

기저세포암종 탐지를 위한 피셔척도 필터링 기반 공초점 라만 스펙트럼 분류

Confocal Raman Spectrum Classification Using Fisher Measure based Filtering for Basal Cell Carcinoma Detection

민소희¹, 김진영¹, 백성준¹, 나승유¹, 주재범²

¹ 광주광역시 전남대학교 전자공학과
E-mail: minsh@chonnam.ac.kr
beyondi@chonnam.ac.kr
tozero@chonnam.ac.kr
syna@chonnam.ac.kr

² 안산시 한양대학교 응용화학과
E-mail: jbchoo@hanyang.ac.kr

요 약

본 논문에서는 공초점 라만 스펙트럼을 이용하여 기저세포암종(basal cell carcinoma, BCC)를 검출하는 문제에 대하여 다룬다. 특히, 라만 스펙트럼 측정시 발생하는 잡음 성분 및 변별력이 없는 성분을 제거하기 위한 방법으로 Fisher의 척도를 활용한 스펙트럼 필터링 방법을 제안한다. 특징벡터의 차원을 줄이기 위하여 주축분석방법을 사용하였고, 분류방법으로는 MAP(maximum a posteriori) 방법을 이용하였다. 실험결과 제안한 방법이 특징벡터의 차원을 줄일 수 있을 뿐만 아니라, 검출 결과도 향상됨을 확인하였다

Abstract

This paper deals with a problem of detecting BCC using confocal raman spectrum. Specially, we propose Fisher measure based filtering for rejection of frequency components being noisy or non-discriminative. we use PCA (principal component analysis) for reduction of feature space dimension. Also, we apply MAP detector for classification of BCC raman spectrum. The experimental results shows that our proposed method can reduce the feature dimension and also raise the detection ratio.

Key Words : BCC, Fisher measure, confocal Raman spectrum, PCA

1. 서 론

피부암은 그 자체로는 치명적이지 않지만, 치료하지 않을 경우 인체의 타 기관으로 전이되어 인간의 생명을 위협하는 암이다. 피부암 중에서 basal cell carcinoma (BCC)와 squamous cell carcinoma(SCC)는 흔히 별건되는 피부암의 일종이다. 특히 BCC는 정상조직(Normal tissue; NOR)과 육안으로 구분이 매우 어렵기 때문에, 복잡한 조직검사를 통하여 판별되어진다. 물론, 생검을 통한 조직검사를

하더라도 100% 정확한 진단을 하는 것은 쉬운 일이 아니다.

근래에 라만 스펙트럼을 이용하여 각종 암을 진단하는 시도가 꾸준히 이루어지고 있다[1-6]. 과거에는 FT 라만 스펙트럼을 이용하였으나, 이 경우 형광현상을 줄이기 위해 긴 파장의 레이저를 사용하여 SNR이 좋지 못해 배경잡음 제거가 중요한 문제였다. 최근에는 짧은 파장을 이용한 공초점 라만 기법을 이용하여 BCC를 검출하고자 하는 방법이 제안되었다. 연구결과 공초점 라만 분광기가 배경잡음의 영향을 최소화 할 수 있어, BCC와 NOR 판별이 수월해짐을 보였다.

그러나 여전히 신호의 측정시 발생하는 잡음과, 변별력 없는 주파수 성분들은 BCC와 NOR 검출을 어렵게 하는 원인 중 하나이다. 본 연구에서는 이러한 변별력 없는 성분을 제거하기 위한 방법으로 Fisher 척도 기반의 주파수 필터링 방법을 제안한다. 일반적으로 특징벡터 공간의 차원을 줄이기 위하여 PCA를 이용하게 되는데, 본 연구에서는 전처리 방법으로 Fisher 척도기반 필터링을 도입하여, 변별력 없는 주파수 성분이 PCA 분석에 미치는 영향을 최소화하도록 하였다.

