

적응형 바이오 스캐너용 생물정보 분석 프로그램 개발

The development of Analysis Program of Bio Infometrics for Bio-scanner

김용민, 이성운, 구기현, 김진용, 이병돈, 문지영

스타 브이-레이 주식회사

ymkim@starvray.com

본 사업에서 개발하게 될 분석 소프트웨어 역시 bio-scanner와 마찬가지로 여러 칩의 data 분석에 적용할 수 있는 적응적 분석기법을 최적화하는데 있다. 현재까지는 각 칩의 종류에 따라 다양한 통계적 분석기법이 사용되고 있기 때문에 칩의 종류에 관계없이 적용할 수 있는 자동화, 적응적 개념의 분석시스템 개발이 필요한 실정이다. 이 사업에서는 최신경향의 통계적 분석기술에 '자동화', '적응적'이라는 개념을 도입한 분석시스템을 개발함으로써 동시 수행 계획인 Bio-scanner 시스템을 최적화 할 수 있도록 한다.

보편적으로 사용되는 cDNA칩 분석 툴은 대부분 수입에 의존하고 있는 것이 현실이며, 개인이 구입하기에는 가격대가 고가이다. 본 개발에서 수행하는 DNA Chip 분석 소프트웨어 개발은, 기존의 분석 툴이 가지고 있는 gridding, segmentation, normalization 기능을 포함 할 뿐만 아니라 특히, BAC 분석 부분에서 (cancer 진단용 칩) 유전자 copy number를 기준으로 분석하는 방법을 적용함으로써 기존에 사용되는 array 분석용 소프트웨어와 차별성을 가진다. 또한 분석결과를 재가공 없이 의료분야의 임상자료로 이용할 수 있는 GUI (graphical user interface)를 구현함으로써 독창적인 기능까지 보유한다. 즉, 국산화 개발 시 기존의 툴이 가지고 있는 기능을 유지하면서 개발에서 수행하려는 임상진단 DNA chip 분석에 최적화된 기능을 포함한다는 점과, 저렴한 가격으로 국내 연구시장에 범용적으로 활용 될 수 있는 중요성을 가진다.

Bio-scanner 운영 소프트웨어는 하드웨어에서 영상을 얻어내는 image grabbing 역할을 한다. 지금까지는 DNA chip 소프트웨어는 이 운영 소프트웨어와 다음에 설명할 분석 소프트웨어가 따로따로 만들어져 판매되고 있었다. 하지만 본 과제에서는 이 두 가지 소프트웨어가 통합되게 함으로써 사용에 편리한 환경을 주도할 예정이다. 즉 운영 소프트웨어의 Image grabbing module에서 얻어진 data 값은 바로 분석 소프트웨어의 web 연동 database와 연결이 되어 진단결과를 얻을 수 있고 이를 bio-scanner 운영 소프트웨어에 저장할 수 있다.

본 사업에서 개발하게 될 bio-scanner는 칩의 종류와 dye의 종류에 구애받지 않는 적응형 bio-scanner가 될 것이며 따라서 이를 운영할 수 있는 소프트웨어도 어떠한 환경에도 적용할 수 있어야 한다.

생물정보 분석 프로그램은 바이오 스캐너를 생물학적 영상을 획득하기 위하여 하드웨어를 설정하고 제어하는 부분과 영상을 취득한 후에 획득한 영상에서 생물학적 정보를 추출하는 두 가지 부분을 통합하였다.

하드웨어를 컨트롤하는 부분은 램프의 on/off 제어를 비롯하여 스테이지 컨트롤, 카메라 컨트롤 등 실제 영상을 얻는 과정에서 사용되며 별도의 DLL 형태로 짜여져 있으며, bio-scanner에서 영상을 획득하

기 이전에 하드웨어의 특성에 기인한 여러 가지 불필요한 특성들에 대한 보정 작업을 수행한다. 이러한 과정은 pre-processing이라하며 영상 처리 및 분석의 전단계로서 dark noise correction이나 flat field correction등이 이 과정에서 수행 된다.

