

2D-PAGE 영상 처리 및 분석 기술

원 용관 박사 (전남대학교 정보통신공학부/의료정보학과 교수)
(017-633-3555, ykwon@chonnam.ac.kr)

Abstract:

2D-PAGE/MALDI-TOF는 프로테오믹스 연구의 중요한 실험 기법중의 하나이다. 이는 단백질 발현 분석을 위한 방법으로, 2D-PAGE 결과로 얻어진 영상 데이터의 분석에 대한 정확도가 단백질 발현에 대한 분석 결과의 질을 결정하는 중요한 요인으로 작용한다.

2D Electrophoresis에 의한 Gel Protein Database는 현재 많은 연구자들에 의해 생산되고 있으며, 대단히 많은 데이터들이 인터넷을 통하여 접근이 가능하다. 이러한 대량 정보의 Database 활용이 가능한 상황은 2D-PAGE에 의해 생산된 Gel Image의 상호 비교에 대한 요구를 도출하였다.

본 발표에서는 영상처리 및 형태인식 기술과 2D-PAGE 연구의 결합을 주제로 하여, 2D-PAGE Gel 영상 처리 및 비교에 관련되는 전처리 (preprocessing), spot detection, feature extraction, spot matching 및 image comparison 기술을 소개한다.

약력

1995 미주리주립대학 컴퓨터공학 박사
1995 - 1996 한국전자통신연구원
1996 - 1999 한국통신 통신망연구소
1999 - 현재 전남대학교 정보통신공학부/의료정보학과 조교수

주요 연구 분야 :

- 영상처리, 형태인식, 데이터마이닝
- 바이오인포메틱스 (프로테오믹스)

주요 BI 연구 분야 :

- 바이오 정보의 고속 통합 검색
- 바이오 영상처리 및 분석

2D Poly-Acrylamide Gel Electrophoresis(PAGE) Image Analysis

전남대학교
공과대학 정보통신공학부
의과대학 의료정보학과

원 용 관



발 표 내 용

- Overview of 2D Gel Image Analysis
- General Steps
- Spot Detection
 - ☞ Problem Definition
 - ☞ Watershed methodology
- Spot Comparison (Spot Matching)
 - ☞ Pre-processing
 - ☞ Techniques & Algorithms
- Local Matching (Gel Matching)
 - ☞ Granulometry
 - ☞ Delaunay triangulation
- Aspects of Data Analysis
- Challenges

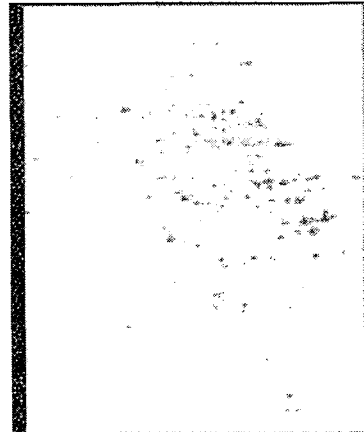


Overview: 2D Gel Electrophoresis

- ❑ Gene Expression Studies
- ❑ Medical Applications
- ❑ Sample Experiments

Overview: 2D Gel Image Analysis

- ❑ 2D G. E. includes thousands of protein spots.
- ❑ Wide variation
 - ☞ Size & Shape
 - ☞ Intensity
 - ☞ Location
 - ☞ Background
- ❑ Why Computer-based?
 - ☞ Large amount Data
 - ☞ Objective analysis
 - ☞ Quantitative analysis



General Steps

- Data Acquisition : Scanner
- Image Processing : Pre-Processing
 - ☞ Noise filtering & Smoothing
 - ☞ Contrast enhancement
 - ☞ Edge detection & Linking
 - ☞ Background subtraction
- Spot Detection *
- Spot Matching *
- Gel/Image Matching *
- Analysis**
- 2D Protein Database Construction



Spot Detection

- Problem Definition
- Spot Segmentation
 - ☞ Gaussian Fitting
 - ☞ Laplacian of Gaussian (LOG)
 - ☞ Watershed
- Post-Processing
 - ☞ Split and Merge
- Spot quantitation
 - ☞ Coordination, intensity, shape, etc.



Problem Definition

□ Objectives

- ☞ Detection of Protein Spots
 - ✓ Yielding Location information $x(s)$, $y(s)$
 - ✓ Calculation of Spot intensity $i(s)$

□ Procedures (All or Some)

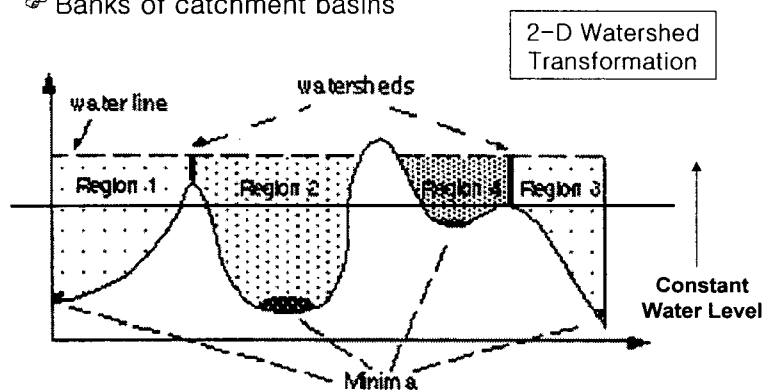
- ☞ Noise Removal, Distortion, Size(Normalization)
- ☞ Background Subtraction
- ☞ Gradient image
- ☞ Edge Detection
- ☞ Segmentation
- ☞ Split and Merge
- ☞ Feature extraction



