

Locality Preserving Projection을 이용한 공초점 라만스펙트럼의 분석

*장근원, 김동국, 민소희, 백성준, 김진영
전남대학교 전자공학과 (dkim@chonnam.ac.kr)

1. 서론

Raman spectroscopy는 의학이나 생물학적으로 관측되는 샘플들의 분자적 구조들에 대한 정보를 제공하는 매우 유용한 spectroscopic 기술이다. Raman 스펙트럼은 근사적으로 $600-1800\text{ cm}^{-1}$ 영역에서 특정 분자들의 성질을 나타내는 일련의 두드러진 대역들을 포함하며, 이는 복잡한 조직 샘플들의 화학적 구성성분을 구별하는데 이용된다. 그러므로 Raman spectroscopy는 의학적인 분석과 피부암과 같은 암진단을 위해 정상 조직과 암 조직을 구별할 수 있는 매우 유용한 도구로써 여겨지고 있다.

최근 들어 많은 연구자들에 의해 라만 스펙트럼을 이용한 BCC (Basal Cell Carcinoma)와 같은 피부암 검출에 관한 연구가 활발히 진행되었다. 근래에는 짧은과장을 이용한 공초점 라만 기법이 BCC 검출에 많이 사용되고 있는데, 이는 배경잡음의 영향을 최소화 할 수 있고 복잡한 처리 없이 BCC와 정상조직을 구별할 수 있는 정보를 제공하여 주고 있다[1].

본 논문에서는 공초점 라만스펙트럼을 사용하여 BCC를 탐지하기 위한 전처리과정들을 비교 분석하였다. 기존에 데이터의 차원을 감소시키기 위해 많이 사용되었던 principal component analysis (PCA) 기법에 더하여 최근에 제안된 locality preserving projection (LPP)기법을 라만스펙트럼의 분석을 위해 이용하였다. 이러한 기법에 근거하여 여러 검출알고리즘을 이용하여 각각의 성능을 비교 실험하였다.

2. 요약

PCA는 데이터의 전체구조를 유지하면서 차원을 감소시킨다. 그러나 실제 많은 분야에서는 데이터

의 local 구조가 매우 중요한 경우가 있다. 이를 위해 locality preserving subspace을 학습하기 위한 LPP알고리즘이 제안되었다. LPP는 데이터의 본래의 구조와 local 구조를 유지하면서 데이터의 차원을 감소한다[2].

라만스펙트럼 신호를 분석하기 위해 10명의 환자로부터 216개를 채취하여 얻은 BCC 조직과 정상조직의 데이터를 실험 샘플로 사용하였다. $20\mu\text{m}$ 로 자른 조직을 공초점 라만 분광기를 이용하여 조직 샘플로부터 깊이방향으로 $30\sim 40\mu\text{m}$ 간격으로 측정하였다. 데이터는 $443-1757\text{cm}^{-1}$ 사이에 분포하였고 1315개의 스펙트럼 데이터를 갖도록 처리하였다. 아래 그림은 BCC와 정상조직에 대한 대표적인 Raman 스펙트럼 신호이다. 여러 스펙트럼에서 정상과 BCC의 차이가 있음을 알 수 있다.

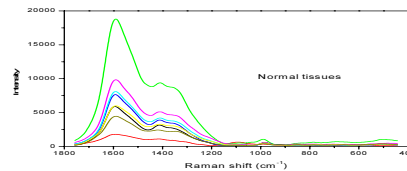


그림 1. 정상조직에 대한 Raman 스펙트럼.

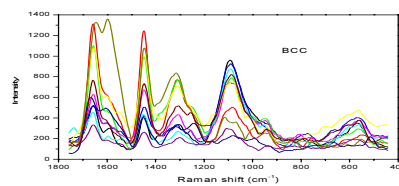


그림 2. BCC에 대한 Raman 스펙트럼.

일반적으로 스펙트럼 데이터는 다른 baseline 값과 다른 dynamic 영역을 갖기 때문에 이를 보정하기 위해 스펙트럼 값을 $[-1,1]$ 사이값으로 정규화 하였다. 정규화된 데이터의 분석을 위해 PCA와 LPP기법들을 적용하여 낮은 차원에서 BCC와 정상조직들의 특징들을 분석하였다. 성능 평가를 위해 cross-validation 과정을 사용하여 실험을 수행하였다. 즉, 학습을 위해 10명의 환자 중 9명은 학

* 본 연구는 산업자원부 지방기술혁신사업 (RT104-03-03) 지원으로 수행되었음.

습을 위해 사용하며, 나머지 1명은 테스트를 위해 사용하고, 위 과정을 10번 반복하였다. 위 과정에서 분석된 데이터는 여러 분류 알고리즘의 입력으로 사용하여 분류성능을 비교하였다.

3. 실험 결과

그림 3과 그림 4는 정규화된 Raman 스펙트럼 신호에 대해 PCA와 LPP를 적용하여 얻어진 1-2번째 주요성분에 대해 결과를 나타낸다. 그림과 같이 PCA에 비해 LPP는 변별력과 local 성질들을 유지하면서 Raman 스펙트럼을 낮은 차원으로 projection 시키므로 해석과 탐지에 매우 유용한 기법임을 알 수 있다.

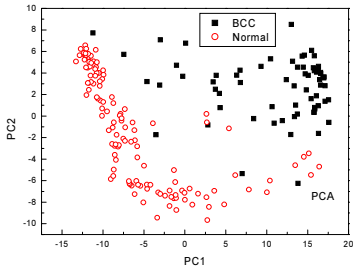


그림3. PCA에 의해 처리된 스펙트럼의 처음 두개의 성분에 대한 분포.

Raman 신호를 분류하기 위한 알고리즘으로 인공 신경망, support vector machine, k-nearest neighbor (kNN)와 알고리즘이 사용되어 진다. 이 실험에서는 kNN을 기법을 이용하여 BCC와 정상에 대한 탐지를 수행하였다.

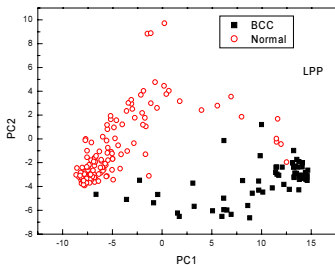


그림4. LPP에 의해 처리된 스펙트럼의 처음 두개의 성분에 대한 분포.

실험결과 PCA와 LPP의 경우는 차원이 각각 8, 10 이고 kNN의 k의 값이 6, 8일때 가장 좋은 성능을

나타내었다. PCA에 비해 LPP를 사용하는 경우 false positive ratio가 3.6%에서 4.3%으로 약간 증가하였으나, 민감도는 93.4%에서 96.1%으로 크게 증가함을 알 수 있었다. 위의 결과로서 LPP가 PCA에 비해 약간의 성능이 향상됨을 알 수 있었다.

표1. PCA와 LPP를 사용한 경우 분류 결과

	PCA		LPP	
	BCC	NOR	BCC	NOR
BCC	93.4%	6.6%	96.1%	3.9%
NOR	3.6%	96.4%	4.3%	95.7%
Dim	8		10	
k	6		8	

참고문헌

[1] J. Choi and et. al., "Direct Observation of Spectral Differences Between Normal and Basal Cell Carcinoma (BCC) Tissues Using Confocal Raman Microscopy," *Biopolymers*, vol. 77, pp. 264-272, 2005.

[2] X. He, S. Yan, Y. Hu, P. Niyogi and H.-J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp.328-340, March 2005.