

논문 2014-51-2-19

# 컴퓨터 타임 기반의 다양체 임베딩을 이용한 파형 신호 인식에 관한 연구

( A Study on Classification of Waveforms Using Manifold Embedding  
Based on Commute Time )

한 희 일\*

( Hee-il Hahn<sup>©</sup> )

## 요 약

본 논문에서는 파형 신호에서 패치를 추출하고 이를 패치 그래프로 구성한 다음, 이로부터 대표적인 다양체 임베딩 방식인 컴퓨터 타임 임베딩 기법을 구현하고, 이의 특성을 분석한다. 특히 음성 신호나 악기 음 등, 시간에 따라 스펙트럼이 가변적인 신호를 임베딩하면 스펙트럼의 변화에도 불구하고 그 신호 고유의 기하 구조를 생섬함을 실험으로 확인한다. 다양체 임베딩은 비선형 공간에 놓여 있는 고차원 데이터를 저차원 공간으로의 효율적인 맵을 가능하게 하지만 그래프 구성에 이용된 데이터에 대한 정보만 알 수 있고 그렇지 않은 데이터(out-of-sample data)에 대해서는 정보를 얻기 어렵다. 따라서 다양체 임베딩은 데이터 클러스터링에 적절히 적용 가능하지만, 훈련 과정을 통해 얻은 정보를 기초로 유추 기능이 요구되는 인식 등에는 응용하기 어려운 제약이 따른다. 이를 해결하기 위하여 본 논문에서는 다양체 임베딩이 인식 분야에도 적용 가능하도록 새로운 알고리즘을 제안하고 악기 음 분류 실험을 통하여 그 특성을 분석한다.

## Abstract

In this paper a commute time embedding is implemented by organizing patches according to the graph-based metric, and its properties are investigated via changing the number of nodes on the graph.. It is shown that manifold embedding methods generate the intrinsic geometric structures when waveforms such as speech or music instrumental sound signals are embedded on the low dimensional Euclidean space. Basically manifold embedding algorithms only project the training samples on the graph into an embedding subspace but can not generalize the learning results to test samples. They are very effective for data clustering but are not appropriate for classification or recognition. In this paper a commute time guided transform is adopted to enhance the generalization ability and its performance is analyzed by applying it to the classification of 6 kinds of music instrumental sounds.

**Keywords** : Manifold learning, Commute time, Manifold embedding, Commute time embedding,  
Commute time-guided transform

## I. 서 론

패턴인식, 인공지능, 추론, 데이터 마이닝 등의 기계

학습 분야에서 처리할 데이터 세트를 고차원 공간의 데이터로 간주하고 이를 저차원 공간의 점으로 표현하여 차원을 줄이는 작업(dimensionality reduction)은 기계 학습에서 가장 중요하고 필수적으로 해결하여야 할 분야에 속한다<sup>[1,3,7]</sup>. 데이터의 차원을 줄이기 위한 가장 보편적이고 고전적인 방법인 PCA(principal component analysis)는 대부분의 데이터가 선형 부공간(linear subspace)에 존재할 때 우수한 성능을 보이지만 그렇지 않으면 성능이 급격히 저하되는 단점이 있다. 이를 해

\* 평생회원, 한국의국어대학교 정보통신공학과  
(Department of Information and Communications Engineering, Hankuk University of Foreign Studies)

© Corresponding Author(E-mail: hihahn@hufs.ac.kr)

※ 이 논문은 2013년도 한국의국어대학교 교내 학술연구 지원에 의하여 연구되었음.

접수일자: 2013년11월18일, 수정완료일: 2014년2월3일

결하기 위하여 데이터를 비선형 다양체 위의 점으로 간주하여 이를 저차원 공간으로 임베딩함으로써 기하학적으로 차원을 줄이려는 다양체 학습(manifold learning) 기법이 최근 들어 활발히 연구되고 있다. 다시 말해서, PCA가 선형 공간에서 잘 동작하도록 최적화되어 있듯이 다양체 학습은 비선형 공간에서 데이터의 차원을 잘 줄이도록 설계된 것으로 볼 수 있다. 대표적인 다양체 학습 기법으로는 ISOMAP<sup>[12]</sup>, LLE(locally linear embedding)<sup>[9]</sup>, hLLE (Hessian LLE)<sup>[6]</sup>, 라플라시안 아이젠맵(Laplacian eigenmap : 이하 LE라고 부름)<sup>[11]</sup> 등이 있는데, 특히 LE는 그래프 라플라시안 행렬에서 구한 고유값(eigenvalue)과 고유벡터(eigenvector)를 이용하여 그래프의 구조적인 특성을 분석한다.

최근 들어서는 그래프에서 랜덤워크 기반으로 유사도를 측정하여 클러스터링이나 임베딩 등에 적용하는 연구가 활발히 진행되고 있다. 여기서 가장 핵심적인 부분은 그래프의 각 노드를 연결하는 에지를 따라 정보가 어떻게 확산되는지를 분석하는 일이다. 그래프에서 최단 거리(즉지 거리)와는 달리, 두 노드 간의 컴뮷트 타임은 두 노드를 연결하는 경로가 많을수록 감소한다. 따라서, 동일한 클러스터에 속한 노드 간의 컴뮷트 타임은 작은 값을 갖는 반면에, 서로 다른 클러스터에 속한 노드 사이에는 매우 큰 값을 가지는 특성이 있다. Saerens<sup>[10]</sup>는 컴뮷트 타임이 거리함수로 이용될 수 있고 이에 기반한 임베딩은 PCA로 해석될 수 있음을 보여 주었다. Taylor<sup>[11]</sup>는 이미지나 파형 신호에서 추출한 패치를 이용하여 그래프를 구성하고 각 패치에서 구한 그레이던트 벡터의 크기에 따라 패치를 두 개의 군으로 분류한 다음 각 군 내에서와 군 간의 컴뮷트 타임의 통계적 특성을 분석하였다. 컴뮷트 타임의 이러한 특성을 이용하여 Qiu<sup>[8]</sup>는 그래프 라플라시안이 이산 그린 함수의 역임을 증명하고 이를 이용하여 컴뮷트 타임을 계산하는 방법을 유도하였으며, 영상 분할과 다 물체 움직임 추적 등에 적용하기 위한 기법을 제안하였다.