Watershed

□ Principle

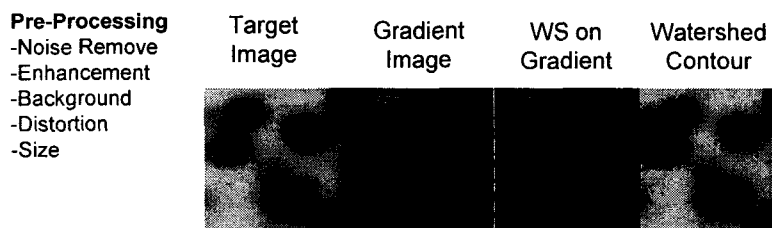
- ☞ Image Segmentation by an immersion process
- ☞ Banks of catchment basins



Segmentation by Watershed

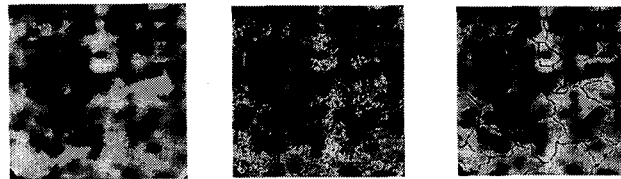
□ 3D Watershed Transform

- ☞ Gray-scale image : Considered as a topographic surface
 - ☞ WST flooding analogy : Demo
- ☞ Gradient image
 - ✓ Valleys correspond to the requested regions
 - ✓ Ridges(*watersheds*) define the optimal contour of a region



Over / Under-segmentation

□ Over/Under-Segmentation



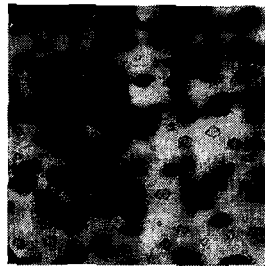
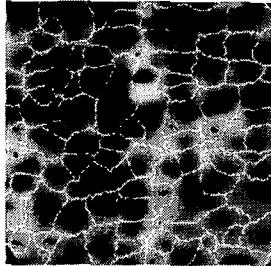
- Split and Merge
 - ☞ Divide an image into a set of disjoint regions
- Region Grown
 - ☞ Start from a seed pixel to collect coherent(similar) pixels.
- How to select *measurements* for :
 - ☞ Similarity, Threshold for “Similar”
 - ☞ Number of initial seed pixels



Marker-controlled Watershed

❑ Reduce Over-segmentation

- ☞ Pre-selected Markers
- ☞ Marker-driven segmentation Demo



- ☞ Marker finding problem : How to select markers?



Quantitation of Spots: Features

❑ Location

- ☞ Coordinate of the spots : Center of gravity of the spot

❑ Shape

- ☞ Degree of fitting to a shape (circle, rectangle, ellipse, etc)
- ☞ Curvature : Convex & Concave

❑ Intensity

- ☞ Sum of all gray-values in the spot
- ☞ Average gray-value of the spot
- ☞ Variance of gray-values of the spot



Spot Matching

□ Problem Definition

- ☞ Comparing/Matching identical spots over series of gels

□ Purposes

- ☞ To observe how a *protein profile changes* under various experimental conditions.
- ☞ To extract *biologically meaningful information*, the output images of different 2D-PAGE experiments have to be compared.

□ Variations over images

- ☞ difference in *sample preparation*
- ☞ chemical and physical *variations of gels*
- ☞ variations in *experimental conditions*
- ☞ *unequal mobility of proteins* in different gel regions



Pre-processing

□ Landmark Spots

- ☞ Used as *tie* points in multiple gel studies
 - ✓ to normalize spot patterns (points of reference)
 - ✓ to compensate for any gel distortion during the electrophoresis process
- ☞ Manual or Automatic Selection

□ Distortion

- ☞ Two identical spots in two gels have different coordination
- ☞ Should be eliminated or acceptably minimized



Techniques

□ Approaches

- ☞ Point/Spot matching
- ☞ Local matching
- ☞ Global matching

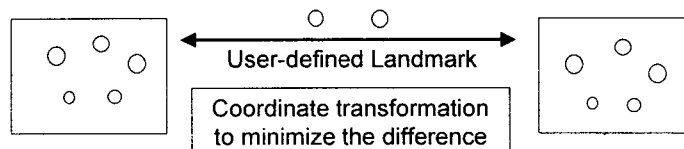
□ Methodologies for Gel matching (Features)