2장에서는 BCC 샘플 및 공초점 라만 측정, 3장에서는 전처리 및 자동식별, 4장에서는 Fisher 척도기반 주파수 필터링 그리고 5장에서는 실험결과에 대하여 설명한다.

2. 실험 샘플 및 공초점 라만 측정

조직 샘플은 일반적인 처리 방법에 따라 준비되었다. BCC 조직은 모두 10명으로부터 일반적인 생검에 의하여 추출되었다. 20 μ m 단면적 조직은 -20 $^{\circ}$ C에서 마이크로톰으로 절단되어, 액체질소 통에 저장되었다. 두 개의 얇은 면으로 절단하였는데, 하나는 자동검출 분류실험을 위하여 사용되었으며, 다른 하나는 조직 검사를 실시하여 BCC와 NOR의 범위를 표시한 후 라만 스펙트럼의 기준으로 사용하였다.

라만스펙트럼 측정을 위하여 Renishaw 2000 공초점 라만 마이크로스코프, 514.5nm의 아르곤 레이저 그리고 20mW의 파워를 사용하였다. 피부 샘플에 대한 공초점 라만스펙트럼의 예를 그림 1에 보였다.

그림 1A에서는 BCC와 NOR과의 확실한 구분을 확인할 수 있으며, 그림 1B의 BCC 스펙트럼은 뚜렷한 특징은 보이지 않지만 BCC로 분류가 가능하다.

그러나 그림 1C의 경우는 BCC와 NOR사이의 구분이 쉽지 않아 보인다. 피부 생검은 피부의

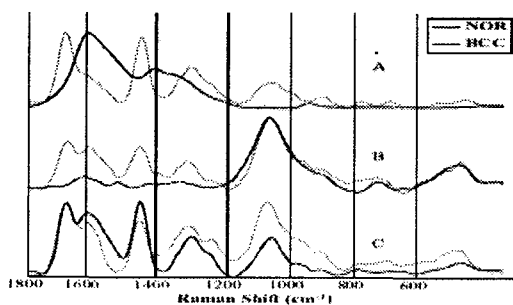


그림 1. 세 명의 환자로부터 얻어진 공초점 라만 스펙트럼 예

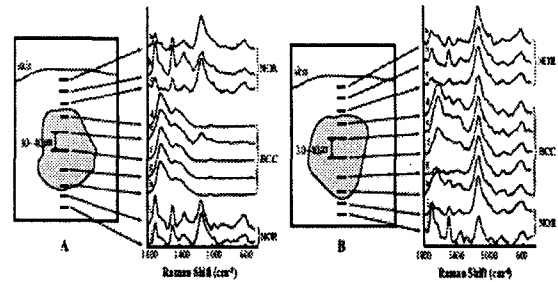


그림 2. 30~40 μ m 간격으로 측정된 공초점 라만 스펙트럼의 프로파일

겉면으로부터 수직방향으로 이루어진다. 그림 2는 30~40 μ m간격으로 라만 스펙트럼측정 값을 보여주고 있다. 이와 같은 방식으로 총 10명의 환자로부터 216개의 라만 스펙트럼이 얻어졌다.

3. 전처리 및 자동검출 방법

본 논문에서 사용하는 BCC 자동검출 방법은 그림 3과 같다. 즉, 측정된 라만 스펙트럼은 신호의 크기를 정규화 하여 비슷한 영역으로 변환한 뒤, PCA를 통하여 특징 벡터 공간의 차원을 감소시킨다. 그리고 BCC 및 NOR 데이터에 대하여 확률 모델을 구축한 후, 새로운 라만 스펙트럼이 들어오면 MAP 분류기에 의하여 NOR과 BCC로 분류하게 된다.

측정되는 라만스펙트럼 신호의 크기가 레이저의 강도, 라만 스펙트럼을 측정하여 적분하는 시간, 그리고 피부의 물리화학적 구성에 따라 강도가 달라지기 때문에 신호의 정규화가 필요하다. 본 논문에서는 정규화를 신호의 최대값으로 나누어 모든 신호가 -1에서 1사이의 값을 갖도록 하는 방법을 이용하였다.