본 논문에서는 주어진 파형 신호에서 서로 겹치도록 패치를 추출하고 이들을 그래프로 구성한 다음 컴뮷트 타임 임베딩(commute time embedding : 이하 CTE라 부름)을 수행하면 그 신호에 고유한 기하구조를 생성함을 보이고 이들의 특성을 분석한다. 특히, 다양체 임베딩은 그래프를 구성하는데 이용된 데이터에 대해서는 이들 간의 거리, 연결 관계 등의 정보를 구할 수 있으나, 그렇지 않은 데이터(out-of-sample-data)에 대한 정보는 전혀 알 수 없어서 클러스터링에는 널리 이용되지

만 인식 등에는 응용하기 어려운 한계가 있다. 본 논문에서는 컴뮷트 타임 유도 변환(commute time guided transform)<sup>[5]</sup>을 채택하여 이러한 문제점을 해결하는 방안을 제시한다.

본 논문의 구성은 다음과 같다. II장에서는 대표적인 다양체 임베딩 방식인 CTE를 리뷰하고 III장에서는 파형 신호에서 패치를 추출하고 이로부터 패치 그래프를 구성하는 방법을 설명한다. 패치 그래프를 통하여 CTE를 수행하여 이들이 신호 고유의 기하구조를 생성함을 IV장에 제시하고, CTE가 파형 인식에도 응용될 수 있도록 새로운 알고리즘을 제안하고 실험을 통하여 그 가능성을 확인한다. 마지막으로 V장에서는 결론을 맺고 향후 연구 진행방향에 대하여 논의한다.

## II. 그래프 라플라시안을 이용한 다양체 임베딩

거리공간  $X$ 에서 거리공간  $Y$ 의 부공간  $S$ 로 매핑하는 위상동형(homeomorphism)이 존재하면 이를 임베딩이라고 부르는데, 여기서는 다양체 위의 점(그래프의 노드)을 유클리드 공간의 점으로 좌표화하는 작업을 말한다. 이를 위해서는 그래프 라플라시안의 고유값과 고유벡터들을 구하여야 하는데, 이 고유값들을 그래프의 스펙트럼이라고 부른다. 다음에서는 그래프 라플라시안과 이를 이용한 임베딩 방식에 대하여 설명한다.

### 1. 그래프 라플라시안

그래프는 기본적으로 무향(undirected)이고 가중 행렬  $W$  대칭 즉,  $w(u,v)=w(v,u)$ 이라고 가정한다. 그래프 라플라시안 행렬  $L$ 은  $D-W$ 로 정의되는데, 여기서  $D$ 는 각 원소가  $d_u = \sum_v w(u,v)$ 인 대각행렬이다.  $L$ 은 positive semidefinite이고 그 고유값은 항상 음이 아닌 값을 갖는다. 또한, 그래프가 완전 연결일 때, 0의 고유값이 존재하고 이에 해당되는 고유 벡터는  $\mathbf{1}=[1, \dots, 1]^T$ 이다. 일반적으로 행렬  $L$ 을 다음과 같은 두 가지 방법으로 정규화하기도 한다.

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$$

$$L_{rw} = D^{-1} L = I - D^{-1} W$$

$L_{rw}$ 의 고유값과 고유벡터가 각각  $\lambda, v$  즉,  $L_{rw}v = \lambda v$ 이면  $L_{sym}$ 의 고유값은  $L_{rw}$ 의 것과 동일한  $\lambda$ 이고 해당 고유벡터는  $D^{1/2}v$  즉,  $L_{sym}D^{1/2}v = \lambda D^{1/2}v$ 이 성립한다. 또한

$Lv = \lambda Dv$ 이면  $L_{rw}$ 의 고유값과 고유벡터는 각각  $\lambda, v$ 이다. 이러한 특성을 이용하면 정규 또는 비정규 라플라시안 행렬을 이용하여 그래프를 다양하게 해석할 수 있는 장점이 있다.

## 2. 컴퓨터 타임 임베딩

CTE는 데이터  $x_1, \dots, x_n \in R^p$ 가 주어질 때, 이들의 정보손실을 최소한으로 줄이면서 보다 낮은 차원의 점  $y_1, \dots, y_n \in R^q$ 로 변환시키기 위하여, 스펙트럴 그래프 이론을 이용하여 적절한 맵을 구한다<sup>[1,3]</sup>. 여기서 데이터  $x_1, \dots, x_n$ 는  $R^p$ 에 임베딩되어 있는 비선형 다양체 위의 점이라고 가정한다. 그런데 이 다양체의 메트릭은 대부분 알지 못하기 때문에 각 점 간의 거리를 구하기 어렵다. 하지만 인근 점들 간의 거리는 유클리드 거리로 근사화할 수 있으므로 인근 점들의 연결 구조를 이용하여 서로 멀리 떨어져 있는 점들 간의 거리도 구할 수 있다. 즉, 두 점  $x_i$ 와  $x_j$ 가 서로 유사하다면 각각 이에 대응되는 다양체 위의 점  $y_i$ 와  $y_j$ 는 서로 근거리에 놓여 있다고 볼 수 있다. 여기서 각 점 간의 유사도는 적절히 정의되어 가중치 행렬  $W$ 의 원소로 정량화되어 있다고 가정한다.

