

디지털 병리학 기술현황과 발전방향

I. 서론

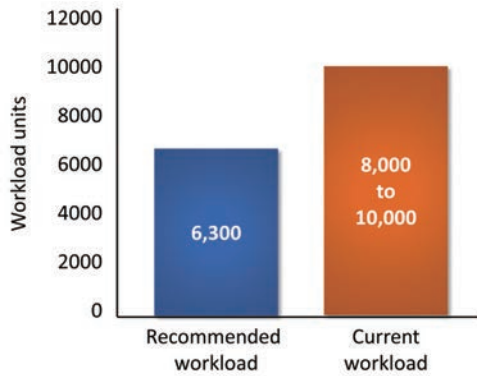
병리학은 질병의 본질적 성질을 연구하는 의학의 한 분야로서, 질병의 원인(병인), 질병 상태의 발전 과정(발병기전), 질병으로 인한 세포, 기관, 전신의 생화학적, 구조적, 기능적 변화, 그리고 이러한 변화의 결과를 연구하는 학문이다. 일반적으로 인체로부터 적출된 세포, 조직, 장기의 표본 샘플을 육안 또는 현미경으로 관찰하여 질병의 유무, 병기 또는 표현형을 판단하게 되며 이는 환자와 질병의 관리 및 치료 방법을 결정하는데 있어 결정적인 근거를 제공하게 된다.

일반적으로 병리학자가 질병이 의심이 되는 환자의 샘플을 검사하고 최종 판단을 내리기까지는 많은 과정을 거치게 된다. 우선, 조직 검사 등을 통해 의심되는 부위의 세포 조직 샘플을 채취하고, 현미경관찰이 가능한 조직표본 제작과정을 거친다. 조직표본 제작은 일련의 화학적 처리과정을 포함하는데 세포 조직 고정(fixation), 탈수(dehydration), 탈 지방 과정, 파라핀 등을 이용한 포매(embedding)와 박절(sectioning)과정 등을 통해 얇은 조직 단면을 얻게 되고 이를 유리 슬라이드에 올려놓고 염색하는 과정을 거치게 된다. 병리학자는 이렇게 준비된 조직표본을 현미경으로 면밀히 관찰하여 질병을 판단하게 한다. 이러한 병리 검사의 과정은 의료기술의 비약적인 발전에도 불구하고 지난 수세기 동안 큰 변화 없이 이어져 오고 있다.

질병의 정확한 진단은 질병의 치료와 관리에 있어 큰 영향을 끼치는데, 병리학적 소견은 병리학자의 정성적, 주관적 판단에 주로 의존하게 된다. 특히, 암 진단에 있어 병리학적 판단은 유일한 진단 수단이지만, 그 안정성에 있어 논쟁의 여지가 있다. 예를 들어, 여러 연구 결과에 따르면 암 진단에 있어 병리학자 관찰자내 (intra-observer) 일치도가 43-78%, 관찰자간 (inter-observer) 일치도가 36-81% 정도



곽진태
세종대학교 컴퓨터공학



〈그림 1〉 병리학자 업무량: 권고 vs. 현재

로 큰 오차를 보이고 있다. 또한, (노령)인구의 증가와 더불어 암 발생건수가 꾸준히 증가하고 있으며^[1-2], 이는 병리학자의 수적 결핍 현상을 불러올 것으로 예상되고 있다 (〈그림 1〉)^[3]. 그리고, 많은 업무량으로 인한 병리학자의 피로는 진단의 정확성과 효율성을 떨어뜨리고 있다.

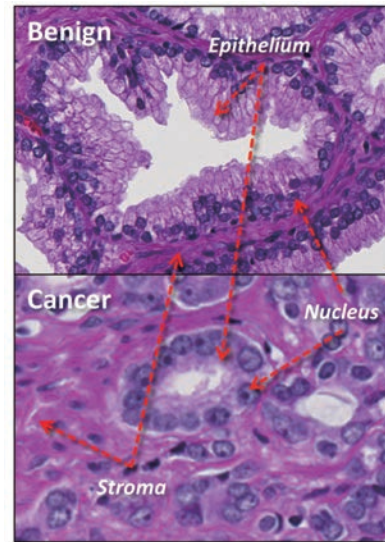
이러한 가운데 디지털 병리학은 기존 병리학을 개선하고 발전시킬 대안으로 각광받고 있다. 디지털 병리학은 세포 조직 표본샘플 영상의 디지털화, 디지털화된 영상의 관찰 및 주석, 관리, 처리, 분석 및 해석 그리고 분석된 결과와 정보의 통합과 공유에 이르게 까지 병리학의 전 범위에 걸친 변화와 변혁을 의미한다. 이는 기존 병리학의 근본적이고 혁명적 변화를 가져올 뿐만 아니라, 관련 의료 서비스의 질 향상에도 크게 기여할 것으로 기대된다.

