

특이값 분해를 이용한 라만 스펙트럼 고속 탐색 알고리즘

서유경¹, 백성준^{1*}, 고대영¹, 박준규¹, 박아론¹
¹전남대학교 전자컴퓨터공학과

A Fast Search Algorithm for Raman Spectrum using Singular Value Decomposition

Yu-Gyung Seo¹, Sung-June Baek^{1*}, Dae-Young Ko¹, Jun-Kyu Park¹, Aaron Park¹

¹Electronics and Computer Engineering, Chonnam National University

요약 본 논문에서는 라만 스펙트럼의 고속 탐색을 위해 특이값 분해(SVD, Singular Value Decomposition)를 이용한 새로운 탐색 알고리즘들을 제안한다. 제안 알고리즘에서는 SVD를 통해 얻은 특이벡터를 중요도에 따라 선별하여 실험에 사용함으로써 계산량 단축을 도모한다. 파일럿 테스트(Pilot test)를 수행하여 일부 데이터들을 미리 탐색 대상에서 제외시키고 부분탐색법(PDS, Partial Distance Search)을 적용하여 탐색을 수행함으로써 큰 폭으로 계산량을 감소시킨다. 실험에 사용한 데이터 베이스는 총 14,032종의 화학 물질 라만 스펙트럼으로 구성하였으며, 기존의 탐색 방법인 전체탐색법(Full Search), PDS와 평균피라미드탐색법(MPS, Mean Pyramid Search)를 1차원공간상의 신호에 적용하기 적절하게 변형한 1DMPS에 PDS를 적용한 실험(1DMPS+PDS), 데이터의 분산을 내림차순 정렬하여 1DMPS와 PDS를 적용한 실험(1DMPS Sort with Variance+PDS), 데이터의 250차원 성분만 SVD 변환하여 PDS를 적용한 실험(250SVD+PDS), 그리고 제안 알고리즘 PSP(Partial SVD with PDS)와 PSSP(Partial SVD with Sorted Pilot test)을 적용한 실험을 비교 분석하였다. 각 알고리즘의 성능은 곱셈 및 덧셈의 연산량 비교를 통해 이루어졌는데, 실험 결과에 따르면 250SVD+PDS에 비해 제안알고리즘 PSP는 15.7%, PSSP에서는 64.8%의 계산량 감소를 확인하였다.

Abstract In this paper, we propose new search algorithms using SVD(Singular Value Decomposition) for fast search of Raman spectrum. In the proposed algorithms, small number of the eigen vectors obtained by SVD are chosen in accordance with their respective significance to achieve computation reduction. By introducing pilot test, we exclude large number of data from search and then, we apply partial distance search(PDS) for further computation reduction. We prepared 14,032 kinds of chemical Raman spectrum as the library for comparisons. Experiments were carried out with 7 methods, that is Full Search, PDS, 1DMPS modified MPS for applying to 1-dimensional space data with PDS(1DMPS+PDS), 1DMPS with PDS by using descending sorted variance of data(1DMPS Sort with Variance+PDS), 250-dimensional components of the SVD with PDS(250SVD+PDS) and proposed algorithms, PSP and PSSP. For exact comparison of computations, we compared the number of multiplications and additions required for each method. According to the experiments, PSSP algorithm shows 64.8% computation reduction when compared with 250SVD+PDS while PSP shows 157% computation reduction.

Keywords : Fast search algorithm, partial distance search, Raman spectrum, singular value decomposition.

1. 서론

라만 분광법은 미량의 물질로도 해당 물질의 생화학

적 정보를 획득할 수 있다. 또한 유리나 플라스틱 병과 같이 밀폐된 용기 안에 있는 물질들에 대해서도 비 파괴적 측정 방법으로 물질의 화학적 구성에 관한 정보를 얻

이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2013K2A2A)

*Corresponding Author : Sung-June Baek(Chonnam National Univ.)

Tel: +82-62-530-1795 email: tozero@chonnam.ac.kr

Received August 28, 2015

Revised (1st September 14, 2015, 2nd September 23, 2015)

Accepted December 4, 2015

Published December 31, 2015

을 수 있다[1]. 이에 따라 라만 분광법의 사용 범위가 넓어지면서 라만 스펙트럼을 비교 판별하는 고속 탐색 알고리즘에 대한 수요도 점차 증가하고 있다.