그래프 상의 랜덤 워크에서 평균 히팅 타임(average hitting time)  $h(x_i, x_j)$ 는 노드  $x_i$ 에서 출발하여 노드  $x_j$ 로 도달하는데 걸리는 평균 시간을 나타낸다. 두 노드  $x_i$ 와  $x_j$ 간의 컴퓨터 타임  $c(x_i, x_j)$ 은 랜덤 워크가  $x_i$ 에서  $x_j$ 로 이동한 다음 다시  $x_i$ 로 되돌아오는데 소요되는 평균시간으로 정의된다. 즉,  $c(x_i, x_j) = h(x_i, x_j) + h(x_j, x_i)$ 이다. 그래프에서 최단 거리(측지거리)와는 달리, 두 노드 간의 컴퓨터 타임은 두 노드를 연결하는 경로가 많을수록 감소한다. 따라서, 동일한 클러스터에 속한 노드 간의 컴퓨터 타임은 작은 값을 갖는 반면에, 서로 다른 클러스터에 속한 노드 사이에는 매우 큰 컴퓨터 타임을 갖는 특성이 있다.

컴퓨터 타임은 다음과 같이 주어지는 확산거리(diffusion distance)<sup>[4]</sup>와 밀접한 관련이 있다.

$$D_i^2(x_i, x_j) = \|p_i(x_i, \cdot) - p_i(x_j, \cdot)\|^2 = \sum_{k=2}^n \gamma_k^{2t} (v_{ik} - v_{jk})^2$$

여기서  $v_{ik}$ 는  $v_k$ 의  $i$  번째 원소이고  $\gamma_k$ 와  $v_k$ 는  $P = D^{-1}W$ 의  $k$  번째 고유값과 고유벡터 즉,  $Pv_k = \gamma_k v_k$ 이다. 이는  $L_{rw}v_k = (1 - \gamma_k)v_k$ 를 만족시킨다. 확산거리가 주어지면, 두 노드  $x_i$ 와  $x_j$  간의 컴퓨터 타임  $c(x_i, x_j)$ 는 다음과 같이

구할 수 있다<sup>[8, 11]</sup>.

$$\begin{aligned} c(x_i, x_j) &= \text{vol} \sum_{t=0}^{\infty} D_{t/2}^2(x_i, x_j) \\ &= \text{vol} \sum_{k=2}^n \frac{1}{1 - \gamma_k} (v_{i,k} - v_{j,k})^2 \\ &= \text{vol} \sum_{k=2}^n \frac{1}{\hat{\lambda}_k} \left( \frac{\phi_{i,k}}{\sqrt{d_i}} - \frac{\phi_{j,k}}{\sqrt{d_j}} \right)^2 \end{aligned} \quad (1)$$

여기서  $\phi_{i,k}$ 는  $L_{sym}$ 의 고유벡터  $\phi_k$ 의  $i$  번째 원소이고  $\text{vol} = \sum d_i$ ,  $\hat{\lambda}_k = 1 - \gamma_k$ 이다. 또한, 고유값  $\hat{\lambda}_k$ 는 다음과 같이 정렬되어 있다고 가정한다.

$$0 = \hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \dots \leq \hat{\lambda}_n < 2 \quad (2)$$

컴퓨터 타임은 데이터를 임베딩하는데 활용할 수 있다. 위 식 (1)을 관찰하면  $\sqrt{c(x_i, x_j)}$ 는  $R^{n-1}$  상에서 두 벡터 간의 거리로 해석될 수 있다. 다시 말해서  $x_i$ 를 다음과 같이 임베딩하면

$$\Psi: x_i \rightarrow \sqrt{\text{vol}} \left( \frac{\phi_{k,2}}{\sqrt{\hat{\lambda}_2 d_i}}, \dots, \frac{\phi_{k,n}}{\sqrt{\hat{\lambda}_n d_i}} \right) \quad (3)$$

$\sqrt{c(x_i, x_j)}$ 는 그래프의 두 노드  $x_i$ 와  $x_j$ 간의 유클리드 거리로 간주될 수 있는데, 이를 컴퓨터 타임 거리라고 부른다. 여기서,  $L_{sym}$ 의 고유벡터  $\phi_k$ 의 인덱스를 2 부터 시작하는 이유는  $\phi_1$ 에 해당되는 고유값  $\hat{\lambda}_1$ 가 0의 값을 갖기 때문이다. 그런데 임베딩 공간의 차원은 노드 수에 따라 증가하므로 어느 정도의 오차를 허용하는 범위 내에서 차원을  $q$ 로 줄이면 다음과 같이 CTE를 정의할 수 있다<sup>[11]</sup>.

$$\Phi: x_i \rightarrow \sqrt{\text{vol}} \left( \frac{\phi_{k,2}}{\sqrt{\hat{\lambda}_2 d_i}}, \dots, \frac{\phi_{k,q+1}}{\sqrt{\hat{\lambda}_{q+1} d_i}} \right) \quad (4)$$

CTE는 다음과 같은 목적함수를 최소화하는 매핑을 구하는 것이다<sup>[3, 8]</sup>.

$$\epsilon = \frac{\sum_{i,j} \|y_i - y_j\|^2}{\sum_{i,j} y_i^T y_j d_i} = \text{tr} \left( \frac{Y^T L Y}{Y^T D Y} \right)$$

임베딩 문제는 다음 식을 만족하는 해를 구하는 것으로 귀결된다.

