

잡음 민감성이 향상된 주성분 분석 기법의 비선형 변형

허경용*, 서진석**, 이임건***

A Non-linear Variant of Improved Robust Fuzzy PCA

Gyeongyong Heo *, Jinseok Seo **, Imgeun Lee ***

요 약

주성분 분석(PCA)은 데이터의 차원을 줄이면서 최대의 데이터 변이를 보존하는 기법으로 차원 축소나 특징 추출을 위해 널리 사용되고 있다. 하지만 PCA는 잡음에 민감하며 가우스 분포에 대하여만 유효하다는 단점이 있다. 잡음 민감성의 개선을 위해 다양한 방법이 제시되었고 그 중 퍼지 소속도를 이용한 반복적 최적화 기법인 RF-PCA2가 다른 방법에 비해 우수한 성능을 보였다. 하지만 RF-PCA2는 가우스 분포에만 사용할 수 있는 선형 알고리즘이라는 한계가 있다. 이 논문에서는 RF-PCA2와 커널 주성분 분석(kernel PCA, K-PCA)을 결합하여 가우스 분포 이외의 분포들도 다룰 수 있는 비선형 알고리즘인 improved robust kernel fuzzy PCA (RKF-PCA2)를 제안한다. RKF-PCA2는 RF-PCA2 알고리즘의 잡음 강건성과 K-PCA의 비선형성을 통해 기존 알고리즘에 비해 잡음 민감성이 적으며 가우스 분포 한계를 효과적으로 극복할 수 있다. 이러한 사실은 실험 결과를 통해 확인할 수 있다.

▶ 키워드 : 주성분 분석, 잡음 민감성, 커널 기법, 비선형성, 커널 주성분 분석

Abstract

Principal component analysis (PCA) is a well-known method for dimensionality reduction and feature extraction while maintaining most of the variation in data. Although PCA has been applied in many areas successfully, it is sensitive to outliers and only valid for Gaussian distributions. Several variants of PCA have been proposed to resolve noise sensitivity and, among the variants, improved robust fuzzy PCA (RF-PCA2) demonstrated promising results. RF-PCA, however, is still a linear algorithm that cannot accommodate non-Gaussian distributions. In this paper, a non-linear algorithm that combines RF-PCA2 and kernel PCA (K-PCA), called improved robust kernel fuzzy PCA (RKF-PCA2), is introduced. The kernel methods make it to accommodate non-Gaussian distributions. RKF-PCA2 inherits noise robustness from RF-PCA2 and non-linearity from K-PCA. RKF-PCA2 outperforms previous methods in handling non-Gaussian distributions in a noise robust way. Experimental results also support this.

▶ Keyword : Principal Component Analysis, Noise Sensitivity, Kernel Trick, Non-linear Property, Kernel Principal Component Analysis

• 제1저자 : 허경용, 교신저자 : 이임건

• 투고일 : 2011-01-08, 심사일 : 2011-01-29, 게재확정일 : 2011-02-09

* 동의대학교 영상미디어센터(Visual Media Center, Dongeui University)

** 동의대학교 게임공학과(Dept. of Game Eng., Dongeui University)

*** 동의대학교 영상정보공학과(Dept. of Visual Information Eng., Dongeui University)

※ 이 논문은 한국콘텐츠진흥원 2010년도 문화기술 공동연구센터 사업의 지원에 의해 연구되었음.

I. 서론

국주성분 분석(principal component analysis, PCA)은 차원 축소를 위해 널리 사용되는 방법 중 하나로, 지승 오차 (squared error) 또는 재건 오류(reconstruction error)의 합을 최소화하는 직교 변환이다[1]. PCA는 여러 영역에 성공적으로 사용되었지만, 지승 오차 합을 사용함으로써 잡음에 민감하며 가우시안 분포만을 다룰 수 있다는 문제점이 있다. PCA의 잡음 민감성을 줄이기 위해 여러 가지 방법이 제안되었고, 이들은 부분 집합을 이용하는 방법과 퍼지 소속도(membership)를 이용하는 방법의 두 가지로 크게 나누어볼 수 있다. 부분 집합을 이용하는 방법은 하나 이상의 데이터 부분 집합에 대해 PCA를 수행하고 이들 결과를 이용하여 전체 데이터의 주성분을 구해낸다[2-4]. 퍼지 소속도를 이용하는 방법은 전형적인 데이터 포인트에만 큰 소속도 값을 할당함으로써 잡음 민감성을 줄인다[5-8]. 이들 PCA의 변형 중 improved robust fuzzy PCA (RF-PCA2)는 소속도와 주성분을 반복적으로 최적화하는 과정을 통해 잡음에 민감하지 않은 주성분을 구해내는 방법으로 다른 방법들에 비해 우수한 성능을 보여주었다. 하지만 RF-PCA2 역시 선형 알고리즘으로 가우시안 분포만을 다룰 수 있는 문제점이 있다[8].

