



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0021328
(43) 공개일자 2011년03월04일

(51) Int. Cl.

G10L 19/02 (2006.01) G10L 15/08 (2006.01)

(21) 출원번호 10-2009-0079057

(22) 출원일자 2009년08월26일

심사청구일자 2009년08월26일

(71) 출원인

인하대학교 산학협력단

인천 남구 용현동 253 인하대학교

(72) 발명자

장준혁

서울특별시 서초구 서초동 1660-31 B동 301호

김상균

인천광역시 부평구 십정1동 324-24 202호

(74) 대리인

김건우

전체 청구항 수 : 총 4 항

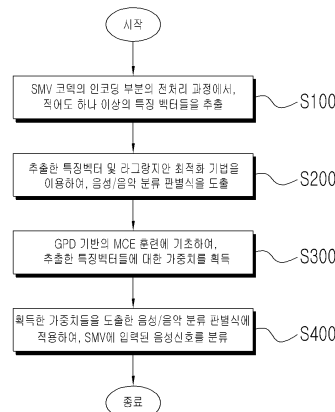
(54) 에스브이엠의 입력벡터에 최적화된 가중치를 적용하여 에스엠브이 코덱의 음성/음악 분류 성능을 향상시키는 방법

(57) 요약

본 발명은 SVM의 입력벡터에 최적화된 가중치를 적용하여 SMV 코덱의 음성/음악 분류 성능을 향상시키는 방법에 관한 것으로서, 보다 구체적으로는 (1) SMV 코덱의 인코딩 부분의 전처리 과정에서, 적어도 하나 이상의 특징벡터를 추출하는 단계, (2) 상기 단계 (1)에서 추출된 특징벡터 및 라그랑지안 최적화(Lagrangian Optimization) 기법을 이용하여, 음성/음악 분류 판별식을 도출하는 단계, (3) GPD(Generalized Probabilistic Descent) 기반의 MCE(Minimum Classification Error) 훈련에 기초하여, 상기 단계 (1)에서 추출된 특징벡터에 대한 가중치를 획득하는 단계, 및 (4) 상기 단계 (3)에서 획득된 가중치를, 상기 단계 (2)에서 도출된 음성/음악 분류 판별식에 적용하여, SMV에 입력된 음성신호를 분류하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명의 SVM의 입력벡터에 최적화된 가중치를 적용하여 SMV 코덱의 음성/음악 분류 성능을 향상시키는 방법에 따르면, 특징벡터들이 음성/음악 분류에 영향을 미치는 정도에 따라, SVM에 입력되는 특징벡터에 대한 가중치를 변별적으로 적용함으로써, SVM 기반 음성/음악 분류 성능을 더욱 향상시키는 것이 가능해진다.

대표도 - 도3



이 발명을 지원한 국가연구개발사업

과제고유번호 2008-F-045-01

부처명 산업자원부

연구관리전문기관

연구사업명 IT핵심기술개발사업

연구과제명 장애인 및 고령자를 위한 Digital Guardian 기술개발

기여율

주관기관 한국과학기술연구원

연구기간 2009.03-2010.02

특허청구의 범위

청구항 1

SVM(Support Vector Machine)의 입력벡터에 최적화된 가중치를 적용하여, SMV(Selectable Mode Vocoder) 코덱의 음성/음악 분류 성능을 향상시키는 방법에 있어서,

- (1) SMV 코덱의 인코딩 부분의 전처리 과정에서, 적어도 하나 이상의 특징벡터를 추출하는 단계;
- (2) 상기 단계 (1)에서 추출된 특징벡터 및 라그랑지안 최적화(Lagrangian Optimization) 기법을 이용하여, 음성/음악 분류 판별식을 도출하는 단계;
- (3) GPD(Generalized Probabilistic Descent) 기반의 MCE(Minimum Classification Error) 훈련에 기초하여, 상기 단계 (1)에서 추출된 특징벡터에 대한 가중치를 획득하는 단계; 및
- (4) 상기 단계 (3)에서 획득된 가중치를, 상기 단계 (2)에서 도출된 음성/음악 분류 판별식에 적용하여, SMV에 입력된 음성신호를 분류하는 단계

를 포함하는 것을 특징으로 하는, SVM의 입력벡터에 최적화된 가중치를 적용한 SMV 코덱의 음성/음악 분류 성능 향상 방법.

