◎ 논 문

딕셔너리 러닝을 이용한 음파 신호 분류기 설계

박성민* · 사성진* · 오광명* · 이희승*

Acoustic Signal Classifier Design using Dictionary Learning

Park, Sung Min*, Sah, Sung Jin*, Oh, Kwang Myung*, Lee, Hui Sung*

Key Words : Dictionary Learning(딕셔너리 러닝), Sparse Coding(희박 부호화), Sparsity(희박도), Acoustic Signal Recognition(음파 신호 인식)

ABSTRACT

As new car technology is developing, temporal interaction is needed in automotive. Rhythmic pattern is one of the practical examples of temporal interaction in vehicle. To recognize rhythmic pattern and its input medium, dictionary learning is applicable algorithm. In this paper, performance and memory requirement of the learning algorithm is tested and is sufficiently good for use this acoustic sound.

1. 서 론

최근 자동차는 전자화가 빠르게 진행됨에 따라 자동차에 다양한 안전 및 편의 신기술이 도입되고 있다. 이런 신기술의 일부 기능은 수동으로 입력이 필요하다. 이 같은 기능을 제어하기 위해서 차량 조작계는 더 많은 버튼이 필요해졌다. 차량이라는 한정된 공간에 더 많은 버튼이 생기게 되면 다음과 같은 현상이 발생한다. 버튼은 크기가 줄어들거나 전체 시스템과 어울리지 않는 곳에 배치된다. 두 현상 모두 운전자가 조작하기 어렵다는 문제가 발생한다. 이 현상은 버튼이 물리적 공간을 이용하는 공간 상호작용(spatial interaction)을 이용한 조작계이기 때문이다. 이 문제를 해결하는 한 방법은 시간적 상호작용(temporal interaction)을 이용한 조작계로 차량 내 물리적 버튼을 대체하는 것이다.

이런 연구의 한 방법론으로 리드믹 패턴(rhythmic pattern)을 이용하여 차량 내 장치를 조작하는 AUI(acoustic user interface) 개발이 진행되고 있다. 이를 간단히 설명하자면 손이 잘 닿는 위치에 음파 패드가 있고 이 패드를 특정한 패턴으로 두드리면 이를 인식하여 차량의 여러 명령을 내리는 장치이다. 이때 하나의 AUI 패드로 여러 명령을 수행하려면 명령어의 개수를 늘려야 한다. 명령어를 늘리는 간단한 방법은 박자를 늘리는 것이다. 그러나 박자를 늘리는 것은 두드림 동작이 복잡해지고 사용자가 기억하기 어렵다는 문제가 있다. 이를 해결하기 위하여 다른 입력 매체로 한 박자를 만들어 내는 방법을 도입하였다. 기존에는 한 박자가 하나의 명령이었는데 이 한 박을 손가락, 손톱, 너클, 손 바닥으로 바꿔주면 한 박자를 표현이 4개로 늘어난다. 그러면 같은 두드림 개수를 가져도 명령어의 개수는 4의 제곱의 배수로 늘어난다.

이런 기능을 수행하기 위해서는 두드림이 어떤 입력매체에서 발생했는지 구분해야 하는 음파 분류 알고리즘을 구현해야 한다. 구현하는 방법은 여러 가지가 있으나 음파 발생 패드가 결정되지 않은 현재 상황변화에 유연하게 대응하려면 머신러닝(machine learning) 기법[1]을 이용하는 것이 적당하다. 머신러닝 기법은 여러 세부 분야가 있는데 그 중 지도학습(supervised learning)이라는 분야에 분류기(classifier) 설계라는 방법이 있다. 이는 새로 입력 받은 데이터를 정해둔 몇 개의 클래스로 구분하는 기능을 한다. 이런 분류기가 수행하려는 기능은 손가락, 손톱, 너클, 손바닥과 같이 발생 매체가 다른 음파를 받아들여 어떤 매체에서 온 음파인지 구분하려는 것이다.