본고에서는 디지털 병리학의 여러 요소들 중에서 영상 처리와 기계 학습 기술의 동향 및 현안을 알아보고 미래 발전방향에 대해 서술하고자 한다.

II. 본론

1. 디지털 병리학 현황

최근 고성능, 고효율 디지털 스캐너가 보급됨에 따라 많은 병원과 기관에서 조직표본의 디지털화가 일상적으로 이루어지고 있다. 이를 통해 많은 양의 조직표본 디지털 영상 데이터가 확보되고 있으며, 이를 활용하여 좀 더 나은 병리 검사, 진단을 꾀하고자 하는 시도들이 점차 늘어나고 있다. 초기에는 비교적 작은 크기의 조직표본 샘플



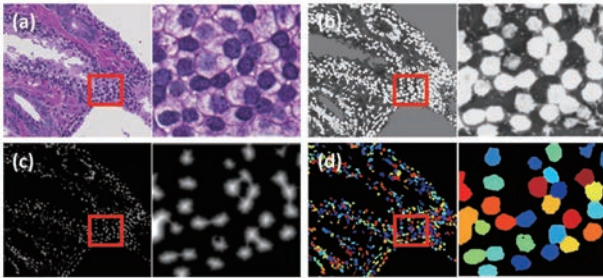
〈그림 2〉 세포 조직 영상: 정상 세포 vs. 암 세포

플의 저배율(〈10배율) 영상을 이용한 연구가 주를 이루었지만, 최근에는 장기 전체 단면에 대한 20배율 또는 40배율의 고배율 디지털 영상을 이용한 연구들이 많이 이루어지고 있다. 이들 연구가 해당 디지털 영상의 크기나 해상도에 영향을 받기도 하지만, 대체로 디지털 병리학에서의 영상 처리 기술과 기계 학습 기술을 활용한 디지털 영상 데이터의 처리 및 분석 과정은 다음과 같이 3단계로 나눌 수 있다.

- 1) 유용한 세포나 조직 개체를 검출하기 위한 영상 분할 (Segmentation) 단계
- 2) 검출된 세포나 개체를 이용한 영상 특징 검출 (Feature Extraction) 단계
- 3) 검출된 영상 특징을 이용한 의사 결정 (Decision Making) 단계

2. 영상 분할 단계

병리 영상 분석에서 가장 기본이 되는 것은 세포나 조직 개체를 분류하는 작업이라 할 수 있다. 상피세포 (Epithelium), 세포기질(Stroma), 세포핵(Nucleus) 등의 개체는 서로 다른 생물학적, 화학적, 기능적 특성을 가지며 이들 개체의 존재 유무, 크기, 모양, 분포 등은 질병의 상태와 성질을 이해하는 데에 핵심이 되는 요소이다 (〈그림 2〉). 예를 들어, 세포핵의 밀도는 암 분화도가 높



〈그림 3〉 Watershed algorithm을 이용한 세포핵 영상 분할

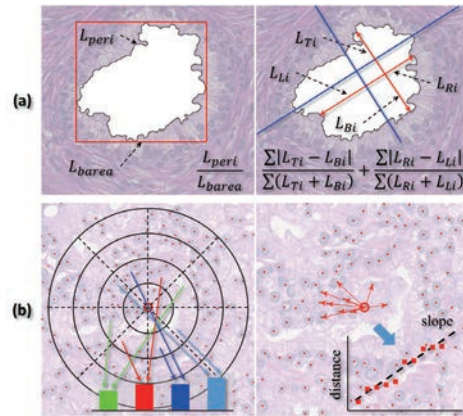
아짐에 따라 증가하는 현상을 보인다. 따라서, 디지털 병리학에서도 많은 경우에 영상 분할을 통해 이들 개체들을 판별해 내는 작업이 선행된다.