청구항 2

제1항에 있어서, 상기 특징벡터에는,

이동평균 에너지, 잡음/목음의 이동평균 반사계수, 부분적 잔류 에너지의 이동평균, 정규화된 피치 상관도의 이동평균, 주기적 계수 및 음악 연속 계수의 이동평균이 포함되는 것을 특징으로 하는, SVM의 입력벡터에 최적화된 가중치를 적용한 SMV 코덱의 음성/음악 분류 성능 향상 방법.

청구항 3

제1항에 있어서, 상기 단계 (2)에서,

상기 음성/음악 분류 판별식($f(Wx)$)은 다음의 수학식으로 정의되는 것을 특징으로 하는, SVM의 입력벡터에 최적화된 가중치를 적용한 SMV 코덱의 음성/음악 분류 성능 향상 방법.

$$f(Wx) = \text{sign}(\sum \hat{\alpha}_i y_i K(Wx_i, Wx_j) + \hat{b})$$

여기서, $\hat{\alpha}_i$ 는 라그랑제 승수들, $\hat{b}$ 는 최적 바이어스를, W 는 특징벡터의 가중치들, $K()$ 는 커널함수들, x_i 와 x_j 는 특징벡터를 의미하며, $y_i \in \{-1, 1\}$ 임.

청구항 4

제3항에 있어서,

상기 특징벡터들에 대한 가중치 $\hat{w}_i$ 는 다음의 수학식으로 갱신될 수 있는 것을 특징으로 하는, SVM의 입력벡터에 최적화된 가중치를 적용한 SMV 코덱의 음성/음악 분류 성능 향상 방법.

$$\tilde{w}_i(n+1) = \tilde{w}_i(n) - \mu \frac{\partial L}{\partial \tilde{w}_i} \bigg|_{\tilde{w}_i = \tilde{w}_i(n)}$$

여기서, μ 는 단순감소 구간의 크기를 의미하며, L 은 손실함수로서 $L = \frac{1}{1 + \exp(-\gamma E)}$, $\gamma > 0$ 을 만족하며, γ 는 시그모이드(sigmoid) 함수의 기울기를, E 는 음성/음악 분류 오류를 각각 나타냄.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 SMV 코덱의 음성/음악 분류 방법에 관한 것으로서, 보다 구체적으로는 SVM의 입력벡터에 최적화된 가중치를 적용하여 SMV 코덱의 음성/음악 분류 성능을 향상시키는 방법에 관한 것이다.

배경 기술

[0002] 최근 들어, 제한된 주파수 대역을 이용하는 이동통신기기에 다양한 멀티미디어 서비스를 제공하기 위해, 주파수 대역을 효과적으로 사용하기 위한 연구가 활발히 진행되고 있다. 예를 들어, 3GPP2(3rd Generation Partnership Project 2)의 표준 코덱인 SMV(Selectable Mode Vocoder)는 제한된 주파수 대역을 효과적으로 사용하기 위하여, 프레임마다 입력되는 음성신호의 특징에 따라 4가지 전송률 중 하나를 결정하고, 결정된 전송률에 기초하여 입력된 음성신호를 부호화하고 있다. 이때, 입력되는 음성신호의 특징에 따라 매 프레임의 전송률을 적절히 조절하는 것은 이동통신기기의 통화음질을 결정하는 중요한 요소이다. 특히, 최근의 이동통신기기는 제한된 주파수를 이용하여 음성뿐만 아니라 음악, 사진, 영상 등과 같은 다양한 멀티미디어 정보들도 전송해야 하므로, 효과적으로 음성과 음악을 분류할 수 있는 방법에 대한 연구가 활발하게 수행되고 있다.

