Data Article

Code and data on the processing of the pulsed-field gel electrophoresis images: A matlab script

Ahmad Mohammadbeigi [a, b], Mohammad Rezaei [c, d, *],
Naser Zohourian Sani [a, f], Nemat Soltani [a], Parviz Mohajeri [e]

[a] Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran
[b] Department of Radiology, Sina Hospital, Tehran University of Medical Sciences, Tehran, Iran
[c] Sleep Disorders Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran
[d] Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran
[e] Department of Microbiology, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran
[f] Bio Medical Engineering Unit, Taleghani Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

ARTICLE INFO

ABSTRACT

Here a matlab script was presented for lane tracking and band detection on the pulsed field gel electrophoresis (PFGE) images. It can also be used as a software tool for automatic analysis of PFGE images. The data consist of several MATLAB codes which collectively have the task of lane tracking, band detecting and pattern recognition on the PFGE images. The lane tracking stage is semi-automatic and the band detection stage is fully automatic. Finally, the pattern of lanes that includes number of, location, width and light intensity level of bands was obtained.

© 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Data

The data consist of several MATLAB codes, which collectively have the task of lane tracking, band detecting and pattern recognition on the pulsed-field gel electrophoresis images. The PFGE is a

---

Specifications Table

Subject	Biomedical Engineering
Specific subject area	image processing in microbiology and biotechnology
Type of data	MATLAB code, image, video
How data were acquired	All source codes written in Matlab software.
Data format	MATLAB code, JPEG, Mp4
Parameters for data collection	All the codes were implemented in MATLAB-R2009a on a system with Intel Core - i5 2430M, quad-core processor overclocked at 3.2 GHz with 8GB of RAM clocked at 1600 MHz. A trial version of GelCompar II version 6.6.11 was used to evaluate and optimize the codes.
Description of data collection	The images were captured using PFGE BIORAD at the Microbiology Laboratory of Kermanshah University of Medical Sciences. The images were provided by two types of bacteria, including Acineto-AF, *Staphylococcus aureus*.
Data source location	Institution: Department of Biomedical Engineering in Kermanshah University of Medical Science City/Town/Region: Kermanshah Country: Iran Latitude and longitude: 34°23′27.9″N 47°06′07.7″E
Data accessibility	- with the article - *The dataset is freely available at* [1] *for any academic, educational, and research purposes.*
Related research article	Repository name: Mendeley Data Data identification number: https://doi.org/10.17632/mcnfncf25t.1 Direct URL to data: https://data.mendeley.com/datasets/mcnfncf25t/1 Author's name: Mohammad Rezaei, Mahmood Amiri, Parviz Mohajeri, Mansour Rezaei Title: A new algorithm for lane detection and tracking on pulsed field gel electrophoresis images Journal: Chemometrics and Intelligent Laboratory Systems DOI: 10.1016/j.chemolab.2016.05.018

Value of the Data
- The provided codes can be used to pulsed-field gel electrophoresis image analysis.
- The Matlab script will allow microbiologist to molecular subtyping.
- This approach can be used to automatic lane tracking, band detection and pattern recognition on PFGE images.

laboratory technique used by researchers and scientists to produce a DNA fingerprint for a bacterial isolate as a group of the same type of bacteria [2–5]. The images were provided by two types of bacteria, including Acinetobacter [6], *Staphylococcus aureus* [7] which were attached with the article. All the codes and data needed for this purpose available in the Mendeley data source [1]. In addition, there is a video showing the performance of the data and Matlab script. The flowchart of implemented algorithm as Matlab script was shown in Fig. 1.

2. Experimental design, materials, and methods

The material used in this dataset includes images and codes. The images were collected using PFGE BIORAD at the Microbiology Laboratory of Kermanshah University of Medical Sciences in "tiff" format. To analyse the images, MATLAB-R2009a [8] on a system with Intel Core - i5 2430M, quad-core processor overclocked at 3.2 GHz with 8GB of RAM clocked at 1600 MHz was used. To evaluate and optimize the codes, a trial version of GelCompar II software was also used.