$$\hat{Y} = \text{argmin} \text{tr} \left( \frac{Y^T L Y}{Y^T D Y} \right)$$

위 식에서  $Z = D^{1/2} Y$ 으로 치환하여 다시 정리하면

$$\frac{Y^T L Y}{Y^T D Y} = \frac{Z^T D^{-1/2} L D^{-1/2} Z}{Z^T Z} = \frac{Z^T L_{sym} Z}{Z^T Z}$$

이므로  $Z$ 가  $L_{sym}$ 의 가장 작은 고유값에 대응되는 고유벡터일 때  $tr\left(\frac{Z^T L_{sym} Z}{Z^T Z}\right)$ 의 최소값을 얻을 수 있다. 다시 말해서  $\hat{Y} = D^{-1/2} \Gamma$  일 때  $\epsilon$ 은 최소값을 갖는다. 여기서  $\Gamma = [\phi_2, \dots, \phi_n]$ 으로  $\phi_m$ 은  $L_{sym}$ 의  $m$  번째 고유값  $\hat{\lambda}_m$ 에 대응하는 고유벡터이다. 이는  $Y = \sqrt{vol} D^{-1/2} \Gamma(\hat{\Lambda}^\dagger)^{1/2}$ 으로 재 정의하여도 동일한 결과를 얻는다. 여기서  $\hat{\Lambda}$ 는  $L_{sym}$ 의 고유값  $\hat{\lambda}_m$ 을 대각 원소로 하는 대각행렬이고  $\hat{\Lambda}^\dagger$ 는  $\hat{\Lambda}$ 의 Moore-Penrose 역으로 다음과 같이 정의된다.

$$\hat{\Lambda}^\dagger = \text{diag} \begin{cases} \frac{1}{\hat{\lambda}_i} & \hat{\lambda}_i \neq 0 \\ 0 & \hat{\lambda}_i = 0 \end{cases}$$

이  $Y$ 는 결국 식 (4)의 벡터 행렬 식과 일치하고,  $\epsilon$ 의 최소값은  $tr(\hat{\Lambda})$ 를 갖는다.

### III. 패치 그래프 구성

본 논문에서는 파형 신호에서 추출한 패치를 그래프의 노드로 이용하기 위하여 다음과 같이 신호  $x[n]$ 에서 추출한 벡터로 정의한다<sup>[11]</sup>.

$$\mathbf{x}_n = (x[n], x[n+1], \dots, x[n+p-1])^T \in R^p$$

여기서 인근 패치는 서로 최대한 겹치게 함으로써 이들 간에 상관관계를 크게 갖도록 한다. 이와 같이 추출한 패치들을 모두 모은 패치세트  $\{\mathbf{x}_n, n=1, \dots, N\}$ 가  $p$  차원 상의 비선형 다양체를 이산화한 것으로 가정할 때, 다양체 위의 두 점이 서로 근방에 위치하면 이들 간의 측지 거리는 유클리드 거리로 근사화시킬 수 있다. 하지만 두 점이 서로 멀리 떨어져 있으면, 다양체의 곡률로 인하여 측지거리와 유클리드 거리 사이에는 큰 오차가 발생하여 측지 거리를 측정하는 것이 사실상 불가능하다. 본 논문에서는 이러한 특성을 반영하여  $k$ -NN( $k$ -nearest neighbor)방식<sup>[2]</sup>을 택한다. 이 방식은  $\mathbf{x}_v$ 가  $\mathbf{x}_u$ 에 가장 가까운  $k$  개의 인근 패치에 속하거나  $\mathbf{x}_u$ 가  $\mathbf{x}_v$ 에 가장 가까운  $k$  개의 인근 패치에 속하면  $\mathbf{x}_u$ 와  $\mathbf{x}_v$ 를 연결시킨다. 이 때, 패치 그래프에서 패치  $\mathbf{x}_u$ 와  $\mathbf{x}_v$  간의 가중치  $w(u, v)$ 는 다음과 같이 구한다<sup>[11]</sup>.

$$w(u, v) = \begin{cases} e^{-\frac{\delta(\mathbf{x}_u - \mathbf{x}_v)^2}{\sigma^2}} & \mathbf{x}_u, \mathbf{x}_v: \text{connected} \\ 0 & \text{otherwise} \end{cases}$$

여기서  $\delta(\mathbf{x}_u - \mathbf{x}_v) = \left\| \frac{\mathbf{x}_u}{\|\mathbf{x}_u\|} - \frac{\mathbf{x}_v}{\|\mathbf{x}_v\|} \right\|$ 이다.

$\delta(\mathbf{x}_u - \mathbf{x}_v)$ 는 그 크기로 정규화된 두 패치  $\mathbf{x}_u$ 와  $\mathbf{x}_v$  간의 거리를 나타내고,  $\sigma$ 는  $\delta(\cdot)$ 에 따른  $w(u, v)$ 의 크기를 제어한다. 패치를 그 크기로 정규화하면  $\mathbf{x}_u \in R^p$ 를  $S^{p-1}$ 으로 매핑함으로써 파형 신호의 진폭 변화에 영향을 받지 않도록 함과 동시에 차원을 1 만큼 줄이는 효과를 얻을 수 있다. 여기서  $S^{p-1}$ 은  $p-1$  차원 공(sphere)을 나타낸다. 또한  $\sigma$ 가  $\delta(\cdot)$ 의 값보다 크면  $w(u, v) \approx 1$ 이지만, 그렇지 않으면 급속히 0으로 감소하게 되므로 결국  $\sigma$ 는 패치 그래프의 확산속도를 제어하는 효과가 있다. 반면에  $\sigma$ 가 매우 작으면  $\delta(\cdot)$ 가 작은 값을 갖더라도  $w(u, v) \approx 0$ 이므로 잡음에 매우 민감한 영향을 받는다. 따라서 본 논문에서는  $\sigma$ 를 충분히 큰 값으로 지정한다.