[0003] 상술한 바와 같이, SMV 코덱은 실시간으로 입력된 음성신호를 음성 혹은 음악으로 분류하고, 이에 기초하여 매 프레임의 전송률을 가변시키고 있다. 상기 SMV 코덱의 음성/음악 분류 성능을 향상시키기 위하여, SMV 코덱 인코딩 부분의 전처리 과정에서 추출한 파라미터 중 통계적 학습 분류 성능이 우수한 것들을 특징벡터로 선택한 다음, 선택한 특징벡터를 SVM(Support Vector Machine)에 적용하여, 입력된 음성신호를 음성 혹은 음악으로 분류하는 SVM 기반 음성/음악 분류 방법이 제안된 바 있다. 그러나 기존에 제안된 SVM 기반의 음성/음악 분류 방법은 SVM에 입력되는 특징벡터들에 동일한 가중치를 적용하고 있는데, 일반적으로 음성/음악을 분류하는 과정에서 각각의 특징벡터가 미치는 영향이 서로 다르기 때문에, 기존에 제안된 SVM 기반의 음성/음악 분류 방법은 실제 상황을 고려하고 있지 못한 문제점이 있다.

발명의 내용

해결 하고자하는 과제

[0004] 본 발명은 기존에 제안된 방법들의 상기와 같은 문제점들을 해결하기 위해 제안된 것으로서, 특징벡터들이 음성/음악 분류에 영향을 미치는 정도에 따라 SVM에 입력되는 특징벡터에 대한 가중치를 변별적으로 적용함으로써, SVM 기반 음성/음악 분류 성능을 더욱 향상시킬 수 있는 방법을 제공하는 것을 그 목적으로 한다.

과제 해결수단

[0005] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른, SVM의 입력벡터에 최적화된 가중치를 적용하여 SMV 코덱의 음성/음악 분류 성능을 향상시키는 방법은,

[0006] (1) SMV 코덱의 인코딩 부분의 전처리 과정에서, 적어도 하나 이상의 특징벡터를 추출하는 단계,

[0007] (2) 상기 단계 (1)에서 추출된 특징벡터 및 라그랑지안 최적화(Lagrangian Optimization) 기법을 이용하여, 음성/음악 분류 판별식을 도출하는 단계,

[0008] (3) GPD(Generalized Probabilistic Descent) 기반의 MCE(Minimum Classification Error) 훈련에 기초하여, 상기 단계 (1)에서 추출된 특징벡터에 대한 가중치를 획득하는 단계, 및

[0009] (4) 상기 단계 (3)에서 획득된 가중치를, 상기 단계 (2)에서 도출된 음성/음악 분류 판별식에 적용하여, SMV에 입력된 음성신호를 분류하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

[0010] 바람직하게는, 상기 특징벡터에는,

[0011] 이동평균 에너지, 잡음/목음의 이동평균 반사계수, 부분적 잔류 에너지의 이동평균, 정규화된 피치 상관도의 이동평균, 주기적 계수 및 음악 연속 계수의 이동평균이 포함될 수 있다.

[0012] 바람직하게는, 상기 단계 (2)에서,

[0013] 상기 음성/음악 분류 판별식($f(Wx)$)은 다음의 수학식으로 정의될 수 있다.

$$f(Wx) = \text{sign}(\sum \hat{\alpha}_i y_i K(Wx_i, Wx_j) + \hat{b})$$

[0015] 여기서, $\hat{\alpha}_i$ 는 라그랑제 승수를, $\hat{b}$ 는 최적 바이어스를, W 는 특징벡터의 가중치를, $K()$ 는 커널함수를, x_i 와 x_j 는 특징벡터를 의미하며, $y_i \in \{-1, 1\}$ 이다.