염색된 조직표본 샘플 영상에서 상피세포, 세포 기질, 세포핵 등은 서로 다른 색상을 나타낸다. 초기에는 이러한 성질을 이용해서 각 개체의 컬러 차이나 세기를 이용한 영상 분할방법이 주를 이루었는데, 이들에 thresholding, adaptive thresholding, clustering, region-growing^[4] 과 같은 방법들이 있다. 이들 방법들은 영상 데이터 간의 컬러 차이에 크게 영향을 받아 일관된 결과를 보이지 못한다. 이외에 watershed algorithm^[5], active contours/level set^[6] 와 같은 방법들이 적용되었다.

여러 세포 조직 중에서 세포핵은 여러 개체들이 겹쳐 보이는 경우가 많아 각각의 개체를 분할하는 것에 많은 연구가 집중되고 있다. 이를 위해 Voting^[7], Sliding Band Filter^[8]를 이용한 방법, 경계선의 오목한 점 (concavity point)을 이용하는 방법^[9] 등 많은 방법들이 제안되었다. 하지만, 영상 데이터 간의 컬러 차이뿐만 아니라, 세포 조직과 세포핵의 무수히 다양한 형태가 존재하고, 애초에 3차원 구조체에서 얻은 2차원 단면 영상이라는 한계로 인해 영상 분할의 정확도를 높이는 것을 어렵게 하고 있다.

3. 영상 특징 검출 단계

병리학자는 조직표본 샘플을 관찰할 때에 세포나 조직 개체들의 형태적 특성과 변화 상태를 관찰하고 이를 기반으로 질병을 판단한다. 디지털 병리학에서도 이러한 기존 병리학의 질병 진단 과정에 기인하여 질병의 탐지와 진단



〈그림 4〉 영상 특징 검출: (a) 형태적 특징 (b) 관계형 특징

에 유용한 영상 특징들을 검출한다. 디지털 병리학에서 쓰이는 영상 특징은 크게 3가지로 나눌 수 있으며, 이들을 얻기 위해서는 세포 조직 개체의 정확한 영상 분할이 필수적이다.

- 1) 세포 조직 개체의 크기와 형태를 이용한 특징 : 단순히 개체의 크기 변화를 이용하는 것에서부터 개체의 형태가 변화하는 양상을 나타내는 다양한 방법들이 사용되고 있다. 대표적으로 개체의 경계선을 이용하는 방법, 개체에 근접한 convex hull 또는 bounding box를 이용하는 방법, 개체의 대칭성을 이용하는 방법 등이 있다^[10]. (〈그림 4〉)
- 2) 텍스처 분석기법을 사용하는 특징 : 동시발생 매트릭스(Co-occurrence matrix), 프랙탈 차원(fractal dimension), LBP(local binary pattern), Wavelet 변환 등 다양한 텍스처 분석 방법들이 이용되고 있다^[5].
- 3) 관계형 특징 : 관계형 특징은 검출된 개체들 간의 공간적 관계성을 나타내는 특징으로, 주로 세포핵의 특성을 파악하는데 사용되고 있다. 개체들 간의 거리에 기반을 두어 상관관계를 파악하거나 그래프 이론에 기반을 둔 방법들이 많이 사용되고 있는데, 이는 세포나 조직의 구조적인 특성을 파악하는 데 유용한 정보를 제공해 주고 있다^{[5][10]}. (〈그림 4〉)

4. 의사결정 단계

디지털화된 조직표본 샘플 영상에서 영상 특징들을 검출하고 이들을 이용해서 표본 샘플의 질병 상태나 정도

를 판단하기 위해 여러 기계 학습 방법들이 적용되어 왔다. 그 중 Naïve Bayes, Neural networks, k-nearest neighbor, logistic regression, random forest, support vector machine이 대표적이다.