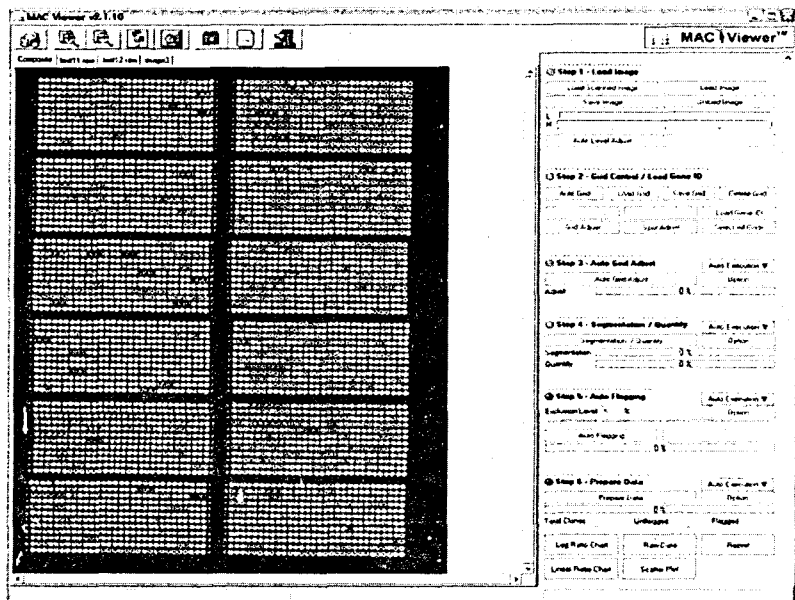
또한 bio-scanner에서 영상을 획득할 때, 최적의 상태를 유지하기 위하여 auto focusing 알고리즘 구현을 구현하였다. 사용된 광학계의 초점심도가 30um 정도로 작기 때문에 auto focusing 중에는 1um 이상의 정밀도를 가지고 조절된다. DNA chip의 경우 스캔하고자 하는 면은 평면에 해당되지만 초점심도를 고려하면 최적화된 영상을 얻기 위해서는 3차원의 제어가 필요하다. 이 초점 정렬은 Cy3 염료(570nm)와 Cy5 염료(670nm)에 대해서 색수차를 고려하여 독립적으로 수행되어야 하며 auto focus의 경우는 자동으로 구현되도록 하였다. 이러한 작업은 스캔이 이루어지기 전에 사전 작업으로 진행되는데 촬영된 영상의 edge 정보를 가지고 가장 sharp한 지점의 encoder 값을 읽은 뒤에 최적화된 3차원의 평면의 방정식을 얻음으로써 이루어 진다.

획득한 영상을 분석하여 생물학 정보를 추출하는 과정에서 scan한 영상에서 spot의 정보를 읽어 내기 전에 사용자가 spot이 있을 만한 영역을 설정하는 것을 grid generation이라 한다. Grid는 그 모양이 다양할 수 있기 때문에 어느 정도의 자유도를 가지고 있다.

형성된 grid에서 spot을 찾아내는 과정을 spot finding이라 한다. 이 과정을 내부 프로그램의 알고리즘에 의해 자동으로 이루어진다. 이러한 알고리즘은 estimation maximization (EM) 알고리즘과 seeded region growing (SRG) 알고리즘이 있는데, 본 연구에서는 이 두 가지를 적절하게 조합하는 방법을 사용하였다.

EM 알고리즘과 SRG 알고리즘은 서로 상반되는 특성을 가지고 있는데, EM 알고리즘은 통계적으로 입증된 방법으로써, 초기 grid 의 중심점이 spot 내에 있지 않아도 spot을 찾아주면서 혹시 spot이 쪼개져 있거나 다른 이물질 등 때문에 spot이 연결되어 있어도 통계적인 확률에 의해 적절히 spot을 찾아 준다. 하지만 계산량이 많고 이물질에 의한 반응점이 클 경우 확률적으로 이 점을 실제 반응점으로 판단할 가능성이 있으며 iterate를 반복하는 과정에서 spot 의 중심점이 초기 지점에서 벗어나면서 전체 grid 의 모양이 흐트러질 가능성이 있다.

Segmentation 및 quantifying 알고리즘은 Pure space-based signal segmentation, Pure intensity-based signal segmentation 등 많은 방법들이 사용되고 있는데, 본 연구에서는 알고리즘의 효율성과 속도가 가장 우수한 Mann-Whitney segmentation과 Trimmed measurement method를 적절히 혼합하여 최상의 알고리즘을 개발하였다. 또한 추후에 실험적인 경험을 바탕으로 중요한 인자에 대해서는 가변 가능하게 함으로써 실험적인 요소를 가미할 수 있도록 하였다.



※본 연구는 산업자원부 차세대 산업기술 개발사업의 지원으로 수행되었으며 지원에 감사드립니다.