한편 측정된 라만 신호의 차원은 1315로서 매우 크다.

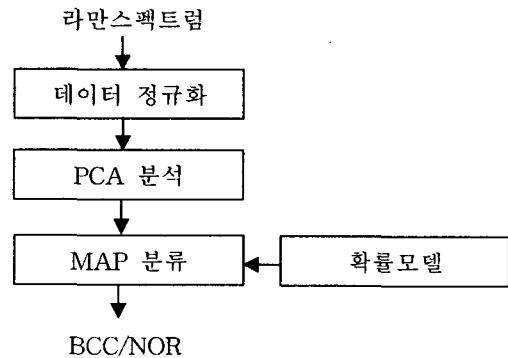


그림 3. BCC 검출 흐름도

이는 300~1800cm⁻¹ 사이의 스펙트럼을 측정 한 것이다. 측정된 라만 신호의 차원이 매우 크기 때문에, 일반적으로 신호의 유효한 축을 얻기 위하여 사용되는 PCA를 신호에 적용하였다. 그런데, 앞에서 기술한 바와 같이 우리는 총 216개의 스펙트럼을 가지고 있으며, 신호의 차원은 1315개 이므로, 공분산 행렬을 이용하는 일반적인 PCA 분석을 적용하기는 힘들다. 따라서 본 논문에서는 SVD(singular value decomposition)을 이용하여 PCA 분석을 수행하였다. PCA 분석 절차는 다음과 같다.

- 1) 모든 스펙트럼의 평균 스펙트럼을 구한다.
- 2) 모든 스펙트럼으로부터 평균 스펙트럼을 제거한다.
- 3) 평균 스펙트럼이 제거된 스펙트럼 신호 행렬로부터 SVD를 적용하여 주축을 얻는다.

다음의 그림 4는 주요한 세 개의 주축을 이용하여 PCA 변환을 한 결과를 보여주고 있다.

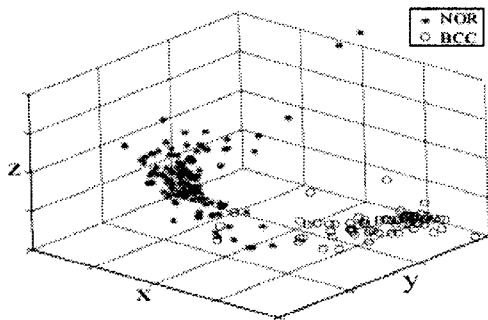


그림 4. PCA 변환 후 NOR과 BCC 분포 (3개의 주축을 이용)

PCA 변환 후 NOR 및 BCC 스펙트럼은 가우시안 확률분포로 모델링 되는데 가우시안 모델 기반 MAP 분류기를 이용한 BCC 검출은 다음과 같다.

Decide w_1 if $g_1(\mathbf{x}) \geq g_2(\mathbf{x})$, where,

$$g_i(\mathbf{x}) = -(1/2) \mathbf{x}^T \Sigma_i^{-1} \mathbf{x} + \Sigma_i^{-1} \mu_i + r_i$$

$$r_i = -(1/2) \mu_i^T \Sigma_i^{-1} \mu_i - (1/2) \ln |\Sigma_i^{-1}|$$

위 식에서 μ_i 와 Σ_i 는 각 BCC와 NOR의 변환 스펙트럼의 평균과 공분산행렬이다.

4. Fisher 척도 기반 주파수 필터링

3장에서 설명한 바와 같이 특징벡터 공간의 차원을 줄이고, 불필요한 신호의 성분을 제거하기 위하여 PCA 분석을 사용하였다. 그런데, PCA 분석은 신호에 부가된 잡음과 변별력 없는 스펙트럼 성분에 의하여 그 성능이 열화될 수 있다. 물론 PCA 자체가 잡음에 해당하는 성분(고유치가 작은 성분들)을 제거하는 기능을 가지고 있기는 하지만, 최적의 성능을 보장하지는 못한다.