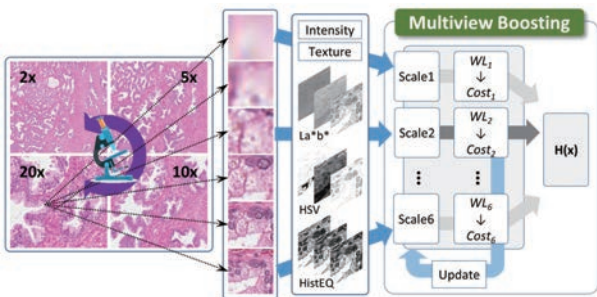
앞서 언급한 영상 특징 검출과 의사결정 결과는 해당 조직표본 샘플을 디지털화 할 때에 설정된 영상의 해상도 또는 배율에 기초하게 된다. 즉, 설정된 해상도에 따라 영상 처리, 분석, 학습이 이루어진다. 하지만, 병리학자는 현미경으로 조직표본 샘플을 관찰할 때에 여러 배율에서 샘플을 관찰하고 그 성질을 파악하고자 한다. 이를 테면, 비교적 낮은 배율에서 세포 조직의 구조적인 특성을 파악하고, 높은 배율에서는 각 개체의 세밀한 변화를 파악한다. 병리학자는 이렇게 여러 배율에서 얻어진 정보들을 종합해서 최종적인 판단을 내리게 된다. 이러한 점을 보완하기 위해 다중 해상도(multi-resolution) 방법들이 고안되었다^[11]. 예를 들어, 고배율에서 얻어진 디지털 영상에서 낮은 배율로 변환해가며 영상 특징들을 검출하고,

고배율 → 저배율 또는 저배율 → 고배율로 진행하며 영상 분류를 하거나, 여러 배율에서 얻어진 영상 특징들을 상호 보완적으로 이용하기 위해 boosting 방법을 사용하기도 한다(〈그림 5〉).

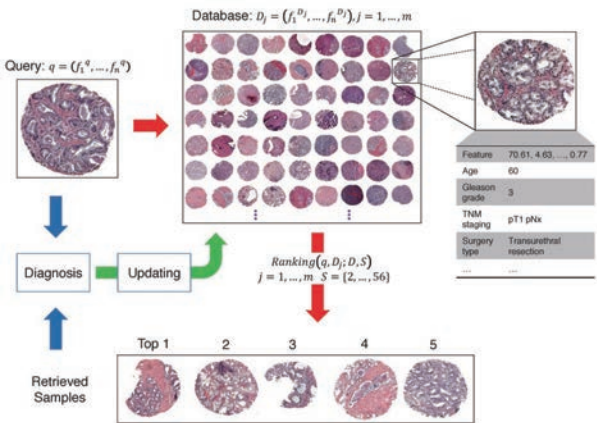
이와 같이 기계 학습 방법을 이용해서 질병의 유무와 상태를 판단할 뿐만 아니라, 치료 후 변화 상태를 예측하거나 (prognosis)^[12] 질병 진단에 도움이 되는 유용한 영상 정보를 검색하거나 추천해주는 내용 기반 세포 조직 영상 검색 (content-based tissue image retrieval) (〈그림 6〉)^[13]과 같은 분야도 활발히 연구가 이루어지고 있다.

III. 문제점

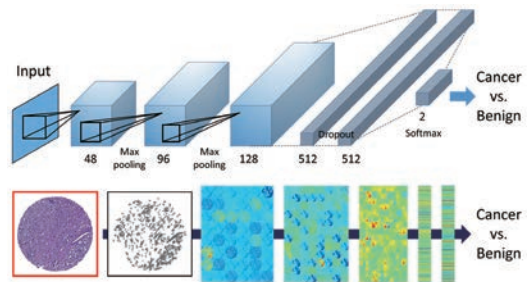
오늘날 디지털 병리학이 직면하고 있는 문제점들은 다음과 같다. 첫째, 디지털화된 영상 데이터간의 차이와 변동성. 종래의 병리학에서 필수적으로 쓰이는 조직표본 샘플의 염색과정을 거치면서 발생하는 병원 간 또는 기관 간 염색의 질적 차이, 업체 간 디지털 스캐너의 차이, 영상 획득 환경의 차이 등으로 인한 영상 데이터의 질적 차이를 보이고 있다. 둘째, 고성능 영상 처리 기술. 최근 들어 장기의 전체 단면 영상을 디지털화하는 whole slide imaging이 빠르게 보급되고 있는데, 이는 각 단면 당 수 기가바이트의 영상 데이터를 발생하고 있다. 이러한 대량의 데이터를 효율적, 안정적으로 처리할 수 있는 기술의 발전이 필요하다. 셋째, 영상 형식과 저장의 표준화. 업체 간 또는 기관 간 영상 데이터의 저장 형식과 방법의 차이로 영상 데이터에 대한 접근성과 활용성을 떨어뜨리고 있다. 넷째, 벤치마크 데이터의 결여. 대규모의 공공 데이터베이스의 부재로 새롭게 개발되는 디지털 병리학 시



〈그림 5〉 다중 해상도 boosting을 이용한 암 세포 검출



〈그림 6〉 세포 조직 영상 검색



〈그림 7〉 암 검출을 위한 Deep Convolutional Neural Networks

스탬의 체계적이고 정확한 검증 및 평가를 어렵게 하고 있다.