- ☞ Geometric properties
- ☞ Statistical characteristics
- ☞ Graph theory based approaches
- ☞ Pattern recognition techniques



Algorithms

□ Coordinate Transformation



- ☞ Overlay & find the closest spots

□ Gabriel graph

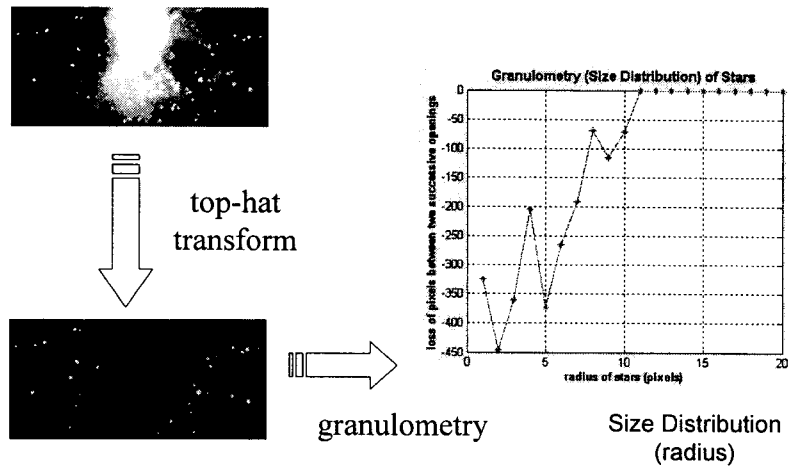
- ☞ Spots become nodes of the graph → Comparing graphs

□ Iterative method

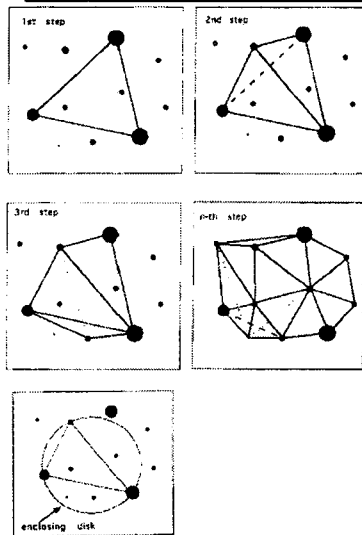
- ☞ Start with user-defined landmarks (initial match)
- ☞ Next match on neighborhood spots until no good match



Local Matching : Granulometry



Local Matching : Delaunay triangulation



- Finding star constellation
- Concept of *intensive edge*
 - ☞ Computing tentative matching pattern locations
 - ☞ Verifying tentative local matching locations
- Distortion problem
 - ☞ Distort iteratively the source pattern
 - ☞ Search distorted pattern in target image

Aspects for Data Analysis

- Quality estimation
 - ☞ Number of valid spots
 - ☞ Number of spots matched to every gel
- Qualitative changes
 - ☞ Appearance/Disappearance of (a) spot(s)
- Quantitative changes
 - ☞ Increase/Decrease spot intensity



Challenges

- Full Automation
 - ☞ Spot Segmentation
 - ☞ Gel matching : Image registration
- Search
- Image DB Construction

Image Analysis : many algorithms, some programs
Protein Identification : many algorithms, few programs
Identification of post-translational modification:
few algorithms, very few programs
Sequence alignment/comparison:
Many algorithms, many programs

Bioinformatics Group @ IGH



References

- ❑ R. Appel et al. Melanie II – a third generation software ... : I. Feature and Interface, *Electrophoresis*, 18, 2724–2748, 1997
- ❑ R. Appel et al. Melanie II – a third generation software ... : II. Algorithms, *Electrophoresis*, 18, 2724–2748, 1997
- ❑ G. Horgan, C. Glasbey, Use of digital image analysis in electrophoresis, *Electrophoresis*, 16, 298–305, 1995
- ❑ J. Panek and Vohradsky, Point pattern matching in the analysis of two dimensional gel electropherograms, *Electrophoresis*, 20, 3484–3491, 1999
- ❑ J. Vohradsky, Adaptive classification of two-dimensional gel electrophoresic spot patterns by neural networks and cluster analysis, *Electrophoresis*, 18, 2749–2754, 1997
- ❑ Y. Watanabe and K. Takahashi, A fast structural matching and its application to pattern alanalysis od 2D electrophoresis images, *Proc. Of Intl. Conf. On Image Processing*, 804–808, IEEE Comp. Soc, 1998
- ❑ S-H Chang et al, Fast algorithm for point pattern matching: invariant to translations, rotations and scale changes, *Pattern recognition*, 30(2), 311–320, 1977
- ❑ L. Brown, A Survey of Image Registration Techniques, *ACM Computing Survey*, 24(4), 1992
- ❑ C. Glasbey and K. Mardia, A review of image warping methods, *Journal of Applied Statistics*, 25, 155–171, 1998
- ❑ J. Maintz and M. Viergever, A survey of medical image registration, *Medical Image Analysis*, 2(1), 1–36, 1998