* 현대자동차, 벤처기술개발팀
E-mail : hr16k@hyundai.com

분류기에는 여러 종류가 있지만[(1)~(6)] 일반적인 상황에서 좋은 성능을 낸다고 알려져 있는 SVM(데이터 경계를 최대로 만드는 비확률적 이진 선형 분류 기법)[(7)]을 검토해 보아SVM이 PC상에서 음파 분류 성능을 신뢰할 만한 수준임을 확인하였다. SVM알고리즘이 PC에서는 적합했으나 PC에 비해 CPU클럭이나 메모리 자원이 현격히 작은 MCU(micro controller unit) 환경에서는 적용하기에 문제가 있다. MCU 에서 분류기를 사용하기 위해 SVM보다 적은 복잡도를 가지고 메모리 사용이 적은 알고리즘이 필요하다.

이 문제를 해결하기 위하여 본 논문에서는 사전학습(dictionary learning, 데이터에 dictionary를 도입하여 의미 있는 정보들의 희박 표현으로 변경시켜 분류함으로써 성능과 데이터 공간을 모두 좋게 하는 기법)[(8)~(9)]이라는 기법을 이용하여 처리 속도와 연산 시간을 확인해 보기로 방향을 잡고 이 알고리즘을 실제 구현하고 MCU에서 동작가능 여부와 인식 성능을 확인해 보았다.

2. 본 론

2.1. 사전 학습(dictionary learning) 분류기를 선정한 이유

SVM 방법은 분류기 중에서 우수한 성능을 가지고 있지만 연산 복잡성 부분과 데이터 저장 용량 부분에서 큰 단점을 가지고 있다. 이런 특징은 MCU 처럼 자원이 부족한 환경에서 구동하기에는 어려움이 있다. 이러한 문제를 해결하기 위하여 희소 코딩(sparse coding) 기법을 도입하여 처리할 데이터의 차원을 줄이는 방법이 연구되어 왔다. 분류기를 그대로 사용하면서 특징점(feature)만 희소코딩으로 바꿔 사용하는 방법으로도 데이터 처리에 도움이 되겠지만 희소코딩을 분류하기 위해 설계된 분류기를 이용하는 것이 데이터 처리 성능 개선에 가장 효과적인 방법일 것이다. 사전학습(dictionary learning)은 이런 희소코딩을 처리하기 위해 만들어진 알고리즘이다. 그래서 본 과제에서는 이 사전학습을 이용하여 분류기를 만들고 음파 분류 성능을 확인하고자 한다.

2.2. 음파 신호 분류 알고리즘

2.2.1. 과제의 상세 목표

그림 Fig. 1은 음파를 발생시키는 플라스틱 재질의 오

Fig. 1 음파 발생 패드

버헤드 콘솔 모양의 패드이다.

위의 패드를 손의 4가지 부위(손톱, 손가락 끝, 손가락 관절, 손바닥)로 두드려서 발생하는 음파를 패드에 밀착된 마이크로폰으로부터 관측하여 어떤 부위로 두드렸는지를 구분한다. 정의된 4가지 패턴 이외의 음파 신호가 들어올 경우에는 이상점(outlier class)으로 인식한다. 해당 클래스에 대한 정보는 Fig. 2와 같다.

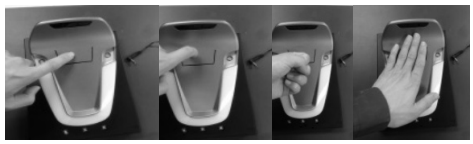

Fig. 2 클래스 구분

위와 같은 신호 분류를 위해 개발될 지도학습 시스템의 블록 선도(block diagram)는 Fig. 3과 같다.