가장 기본적인 탐색 방법은 데이터베이스에 있는 모든 데이터와 입력 데이터의 유클리드 거리를 비교해서 가장 적은 거리를 갖는 데이터를 찾아내는 전체탐색법(Full Search)이다. 이 방법은 구형이 단순하지만 비교가 전체 데이터와 이루어지기 때문에 오랜 계산 시간을 필요로 하여 판별 시스템의 응답이 느려지게 된다. 이를 극복하기 위해 고속 탐색 방법에 대한 요구가 생겨났고, 그 결과 다음과 같은 여러 방법이 사용되었다[2-7].

가장 일반적으로 사용되는 고속 탐색 방법으로는 부분탐색법(PDS, Partial Distance Search)과 평균피라미드 탐색법(MPS, Mean Pyramid Structure or Search)이 있다. PDS알고리즘은 매 차원 거리 계산을 할 때마다 현재 계산된 최단거리와 비교하여 이보다 큰 값이 나오는 순간, 계산을 멈추고 다음 데이터로 넘어가는 방법이다. 이는 곱셈의 계산량을 줄여주어 전체적인 계산량을 감소시키지만 그 감소 정도가 완만하다는 한계점을 지지고 있다[2,4].

MPS알고리즘은 입력 데이터의 평균값과 데이터베이스 안에 있는 데이터들 간의 평균값을 비교하여, 이 차이와 최단 거리 사이에 성립하는 관계식을 이용하여 입력 데이터의 평균값이 일정한 값보다 크면 이를 탐색 대상에서 제외시켜 계산량을 줄이는 방법이다. 이 방법은 일반적으로 PDS보다 좋은 성능을 보이는 것으로 알려져 있다.

본 논문에서는 이 방법들을 보다 발전시키기 위해 특이값 분해(SVD, Singular Value Decomposition)를 고속 탐색에 도입하였다. 제안된 방법은 SVD를 PDS와 결합하고 파일럿 검사(Pilot Test)를 통해 불필요한 거리 계산을 크게 감소시키는 방법이다. 실험 결과에 따르면 SVD 변환을 도입하지 않고 PDS와 MPS만을 적용하였을 때보다 제안된 방법이 매우 큰 폭으로 계산량을 감소시킨다는 것을 확인할 수 있었다.

2. 본론

2.1 데이터베이스와 전처리

본 연구에서는 라만 스펙트럼으로 구성된 데이터베

이스를 사용하였다. 데이터베이스는 라만 스펙트럼 14,032개로 이루어져 있으며 이는 Thermo Fisher Scientific Inc.의 라만 스펙트럼 라이브러리로부터 얻은 것이다.

라만 스펙트럼은 일반적으로 측정 환경에 따라 고주파 성분인 가산 잡음(additive noise)과 자가 형광 및 낮은 신호 대 잡음비에 의한 배경 잡음을 포함한다[8]. 이런 배경 잡음은 분류 성능에 직접적인 영향을 미치므로 전처리를 통해 이를 제거하는 것이 필수적이다[9]. 본 연구에서는 기존에 발표된 평활화된 도함수법(Smoothed derivative)[10]을 적용하여 배경 잡음을 제거하였다.

잡음을 제거한 스펙트럼은 본래 스펙트럼과의 편차를 최소화하기 위하여 스펙트럼을 하나의 열벡터로 간주하고 정규화(normalization)를 수행한다. \mathbf{x} 를 잡음이 제거된 스펙트럼이라고 하면 정규화한 단위 벡터 $\hat{\mathbf{x}}$ 는 다음과 같다.

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|} = \frac{\mathbf{x}}{\sqrt{x_1^2 + \dots + x_N^2}} \quad (1)$$

이 때, $\mathbf{x} = \{x_1, x_2, \dots, x_N\}^T$ 는 스펙트럼 행렬의 열벡터이고 N 는 라만 스펙트럼의 차원이다[9]. 본 연구에서는 200-3500 cm^{-1} 에 해당하는 3300차원의 라만 스펙트럼을 사용하였다.