가우시안 이외의 분포를 다룰 수 있도록 하기 위한 방법으로는 유클리드 거리가 아닌 다른 거리 척도를 사용하는 방법과 커널을 사용하는 방법이 있다. 유클리드가 아닌 대표적인 거리 척도로는 geodesic distance가 있다[9]. 하지만 이는 거리 계산에 많은 연산이 필요하며 잡음에 민감한 단점이 있다. 커널 기법은 주어진 데이터를 고차원의 특징 공간(feature space)으로 사상하고 사상된 공간에서 선형 알고리즘을 적용함으로써 본래의 입력 공간(input space)에서 비선형 알고리즘을 적용한 결과를 얻는 방법이다[10]. 커널 기법은 support vector machine(SVM)의 성공에 힘입어 다양한 선형 알고리즘을 비선형 알고리즘으로 변환하기 위해 사용되어 왔고[10], PCA 역시 커널 기법을 도입한 kernel PCA (K-PCA)가 존재한다[11]. 하지만 K-PCA는 잡음에 민감한 PCA의 성질을 그대로 가지고 있다.

이 논문에서는 RF-PCA2와 K-PCA를 결합하여 잡음에 강건하면서 비선형인 improved robust kernel fuzzy PCA (RKf-PCA2) 알고리즘을 제안한다. RKf-PCA 알고리즘이 이미 소개되었지만[5] 이는 RF-PCA를 확장한 것으로 RF-PCA가 그런 것처럼 국부 최적해에 수렴함을 보장할 수 없다[8]. 하지만 RKf-PCA2는 RF-PCA2의 확장으로 하나

의 공통된 목적함수를 반복적으로 최적화함으로써 국부 최적해에 수렴함을 보장한다. 또한 RKf-PCA2는 RKf-PCA에 비해 수렴 속도가 빠르며 K-PCA에 비해 잡음의 영향을 적게 받는 등 RF-PCA2의 장점을 그대로 가진다.

이 논문의 구성은 다음과 같다. 다음 장에서는 RF-PCA2를 간략히 소개하고 3장에서는 RF-PCA2와 K-PCA를 이용하여 RKf-PCA2를 유도한다. 실험 결과는 4장에서 보이며 결론 및 향후 연구 방향은 5장에서 언급한다.

II. 향상된 Robust Fuzzy PCA

RF-PCA2는 잡음에 강건한 k 개의 정규직교(orthonormal) 벡터 즉, 주성분(principal component)을 찾아내기 위한 반복 알고리즘이다. 데이터 집합 $X = \{x_1, \dots, x_N\}$ 가 주어진 경우 RF-PCA2의 목적 함수는 식 (1)과 같이 나타낼 수 있다[8].

$$\begin{aligned} \operatorname{argmin}_W J &= \sum_{i=1}^N u_i \| (x_i - \mu_R) - WW^T(x_i - \mu_R) \|^2 \\ &\quad + \sigma^2 \sum_{i=1}^N (u_i \log u_i - u_i) \\ &= \sum_{i=1}^N u_i e(x_i) + \sigma^2 \sum_{i=1}^N (u_i \log u_i - u_i) \end{aligned}$$

s.t. $W^T W = I$ (1)

이 때 u_i 는 데이터 포인트 x_i 의 소속도를, μ_R 은 가중치 평균을, $e(x_i) = \| (x_i - \mu_R) - WW^T(x_i - \mu_R) \|^2$ 는 재건 오류[12]를, W 는 k 개 단위 직교 벡터를 열로 가지는 행렬을, I 는 단위 행렬을 나타내며 σ 는 regularization 상수이다. k 는 재건에 사용될 직교 벡터의 개수를 나타내며 알려진 것으로 가정한다. 식 (1)은 두 개의 항으로 구성되어 있다. 첫 번째 항은 소속도를 가중치로 가지는 재건 오류의 합을 나타내고 두 번째 항은 주성분의 잡음 민감성을 줄이기 위한 regularization 항이다.