[0016] 더욱 바람직하게는, 상기 특징벡터들에 대한 가중치 $\tilde{w}_i$ 는 다음의 수학식으로 갱신될 수 있다.

$$\tilde{w}_i(n+1) = \tilde{w}_i(n) - \mu \frac{\partial L}{\partial \tilde{w}_i} \bigg|_{\tilde{w}_i = \tilde{w}_i(n)}$$

[0017]

[0018] 여기서, μ 는 단순감소 구간의 크기를 의미하며, L 은 손실함수로서 $L = \frac{1}{1 + \exp(-\gamma E)}$, $\gamma > 0$ 을 만족하며, γ 는 시그모이드(sigmoid) 함수의 기울기를, E 는 음성/음악 분류 오류를 각각 나타낸다.

효 과

[0019] 본 발명에서 제안하고 있는, SVM의 입력벡터에 최적화된 가중치를 적용하여 SMV 코덱의 음성/음악 분류 성능을 향상시키는 방법에 따르면, 특징벡터들이 음성/음악 분류에 영향을 미치는 정도에 따라 SVM에 입력되는 특징벡터에 대한 가중치를 달리 적용함으로써, SVM 기반 음성/음악 분류 성능을 더욱 향상시키는 것이 가능해진다.

발명의 실시를 위한 구체적인 내용

[0020] 이하에서는 첨부된 도면들을 참조하여, 본 발명에 따른 실시예에 대하여 상세하게 설명하기로 한다.

[0021] SMV 코덱 인코딩 부분의 전처리 과정에서 추출한 파라미터 중, 많은 음성과 음악 데이터에 대해 수동으로 히스토그램을 이용하여 통계적 학습 분류 성능을 조사한 결과, 음성/음악 분류 성능이 우수한 아래의 수학식 1 내지 수학식 6으로 정의되는 6개의 파라미터를 특징벡터로 선택할 수 있다.

수학식 1

[0022] 이동평균 에너지($\bar{E}$): $\bar{E} = 0.75 \cdot \bar{E} + 0.25 \cdot E$

[0023] 여기서, E는 프레임 에너지이다.

수학식 2

[0024] 잡음/목음의 이동평균 반사계수($\overline{k_N(i)}$)

$$\overline{k_N(i)} = 0.75 \cdot \overline{k_N(i)} + 0.25 \cdot k_1(i), \quad i = 1, \dots, 10$$

[0026] 여기서, i는 차수이다.

수학식 3

[0027] 부분적 잔류 에너지의 이동평균($\overline{E_N^{res}}$): $\overline{E_N^{res}} = 0.9 \cdot \overline{E_N^{res}} + 0.1 \cdot E^{res}$

[0028] 여기서, $\overline{E_N^{res}}$ 는 상기 $\overline{k_N(i)}$ 에 따라서 값이 갱신된다.

수학식 4

[0029] 정규화된 피치 상관도의 이동평균($\overline{corr_P}$)

$$\overline{corr_P} = 0.8 \cdot \overline{corr_P} + 0.2 \cdot \left(\frac{1}{5} \cdot \sum_{i=1}^{i=5} corr_P^B(i) \right)$$

[0030]

[0031] 여기서, $corr_P^B(i)$ 는 이전 프레임의 피치 상관도이다.

수학식 5

[0032] 주기적 계수($\overline{c_{pr}}$): $\overline{c_{pr}} = \alpha \cdot \overline{c_{pr}} + (1 - \alpha) \cdot c_{pr}$

[0033] 여기서, α 는 c_{pr} 에 따라 값이 변경되는 미리 정해진 가중치이다.

수학식 6

[0034] 음악 연속 계수의 이동평균($\overline{c_M}$): $\overline{c_M} = 0.9 \cdot \overline{c_M} + 0.1 \cdot c_M$

[0035] SMV의 VAD에서는 상기 수학식 1 내지 수학식 5를 통해 획득한 결과들을 정해진 문턱 값과 비교하여 음성의 유무를 판단하며, $\overline{c_{pr}} \geq 18$ 또는 $\overline{c_M} > 200$ 이면 음악으로 판단한다.