따라서 본 논문에서는 Fisher 척도에 기반한 주파수 필터링 방법을 제안하여 변별력 없는 성분을 PCA 분석을 적용하기 이전에 제거하는 방법을 사용하였다.

Fisher 척도를 이용한 두 집합사이의 변별력 있는 특징벡터를 구하는 방법을 LDA (linear discriminant analysis)라고 한다. LDA의 주 개념은 특징 벡터를 변환시킨 후 집합 내의 편차(scatter)는 최대한 줄이고 집합간 거리는 최대한 하는 것이다. 일반적으로 집합 내 편차는 각 집합에서의 공분산 행렬로 표현되며, 집합간 거리는 두 집합간의 평균값의 차이로 표현되어 진다[7]. 즉, 특징 벡터 \mathbf{x}_i 를 변환벡터 \mathbf{a} 에 의하여 1차원 특징으로 변환한다고 하면,

y_i 는 $\mathbf{a}^T \mathbf{x}_i$ 로 표현된다. 분류해야할 두 집합의 변환된 특징을 y_i^j 라고 하자. 여기서 j는 1또는 2이다. $\{y_i^1\}$ 의 평균을 \bar{m}_1 , 표준편차를 \bar{s}_1 이라하고 $\{y_i^2\}$ 의 평균과 표준편차를 \bar{m}_2 및 \bar{s}_2 라고 하자. 그러면 LDA의 문제는 다음의 cost 함수를 최소화하도록 행렬 \mathbf{A} 를 구하는 것이다.

$$J(\mathbf{a}) = \frac{|\bar{m}_1 - \bar{m}_2|^2}{\bar{s}_1^2 + \bar{s}_2^2} \quad (1)$$

본 논문에서는 위 식 (1)을 Fisher 척도라고 정의하였다. 결국 식 (1)로 정의된 cost 함수는 변환된 특징 파라미터 y_i 가 얼마나 변별력 있는지를 나타내주는 척도라고 할 수 있다.

본 연구에서는 위 cost 함수를 각 스펙트럼 성분의 변별력을 계산하는 파라미터로서 사용하였다. 측정된 스펙트럼을 x_i 라고 하자. 그러면 우리는 각 스펙트럼 성분 x_i 에 대하여 cost 함수 $J(x_i)$ 를 계산할 수 있다. 여기서 x_i 는 물론 1차원 값이다. 그러면 $J(x_i)$ 를 Fisher 척도 기반 변별력이라고 정의할 수 있다.