먼저 식 (1)을 u_i 에 대해 편미분하면 식 (2)를 얻을 수 있다.

$$\frac{\partial J}{\partial u_i} = e(x_i) + \sigma^2 \log u_i = 0 \dots\dots\dots (2)$$

식 (2)를 u_i 에 대해 정리하면 소속도에 대한 update equation인 식 (3)을 얻을 수 있다.

$$u_i = \exp\left(-\frac{e(x_i)}{\sigma^2}\right) \dots\dots\dots (3)$$

유사하게, 식 (1)을 μ_R 에 대해 편미분하면 식 (4)를 얻을 수 있다. 이 때 $I - WW^T$ 는 역행렬이 존재하여야 한다.

$$\frac{\partial J}{\partial \mu_R} = \sum_{i=1}^N u_i (-x_i + \mu_R) = 0 \dots\dots\dots (4)$$

식 (4)로부터 가중치 평균에 대한 update equation인 식 (5)를 얻을 수 있다.

$$\mu_R = \frac{\sum_{i=1}^N u_i x_i}{\sum_{i=1}^N u_i} \dots\dots\dots (5)$$

가중치 평균을 이용하여 주어진 데이터 집합은 영의 가중치 평균을 가지도록 만들 수 있고, $x'_i = x_i - \mu_R$, 이 때 식 (1)을 최적화하는 W 를 구하는 것은 식 (6)을 최적화하는 것과 같아진다.

$$J' = \sum_{i=1}^N u_i \|x'_i - WW^T x'_i\|^2 \dots\dots\dots (6)$$

$W^T W = I$ 의 제약 조건 하에서 식 (6)을 최소화하는 W 는 식 (7)로 정의되는 가중치 분산 행렬(weighted covariance matrix)에서 k 개의 최대 고유치를 가지는 고유 벡터로 구성할 수 있다.

$$C_R = \frac{1}{N} \sum_{i=1}^N u_i x'_i x'^T_i \dots\dots\dots (7)$$

$$= \frac{1}{N} \sum_{i=1}^N u_i (x_i - \mu_R)(x_i - \mu_R)^T$$

이에 관한 증명은 참고문헌 [8]을 참조하면 된다. RF-PCA2 알고리즘은 그림 1과 같이 요약할 수 있으며, 이 때 ϵ 은 미리 정해지는 상수이고 t_{\max} 는 최대 반복 횟수를 나타낸다.

1. 소속도 $U_0 = [u_1, \dots, u_N]^T = [1, \dots, 1]^T$ 와 카운터 $t = 0$ 를 초기화한다.
2. **repeat**
3. $t \leftarrow t + 1$
4. 평균 μ_R 과 분산 행렬 C_R 을 구한다.
5. C_R 의 고유 벡터로 W_t 를 구성한다.
6. 식 (3)으로 소속도 U_t 를 구한다.
7. **until** $J^{t-1} - J^t < \epsilon$ 또는 $t > t_{\max}$
8. **return**

그림 1. RF-PCA2 알고리즘
Fig. 1. RF-PCA2 algorithm

III. 향상된 Robust Kernel Fuzzy PCA

RF-PCA2는 선형 알고리즘으로 가우시안 분포에 대해서만 유효하다는 단점이 있다. 따라서 이 장에서는 kernel PCA를 이용하여 RF-PCA2를 비선형 알고리즘인 improved robust kernel fuzzy PCA (RKF-PCA2)로 확장한다.

$X_\phi = \{\phi(x_1), \dots, \phi(x_N)\}$ 를 특징 공간으로 사상된 데이터의 집합이라고 하자. 특징 공간에서의 데이터 평균과 공분산 행렬은 식 (8)과 (9)로 나타낼 수 있다.

$$\mu_\phi = \frac{\sum_{i=1}^N u_i \phi(x_i)}{\sum_{i=1}^N u_i} \dots\dots\dots (8)$$

$$= \frac{1}{N'} \sum_{i=1}^N u_i \phi(x_i)$$

$$C_\phi = \frac{1}{N'} \sum_{i=1}^N u_i (\phi(x_i) - \mu_\phi)(\phi(x_i) - \mu_\phi)^T \dots\dots\dots (9)$$