### IV. 실험 및 토론

본 논문에서는 파형 신호에서 패치를 추출하고 이로부터 패치그래프를 구성하여 실험한다. 모든 패치의 차원은 25 차로 고정한다. 본 논문에서는 시각화가 가능하도록 임베딩 차원  $q=3$ 으로 고정하여 임베딩한다.

#### 1. 노드 수에 따른 커뮤니티 타임 임베딩 특성

우선, 패치의 개수가 임베딩에 어떠한 영향을 주는지 알아보기 위하여 chirp 신호를 선정하여 파형 길이를 700, 800, 900, 1,000, 1,400 등으로 발췌하여 각 신호에서 패치를 676, 776, 876, 976, 1,376 개씩 구한 다음 식 (4)의 방식으로 CTE를 수행하여 그 결과를 그림 1에 제시하였다.

이 그림에서는 각 패치에서 그레이던트 벡터의 크기를 각각 구하고 이를 오름 차순으로 정렬하여 중간 값보다 작으면 파란 점으로, 크면 빨간 점으로 나타낸다. 즉, 파란 점은 부드러운 저주파 영역의 패치에 해당되는 반면, 빨간 점은 고주파 영역을 나타낸다. 이 그림에서 알 수 있듯이 패치의 개수가 충분히 많지 않아 패치들 간의 거리가 랜덤하게 분포하면 이에 비례하여 임베딩 공간에서 매핑되는 점들 또한 흐트러지는 현상이 발생한다. 하지만 패치의 수가 증가하여 패치들 간의 거

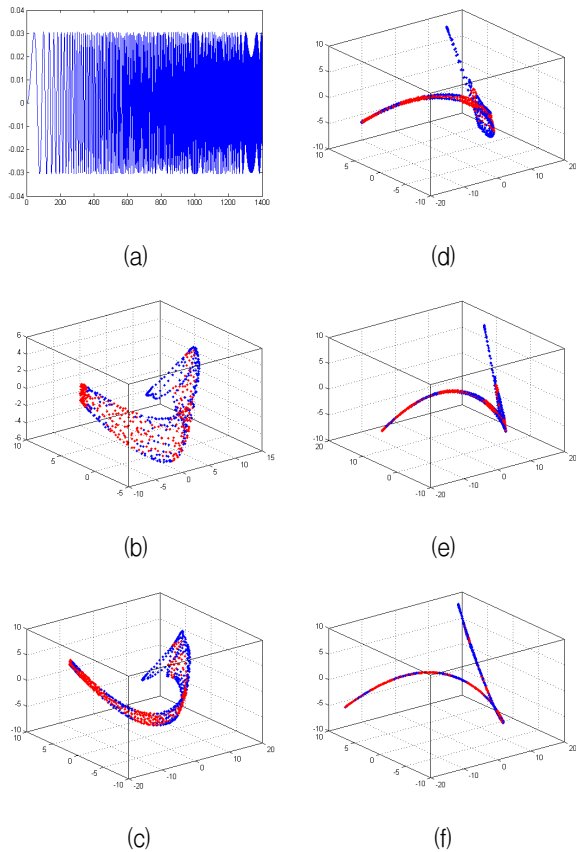


그림 1. chirp 신호의 패치 수에 따른 컴퓨터 타임 임베딩의 변화. (a) chirp 신호 (b) 676 개, (c) 776 개, (d) 876 개, (e) 976 개, (f) 1,376 개

Fig. 1. Evolution of a commute time embedding as the number of patches extracted from a chirp signal increases. (a) a chirp signal, The number of patches are (b) 676, (c) 776, (d) 876, (e) 976, (f) 1,376.

리 분포가 조밀해짐에 따라 임베딩 공간에서는 부드러운 곡선으로 나타나게 된다. 이는 파형 신호를 패치화 함으로써 삼차원 공간 상의 곡선 즉, 그 신호에 고유한 기하 구조로 표현 가능함을 보여 준다.

Luxburg<sup>[13]</sup>의 연구 결과에 의하면 그래프의 노드 수가 증가함에 따라 컴퓨터 타임  $c(v_i, v_j)$ 는  $vol(1/d_i + 1/d_j)$ 으로 수렴한다. 즉,  $c(v_i, v_j)$ 는 노드  $v_i$ 와 노드  $v_j$ 의 연결 특성만을 반영하고 그 외의 노드 간 연결 구조에는 거의 영향을 받지 않기 때문에 그래프의 대역적 특성을 반영하지 못하는 문제점을 드러낸다. 그 그래프에 따라 가변적일 수 있으나 노드 수가 1,000 ~ 2,000 개 이상이면 이러한 현상이 나타나기 시작하는 것으로 알려져 있다. 이와 같은 문제를 피하기 위해 노드의 수를 적절한 수준으로 정하여야 하는데, 본 논문에서는 노드 수를 1,500 개 이하로 제한하였다.

## 2. 임베딩을 이용한 파형 신호 인식 실험

다량체 임베딩을 통하여 파형 신호를 기하 구조로 변환할 수 있음을 확인하였다. 그 기하 구조의 내재적 특성을 파악하기 위하여 바이올린과 플룻 악기 음에서 1,500 개 샘플 길이의 파형 세그먼트를 여러 개 추출하여 이들에 대한 임베딩의 변화 과정을 그림 2와 그림 3에 각각 제시하였다. 이 임베딩 결과는 악기 음의 스펙트럼 변화에도 불구하고 그 신호 고유의 기하 구조는 일정한 패턴을 갖고 있음을 보여 준다. 기본적으로 다량체 임베딩은 비선형 공간에 놓여 있는 고차원 데이터를 저차원 공간으로의 효율적인 맵을 가능하게 하지만 그래프 구성에 이용된 데이터에 대한 정보만 알 수 있고 그렇지 않은 데이터(out-of-sample data)에 대해서는 정보를 얻기 어렵다. 따라서 데이터 클러스터링에는 적절히 적용 가능하지만, 훈련 과정을 통해 얻은 정보를 기초로 유추 기능이 요구되는 인식 등에는 응용하기 어려운 제약이 따른다.