2.1. Algorithm for the lane tracking

A program, "*lane_tracking.m*", is responsible for lane tracking. First, image was converted to gray-scale from RGB format. The desired area of the image including lanes was cropped then was resized to 500 × 500 pixels. Next, image segmented to sub-images to calculate vertical projection profile (VPP).

To detect local of each lane in the sub-images, local maxima from signal of VPP was detected. Using the matlab codes, "*widthfind.m*" and "*Remov_locmax.m*", false-positive local maxima in the background was removed. After detecting the center of lanes in each of sub-images, the centers associated with each lane are given to the function "*createFit.m*" to fit the patch of lane. In detail, the proposed algorithm is described in Ref. [9]. Fig. 2 depicts the lane tracking.

2.2. Algorithm for the band detection

The band detection as one of the stages consists of two phases. First, a matlab code, "*bandremovebackgroundm.m*", is responsible for removing background noise from lanes image. To this end, the lane image was smoothed by an adaptive median filter on lane images. Then, as a completion step, lower envelope of VPP of lane was subtracted. The source code for this function is given bellow.

```matlab
function [II] = bandremovebackgroundm(I)
%>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
L_vpp=sum(I,1);
L_vpp=500*(L_vpp-min(L_vpp))/(max(L_vpp)-min(L_vpp));
P=4; % P Equal to half of Averaged Range - width of Lane is 40 pixels
L_vpp1=L_vpp;
for i=P+1:size(L_vpp,2)-P
    L_vpp(i)=sum(L_vpp1(i-P:i+P))/9;
end
%>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
L_valley_vpp=-1*(L_vpp-max(L_vpp));
SlopeThreshold=0.005;
SlopeThreshold_valley=0.00001;
AmpThreshold=0;
smoothwidth=2;
peakgroup=5;
smoothtype=5;
P=peakfind([1:size(L_vpp,2)],L_vpp,SlopeThreshold,AmpThreshold,smoothwidth,pe
akgroup,smoothtype);
P_valley=peakfind([1:size(L_valley_vpp,2)],L_valley_vpp,SlopeThreshold_valley
,AmpThreshold,smoothwidth,peakgroup,smoothtype);
P_valley(:,3)=-1*(P_valley(:,3)-max(L_vpp));

P_valley(P_valley(:,2)>500,:)=[];
P_valley(P_valley(:,2)<0,:)=[];

L_vpp(1:ceil(P_valley(1,2)))=0;
L_vpp(fix(P_valley(end,2)):end)=0;
%>>>>>>>>>>>>>>>>>>>>>>>>
for i=1:size(P_valley,1)-1
    m=(P_valley(i+1,3)-P_valley(i,3))/(P_valley(i+1,2)-P_valley(i,2));
    L_vpp_envelope=m*([round(P_valley(i,2)):round(P_valley(i+1,2))-1]-...
        P_valley(i,2))+P_valley(i,3);
    L_vpp(round(P_valley(i,2)):round(P_valley(i+1,2))-1)=L_vpp(round(...
        P_valley(i,2)):round(P_valley(i+1,2))-1) - L_vpp_envelope;
end

for j=1:numel(L_vpp)
    if L_vpp(j)<10
    L_vpp(j)=0;
    end
end
%>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
II=repmat(L_vpp,75,1);
```
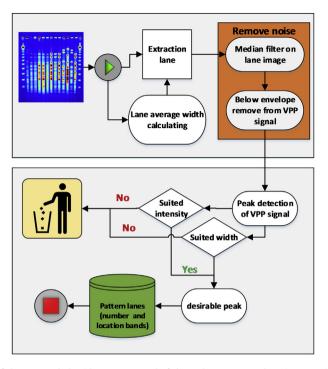
Fig. 1. The flowchart of the proposed algorithm. It composed of three phases:1- Lane detection 2- Noise reduction and band extraction 3- Pattern recognition.
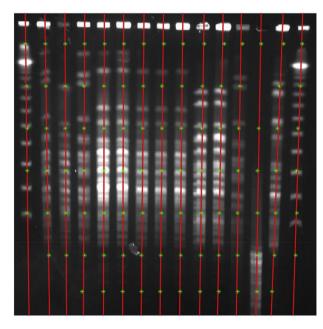

Fig. 2. A PFGE image with lanes which tracked using the matlab script "*lane_tracking.m*": The red lines demonstrate lanes and the green stars are center of the lanes.