이 때 $N' = \sum_{i=1}^N u_i$ 로 정의된다. 특징 공간에서의 가중치 평균 μ_ϕ 가 영이라고 가정하면 식 (9)는 식 (10)과 같이 간단히 나타낼 수 있다.

$$\begin{aligned}
 C_\phi &= \frac{1}{N} \sum_{i=1}^N u_i \phi(x_i) \phi(x_i)^T \quad \dots\dots\dots (10) \\
 &= \frac{1}{N} \sum_{i=1}^N (\sqrt{u_i} \phi(x_i)) (\sqrt{u_i} \phi(x_i))^T \\
 &= \frac{1}{N} \sum_{i=1}^N \psi(x_i) \psi(x_i)^T
 \end{aligned}$$

이 때 $\psi(x_i) = \sqrt{u_i} \phi(x_i)$ 는 소속도의 제곱근을 가중치로 곱한 특징 공간에서의 데이터 포인트이다. RF-PCA2에서와 마찬가지로 RKF-PCA2에서도 C_ϕ 의 고유벡터를 구함으로써 잡음에 강건한 주성분을 구할 수 있으며 이는 식 (11)을 만족시키는 w 를 구하는 것이다.

$$C_\phi w = \lambda w \quad \dots\dots\dots (11)$$

하지만 특징 공간은 입력 공간에 비해 차원이 높고 종종 무한 차원으로 여겨지므로 행렬 C_ϕ 는 $\infty \times \infty$ 크기의 행렬에 해당한다. 따라서 식 (11)을 직접 계산할 수는 없으며 커널 트릭(kernel trick)을 사용하여 간접적으로 계산하여야 한다. 커널 이론에 따르면 벡터 w 는 식 (12)와 같이 나타낼 수 있다.

$$w = \sum_{i=1}^N \alpha_i \psi(x_i) = \sum_{i=1}^N \alpha_i \sqrt{u_i} \phi(x_i) \quad \dots\dots\dots (12)$$

이 때 $\alpha = [\alpha_1, \dots, \alpha_N]^T$ 는 가중치 벡터이다. $N \times N$ 커널 행렬 K' 의 각 성분을 식 (13)과 같이 정의하자.

$$\begin{aligned}
 K'_{ij} &= \psi(x_i) \cdot \psi(x_j) \quad \dots\dots\dots (13) \\
 &= \sqrt{u_i u_j} \phi(x_i) \cdot \phi(x_j) \\
 &= \sqrt{u_i u_j} K_{ij}
 \end{aligned}$$

이 때 $K_{ij} = \phi(x_i) \cdot \phi(x_j)$ 는 $N \times N$ 커널 행렬 K 의 요소를 나타낸다. K-PCA의 유도 과정을 통해 식 (11), (12), (13)으로부터 식 (14)를 얻을 수 있다.

$$K' \alpha = \lambda \alpha = \lambda' \alpha \quad \dots\dots\dots (14a)$$

$$UKU = \lambda' \alpha \quad \dots\dots\dots (14b)$$

이 때 U 는 $N \times N$ 대각 행렬로 대각 성분은 $U_{ii} = \sqrt{u_i}$ 의 값을 가진다. 식 (11)과 달리 식 (14)는 $N \times N$ 행렬의 고유벡터를 구하는 문제로 직접 계산이 가능하며 고유벡터 $\alpha^k = [\alpha_1^k, \dots, \alpha_N^k]^T$ ($1 \leq k \leq N$)를 얻을 수 있다. 고유벡터 α^k 가 주어진 경우 테스트 포인트 x 의 고유벡터 w^k 로의 사상은 식 (15)와 같이 구할 수 있다.

$$\begin{aligned}
 w^k \cdot \phi(x) &= \sum_{i=1}^N \alpha_i^k \psi(x_i) \cdot \phi(x) \quad \dots\dots\dots (15) \\
 &= \sum_{i=1}^N \alpha_i^k \sqrt{u_i} \kappa(x_i, x)
 \end{aligned}$$

앞서 데이터는 특징 공간에서 영의 가중치 평균을 가지는 것으로 가정하였다. 이를 만족하지 않는 경우에는 식 (16)을 통해 영의 가중치 평균을 가지도록 해주어야 한다.

