents or equipment).

Application enables searching and viewing this data by patching predicates on all defined parameters (Fig.7). Searching engine supports regular expressions. Some predicates were also predefined and may be linked with logical condition (AND, OR). The development of search subsystem supports databases with hundreds of patients and thousands of examinations' data. This approach guarantees possibility to extend ChromSee with reasoning mechanism in the future by giving rapid access to whole data filtered by all components. It will be very useful during searching complex dependences or finding similar cases.

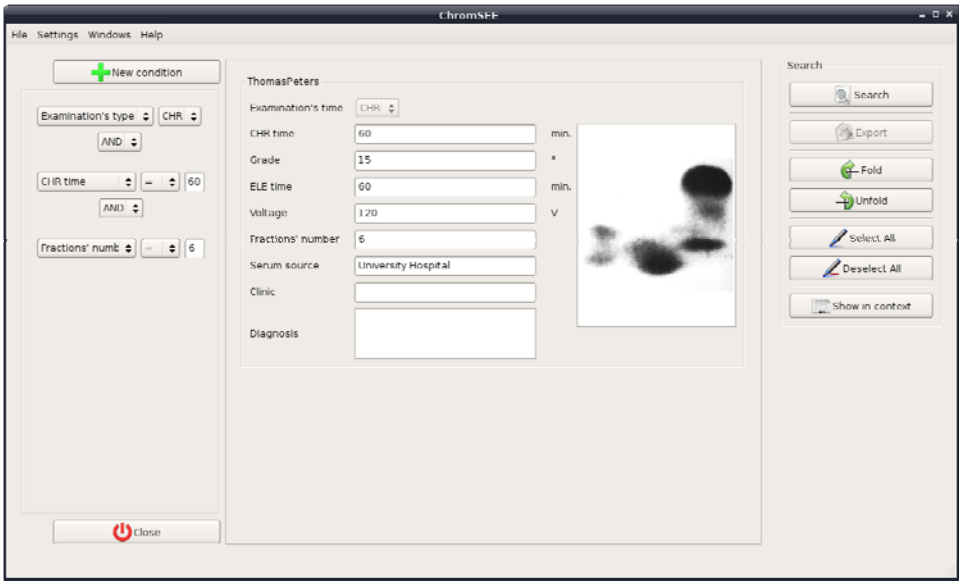

Fig.7 ChromSee search module, note possible search parameters to be set on the left

All information entered into database is validated by client application and at database's logic alike. Deleted data becomes hidden but is still stored inside the system and available to administrator. Special mechanisms assure no performance loss after deletion. During the testing phase, all of the mentioned mechanisms decreased the risk of errors and lack of cohesion.

4. CONCLUSION

After a year of cooperation between IT specialists and medical personnel, ChromSee became a functional and user-friendly application dedicated to chromatoelectrophoretic image storage, analysis and medical research team cooperation. The matter of utmost concern in future would be providing the ChromSee database with secure access to the data from different, remote locations, as almost all the data stored are personal.

Possible future modifications of an algorithm include adding H-dome transformation or Gaussian and diffusion-based parametric spot models [5]. Local adaptive thresholding techniques are also taken into account – such as Sauvola and Pietaksinen or Bernsen [8]. In parallel, agent system approach is being developed.

The most difficult challenge for future development is to expand application by data mining module which enables automatic fractions recognition and preparing initial diagnosis. Today it is rather difficult because of insufficient number of examinations inside ChromSee database and only partial assay for reasoning mechanism which blocks even a theoretical deliberation about this functionality.

BIBLIOGRAPHY

[1] BOGDANIKOWA B.: Klinika białek osocza. Polski Związek Wydawnictw Lekarskich 1963
[2] KILLINGSWORTH LM: CRC critical reviews in clinical laboratory science. 1979 Aug; 11(1): 1-30
[3] LUTZ W: Perspectives in the study of plasma proteins and their value in clinical diagnosis Polski Tygodnik Lekarski. 1980 Oct;35(43):1667-70.
[4] MCMASTER ML, CSAKO G. Protein electrophoresis, immunoelectrophoresis and immunofixation electrophoresis as predictors for high-risk phenotype in familial Waldenström macroglobulinemia. International journal of cancer. 2008 Mar 1;122 (5):1183-8.

[5] PEDERSEN L. Analysis of two-dimensional electrophoresis gel images, Informatics and Mathematical Modelling Ph.D. Thesis No. 96 Kgs. Lyngby 2002
[6] PUSHKAREV IA, BORISOV LL, PETERS VH, STORCH H, GUSEVA LN. Use of agarose gel and cellulose acetate electrophoresis for detecting and identifying monoclonal immunoglobulin. Laboratornoe delo. 1979;(10):606-9.
[7] RIGHETTI PG: Electrophoresis: The march of pennies, the march of dimes. Journal of Chromatography A, 1079 (2005) 24–40
[8] SEZGIN M., SANKUR B., Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging 13(1), 146–165, 2004
[9] TADEUSIEWICZ R., KOROHODA P., Komputerowa analiza i przetwarzanie obrazów. Wydawnictwo Postępu Telekomunikacji, Kraków 1997
[10] WAŚKO M., KONIECZNY L., ROTERMAN I., Chromatoelectrophoresis. A method disclosing molecular weight and charge of serum proteins, Eur J Med Res 11 (Supplement II) I-XVIII, 7, 2006
[11] http://dev.mysql.com/doc/refman/5.1/en/backup-and-recovery.html
[12] http://dev.mysql.com/doc/refman/5.1/en/index.html
[13] http://opencvlibrary.sourceforge.net/
[14] http://trolltech.com/products/qt/