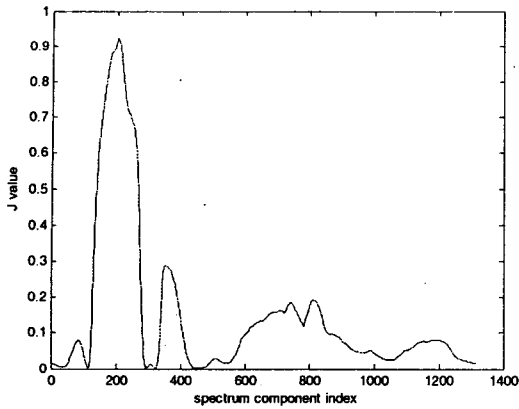


그림 5. 주파수 성분별 변별력

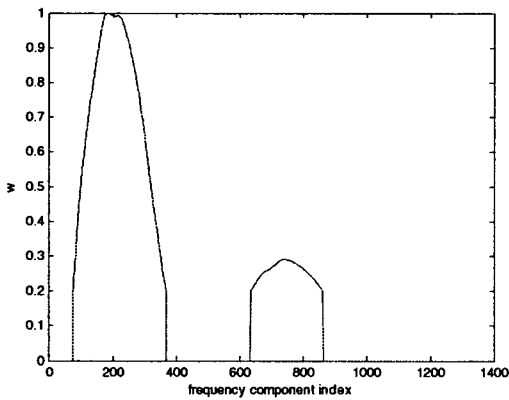


그림 6. 임계값이 0,3인 경우 가중함수

다음의 그림 5는 주파수 성분별로 cost 함수를 계산하여 보인 것이다. 그림 5에서 $J(x_i)$ 이 클수록 NOR과 BCC를 구분하기에 좋은 주파수 성분이라는 것을 의미한다. 계산된 $J(x_i)$ 의 변별력 함수를 이용하여 우리는 원래의 라만 스펙트럼 신호를 가중하는 다음과 같은 방법을 제안하였다. 즉,

$$x'_i = x_i * w_i$$

이때 w_i 는 i 번째 주파수 성분의 가중값이다.

가중값 w_i 는 $J(x_i)$ 의 함수로 표현되는데, 본 논문에서는 다음과 같은 가중 함수를 사용하였다.

$$w_i = \begin{cases} J(x_i)/J_{\max}, & \text{if } J(x_i) > \text{threshold} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

위 식 (2)는 다양한 가중함수를 적용하여 본

후, 가장 성능이 좋은 함수를 선택한 것이다. J_{\max} 는 $J(x_i)$ 의 최대값이다. 다음 그림 6은 임계값이 0.3인 경우의 가중 함수를 보여준다.

5. 실험 결과

본 논문에서 제안한 Fisher 척도 가중함수를 이용한 주파수 성분 필터링의 결과는 PCA를 수행할 경우 고유치의 누적분포를 살펴봄으로써 쉽게 관찰할 수 있을 것이다. 그림 7은 고유치 인덱스 축에 대한 누적 고유치 전력을 나타낸 것이다.

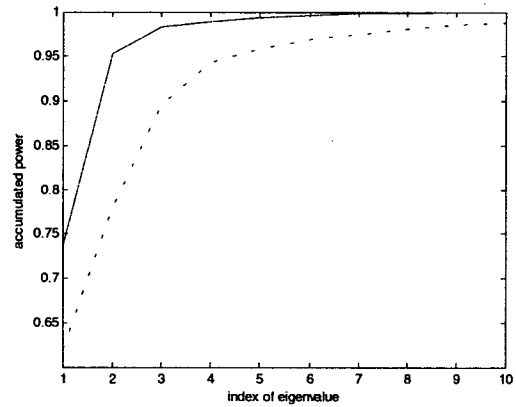


그림 7. Fisher 척도 필터링의 효과

(dotted line : 필터링 전, solid line : 필터링 후)

그림에서와 같이 Fisher 척도 기반 필터링을 적용하는 경우, 분석 시 적은수의 주축이 필요함을 알 수 있다.

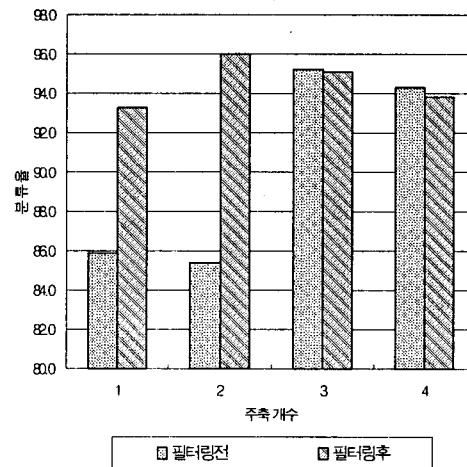


그림 8. 주축개수에 따른 성능비교

그림 8은 Fisher 척도 기반 필터링을 적용하기 전후의 BCC와 NOR의 분류 성능을 보여주고 있다.