또한, 그림 2와 그림 3의 각 임베딩은 악기 음 별로 유사한 고유 구조를 갖고 있으나 이들 간의 유사도를

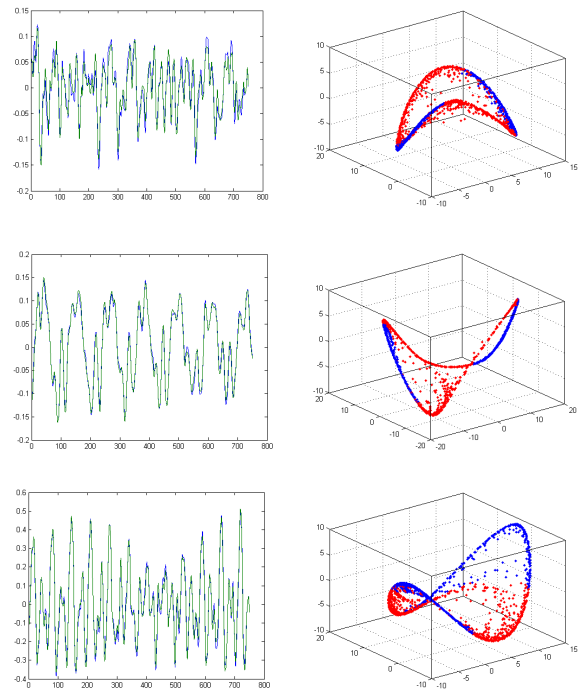


그림 2. 바이올린 악기 음의 스펙트럼 변화에 따른 컴퓨터 타임 임베딩의 변화  
좌열: 파형 신호, 우열: 컴퓨터 타임 임베딩

Fig. 2. The variation of commute time embedding as the spectrum of violin sounds changes. Left: Violin instrumental sounds, Right: Commute time embedding.

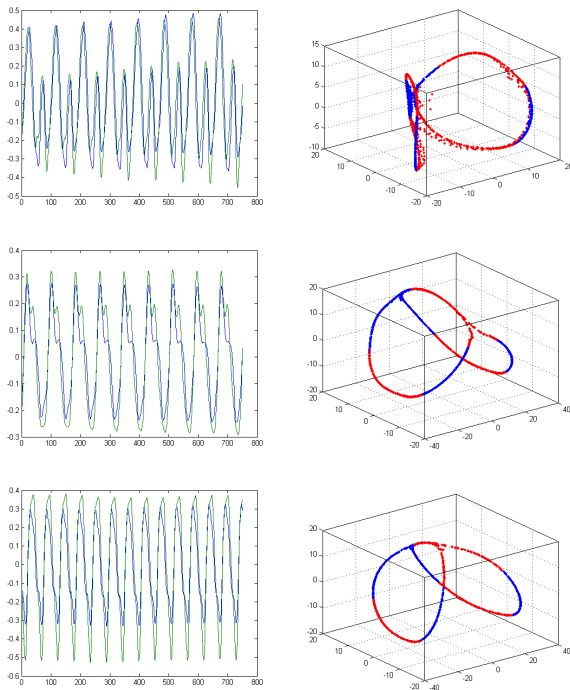


그림 3. 플룻 악기 음의 스펙트럼 변화에 따른 commute 타임 임베딩의 변화

좌열: 파형 신호, 우열: commute 타임 임베딩

Fig. 3. The variation of commute time embedding as the spectrum of flute sounds changes. Left: Violin instrumental sounds, Right: Commute time embedding.

측정하거나 분류하는 것은 매우 어려운 일이다. 이를 해결하기 위하여 본 논문에서는 CTG(commute time guided) 변환<sup>[5]</sup>을 이용하여 다양체 임베딩이 인식 분야에도 적용 가능함을 보인다. 우선, 패치 노드

$$\mathbf{x}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|} \in S^{p-1}$$

를 또 다른 임베딩 공간의 점  $\mathbf{y}_i \in R^q$  으로 매핑하기 위하여 사영 행렬(unitary projection matrix)  $\Omega = R^{p \times q}$  를 정의한다. 즉,  $\mathbf{y}_i = \Omega^T \mathbf{x}_i$  이다. 이 때 최적의  $\Omega$  는

$$\hat{\Omega} = \underset{\Omega^T \mathbf{X} \mathbf{A} \mathbf{X}^T \Omega = I}{\operatorname{argmin}} \sum_{i,j} \frac{\|\Omega^T \mathbf{x}_i - \Omega^T \mathbf{x}_j\|^2}{c(\mathbf{x}_i, \mathbf{x}_j)}$$

으로 주어지는데, 이는 다음 식(generalized eigen decomposition)을 만족하는 고유벡터에 해당된다<sup>[5]</sup>.

$$\mathbf{X}(\mathbf{A} - \mathbf{G})\mathbf{X}^T \Omega = \mathbf{X} \mathbf{A} \mathbf{X}^T \Omega \Lambda$$

여기서,  $\mathbf{X}_i = [\chi_1^i, \dots, \chi_n^i]$ ,  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_K]$ ,  $\mathbf{G} = [g_{ij}] = [1/c(x_i, x_j)]$ ,  $\mathbf{A} = \operatorname{diag} \left[ \sum_j 1/c(x_i, x_j) \right]$  으

로 정의되고,  $\Lambda$ 는 고유값을 원소로 하는 대각 행렬이며  $\chi_k^i$ 의 위 첨자는 각 악기 별 훈련 데이터의 인덱스를, 아래 첨자는 각 훈련 데이터에서 추출한 패치 인덱스를 각각 나타낸다.