[0036] 이전 패턴 분류에 뛰어난 성능을 보이는 SVM은 패턴을 고차원 특징공간으로 사상(Mapping)시킬 수 있어, 선형분리가 불가능한 패턴의 선형분리를 가능하게 하고, 알려지지 않은 확률 분포를 갖는 데이터에 대해 잘못 분류하

는 확률을 최소화하는 구조적인 위험 최소화(Structural Risk Minimization) 방법에 기초하고 있다. SVM은 선형적으로 분류 가능한 데이터에 대한 이진 분류에 있어, 두 개의 클래스를 분류할 수 있는 많은 초평면(Hyperplane) 중 클래스의 가장 가까운 점들과 마진이 최대가 되는 최적의 초평면을 구하는 과정으로 귀결될 수 있다.

[0037] 구체적으로, 학습 데이터가 $(x_1, y_1), \dots, (x_n, y_n)$, $y \in \{-1, 1\}$ 과 같이 주어질 때, 초평면의 방정식은 $w^T x_i + b = 0$ 이다. 여기서, w 는 초평면에 대한 단위 법선 벡터이고, b 는 중심에서 초평면까지의 거리이다. 이때, SVM의 학습능률을 높이기 위해서는 마진을 최대화하는 최적의 초평면을 구해야 하며, 상기 최적의 초평면을 구하기 위한 조건은 다음의 수학적 7과 같다.

수학적 7

$$\text{Minimize : } \mathcal{J}(w) = \frac{\|w\|^2}{2}$$

[0038] Subject to : $y_i(w^T x_i + b) - 1 \geq 0$, for $i = (1, \dots, N)$

[0039] 최적의 $\hat{w}$ 와 $\hat{b}$ 는 라그랑지안 최적화(Lagrangian Optimization)를 이용하여, 목적식과 제약식을 결합한 후, 라그랑주 승수 α_i 를 포함하는 다음의 수학적 8로부터 획득할 수 있다.

수학적 8

$$\mathcal{J}(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m \alpha_i (y_i \cdot [w^T x_i + b] - 1),$$

$$(\alpha_i \geq 0, i = 1, 2, \dots, N).$$

[0040]

[0041] 마진 최대화에 Karush-Kuhn-Tucker(KKT) 조건을 적용하면, 다음의 수학적 9와 같이 정의되는 최적의 $\hat{w}$ 와 $\hat{b}$ 을 도출할 수 있다.

수학적 9

$$\hat{w} = \sum_{i=1}^{N_S} \alpha_i y_i x_i$$

$$\hat{b} = \frac{1 - y_i \hat{w}^T x_i}{y_i} \rightarrow \hat{b} = 1 - \hat{w}^T x_i, y_i = 1$$

[0042]

[0043] 최종적으로 임의 패턴 x 가 주어질 때, 상기 수학적 9를 통해 획득한 $\hat{w}$ 와 $\hat{b}$ 을 이용하여 음성/음악 판별식을 다음의 수학적 10과 같이 정의할 수 있다.

수학식 10

$$f(x) = \text{sgn}(\sum_{i=1}^N \hat{\alpha}_i y_i x_i^T x + \hat{b})$$

[0044]

[0045] 음성신호는 명확하게 선형분리가 되지 않기 때문에, 비선형 변환함수를 이용하여 고차원의 공간으로 사상시킴으로써, 선형분리를 가능하게 할 수 있다. 이때, 비선형 변환함수에 의해 사상된 공간에서도 원 공간에서의 거리 관계를 어느 정도는 보존시켜야만 한다. 상기 비선형 변환함수인 사상함수 $\Phi(x)$ 를 이용하여, 다음의 수학식 11과 같은 커널함수(Kernel Function)를 정의할 수 있다.