2.1 기존 고속 탐색 알고리즘

2.2.1 PDS(Partial Distance Search)

PDS는 모든 데이터를 탐색하되 데이터의 전체 성분 중 일부만을 계산함으로써 계산량을 줄이는 방법이다. 이 방법은 입력 $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$ 와 데이터베이스에 있는 처음 데이터 \mathbf{y}_1 사이의 유클리드 거리의 제곱 $d^2(\mathbf{x}, \mathbf{y}_1)$ 을 계산하여 d_{\min}^2 으로 설정한다. 다음 데이터 $\mathbf{y}_i = \{y_{i1}, y_{i2}, \dots, y_{iN}\}, i=2$, 에 대해서는 처음부터 K 차원까지의 유클리드 거리의 제곱 누적 합이 d_{\min}^2 보다 크면, j 가 N 에 이르지 않았더라도 거리 계산을 멈추고 다음 데이터로 넘어간다.

$$\sum_{j=1}^K (x_j - y_{ij})^2 \geq d_{\min}^2, 1 \leq K \leq N \quad (2)$$

만일 입력벡터의 N 차원까지 모두 계산한 유클리드 거리의 제곱 누적 합이 d_{\min}^2 보다 작을 경우, 이 값으로 d_{\min}^2 을 갱신한다. 따라서 d_{\min}^2 은 모든 차원의 데이터를

다 계산하지 않고도 항상 최솟값을 지니게 된다[2].

2.2.2 MPS(Mean Pyramids Structure or Search)

이 방법에서는 먼저 입력 데이터 \mathbf{x} 의 평균 피라미드를 구성한다. 가령 입력데이터가 8 차원이라면 2차원씩 평균한 4개의 평균값, 그 다음에는 이를 다시 2개씩 평균한 2개의 평균값을 구하고 최종적으로 이 2개를 평균한 최종 평균값을 구한다. 이때 평균 피라미드는 8-4-2-1로 구성되는데 데이터베이스에 있는 모든 데이터들에 대해서도 마찬가지로 이를 구성한다.

이때 제일 상단 피라미드에 있는 평균값의 유클리드 거리를 $d_0^2(\mathbf{x}, \mathbf{y}_i)$, 그리고 n 번째 층에 있는 평균값의 유클리드 거리를 $d_n^2(\mathbf{x}, \mathbf{y}_i)$ 으로 나타내면 이들 간에는 다음과 같은 관계식이 성립한다.

$$d^2(\mathbf{x}, \mathbf{y}_i) \geq 4d_{n-1}^2(\mathbf{x}, \mathbf{y}_i) \geq \dots \geq 4^n d_0^2(\mathbf{x}, \mathbf{y}_i) \quad (3)$$

거리를 계산할 때 위의 관계식을 적용하면, 만약 $d_n^2(\mathbf{x}, \mathbf{y}_i)$ 이 현재까지 계산된 최소 거리 d_{\min}^2 보다 크다면 $d^2(\mathbf{x}, \mathbf{y}_i)$ 를 계산하지 않고도 이 데이터가 최소 거리를 갖는 데이터가 될 수 없다는 사실을 알 수 있다. 따라서 최하위 레벨까지 차례대로 검사를 실시하여 만약 $d^2(\mathbf{x}, \mathbf{y}_i) < d_{\min}^2$ 이면 그때에만 현재 얻은 $d^2(\mathbf{x}, \mathbf{y}_i)$ 으로 d_{\min}^2 을 갱신하고 그렇지 않은 경우에는 모든 차원의 거리 계산을 다 하지 않고 도중에 거리 계산을 종료하게 되어 계산량에서 이득을 보게 되는 것이다[6,7]. 이 방법은 일반적으로 이미지 코딩과 같은 2차원 공간상의 신호에 적용된다. 본 연구에서는 1차원 공간상의 신호인 라만 스펙트럼을 사용하므로 이에 적합한 1DMPS(1-Dimensional MPS)를 사용한다. 1DMPS는 데이터가 가진 차원의 약수에 한하여 세그먼트 수를 결정한다.

2.3 제안 고속 알고리즘

2.3.1 PSP(Partial SVD with PDS)

앞서 언급했듯이 데이터베이스는 각각 3300차원을 지닌 14032개의 라만 스펙트럼으로 구성되어있다. 따라서 데이터의 모든 차원을 탐색에 사용할 경우에는 그 계산량이 방대할 것이 자명한 일이다. 본 연구에서는 이를 해결하기 위해 데이터의 일부 차원만을 사용하여 탐색이 가능한 방법을 찾았는데 그 중 차원 축소에 적합한 방법

으로 알려진 SVD를 이용하고 이를 기존 고속 탐색 방법인 PDS와 접목시켜 사용하기로 하였다.