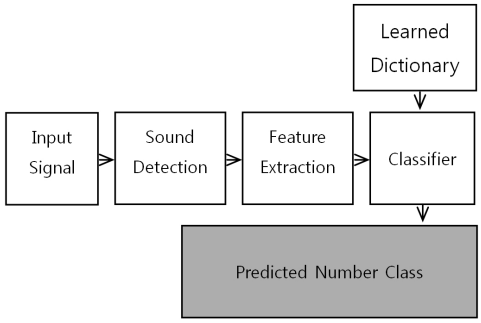

Fig. 3 알고리즘 전개도

2.2.2. 시간 영역에서의 음파 신호 특징 분석

음파 신호는 언제 들어올지 모르기 때문에 마이크는

항상 켜져 있는 상태로 있으므로 음파 신호가 계속 들어오지만 모든 음파에 대해 항상 신호 분석을 하는 것은 mcu에 부담을 주기 때문에 의미 있는 신호가 들어왔을 때 신호 분석을 하는 것이 합리적이다. 이를 위해 신호의 시작과 끝을 결정해 의미 있는 신호를 잘라야 할 필요가 있다. 위에서 정의한 4가지 부위별 두드림 음파들을 시간 영역(time domain)에서 신호의 개략적인 형태를 분석해야 할 필요가 있다. 실제 실험을 통해 알아보기 위하여4가지 패턴별로 20번씩 음파를 발생시켜 신호의 형태를 그려 보면 Fig. 4~7과 같다.

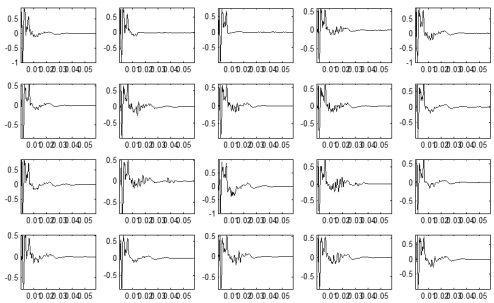

Fig. 4 손톱 두드림 신호

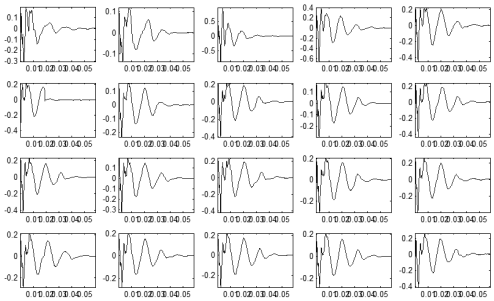

Fig. 5 손가락 두드림 신호

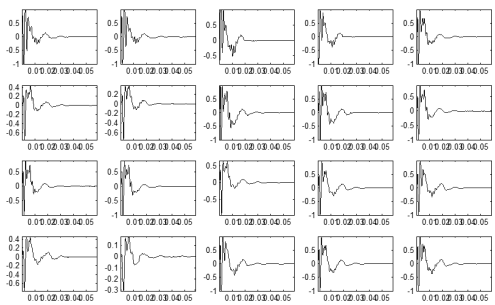

Fig. 6 너클 두드림 신호

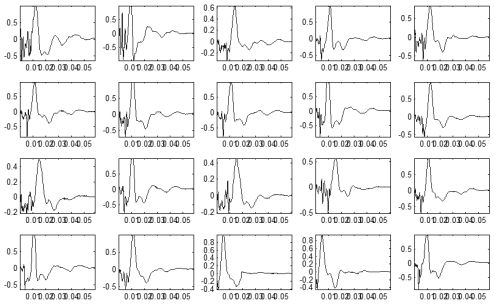

Fig. 7 손바닥 두드림 신호

위의 그림들은 0~60ms 길이에 해당하는 신호의 그래프를 클래스별로 그린 것이다. 손톱과 손가락 관절을 두드려서 발생하는 음파의 경우 신호의 파워가 앞부분에 몰려있어서 30~60ms에 해당하는 부분에는 거의 파워가 없었고, 그에 비해 손가락 끝살과 손바닥으로 두드려서 발생하는 음파의 경우에는 신호의 파워가 전체적으로 고루 퍼져있으나 60ms 이내에 거의 모든 에너지가 들어옴이 관측되었다.

2.2.3. 음파 신호 검출(Sound detection)

현재 시스템에서 음파 신호는 패드에 부착된 마이크로부터 실시간으로 들어오고 있다. 알고리즘이 음파를 처리하려면 계속적으로 들어오는 신호에서 두드림 신호에 대한 정보를 가지고 있는 신호를 잘라와야 한다. 즉 분석하게 될 신호를 특정 길이만큼 잘라낸 뒤에 처리를 해주어야 한다. 본 논문에서 사용한 구체적인 방법은 아래와 같다.