본 실험에서는 데이터의 개수가 많지 않기 때문에 9명의 환자로 학습하고 나머지 1명의 환자로 테스트를 하였으며, 총 10번의 실험을 반복하였다. 그림에서 살펴 볼 수 있는 바와 같이 최대성능은 필터링 전, 주축의 개수가 3개 일 때 나타났지만, 필터링 후에는 주축개수 2개에서 최대의 성능을 얻을 수 있었다. 또한 최대 분류율도 95.2%에서 96%로 향상되었다. 한 가지 더 살펴볼 수 있는 것은 주축수 필터링 전에는 사용 주축의 개수에 따라 분류 성능이 크게 변화하였지만, 필터링 후에는 상대적으로 안정화되었음을 알 수 있다.

표 1과 2는 가장 좋은 성능을 보인 경우에 대하여 분류결과를 보인 것이다. 필터링 전의 결과를 보면, NOR과 BCR의 분류율의 차이가 3.6%이었지만, 필터링을 적용함으로써, 2.7%로 그 차이를 줄이고 있음을 살펴볼 수 있어, 인식성능의 단순 향상이 아니라, NOR과 BCC의 각각의 분류성능이 비슷해지고 있다.

표 1. 필터링 전 분류 성능

(*는 전문의에 의한 진단)

	NOR	BCC
NOR*	96.4	3.6
BCC*	7.3	92.8

표 2. 필터링 후 분류 성능

(*는 전문의에 의한 진단)

	NOR	BCC
NOR*	97.1	2.9
BCC*	5.7	94.3

6. 결론

본 논문에서는 공초점 라만 스펙트럼을 이용하여 피부암의 일종인 BCC를 검출하는 문제에 대하여 기술하였다. PCA와 MAP 분류기를 이용하는 기본 시스템의 성능 향상을 위하여 Fisher 척도에 기반 한 스펙트럼 필터링 방법을 제안하였으며, NOR과 BCC의 분류 실험을 통하여 제안한 방법의 타당성을 검증하였다. 즉, 제안한 필터링 방법이 변별력이 없는 스펙트럼 성분을 제거함으로써, 분류기의 성능을 향상시켰다.

향후, MAP 방법뿐 아니라, SVM (support vector machine), 신경회로망, kNN 등 다양한 방법을 적용하여, 성능을 평가해야 할 필요가 있으며, 좀 더 많은 데이터를 확보하여 실험을 해야 하겠다.

참 고 문 헌

- [1] A. Jijssen, T. C. B. Schut, F. Heule, P. J. Caspers, D. P. Hayes, M. H. Neumann, and G. J. Puppels, "Discriminating Basal Cell Carcinoma from its Surrounding Tissue by Raman Spectroscopy," *Journal of Investigative Dermatology*, vol. 119, pp. 64-69, July 2002.
- [2] J. Choi, J. Choo, H. Chung, D.-G. Gweon, J. Park, H. J. Kim, S. Park, C.-H. Oh, "Direct observation of spectral differences between normal and basal cell carcinoma (BCC) tissues using confocal Raman microscopy," *Biopolymers*, vol. 77, pp. 264-272, April 2005.
- [3] S. Sigurdsson, P. A. Philipsen, L. K. Hansen, J. Larsen, M. Gniadecka, and H. C Wulf, "Detection of skin cancer by classification of Raman spectra," *IEEE Trans. on Biomedical Engineering*, vol. 51, pp. 1784-1793, Oct. 2004.
- [4] L. O. Nunes, A. A. Martin, L. Silveira Jr, M. Zampieri, E. Munin, "Biochemical changes between normal and BCC tissue: a FT-Raman study," *Proceedings of the SPIE*, vol. 4955, pp. 546-553, July 2003.
- [5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Son, Inc, 2001.
- [6] M. Gniadecka, H. Wulf, N. Mortensen, O. Nielsen, and D. Christensen, "Diagnosis of Basal Cell Carcinoma by Raman Spectra," *Journal of Raman Spectroscopy*, vol.28, pp. 125-129, 1997.
- [7] S.J. Baek and K.-M. Sung, "Fast KNN Search Algorithm for Nonparametric Classification," *IEE Electronics Letters*, vol. 35, no. 24, pp. 2104-2105, Nov. 2000.