본 논문에서는 다양체 임베딩이 인식에 응용될 수 있는지를 확인하기 위하여 바이올린, 첼로, 플룻, 호른, 트롬본, 벨 등 여섯 개의 악기 음을 인식하는 실험을 한다.

각 악기 음 별로  $K$  개의 훈련 데이터가 정해지면  $\mathbf{X}$  는  $1,476K \times 25$  행렬이고  $\Lambda$ 는 25 개의 고유값 중에서 3 개의 최소 고유값만 저장하며  $25 \times 3$  행렬인  $\Omega$ 는 그 고유값에 대응되는 고유벡터를 나타낸다.  $k$  번째 악기 음에서 추출한  $\mathbf{X}_k$ 를 임베딩 공간에 매핑(  $\mathbf{s}_i^k = \Omega_k^T \mathbf{X}_i$  ) 하여 이를  $k$  번째 악기 음의 템플릿으로 이용한다. 이와 같은 훈련 과정이 완료된 후에는 다음 과정을 통하여 입력된 파형 신호를 인식한다. 우선, 입력 파형 신호에서 정규화된 패치 세트  $\mathbf{Y} = [\chi_1, \dots, \chi_N]$ 를 추출하고, 이를 각 악기 음 별 사영 행렬  $\Omega_k$  으로 사영시킨다. 즉, 각  $k$ 에 대하여  $\mathbf{y}_i^k = \Omega_k^T \chi_i$ 를 구하고, 다음 식을 이용하여 템플릿  $\mathbf{s}^k = [\mathbf{s}_1^k, \dots, \mathbf{s}_N^k]$ 와의 오차를 계산한 후, 최소 오차에 해당하는 인덱스  $k$ 를 찾아서 악기 음을 판별한다.

$$\operatorname{argmin}_k \sum_j d_H(\mathbf{y}_j^k, \mathbf{s}^k)$$

여기서,  $d_H(\mathbf{y}_j^k, \mathbf{s}^k) = \min_{i=1, \dots, N} \|\mathbf{y}_j^k - \mathbf{s}_i^k\|$  이다. 본 논문에서는 구현의 편의를 위하여 각 악기 별 훈련 데이터의 수  $K=1$ 로 정하고 시험 데이터를 각각 9 개씩 할당하여 인식 실험을 하였을 때의 실험 결과를 <표 1>에 제시한다.

표 1. 제안 알고리즘의 악기 음 인식률  
Table 1. The classification rate of the proposed method.

입력	바이올린	첼로	플룻	호른	트롬본	벨
바이올린	0.78	0.22				
첼로	0.33	0.56		0.11		
플룻			0.89	0.11		
호른			0.11	0.78	0.11	
트롬본				0.22	0.78	
벨			0.11		0.11	0.78

그림 4는 훈련 바이올린 악기음에서 구한 사영 행렬  $\Omega_{violin}$  를 시험 바이올린 악기 음과 다른 악기 음에 적용 하였을 때의 매핑 결과를 각각 보여 주고 있다. 이 실험 결과를 살펴 보면  $\Omega_{violin}$  를 구하는데 사용되지 않은 바이올린 악기 음에 적용하여도 매핑 결과는 템플리트와 매우 유사한 반면, 플루트이나 호른 등, 다른 악기 음에 적용하면 충분히 구별 가능할 정도로 다른 매핑 결과를 보여 주고 있음을 알 수 있다.

또한, 바이올린 음과 첼로 음에서 오울이 비교적 높게 나타나고 있는데, 이는 그림 5에 제시한 바와 같이, 이 두 악기 음의 기하구조가 매우 유사하기 때문임을 알 수 있다. 즉, CTE는 각 신호를 고유의 기하 구조로 변환시키는 바, 현의 진동에 의해 발생하는 바이올린과

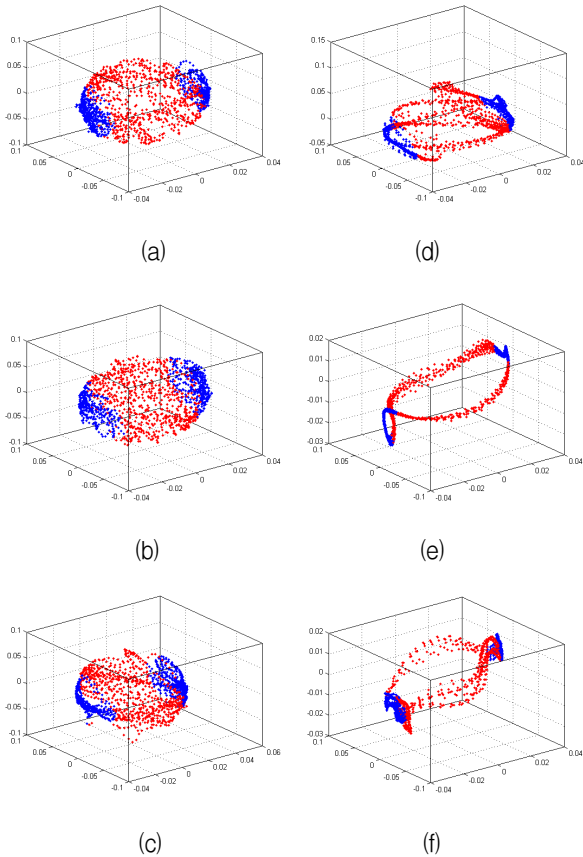


그림 4. 바이올린 악기 음의 사영 행렬  $\Omega_{violin}$  를 여러 시험 악기 음에 적용한 결과  
 (a) 바이올린 템플리트 (b) 바이올린 악기 음 1  
 (c) 바이올린 악기 음 2 (d) 첼로 악기 음  
 (e) 플루트 악기 음 (f) 호른 악기 음  
 Fig. 4. The projection matrix  $\Omega_{violin}$  corresponding to a violin applies to several test waveforms like.  
 (a) a violin template (b) a violin sound 1  
 (c) a violin sound 2 (d) a cello sound  
 (e) a flute sound (f) a horn sound.