SVD에 사용될 데이터 행렬 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ 은 데이터베이스에 있는 모든 데이터를 열벡터로 결합시켜 구한 것이다.

$$\mathbf{X} = \begin{Bmatrix} x_{11}, x_{21}, \dots, x_{M1} \\ x_{12}, x_{22}, \dots, x_{M2} \\ \vdots, \vdots, \dots, \vdots \\ x_{1N}, x_{2N}, \dots, x_{MN} \end{Bmatrix}, \quad \mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (4)$$

이때 $\mathbf{\Sigma}$ 행렬은 SVD의 대각 성분을 나타내는데 특이값 크기순으로 값을 저장하였다. 모든 데이터의 개수 M 은 14032, 각 데이터의 차원 N 은 3300이고. 실험데이터 $\mathbf{z} = \{z_1, z_2, \dots, z_N\}$ 에 특이벡터 \mathbf{U} 를 사용한 다음 식 (5)을 통하면 새로운 변환 데이터 $\mathbf{a} = \{a_1, a_2, \dots, a_K\}$ 를 얻을 수 있다.

$$\mathbf{a} = \mathbf{U}^T \mathbf{z} \quad (5)$$

변환 데이터 \mathbf{a} 는 에너지가 낮은 차수에 집중되어 있으므로 PDS와 결합하면 큰 계산량의 감소를 얻을 수 있다. 이 때 변환 데이터 \mathbf{a} 는 \mathbf{X} 가 아니라 다음과 같이 변환된 $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M\}$ 의 열벡터와 비교하게 된다.

$$\mathbf{B} = \mathbf{U}^T \mathbf{X} \quad (6)$$

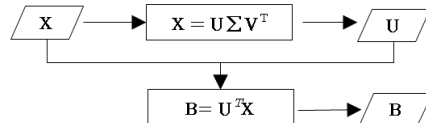


Fig. 1. 변환된 데이터베이스 Flow Chart

이 방법을 적용하면 대부분의 데이터가 매우 낮은 차수에서 탐색을 마치기 때문에 계산량을 크게 줄일 수 있다. 하지만 이 방법은 변환에 소요되는 계산량을 추가로 필요하다는 단점을 지니고 있다. 데이터베이스의 크기가 14032임을 고려할 때, 3300차원의 변환은 전체탐색의 곱셈 계산량에 비해 오버헤드가 약 23%정도이다. 결국 변환에 부가되는 이 추가 계산량으로 인해 이 방법은 기존 탐색 방법인 부분탐색법과 비슷한 수준의 성능밖에 보이지 못하는 것이다.

따라서 보다 나은 성능을 위해서는 데이터 변환 시에 발생하는 오버헤드를 최소화해야 한다. 이는 위 실험을 통해 확인된, 판별성능이 우수한 \mathbf{a} 의 일부 열벡터만으로도 탐색이 가능하다는 사실을 이용하여 해결할 수 있다.

데이터의 차원을 몇 개까지 이용하면 거리 계산이 사

실상 전체 차원의 계산과 거의 같은지를 살펴보기 위해 식(4)의 행렬 Σ 의 특이값을 크기가 큰 순서대로 누적하여 나타내보았다.

$$cumulative\ Ratio(K) = \frac{\sum_{j=1}^K \lambda_j}{\sum_{j=1}^N \lambda_j} \quad (7)$$

식 (7)을 이용하여 나타난 그림 1을 보면 K 가 600번 까까지 특이값의 합이 전체 합의 약 95% 이상을 차지함을 볼 수 있다. 그러므로 식 (4)의 \mathbf{U} 행렬 중 600차원 성분 정도만 사용해도 원래 데이터를 95% 이상 근사할 수 있다는 것을 예상할 수 있다. 실제로 \mathbf{U} 를 600차원부터 점차 차원을 줄여가며 식(5)을 수행하여 변환 데이터 \mathbf{a} 를 얻은 다음, 이에 부분탐색법을 적용한 실험을 진행해보았다. 그 결과 250차원 정도의 성분만 사용하여도 전체탐색을 적용한 경우와 동일한 탐색 결과를 나타냄을 확인하였다. 따라서 본 연구에서는 K 를 250으로 고정하고 \mathbf{U} 의 250차원 성분만을 식 (5)에 적용시켜 얻은 변환 데이터 $\mathbf{a} = \{a_1, a_2, \dots, a_{250}\}$ 를 사용함으로써 데이터 변환 시 발생하는 오버헤드를 최소화하도록 하였다.