$$\begin{aligned}
 \tilde{K}_{ij} &= (\phi(x_i) - \mu_\phi) \cdot (\phi(x_j) - \mu_\phi) \quad \dots\dots\dots (16) \\
 &= K_{ij} - \frac{1}{N'} \sum_{n=1}^N u_n K_{in} \\
 &\quad - \frac{1}{N'} \sum_{m=1}^N u_m K_{mj} + \frac{1}{N'^2} \sum_{m,n=1}^N u_m u_n K_{mn}
 \end{aligned}$$

또한 테스트 포인트 x 도 식 (17)을 이용하여 영의 가중치 평균을 가지도록 해주어야 한다.

$$\begin{aligned}
 \tilde{K}(x, x_i) &= (\phi(x) - \mu_\phi) \cdot (\phi(x_i) - \mu_\phi) \quad \dots\dots\dots (17) \\
 &= K(x, x_i) - \frac{1}{N'} \sum_{n=1}^N u_n K(x, x_n) \\
 &\quad - \frac{1}{N'} \sum_{m=1}^N u_m K(x_m, x) \\
 &\quad + \frac{1}{N'^2} \sum_{m,n=1}^N u_m u_n K_{mn}
 \end{aligned}$$

소속도 계산을 위해서는 재건 오류 역시 특징 공간에서 계산되어야 한다. 하지만 이 역시 차원 문제로 직접 계산될 수 없으며 커널 트릭을 이용하여 간접적으로 계산하여야 한다. 특징 공간에서의 재건 오류는 식 (18)과 같이 계산된다.

$$\begin{aligned}
 e(\phi(x)) &= \kappa(x, x) - \frac{2}{N'} \sum_{i=1}^N u_i \kappa(x_i, x) \\
 &+ \frac{1}{N'^2} \sum_{i,j=1}^N u_i u_j \kappa(x_i, x_j) \\
 &- 2 \sum_{i=1}^k ((\beta^i)^2 - 2\beta^i \gamma^i + (\gamma^i)^2) \\
 &+ \sum_{l,m=1}^k \sum_{i,j=1}^N (\beta^l \beta^m - 2\beta^l \gamma^m + \gamma^l \gamma^m) \xi_i^l \xi_j^m \kappa(x_i, x_j)
 \end{aligned} \tag{18}$$

이 때 $\beta^j, \gamma^j, \xi_i^j$ 는 각각 식 (19), (20), (21)로 정의된다.

$$\beta^j = \sum_{i=1}^N \alpha_i^j \sqrt{u_i} \kappa(x_i, x) \tag{19}$$

$$\gamma^j = \frac{1}{N'} \sum_{i,j=1}^N \alpha_i^j \sqrt{u_i} u_j \kappa(x_i, x_j) \tag{20}$$

$$\xi_i^j = \alpha_i^j \sqrt{u_i} \tag{21}$$

위의 재건 오류 계산식들은 참고문헌 [13]과 유사하지만 약간의 차이점이 있다. 이는 RKF-PCA와 RKF-PCA2의 목적함수 차이에서 연유한다. 자세한 유도 과정은 참고문헌 [13]을 참고하면 된다. 재건 오류가 계산되면 식 (22)와 같이 소속도를 계산할 수 있다.

$$u_i = \exp\left(-\frac{e(\phi(x_i))}{\sigma^2}\right) \tag{22}$$

그림 2는 RKF-PCA2 알고리즘을 요약한 것이다.

- | |
|---|
| <ol style="list-style-type: none"> 1. 소속도 $U_0 = [u_1, \dots, u_N]^T = [1, \dots, 1]^T$, 카운터 $t = 0$, 커널 행렬 K를 초기화한다. 2. repeat 3. $t \leftarrow t + 1$ 4. 식 (16)으로 데이터의 평균 μ_ϕ를 영으로 만든다. 5. 식 (14)를 풀어 고유치와 고유벡터를 구한다. 6. 식 (22)로 소속도를 구한다. 7. until $J^{t-1} - J^t < \epsilon$ 또는 $t > t_{\max}$ 8. return |
|---|