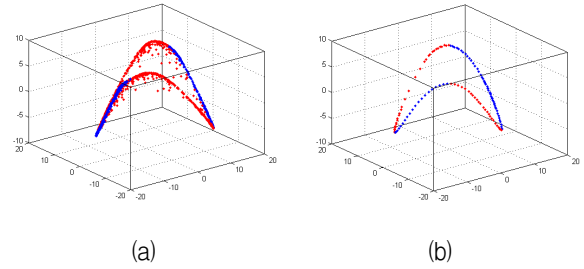


그림 5. 첼로 악기 음과 정현파 신호의 컴퓨터 타임 임베딩 결과 (a) 첼로 악기 음, (b) 정현파  
 Fig. 5. The commute time embedding of (a) a cello instrumental sound, (b) a sinusoidal waveform.

첼로 음은 기본 속성이 정현파적임을 의미한다.

본 논문의 목적은 악기 음 분류의 인식을 제고보다는 클러스터링에 주로 적용되던 다양체 학습 기법을 인식 분야에도 응용할 수 있음을 보이는데 있다는 점에서 본 실험 결과는 그 가능성을 충분히 보여 주고 있다고 판단된다.

#### IV. 결 론

본 논문에서는 파형 신호에서 패치를 추출하고 이를 그래프로 구성한 다음, 이로부터 그래프 라플라시안의 고유벡터를 구하여 각 패치 간의 컴퓨터 타임에 기반한 다양체 임베딩을 구현하고 이의 특성을 분석하였다. 패치의 수를 충분히 확보하면 CTE는 파형 신호를 이에 고유한 기하구조로 변환, 생성함을 실험적으로 보여 주었다. 이 실험 결과를 여러 악기 음에 적용시켜 본 바, 악기 음의 스펙트럼 변화에도 불구하고 임베딩 결과는 그 신호에 고유한 기하 구조를 보존하고 있음을 확인하였다. 하지만, 이들 간의 유사도를 측정하거나 분류하는 것은 매우 어려운 일이다. 이를 해결하기 위하여 본 논문에서는 CTG 변환을 이용하여 다양체 임베딩이 인식 분야에도 적용 가능함을 실험적으로 확인하였다. 인식을 면에서는 타 알고리즘에 비해 성능이 우수하지는 않지만 다양체 임베딩을 통하여 기하학적으로 인식 분야에 활용될 수 있는 가능성을 모색하였다는 점에서 의미가 있다고 판단된다. 향후에는 다양체 임베딩에서 그래프 구성에 참여하지 않은 데이터에 대한 정보 추출 방법을 연구함과 동시에 임베딩 결과를 위상 기하적으로 분류하기 위한 연구를 시행할 계획이다.



## REFERENCES

- [1] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation* 15(6), pp.1373-1396, 2003.
- [2] M. Brito, E. Chavez, A. Quiroz, and J. Yukich, "Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection," *Statistics and Probability Letter*, 1997.
- [3] F. Chung, *Spectral graph theory*, American *Mathematical Society*, 1997.
- [4] R.R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, pp.5-30, 2006.
- [5] Y. Deng, Q. Dai, R. Wang, Z. Zhang, "Commuter time guided transformation for feature extraction," *Computer Vision and Image Understanding* 116(2012) pp. 473-483.
- [6] D.L. Donoho and C. Grimes, "Hessian eigenmaps: new locally linear embedding techniques for high dimensional data," *Proceedings of the National Academy of Sciences*, pp.5591-5596, 2003.
- [7] M. Cho, Y. Lee, and E. Kim, "Dimensionality reduction by manifold learning," *IEEK Fall Conference 2008*.
- [8] H. Qiu and E.R. Hancock, "Clustering and embedding using commute times," *IEEE Trans. PAMI*, Vol. 29, No. 11, Nov., 2007.
- [9] S.T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science* Vol.290, 2000.
- [10] M. Saerens, F. Fouss, L. Yen, and P. Dupont, "The principal component analysis of a graph, and its relationships to spectral clustering," *Machine Learning: ECML 2004*, pp. 371-383.
- [11] K.M. Taylor, "The geometry of signal and image patch-sets," PhD Thesis, University of Colorado, Boulder, Dept. of Applied Mathematics, June 2011.
- [12] J.B. Tenenbaum, V. deSilva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, Vol. 290, pp.2319-2323, 2000.
- [13] U. von Luxburg, A. Radl, and M. Hein, "Getting lost in space: Large sample analysis of the commute distance," *Neural Information Processing Systems (NIPS)*, 2010.

## 저 자 소 개



한 희 일(평생회원)

1984년 서울대학교 제어계측  
공학과 (공학사)

1988년 서울대학교 제어계측  
공학과 (공학석사)

1995년 미국 아리조나 대학교  
전기및컴퓨터공학과  
(공학박사)

1998년~현재 한국외국어대학교 정보통신공학과  
교수.

1987년~1998년 한국전자통신연구원 컴퓨터 연구  
단 선임연구원

<주관심분야 : 신호처리, 컴퓨터비전, 패턴인식>