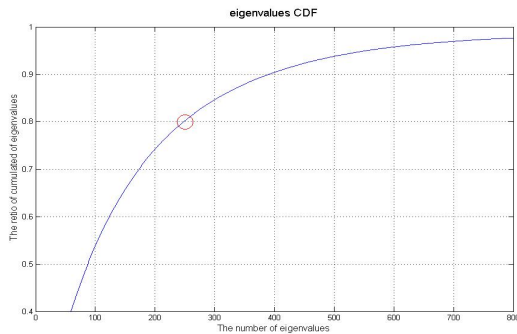


Fig. 2. Cumulative sum of eigenvalues for experimental data

하지만 250차원도 여전히 오버헤드가 크므로 본 연구에서는 다음과 같은 파일럿 테스트를 도입하였다. 먼저 250차원의 20%인 50차원 데이터만을 이용하여 일차 파일럿 테스트를 수행한다. 이 과정에서 거리가 현재 얻은 최소 거리보다 더 먼 거리를 갖는 데이터를 걸러낸다. 여기서 걸러지지 않은 데이터에 한해서 250차원 성분을 사용하여 2차 탐색을 수행한다. 파일럿 테스트와 2차 탐색에는 모두 부분탐색법을 적용하였다. 그리고 계산의 중복을 피하기 위해 2차 탐색은 파일럿 테스트 때 탐색이 수행된 차원 이후부터 수행하도록 하였다.

이 알고리즘은 기존 알고리즘에 비해 파일럿 테스트만으로 탐색이 완료되는 데이터들에 의해 계산량의 감소를 얻을 수 있었으나, 250차원 성분전체를 사용하여 부분탐색법을 적용한 실험과 비교했을 때 감소폭이 크지는 않았다. 이는 250차원에서도 이미 부분탐색 방법이 효과적으로 작용하고 있다는 것을 보여준다. 또한 차원을 250차원으로 제한하게 되면 실험 데이터가 일정 수준 이상의 잡음을 가지고 있는 경우에 탐색 결과가 전체탐색 결과와 상이하게 나타날 가능성이 있어 이를 해결하는 것이 매우 중요한 문제로 대두되었다.

2.3.2 PSSP(Partial SVD with Sorted Pilot test)

본 방법에서는 앞서 도입한 파일럿 테스트와 동일하게 식(5)와 식(6)을 적용해 $\mathbf{a} = \{a_1, a_2, \dots, a_K\}$ 와 $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M\}$ 를 얻어 실험에 사용하였다. 단, K 는 50이고 파일럿 테스트를 다음과 같이 2단계로 분리하였다. 파일럿 테스트 1단계에서는 50차원의 약 20%인 10차원의 데이터만을 이용하여 구한 유클리드거리의 제곱 $d_i^2(\mathbf{a}, \mathbf{b})$ 를 오름차순으로 정렬한다. 파일럿 테스트 2단계에서는 가장 첫 번째로 정렬된 데이터 \mathbf{b}_i 의 변환 전 데이터 \mathbf{x}_i 와 실험 데이터 $\mathbf{z} = \{z_1, z_2, \dots, z_N\}$ 와의 거리를 최초의 최소거리 $d_{\min}^2(\mathbf{a}, \mathbf{b})$ 로 둔다. 이로부터 정렬된 $d_i^2(\mathbf{a}, \mathbf{b})$ 를 순서대로 사용하여 데이터의 50차원에 부분탐색법을 적용하여 거리를 비교한다.

이때 파일럿 테스트를 통과하지 못한 데이터는 최소거리 데이터가 될 수 없으므로 버리고 나머지 데이터에 한해서는 2차 탐색을 진행한다. 이때 앞서의 방법과는 달리 추가로 SVD 변환을 행하지 않고 실험 데이터 $\mathbf{z} = \{z_1, z_2, \dots, z_N\}$ 와 데이터베이스의 \mathbf{x}_i 사이의 $d_i^2(\mathbf{z}, \mathbf{x}_i)$ 를 부분탐색법을 적용하여 비교한다.