수학식 11

$$K(x, x') = \Phi(x)^T \Phi(x')$$

[0046]

[0047] 식별 함수와 최적화 문제에 사상함수 $\Phi(x)$ 를 쓰지 않고 $K()$ 로만 표기하고, 계산을 회피하는 방법을 커널 트릭(Kernel Trick)이라 하며, $\Phi(x)$ 가 존재할 수 있는 커널함수가 주어진 경우에만 유용하다. 아래 표 1은 이용할 수 있는 커널함수들을 정의하고 있으며, 바람직하게는 RBF(Radial Basis Function)를 커널함수로 이용할 수 있다.

표 1

Kernel function	Type of Classifier
Polynomial	$K(x, x') = (x^T x' + 1)^p$
RBF	$K(x, x') = \exp(-\frac{\ x - x'\ ^2}{2\sigma^2})$
Sigmoid	$K(x, x') = \tanh(\beta_0 x^T x' + \beta_1)$

[0048]

[0049] 수학식 11로 정의된 커널함수를 이용하는 경우, 음성/음악 판별식은 다음의 수학식 12로 나타낼 수 있다.

수학식 12

$$f(x) = \text{sign}(\sum \hat{\alpha}_i y_i K(x_i, x_j) + \hat{b})$$

[0050]

[0051] 도 1은 본 발명의 일실시예에 따른 SMV 코덱 인코딩 부분의 전처리 과정에서 추출한 6개의 특징벡터에 대한 음성/음악 분포도이다. 도 1에 도시된 바와 같이, 각각의 특징벡터들은 서로 다른 분포를 가지기 때문에, 음성/음악 분류에 모든 특징벡터가 동일하게 영향을 준다고 보기는 어렵다. 따라서 본 발명에서는 다음 수학식 13에서 정의된 것처럼, 특징벡터들이 음성/음악 분류에 영향을 미치는 정도에 따라, 특징벡터들에 가중치를 변별적으로 부여하는 음성/음악 판별 방법을 제안한다.

수학식 13

$$f(Wx) = \text{sign}(\sum \hat{\alpha}_i y_i K(Wx_i, Wx_j) + \hat{b})$$

[0052]

[0053] 여기서, x 는 상기 수학식 1 내지 수학식 6을 통해 획득할 수 있는 특징벡터이고, $W = \{w_1, w_2, \dots, w_i\}$ 는 가중치 벡

터로서 다음의 수학적 식 14와 같은 제약 조건을 만족한다.

수학적 식 14

$$\sum_{i=1}^6 w_i = 1, \quad w_i \geq 0$$

[0054]

수학적 식 14와 같은 제약조건을 이용하여, 가중치 w_i 는 다음의 수학적 식 15로 변경할 수 있다.

수학적 식 15

$$\tilde{w}_i = \log w_i$$

[0056]

학습 데이터 각각의 프레임에서 음성 $h_s()$ 과 음악 $h_m()$ 을 나누는 두 개의 함수를 다음 수학적 식 16과 같이 정의한다.

수학적 식 16

$$h_s = f(Wx) - \eta$$

$$h_m = -f(Wx) + \eta$$

[0058]

여기서, η 는 음성과 음악을 분류하는 문턱 값에 해당된다.

[0059]

각각의 특징벡터들에 대한 최적의 가중치를 구하기 위해서, 본 발명의 일 실시예에서는 GPD(Generalized Probabilistic Descent) 기반의 MCE(Minimum Classification Error) 훈련 방법을 이용하고 있다. 이를 위해, 학습 데이터의 음성/음악 분류 오류 $E(n)$ 을 다음 수학적 식 17과 같이 정의할 수 있다.

수학적 식 17

$$E(n) = \begin{cases} h_m(n) - h_s(n) & \text{if } h_s \text{ is true class} \\ h_s(n) - h_m(n) & \text{if } h_m \text{ is true class} \end{cases}$$

[0061]

여기서, 음성/음악 분류 결과가 참이면, $E(n)$ 의 값은 음수가 된다.

[0062]

또한, 오류가 발생했을 때, 가중치를 갱신해 주기 위한 손실함수 L 은 다음 수학적 식 18과 같이 정의할 수 있다.