In addition, the program "*band.m*" is responsible for band detection. To this end, after resizing it to 100×500 pixels and calculating the VPP of lane, the location of the bands was detected using the function "peakfind.m". The source code for this step is as follows. Finally, the pattern of lanes was determined by calculating the four parameters, including number of location, width and light intensity level of bands.

```
I=imresize(I,[100,500]); % Changing the image size to 100*500 pixels
%>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
L_vpp=sum(I,1);
L_vpp=500*(L_vpp-min(L_vpp))/(max(L_vpp)-min(L_vpp));
P=5; % P Equal to half of Averaged Range - the width of Lane is 40 pixels
L_vpp1=L_vpp;
for i=P+1:size(L_vpp,2)-P
    L_vpp(i)=sum(L_vpp1(i-P:i+P))/9;
end
%>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
PLans={};
SlopeThreshold=0.000001;
AmpThreshold=0;
smoothwidth=2;
peakgroup=5;
smoothtype=5;
P=peakfind([1:size(L_vpp,2)],L_vpp,SlopeThreshold,AmpThreshold,…
smoothwidth,peakgroup,smoothtype);
```

Acknowledgments

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.105035.

References

[1] M. Rezaei, Matlab Scripts for Lane Tracking and Band Detection on the Pulsed-Field Gel Electrophoresis Images, Mendeley Data, 2019, p. v1.
[2] S. Jeon, N. Lim, S. Park, M. Park, S. Kim, Comparison of PFGE, IS6110-RFLP, and 24-Locus MIRU-VNTR for molecular epidemiologic typing of Mycobacterium tuberculosis isolates with known epidemic connections, J. Microbiol. Biotechnol. 28 (2018) 338−346, https://doi.org/10.4014/jmb.1704.04042.
[3] K.R. Alkharsah, S. Rehman, A. Alnimr, A. Diab, A. Hawwari, S. Tokajian, Molecular typing of MRSA isolates by spa and PFGE, J. King Saud Univ. Sci. 31 (2018) 999−1004, https://doi.org/10.1016/j.jksus.2018.07.018.
[4] B.K. Sharma-Kuinkel, T.H. Rude, V.G. Fowler, Pulse field gel electrophoresis, Methods Mol. Biol. 1373 (2016) 117−130, https://doi.org/10.1007/7651_2014_191.
[5] L. Lopez-Canovas, M.B. Martinez Benitez, J.A. Herrera Isidron, E. Flores Soto, Pulsed field gel electrophoresis: past, present, and future, Anal. Biochem. 573 (2019) 17−29, https://doi.org/10.1016/j.ab.2019.02.020.
[6] L.S. Munoz-Price, R.A. Weinstein, Acinetobacter infection, N. Engl. J. Med. 358 (2008) 1271−1281, https://doi.org/10.1056/NEJMra070741.
[7] R.N. Stijn Blot, K. Vandewoude, F. Colardyn, *Staphylococcus aureus* infections, N. Engl. J. Med. 339 (1998) 520−532, https://doi.org/10.7326/0003-4819-61-4-808_1.
[8] The Mathworks Inc, MATLAB - MathWorks, 2016-11-26. Www.Mathworks.Com/Products/Matlab, 2016.
[9] M. Rezaei, M. Amiri, P. Mohajeri, M. Rezaei, A new algorithm for lane detection and tracking on pulsed field gel electrophoresis images, Chemometr. Intell. Lab. Syst. 157 (2016) 1−6, https://doi.org/10.1016/j.chemolab.2016.05.018.