이 방법의 장점은 파일럿 테스트 1단계에서 정렬이 이루어지므로 현재 $d_{\min}^2(\mathbf{a}, \mathbf{b})$ 가 파일럿 테스트 1단계에서 저장된 거리 값의 크기보다 커지는 시점 이후에 있는 모든 데이터를 계산에서 제외해도 된다는 점이다. 또한, 이 알고리즘은 앞서의 방법과 달리 2차 탐색에서 실험 데이터의 모든 차원을 사용하기 때문에 실험 데이터에 잡음이 많더라도 전체탐색 결과와 항상 동일한 결과를 보여 PSP의 문제를 해결했다.

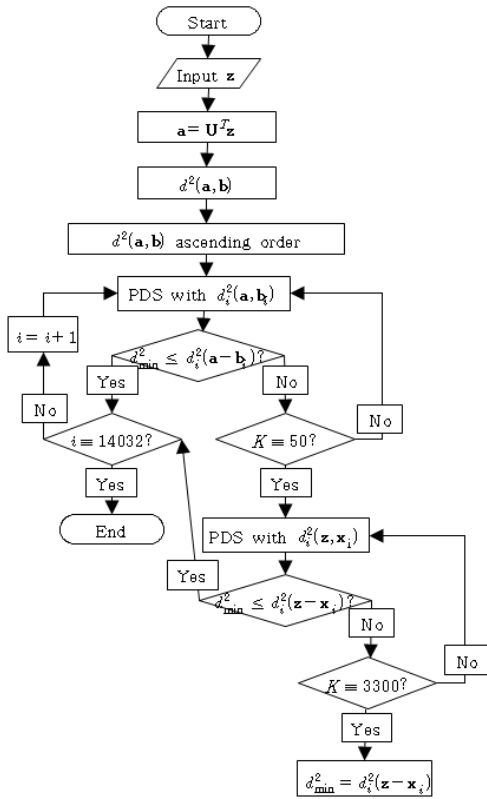


Fig. 3. 알고리즘 PSSP Flow Chart

2.4 실험 방법 및 결과

라만 스펙트럼은 특성상 측정 환경이나 기기에 따라 일정하지 않은 잡음을 포함하게 된다. 이 점을 고려해 실험데이터는 라이브러리에서 500개를 임의로 골라 low noise(40dB), 그리고 high noise(20dB)를 가산한 경우로 준비하였다. 또한, 사용자의 프로세서의 성능에 따라 알고리즘의 속도 측정이 영향을 받을 수 있으므로 정확한 비교를 위해 각 실험별로 필요한 덧셈과 곱셈의 계산량을 비교하였다.

실험에는 총 7가지 알고리즘을 구현하였다. 기존의 전체탐색과 부분탐색 방법, 1DMPS에 부분탐색 방법을 적용한 1DMPS+PDS, 데이터의 분산을 구한 후, 이것을 분할된 데이터의 평균이 큰 값부터 작은 값으로 내림차순 정렬한 1DMPS Sort with Variance+PDS, 250차원 SVD를 사용하여 부분탐색법을 적용시킨 250SVD+PDS, 그리고 제안 알고리즘인 PSP와 PSSP로 구성하였다.

전체탐색방법에 비해 PDS를 적용한 실험과 1DMPS+PDS를 적용한 실험은 각각 19%, 8.7%의 계산

량을 필요로 하였다. 이는 PDS방법이 계산량 감소에 유효하게 작동하고 있다는 점과 1DMPS방법이 계산량을 적절히 감소시킨다는 점을 확인시켜준다. 또한 1DMPS를 보다 효과적으로 적용하기 위해 사용한 1DMPS Sort with Variance+PDS는 1DMPS+PDS에 비해 계산량이 44.8%정도 감소했지만 250SVD+PDS보다는 성능이 좋지 않은 것을 확인할 수 있었다. 따라서 데이터의 분산을 이용하는 것보다 SVD를 통한 차원 축소를 이용하는 것이 고속탐색에 더 적합함을 알 수 있었다.