수학적 식 18

$$L = \frac{1}{1 + \exp(-\gamma E)}, \quad \gamma > 0$$

[0064]

여기서, γ 는 시그모이드(sigmoid) 함수의 기울기이다.

[0065]

상기 수학적 식 15로 정의된 가중치 $\tilde{w}_i$ 는 음성/음악 분류 결과가 참이면, 이전 프레임의 값을 그대로 사용하지만, 결과가 오류로 판명되면 다음 수학적 식 19에 기초하여 가중치를 갱신할 수 있다.

[0066]

수학식 19

$$\widetilde{w}_i(n+1) = \widetilde{w}_i(n) - \mu \frac{\partial L}{\partial \widetilde{w}_i} \bigg|_{\widetilde{w}_i = \widetilde{w}_i(n)}$$

[0067]

[0068]

여기서, μ 는 단순감소 구간의 크기이다.

[0069]

$\widetilde{w}_i$ 을 갱신한 다음, 다음 수학식 20을 이용하여 각각의 특징벡터에 대한 가중치 w_i 를 복원할 수 있다.

수학식 20

$$w_i = \frac{\exp(\widetilde{w}_i)}{\sum_{i=1}^6 \exp(\widetilde{w}_i)}$$

[0070]

[0071]

상기 수학식 20으로부터 획득할 수 있는 각각의 특징벡터에 대한 가중치를 상기 수학식 13에 적용함으로써, 입력되는 음성신호에 대하여 음성/음악 분류를 수행할 수 있다.

[0072]

도 2는 본 발명의 일실시예에 따른 SMV의 특징벡터에 적용된 가중치 분포를 나타내는 도면이다. 도 2에 도시된 바와 같이, 특징벡터들이 음성/음악 분류에 영향을 미치는 정도에 따라, 특징벡터들에 서로 다른 가중치가 부여된 것을 확인할 수 있다.

[0073]

도 3은 본 발명의 일실시예에 따른 SVM의 입력벡터에 최적화된 가중치를 적용하여 SMV 코덱의 음성/음악 분류 성능을 향상시키는 방법의 흐름도이다. 도 3에 도시된 바와 같이, 본 발명의 일실시예에 따른 SVM의 입력벡터에 최적화된 가중치를 적용하여 SMV 코덱의 음성/음악 분류 성능을 향상시키는 방법은, SMV 코덱의 인코딩 부분의 전처리 과정에서 적어도 하나 이상의 특징벡터를 추출하는 단계(S100), 단계 S100에서 추출된 특징벡터 및 라그랑지안 최적화(Lagrangian Optimization) 기법을 이용하여, 음성/음악 분류 판별식을 도출하는 단계(S200), GPD(Generalized Probabilistic Descent) 기반의 MCE(Minimum Classification Error) 훈련에 기초하여, 단계 S100에서 추출된 특징벡터에 대한 가중치를 획득하는 단계(S300), 및 단계 S300에서 획득된 가중치들을, 단계 S200에서 도출된 음성/음악 분류 판별식에 적용하여, SMV에 입력된 음성신호를 분류하는 단계(S400)를 포함할 수 있다.

[0074]

단계 S100에서는, SMV 코덱의 인코딩 부분의 전처리 과정에서 추출한 파라미터 중 음성/음악 분류 성능이 우수한 적어도 하나 이상의 파라미터들을 특징벡터로 선택한다. 바람직하게는, 이동평균 에너지, 잡음/목음의 이동평균 반사계수, 부분적 잔류 에너지의 이동평균, 정규화된 피치 상관도의 이동평균, 주기적 계수 및 음악 연속 계수의 이동평균 중 적어도 하나 이상을 특징벡터로 선택할 수 있다.

[0075]

단계 S200에서는, 단계 S100에서 추출된 특징벡터 및 라그랑지안 최적화(Lagrangian Optimization) 기법을 이용하여, 앞서 설명한 수학식 13으로 정의될 수 있는 음성/음악 분류 판별식을 도출한다.