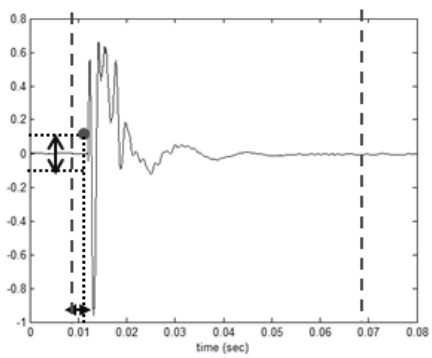

Fig. 8 음파신호 계형

Fig. 8에서 볼 수 있듯이 신호의 절대값이 임의의 문턱값(threshold)보다 높게 나오면 두드림이 발생했다고 생

각한다. 신호의 절대값이 문턱값까지 도달하는데 시간이 걸리고 앞 부분에도 신호가 존재하기 때문에 이 지점부터 약 3ms 앞 지점을 신호의 시작으로 잡는다. 그리고 그 지점부터 60ms 길이의 신호를 잘라내어 분석에 사용한다.

2.2.4. 특징점 추출(Feature extraction)

앞서 Fig. 4~7에서 보듯이 두드림 신호는 시간영역에서 특정한 패턴을 보이지 않고 시간대별로 다른 주파수 특성을 가지고 있다. 이러한 특성을 갖는 신호를 분석할 때는 short time Fourier transform이 일반적으로 널리 사용되고 있다. 이 개념을 그림으로 표현하면 아래 Fig. 9와 같고 이에 대한 수식은 x[n]은 입력 신호이고, w[n]은 window 라고 할 때 아래 (식 1)과 같은 discrete Fourier transform수식이 된다.

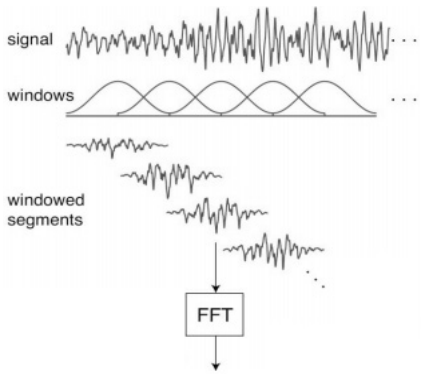

Fig. 9 STFT 개념도

$$X[\tau,\omega] = \sum_{n=-\infty}^{n=\infty} x[n]\omega[n-m]e^{-j2\omega n} \quad (1)$$

위에서 제안한 특징점이 4개의 두드림 클래스에 따라서 서로 다른 값을 가져야 한다. 이를 확인하기 위하여 4가지 두드림 패턴에 대해서 음파 dataset을 수집하고, 수집된 dataset에 대해서 4가지 패턴 별의 특징점이 어떻게 다른지 Fig. 10의 그래프로 나타내 보았다.

손톱과 손가락 관절, 손바닥, 손가락 끝 마다 잘 구분되는 특징이 보이는 것을 알 수 있다. 위 그림에서 알 수 있듯이 제안된 특징점은 클래스를 잘 구분해 내며 이 특징을 사용하여 우수한 분류기를 만들 수 있음을 확인하였다.

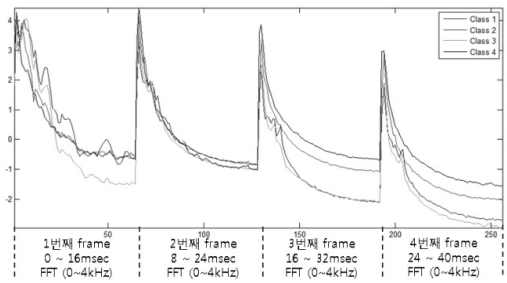

Fig. 10 클래스 별 특징점 경향

2.2.5. 사전 학습 및 분류기 학습(Dictionary learning and Classifier Learning)

앞의 2.1.에서 언급했듯이 사전학습은 SVM에 비해 단순화된 알고리즘이므로 연산 속도가 빠르고 적은 리소스로도 구동이 가능하다. 그러므로 본 과제처럼 MCU에서 신호를 분류하는데 효과적인 알고리즘으로 그 개요는 Fig. 11과 같고 이때 수식은 아래 (식 2)와 같다.