고속 탐색에서 우수한 성능을 나타낸 PDS와 SVD를 적절히 결합한 제안알고리즘 PSP와 PSSP는 250SVD+PDS에 비해 각각 15.7%, 64.8%의 감소폭을 보여 그 성능을 증명하였다. 그리고 PSP의 성능을 보완시킨 PSSP의 결과 값은 PSP에 비해서도 약 58.3%가 감소한 것을 볼 수 있다.

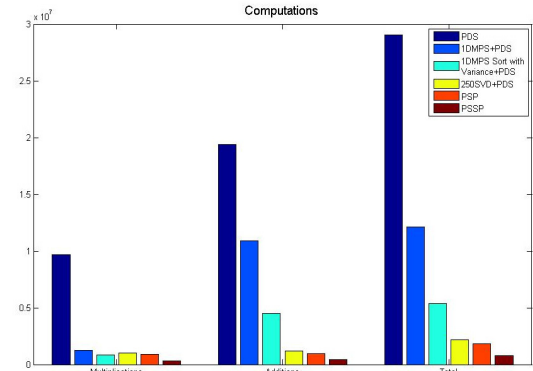


Fig. 4. Average number of multiplications, additions and total

Table 1. Comparison using low noise(40dB) test data

Method	Multiplication	Addition	Total
Full Search	46,553,100	93,092,093	139,645,193
PDS	8,925,861	17,851,721	26,777,583
1DMPS+PDS	1,233,016	10,907,927	12,140,943
1DMPS Sort with Variance+PDS	855,087	4,527,701	5,374,415
250SVD+PDS	994,189	1,163,127	2,157,317
PSP	881,497	937,742	1,819,239
PSSP	312,667	446,177	758,844

Table 2. Computation using high noise(20dB) test data

Method	Multiplication	Addition	Total
Full Search	46,553,100	93,092,093	139,645,193
PDS	9,681,654	19,363,308	29,044,962
1DMPS+PDS	1,233,438	10,908,255	12,141,671
1DMPS Sort with Variance+PDS	855,152	4,528,014	5,383,166
250SVD+PDS	994,970	1,164,690	2,159,660
PSP	882,009	938,767	1,820,776
PSSP	313,316	447,475	760,791

3. 결론

본 논문에서는 파일럿 테스트를 도입함으로써 탐색 알고리즘의 계산량을 큰 폭으로 감소시키는 고속 탐색 알고리즘을 제안하였다. 제안된 알고리즘에서는 제한된 차원의 SVD를 수행하여 얻은 변환 벡터를 이용하여 계산량을 1차적으로 감소시켰다. 변환 벡터를 보다 효과적으로 사용하기 위해 파일럿 테스트를 도입하고 정렬을 통해 불필요한 거리계산을 최소화하였으며 부분탐색 기법을 적용하여 2차 탐색에서도 큰 폭의 계산량 감소를 도모하였다. 실험에 따르면 최종 제안 알고리즘인 PSSP 방법은 전체탐색 방법에 비해 99.5%, 250SVD+PDS방법에 비해서는 58.3%라는 아주 큰 계산량 감소를 얻을 수 있었다.

제안된 알고리즘은 데이터베이스에 SVD를 수행하고 이를 따로 저장하기 위한 추가적인 메모리를 요구한다는 단점이 있지만 동시에 이를 극복할 만큼 큰 계산 감소량을 보여주었다. 제안된 알고리즘은 유클리드 거리가 사용되는 탐색의 경우에는 어디에나 적용될 수 있다는 장점을 지니고 있어, 특히 데이터의 차원이 큰 경우에 유용하므로 라만 스펙트럼 뿐 아니라 IR이나 다른 분광 스펙트럼 분석에도 본 연구에서 제안한 방법을 적용한다면 매우 빠른 계산 결과를 얻을 수 있을 것으로 기대한다.

References

[1] E.L. Izak, "Forensic and homeland security applications of modern potable Raman spectroscopy," *Forensic Sci. Int.* Vol.202, issue 1-3, pp.1-8, Oct, 2010.
DOI: <http://dx.doi.org/10.1016/j.forsciint.2010.03.020>

[2] LINDE, Y., BUZO, A., and GRODY, R.M, "An algorithm for vector quantizer design," *IEEE Trans.*, pp. 84-95, 1980.
DOI: <http://dx.doi.org/10.1109/TCOM.1980.1094577>

[3] GRAY, R.M, "Vector quantization," *Proc. of 2nd IEEE ASSP Mag.*, pp. 4-29, 1984.
DOI: <http://dx.doi.org/10.1109/MASSP.1984.1162229>

[4] S.-J.Baek, D.-R.Lee, B.K.-Jeon, and K.-M.Sung, "A Fast VQ Encoding Algorithm", *The Journal of the Acoustical Society of Korea* v.16, no.2, 95-100, Jan, 1997.