IV. 발전 방향

디지털 병리학에서 영상 처리 기술과 기계 학습 기술은 꾸준히 발전해 왔으며, 최근에는 deep learning을 활용한 기술들이 두드러지고 있다. 특히, convolutional neural networks을 이용한 방법들이 많이 제안되어 왔다(그림 7)^[14-16]. Deep learning을 활용한 기술들은 영상 분할, 영상 특징 검출, 의사 결정 단계에 이르기까지 전 방위적으로 영향을 끼치고 있으며, 이미 몇몇 연구들에서는 종래의 기술들을 상회하는 성과들을 보여주고 있다.

지금까지의 디지털 병리학은 주로 염색된 조직표본 샘플의 현미경 영상을 기반으로 발전되어 왔지만, 앞으로는 새로운 영상 획득 기술의 발전을 통해 더 많은 변화와 발전이 있을 것으로 기대된다. 한 가지 예로, 적외선 분광 영상을 이용하여 세포 조직의 화학적 성질을 분석하고 이를 이용해 염색된 세포 조직 영상을 재현해 낼 수 있다^[17]. 이러한 연구는 기존 임상 병리학, 디지털 병리학의 workflow에 큰 변화를 주지 않으면서 의료 기술과 서비스의 질을 향상시킬 수 있는 기술로써 향후 큰 영향을 줄 것으로 기대된다.

효율적이고 정확한 질병의 진단과 관리를 위해 다양한 의료 영상 및 데이터가 생성되고 이용되고 있다. 그 중에서도 비침습적 방법으로 세포, 조직, 장기의 상태와 변화

를 관찰할 수 있는 방사선 영상들(MRI, CT 등)은 세포 조직 샘플의 채취가 필수적인 디지털 병리학의 영상 데이터와 상화 보완적으로 이용될 수 있다^[18]. 더불어 이들 영상 데이터와 유전자 정보를 결합하려는 시도 또한 이루어지고 있다(그림 8). 이러한 연구와 방향은 가용한 모든 정보 및 데이터를 이용하여 보다 정밀한 진단 및 치료를 하고자하는 정밀의료(Precision Medicine)의 패러다임 속에서 점점 더 확장되고 발전 될 것이다.

V. 결론

디지털 병리학은 단순히 디지털 기술의 차용이 아니라, 기존 병리학의 전 과정에 획기적인 변화와 발전을 가져올 디지털 혁명을 의미한다. 급속도로 발전되는 컴퓨팅 기술과 생명 공학 기술과 함께 종래의 병리학이 가지는 많은 문제점들을 해결하고 단점을 보완하여 환자와 질병에 대한 관리, 진단, 치료의 수준을 한 단계 끌어올릴 수 있을 것이다. 이를 위해서는 영상 처리, 분석, 해석, 통합, 공유에 이르기까지 많은 기술적 발전뿐만 아니라, 공학자, 과학자, 그리고 의료 전문가의 협업이 필수적이다. 향후 디지털 병리학에 대한 인식 확대와 함께 안정적인 연구 생태계 형성 그리고 활발한 협업을 통한 우수한 기술의 발전을 기대해 본다.

참고 문헌

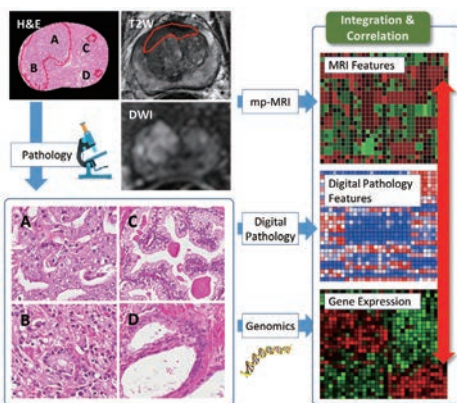
[1] Rebecca L. Siegel, Kimberly D. Miller, and Jemal Ahmedin, "Cancer statistics, 2016", CA: a cancer journal for clinicians, 2016.

[2] 국가 암 정보 센터, 2013년도 통계로 본 암현황, 2013.04.