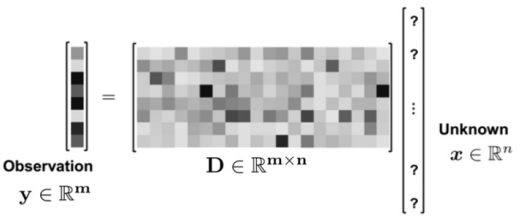

Fig. 11 사전 학습의 개요

$$\min ||x||_0 \quad s.t. \quad Dx = y \quad (2)$$

사전 학습 기술을 사용한 알고리즘 중 빠른 학습속도, C++ 및 임베디드 환경으로 포팅(porting)하기 적합한 알고리즘인 Label Consistent K-SVD의 알고리즘을 사용하였다. 이 알고리즘의 핵심 되는 수식은 아래 (식 3~5)와 같다.

$$<D,X> = \underset{D,X}{\operatorname{argmin}} ||Y - DX||_2^2 + \lambda_1 ||X||_0 \quad (3)$$

$$<W> = \underset{W}{\operatorname{argmin}} L\{H, f(X,W)\} \quad (4)$$

$$<D,X,W> = \underset{D,X,W}{\operatorname{argmin}} ||Y - DX||_2^2 + \lambda_1 ||X||_0 + \alpha ||H - WX||_2^2 \quad (5)$$

이 알고리즘의 특징은 사전(dictionary) D와 분류기 W를 동시에 학습하는 것이다. (식 3)에서 생성된 D는 입력된 신호 Y(Y는 학습을 위해 수집된 데이터)를 최소의 0-norm을 가진 X로 잘 표현하는 방법을 의미한다. 즉 이 수식의 결과로 생성된 사전은 수집된 신호를 한정된 개수의 사전 만을 사용하여 효과적으로 표현하는 과정이다. H를 입력된 신호들의 레이블(Label)이라고 하면 (식 4)는 분류기 설계 과정을 의미한다. 분류기 W는 위의 사전을 통해 생성된 X를 이용하여 입력 신호를 효과적으로 분류하는 것이 목적이 된다. 사전 학습을 이용하여 학습이 끝나면 W는 선형 분류기(Linear Classifier)의 X와의 단순 행렬 연산이 된다. 이때 특징적인 것은 X의 값이 희소(Sparse)하기 때문에 구현 시에 빠르게 연산을 수행할 수 있다. 행렬 연산 WX를 수행 했을 시에 각 열 벡터(column vector)의 값들은 각 클래스의 예측 값이 된다. 이 값 중 가장 높은 값을 가지고 있는 클래스로 입력 신호는 분류되게 된다.

사전 학습에서 D와 W의 특성을 동시에 만들어주기 위하여 식(3~5)를 모아서 한번에 풀어준다. 이때 Y-DX와 H-WX가 X에 대하여 동일한 구조를 가지고 있기 때문에 행렬을 아래와 같이 쌓아 올려 K-SVD 알고리즘을 사용하여 D, W, X를 동시에 구하는 Fig. 12와 같은 방법으로 구한다.

$$\begin{bmatrix} Y \\ H \end{bmatrix} = \begin{bmatrix} D \\ W \end{bmatrix} \begin{bmatrix} X \end{bmatrix}$$

Fig. 12 사전 학습의 학습 방법

학습과정에서 학습된 사전을 사용하여 입력 신호를 분류하기 위해서는 신호를 희소 표현(Sparse Representation)해 주어야 한다. 이 때 분류 성능을 높이기 위해서는 표현(Representation)에 사용된 계수(coefficient)들이 모두 0이상의 값을 유지하면서 적은 개수의 기저(Basis)만 사용하게 하는 비음희소표현(Nonnegative Sparse Representation)을 사용하며 적합한 속도와 성능, 그리고 MCU 환경에서도 성능을 유지할 수 있게 하기 위해서 Nonnegative Orthogonal Matching Pursuit(NOMP)[10]~[11]를 시행한다. NOMP를 수행함으로써 Outlier 신호들이 표현 안되도록 하면서 분류의 에러도 줄일 수 있다.