[5] D.-Y.Ko, S.-J.Baek, J.-K.Park, Y.-K.Seo, and S.-I.Seo, "The Fast Search Algorithm for Raman Spectrum", *Journal of the Korea Academia-Industrial cooperation Society* Vol. 16, No. 5 pp. 3378-3384, 2015.
DOI: <http://dx.doi.org/10.5762/KAIS.2015.16.5.3378>

[6] Lee, C.H and Chen, L.H, "A Fast Search Algorithm for Vector Quantization Using Mean Pyramids of Codewords," *IEEE Trans.*, pp. 1697-1702, 1995.
DOI: <http://dx.doi.org/10.1109/26.380218>

[7] S.-W. Ra and J.-K. Kim, "A Fast Mean-Distance-Ordered Partial Codebook," *IEEE Transactions On Circuits And Systems-II: Analog And Digital Signal Processing*, VOL. 40, NO. 9, SEPTEMBER 1993.
DOI: <http://dx.doi.org/10.1109/82.257335>

[8] Z. Jianhua, L. Harvey, M. David and Z. Haishan, "Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy," *Society for Applied Spectroscopy*, Vol.61, pp.248A-270A, Nov. 2007.

[9] A. Park and S.-J.Baek, "A screening of Alzheimer's disease using basis synthesis by singular value decomposition from Raman spectra of platelet", *Journal of the Korea Academia-Industrial cooperation Society* Vol 14, No. 5 pp. 2393-2399, 2013.
DOI: <http://dx.doi.org/10.5762/KAIS.2013.14.5.2393>

[10] S.-J. Baek, A. Park, A. Shen and J. Hu, "A simple background elimination method for Raman spectra," *Chemometrics and Intelligent Laboratory Systems*, vol 98, issue 1, pp. 24-30, Aug. 2009.
DOI: <http://dx.doi.org/10.1016/j.chemolab.2009.04.007>

서 유 경(Yu-Gyung Seo)

[준회원]



- 2014년 2월 : 전남대학교 전자컴퓨터공학부(공학사)
- 2014년 3월 ~ 현재 : 전남대학교 전자공학과(석사과정)

<관심분야>
디지털 신호처리, 패턴인식

백 성 준(Sung-June Baek)

[정회원]



- 1989년 2월 : 서울대학교 전자공학과(공학사)
- 1992년 2월 : 서울대학교 전자공학과(공학석사)
- 1999년 2월 : 서울대학교 전자공학과(공학박사)
- 2002년 3월 ~ 현재 : 전남대학교 전자공학과 교수

<관심분야>

의료, 통신, 음성관련 디지털 신호처리

박 아 론(Aaron Park)

[정회원]



- 2012년 2월 : 전남대학교 전자공학(공학박사)
- 2012년 3월 ~ 2013년 2월 : 전남대학교 유비쿼터스정보가전사업단 Postdoc
- 2013년 3월 ~ 2014년 2월 : 전남대학교 BIT융합기술기반구축사업단 학술연구교수
- 2014년 3월 ~ 2015년 2월 : UC Davis, Department of Surgery, Research Scholar
- 2015년 4월 ~ 현재 : 전남대학교 BIT융합기술기반구축사업단 학술연구교수

<관심분야>

디지털 신호처리, 패턴인식, 바이오응용 패턴인식

고 대 영(Dae-Young Ko)

[준회원]



- 1999년 2월 : 전남대학교 전자공학과 (공학사)
- 2002년 2월 : 전남대학교 전자공학과 (공학석사)
- 2014년 3월 ~ 현재 : 전남대학교 (박사과정)

<관심분야>

디지털 신호처리, 임베디드 시스템

박 준 규(Jun-Kyu Park)

[정회원]



- 2009년 2월 : 전남대학교 전자컴퓨터공학부(공학사)
- 2009년 3월 ~ 현재 : 전남대학교 전자공학과(석박사통합과정)

<관심분야>

디지털 신호처리, 패턴인식