[3] <http://www.canhealth.com/>

[4] Yahui Peng, Yulei Jiang, Laurie Eisengart, Mark A. Healy, Francis H. Straus, and Ximing J. Yang, "Computer-aided identification of prostatic adenocarcinoma: Segmentation of glandular structures", Journal of Pathology Informatics, 2011.

[5] Scott Doyle, Michael Feldman, Natalie Shih, John Tomaszewski, and Anant Madabhushi, "Cascaded discrimination of normal, abnormal, and confounder classes



〈그림 8〉 다중 영상/데이터 통합과 상관관계 분석



- in histopathology: Gleason grading of prostate cancer”, BMC Bioinformatics, 2012.
- [6] Pingkun Yan, Xiaobo Zhou, and Mubarak Shah, “Automatic segmentation of high-throughput RNAi fluorescent cellular images”, IEEE Transaction on Information Technology in Biomedicine, 2008.
- [7] Bahram Parvin, Qing Yang, Ju Han, Hang Chang, Bjorn Rydberg, and Mary Helen Barcellos-Hoff, “Iterative voting for inference of structural saliency and characterization of subcellular events”, IEEE Transaction on Image Processing, 2007.
- [8] Pedro Quelhas, Monica Marcuzzo, Ana Maria Mendonca, and Aurelio Campilho, “Cell nuclei and cytoplasm joint segmentation using the sliding band filter”, IEEE Transaction on Medical Imaging, 2010.
- [9] Jin Tae Kwak, Stephen M. Hewitt, Sheng Xu, Peter A. Pinto, and Bradford J. Wood, “Nucleus detection using gradient orientation information and linear least squares regression”, SPIE Medical Imaging, 2015, Orlando, FL, USA.
- [10] Jin Tae Kwak, Stephen M. Hewitt, Saurabh Sinha, and Rohit Bhargava, “Multimodal microscopy for automated histologic analysis of prostate cancer”, BMC Cancer, 2011.
- [11] Scott Doyle, Michael Feldman, John Tomaszewski, and Anat Madabhushi, “A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies”, IEEE Transaction on Biomedical Engineering, 2012.
- [12] Andrew H. Beck, Ankur R. Sangoi, Samuel Leung, Robert J. Marinelli, Torsten O. Nielsen, Marc J. van de Vijver, Robert B. West, Matt van de Rijn, and Daphne Koller, “Systematic analysis of breast cancer morphology uncovers stromal features associated with survival”, Science Translational Medicine, 2011.
- [13] Jin Tae Kwak, Stephen M. Hewitt, Andre Balla, Saurabh Sinha, and Rohit Bhargava, “Automated prostate tissue referencing for cancer detection and diagnosis”, BMC Bioinformatics, 2016.
- [14] Geert Litjens, Clara I. Sanchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Irngo Kovacs, Christina Hulsbergen-van de Kaa, Peter Bult, Bram van Ginneken, and Jeroen van der Laak, “Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis”, Scientific Reports, 2016.
- [15] Korsuk Sirinukunwattana, Shan E. Ahmed Raza, Yee-Wah Tsang, David R. J. Snead, Ian A. Cree, and Nasir M. Rajpoot, “Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology”, IEEE Transaction on Medical Imaging, 2016.
- [16] Jun Xu, Xiaofei Luo, Guanhao Wang, Hannah Gilmore, and Anant Madabhushi, “A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images”, Neurocomputing, 2016.
- [17] David Mayerich, Michael J. Walsh, Andre Balla, Partha S. Ray, Stephen M. Hewitt, and Rohit Bhargava, “Stained-less staining for computed histopathology”, Technology, 2015.
- [18] Jin Tae Kwak, Sandeep Sankineni, Sheng Xu, Baris Turkbey, Peter Choyke, Peter Pinto, Maria Merino, and Bradford J. Wood, “Correlation of magnetic resonance imaging with digital histopathology in prostate”, International Journal of Computer Assisted Radiology and Surgery, 2015.



곽진태

- 2005년 2월 고려대학교 전기전자전파공학부(학사)
- 2007년 5월 Purdue University, Electrical and Computer Engineering(석사)
- 2012년 12월 University of Illinois at Urbana-Champaign, Computer Science(박사)
- 2012년 12월~2016년 2월 National Institutes of Health, 박사후연구원
- 2016년 3월~현재 세종대학교, 컴퓨터공학과, 조교수

〈관심분야〉

의료영상처리, 패턴인식, 머신러닝, 딥러닝