2.3. 음파 신호 분류 실험

2.3.1. 음파 데이터 수집 및 실험 설계

실험은 10명의 인원(남자 8명, 여자 2명)에 대해서 한 명당 4개(손톱, 손끝, 너클, 손바닥)의 두드림 패턴, 패턴당 20개(총 800개)의 샘플을 수집하였다. 실험의 샘플링 레이트(Sampling rate)는 8kHz로 진행하였고 총 800개의 수집된 샘플 중 400개를 training으로 하여 사전과 분류기를 만들고 이를 사용하여 나머지 400개를 분류하는 실험을 진행한다.

배경 클래스(background class)로는 주행 중 차량 내부에서 획득한 소리 중 임의로 추출한 400개의 음파신호를 사용하였다. 이렇게 만들어진 800(Target class 400개, background class 400개)에 대하여 실험을 진행하였다.

2.3.2. 매개변수 조정(parameter tuning)

Dictionary size, sparsity tuning에 대해 알아보자. 사전 크기(Dictionary size)가 작아질수록 시스템의 크기와 속도가 빨라지므로 작은 사전을 만드는 것이 유리하다. 하지만 너무 작아지면 분류 성능이 나빠지므로 성능을 유지하며 최대한 작은 크기의 사전을 만드는 것이 유리하다. 그 실험은 Fig. 13과 같으며 여기서 확인할 수 있듯이 약 320의 Dictionary size까지 성능이 유지되는 것을 알 수 있다.

Fig. 13 사전 크기에 따른 성능 변화

Sparsity가 낮을수록 시스템의 속도가 빨라지는 장점을 가지므로 같은 성능에서는 sparsity를 낮게 가져가는 것이 유리하나 너무 낮추면 성능이 떨어지게 된다. 따라서 적당한 크기의 sparsity가 필요하다. 그에 대한 실험

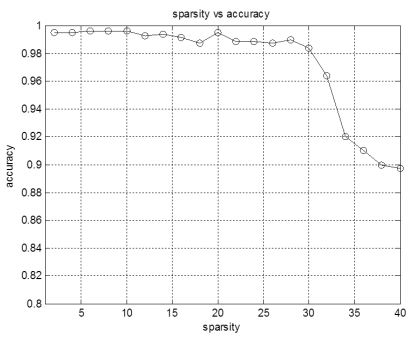

Fig. 14 Sparcity에 따른 성능 변화

은 Fig. 14와 같고 위 그림에서 보듯이 안정성을 위하여 실제 구현 시에는 4의 sparsity를 사용하였다.

Feature dimension tuning에 대해 알아보자. 특징점 추출 시의 STFT의 window size와 frame number에 따라 성능이 달라진다. 시스템의 성능이 떨어지지 않는 한도에서 가장 적은 양의 데이터를 사용하는 것이 유리하다. 그 실험 결과는 Fig. 15~16과 같고 실험 결과 window size로는 128(16ms), frame number로는 5개(1/2 overlap,

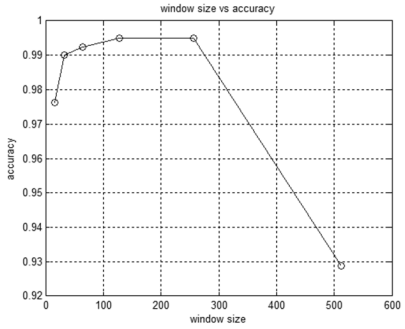

Fig. 15 윈도우 크기에 따른 성능 변화

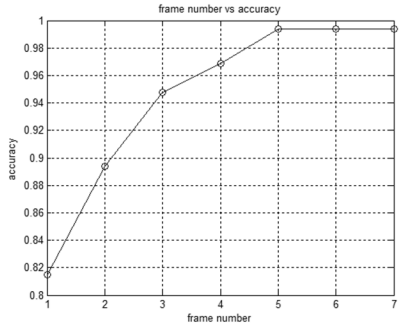

Fig. 16 구간 개수에 따른 성능 변화

총 길이 48ms)를 사용하는 것이 바람직하다.