그림 2 RKF-PCA2 알고리즘
Fig. 2 RKF-PCA2 algorithm

IV. 실험 결과 및 고찰

RKF-PCA2가 RKF-PCA에 비해 나은 결과를 보인다는 것을 확인하기 위해 이 장에서 두 알고리즘을 Matlab으로 구현하여 실험하였다. 실험에 사용한 데이터는 커널 PCA의 원 논문에서 사용된 데이터로 3개의 가우스 분포에서 각기 30개의 데이터를 생성한 90개의 점들로 구성되어 있다[11]. 이 논문에서는 여기에 10개의 잡음을 추가하고 잡음이 없는 데이터(D_1)에 K-PCA를 적용한 결과와 잡음이 존재하는 데이터(D_2)에 RKF-PCA와 RKF-PCA2를 적용한 결과를 비교함으로써 제안한 알고리즘이 잡음에 강건하면서 RKF-PCA보다 우수하다는 사실을 보인다. 잡음은 $-0.5 \leq x \leq 0.5$, $1 \leq y \leq 2$ 범위의 균일 분포에서 생성하였다. 그림 3은 실험에 사용된 데이터의 예를 보인 것이다. 이 장에서의 결과는 그림 3과 같은 데이터를 무작위로 100회 생성하여 실험한 결과를 평균한 값들이다.

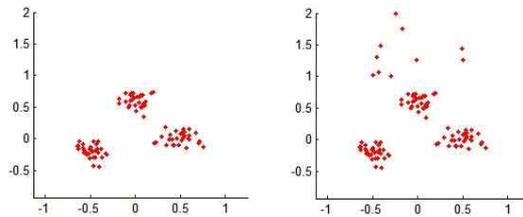


그림 3. 잡음이 없는 데이터와 잡음이 첨가된 데이터
Fig. 3. Noise-free and noisy data

그림 4와 그림 5는 D_2 에 RKF-PCA와 RKF-PCA2를 적용한 경우 목적 함수의 값을 나타낸 것이다. 이 때 최대 반복 회수 t_{\max} 는 20, ϵ 는 0.001로 설정하였다.

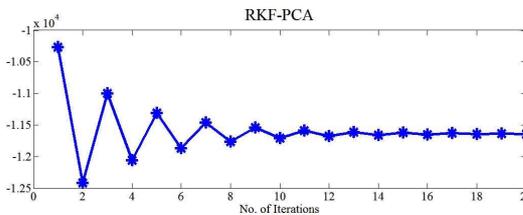


그림 4. RKF-PCA의 수렴
Fig. 4. Convergence of RKF-PCA

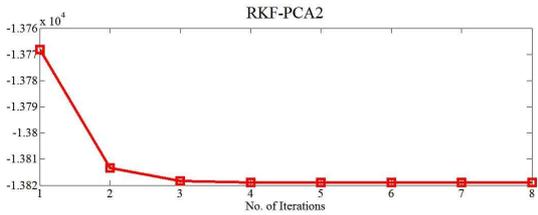


그림 5. RKF-PCA2의 수렴
Fig. 5. Convergence of RKF-PCA2

그림 4와 5에서 알 수 있듯이 RKF-PCA는 RF-PCA와 마찬가지로 두 개의 목적 함수를 사용함으로써 목적 함수가 단조 감소하지 않는다. 반면 RKF-PCA2는 통합된 하나의 목적 함수를 최적화함으로써 단조 감소한다. RKF-PCA2는 또한 빠른 수렴 속도를 보이는 것을 확인할 수 있다. 실험에서 RKF-PCA는 평균 17.82회만에 수렴하였고 RKF-PCA2는 평균 7.31회만에 수렴하였다.

RKF-PCA2는 빠른 수렴 속도뿐만 아니라 찾아진 주성분 역시 잡음의 영향을 적게 받는다. 그림 6은 D_1 에 K-PCA를 적용하여 얻은 첫 번째 및 두 번째 고유 벡터를 나타낸 것이며 그림 7과 8은 D_2 에 RKF-PCA와 RKF-PCA2를 적용하여 얻은 고유 벡터를 나타낸 것이다.

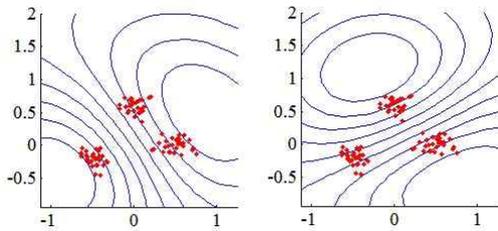


그림 6. K-PCA의 고유 벡터
Fig. 6. Eigen-vectors of K-PCA

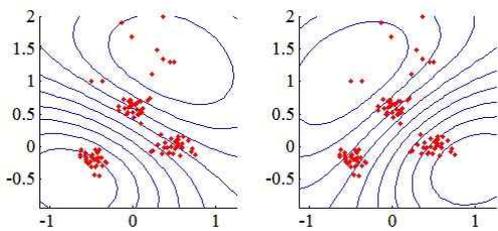


그림 7. RKF-PCA의 고유 벡터
Fig. 7. Eigen-vectors of RKF-PCA

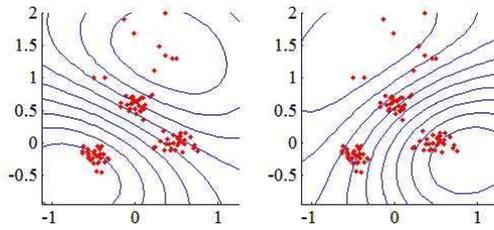


그림 8. RKF-PCA2의 고유 벡터
Fig. 8. Eigen-vectors of RKF-PCA2

그림에서 보아 알 수 있듯이 그림 7과 8은 거의 유사하게 보인다. 이를 정량적으로 평가하기 위해 식 (14)에서의 가중치 벡터 α 를 통해 비교하였다. 비교 방법은 먼저 K-PCA에서 구한 α_0 와 RKF-PCA에서 구한 α_1 의 각도를 식 (23)으로 구한다.

$$\theta_1 = \cos^{-1} \left(\frac{\alpha_0 \cdot \alpha_1}{|\alpha_0| |\alpha_1|} \right) \dots\dots\dots (23)$$

유사하게 RKF-PCA2에서의 α_2 와 α_0 의 각도 θ_2 를 구하고 θ_1 과 θ_2 를 비교함으로써 잡음에 민감한 정도를 비교하였다. 비교 과정에서의 한 가지 문제점은 α_0 과 α_1 이 서로 다른 데이터에서 얻어진 것으로 그 차원이 다르다는 점이다. 따라서 이 논문에서는 α_1 에서 잡음에 의한 부분을 제외하고 α_0 의 차원과 동일하게 수정한 후 계산하였다. 비록 θ_1 과 θ_2 가 절대적인 값을 제공하지는 않지만 각 데이터 포인트들이 고유 벡터를 구성하기 위해 참여하는 정도를 나타내므로 상대적인 비교 값으로는 충분히 활용할 수 있으며 작은 값일수록 좋은 성능을 보이는 것으로 볼 수 있다. 실험에서 θ_1 는 평균 22.142를 θ_2 는 21.875를 보여 RKF-PCA2가 RKF-PCA보다 약간 나은 성능을 보였다. D_2 에 K-PCA를 적용하여 같은 방법으로 비교한 경우 θ 의 평균은 37.232로 RKF-PCA2보다 잡음에 민감한 것을 알 수 있으며 RKF-PCA 역시 K-PCA보다 잡음 민감성이 적은 것을 알 수 있다.

V. 결론

Improved robust fuzzy PCA (RF-PCA2)는 퍼지 소속도를 이용하여 PCA의 잡음 민감성을 줄여주는 반복 알고리즘(iterative algorithm)으로 이전 알고리즘들에 비해 잡음

민감성이 줄어든다는 사실이 입증되었다. 하지만 RF-PCA2는 선형 알고리즘으로 가우시안 분포에 대해서만 유효하다는 단점이 있다. 이 논문에서는 이러한 RF-PCA2의 단점을 보완하기 위해 커널 기법을 도입하여 RF-PCA2를 비선형 알고리즘으로 확장한 improved robust kernel fuzzy PCA (RKF-PCA2) 알고리즘을 제안하였다. RKF-PCA2 알고리즘은 RF-PCA2의 잡음에 강한 성질을 가지면서 kernel PCA를 통해 다양한 분포에 적용할 수 있다는 장점이 있다. 특히 RKF-PCA2는 이전 RKF-PCA에 비해 현저히 빠른 수렴 속도를 보이며 찾아진 주성분 역시 RKF-PCA보다 나은 것을 실험 결과에서 확인할 수 있다.