2.4. 실험 결과

위에서 조정된 매개변수를 이용하여 음파 분류 실험을 진행해보았고 그 결과를 표로 나타내면 Table 1, 2와 같다.

Table 1 음파 분류 실험 결과

Prediction / True	손톱	손끝	너클	손바닥	노이즈
손톱	99	1	0	0	0
손끝	1	99	0	0	0
너클	1	0	99	0	0
손바닥	1	0	0	99	0
노이즈	0	0	0	0	400

Table 2 음파 분류 실험 오인식률

	인식률	오인식률
손톱	99%	0.43%
손끝	99%	0.14%
너클	99%	0%
손바닥	99%	0%
노이즈	100%	0%

800개의 테스트 신호(손톱 100개, 손끝 100개, 너클 100개, 손바닥 100개, 노이즈 400개)에 대하여 실험을 진행한 결과이며, 전체적으로99.5%의 인식률을 나타냈다. 오인식률도 어느 클래스건 0.5%보다 적은 비율로 발생했다. 결과에서 보듯이 MCU에서 구동가능하며 목표 인식 성능에 도달한 알고리즘이 구현되었다.

3. 결 론

본 논문에서는 MCU환경에서 사전학습(dictionary learning)을 분류기로 사용하여 패드에 입력되는 4가지 서로 다른 두드림 매체 인식의 인식률과 수행 속도를 확인해 보았다. 사전학습 알고리즘이 MCU환경에서 무리 없이 동작하면서도 인식률은 충분한 수준임을 확인할 수 있었다. 이제 충분한 양의 음파 DB를 확보하여 희소 표현을 찾고 지금까지 개발한 사전학습 알고리즘을 이용하여 차량 내에서 상기의 4개 음파 분류를 수행하여 차량 응용(application)에 적용할 예정이다.

참고문헌

(1) Shai Shalev-Shwartz and Shai Ben David, 2014, "Understanding Machine Learning from Theory to Algorithms", Cambridge University Press.

(2) M Janvier, X Alameda-Pineda, L Girin, and Radu Horaud, 2012, "Sound Event Recognition with a Companion Humanoid," IEEE International Conference on Humanoid Robotics.

(3) J.D. Krijnders, M.E. Niessen, and T.C. Andringa, 2010, "Sound Event Recognition through Expectancy-based Evaluation of Signal-driven Hypotheses," Pattern Recogntion Letters.

(4) R Mogi, and H Kasai, 2013, "Noise-Robust Environmental Sound Classification Method Based on Combination of ICA and MP features," Artificial Intelligence Research.

(5) E Tsau, S Chachada, and C. J. Kuo, 2012, "Content/Context-Adaptive Feature Selection for Environmental Sound Recognition," IEEE Signal&Information Processing Association Annual Summit&conference.

(6) A Rabaoui, H Kadri, Z Lachiri, and N Ellouze, 2008, "One-Class SVMs challenges in Audio Detection and Classification Applications," EURASIP Journal on Advances in Signal Processing.

(7) Christopher J.C. Burges, 1998, "A Tutorial on Support Vector Machines for Pattern Recognition", Microsoft Research.

(8) J Mairal, F Bach, J Ponce, G Sapiro, and A Zisserman, 2009, "Supervised Dictionary Learning," NIPS.

(9) Z Jiang, Z Lin, and L.S. Davis, 2011, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," CVPR.

(10) S. Jo and C. D. Yoo, 2010, "Psychoacoustically constrained and distortion minimized speech enhancement," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 8, pp. 2099-2110.

(11) A. Dufaux, 2001, "Detection and recognition of Impulsive Sounds Signals," Ph.D. dissertation, Facult e´ des sciences de l'Universite´ deNeuchatel, Switzerland.