RKF-PCA2가 비록 이전 알고리즘들에 비해 나은 성능을 보이지만 개선의 여지는 남아있다. RKF-PCA2는 반복 알고리즘으로 많은 연산량을 필요로 한다. 이의 해결 방안으로는 데이터 집합의 크기를 줄여주는 전처리 과정의 도입이나 incremental algorithm으로의 변환 등이 있을 수 있으며 이는 현재 연구 중에 있다. 또 다른 개선 방안으로는 RKF-PCA2의 초기화 과정이 있을 수 있다. RKF-PCA2는 RF-PCA2와 마찬가지로 균일한 초기 속속도 값으로 시작한다. 하지만 국부적인 최적해만을 보장할 뿐 전역적인 최적해는 보장하지 않는다. 따라서 초기화 방법을 달리 함으로써 전역 최적해에 보다 가까운 해를 얻을 수 있을 것으로 기대되며 이는 향후 연구 과제로 남아 있다.

참고문헌

- [1] I. T. Jolliffe, *Principal Component Analysis*, 2nd Edition, Springer, 2002.
- [2] P. Rousseeuw, "Multivariate estimation with high breakdown point," *Mathematical Statistics and Applications B*, pp. 283 - 297, 1985.
- [3] C. D. Lu, T. Y. Zhang, X. Z. Du, and C. P. Li, "A robust kernel PCA algorithm," *Proceedings of the 3rd International Conference on Machine Learning and Cybernetics*, pp. 3084 - 3087, Aug., 2004.
- [4] C. Lu, T. Zhang, R. Zhang, and C. Zhang, "Adaptive robust kernel PCA algorithm," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. VI 621 - 624, April, 2003.
- [5] G. Heo, P. Gader, and H. Frigui, "RKF-PCA: Robust kernel fuzzy PCA," *Neural Networks*, Vol. 22, No. 5-6, pp. 642 - 650, 2009.
- [6] T. N. Yang and S. D. Wang, "Fuzzy auto-associative neural networks for principal component extraction of noisy data," *IEEE Transaction on Neural Networks*, Vol. 11, No. 3, pp. 808 - 810, 2000.
- [7] T. R. Cundari, C. Sarbu, and H. F. Pop, "Robust fuzzy principal component analysis (FPCA). A comparative study concerning interaction of carbonhydrogen bonds with molybdenum-oxo bonds," *Journal of Chemical Information and Computer Sciences*, Vol. 42, No. 6, pp. 1363 - 1369, 2002.
- [8] Gyeongyong Heo, Seong Hoon Kim, Young Woon Woo, "An Improved Robust Fuzzy Principal Component Analysis," *The Journal of the KIMICS*, Vol. 14, No. 5, pp. 1093 - 1102, 2010.
- [9] P. A. Estevez, A. M. Chong, C. M. Held, and C. A. Perez, "Nonlinear Projection Using Geodesic Distances and the Neural Gas Network," *Proceedings of the 16th International Conference on Artificial Neural networks*, vol. I, pp. 464-473, 2006.
- [10] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2001.
- [11] B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.
- [12] G. Heo and P. Gader, "Fuzzy SVM for noisy data: A robust membership calculation method," *Proceedings of the 2009 IEEE International Conference on Fuzzy Systems*, pp. 431 - 436, Aug., 2009.
- [13] G. Heo, P. Gader, and H. Frigui, "Robust Kernel PCA using Fuzzy Membership," *Proceedings of the 2009 International Joint Conference on Neural Networks*, pp. 1213-1220, June, 2009.

저자 소개



허 경 용

1996년 8월 : 연세대학교 본대학원 전
자공학과 (공학석사)

2009년 12월 Department of Computer
and Information Science and
Engineering, University of
Florida (공학박사)

관심분야 Machine Learning,
Pattern Recognition,
Image Processing

Email : hgycap@hotmail.com



서 진 석

1998년 2월 : 건국대학교 공학사

2000년 2월 : 포스텍 공학석사

2005년 2월 : 포스텍 공학박사

2005년-현재 : 동의대학교 게임공학과
조교수

관심분야 : 컴퓨터 게임, 저작도구, 가
상현실, 증강현실

Email : jsseo@deu.ac.kr

이 임 건



1991년 연세대학교 공학사

1993년 연세대학교 공학석사

1998년 연세대학교 공학박사

2002년 - 현재 : 동의대학교 영상
정보공학과 교수

관심분야 영상복원, 영상신호처리 컴퓨
터비전

Email : iglee@deu.ac.kr