[0076]

단계 S300에서는, GPD(Generalized Probabilistic Descent) 기반의 MCE(Minimum Classification Error) 훈련에 기초하여, 단계 S100에서 추출된 특징벡터에 대한 가중치를 획득한다. 이를 위하여, 단계 S300에서는, 앞서 설명한 수학식 15 내지 수학식 20을 이용할 수 있다.

- [0077] 단계 S400에서는, 단계 S300에서 획득된 가중치들을 단계 S200에서 도출한 음성/음악 분류 판별식에 적용하여, SVM에 입력된 음성신호를 분류한다.
- [0078] 본 발명의 음성/음악 분류 방법의 성능을 평가하기 위해, 기존의 SVM 기반의 음성/음악 분류 방법과 본 발명의 음성/음악 분류 방법을 적용하여 획득한 ROC(Receiver Operating Characteristics) 곡선과 음성/음악 검출 확률(P_d)을 비교하였다. 모의실험에서 사용된 음성 데이터베이스는 8 kHz로 샘플링된 약 6초 정도의 깨끗한 음성으로서, 326명의 남자와 138명의 여자 화자에 의해서 각 10개의 파일이 발음된 TIMIT 데이터베이스가 사용되었다. 또한, 음악 데이터베이스는 CD로부터 여러 장르의 음악을 모바일폰을 통해서 녹음하였고, 8kHz로 다운 샘플링되었으며, 5분 정도의 음악 파일이 사용되었다. 본 발명에서 제안한 음성/음악 분류 방법은 음성파일 4200개와 음악파일 60개(메탈 12개, 재즈 12개, 블루스 12개, 힙합 12개, 클래식 12개)를 이용하여 트레이닝하였다.
- [0079] 이에 더해, 기존의 SVM 기반의 음성/음악 분류 방법과 본 발명에서 제안한 음성/음악 분류 방법의 성능을 객관적으로 평가하기 위하여, 테스트 파일을 만들었다. 동일한 데이터에 의한 성능 향상을 피하기 위해서, 트레이닝에 사용된 음성/음악 데이터는 테스트에 사용되지 않았다. 테스트 파일은 5개 음성 파일(6~12초), 5개 음악 파일(28~32초), 10개 무음(3~15초)을 사용하여 만들었다. 다양한 음악 장르에 대한 음성/음악 분류 성능을 확인하기 위해서, 각 장르별(힙합, 메탈, 재즈, 블루스, 클래식)로 구성된 형태의 테스트 파일 60개와 음악 장르가 혼합된 형태의 테스트 파일 24개, 총 84개의 테스트 파일을 만들었다. 또한 상기 두 가지 방법의 성능을 검증하기 위하여, 테스트 파일의 20ms마다 결과를 0(무음), 1(음성), 2(음악)로 수동으로 작성한 것과 비교하였다.
- [0080] 도 4는 기존의 SVM 기반 음성/음악 분류 방법과 본 발명의 일실시예에 따른 음성/음악 분류 방법의 ROC 곡선을 나타낸 도면이다. ROC 곡선은 SVM에서 이진 분류를 할 때 문턱 값을 -1~1까지 0.01씩 증가시키면서 변화하는 음성/음악의 검출 확률(P_d)을 나타내는 곡선이다. 도 4에서 도시된 바와 같이, 본 발명에서 제안한 음성/음악 분류 방법이 기존의 SVM 기반의 음성/음악 분류 방법에 비하여, ROC 곡선이 우상향 쪽에 위치하므로 음성/음악 분류 성능이 더 우수한 것을 확인할 수 있다.
- [0081] 표 2는 기존의 SVM 기반 음성/음악 분류 방법과 본 발명의 일실시예에 따른 음성/음악 분류 방법의 음성/음악 검출 확률(P_d)을 나타내고 있다. P_e (Probability of Error)는 음성과 음악에 대한 미검출 확률($1-P_d$)에 테스트 데이터의 음성과 음악 비율을 적용하여 정규화한 값을 의미한다. 표 2로부터, 본 발명에서 제안한 음성/음악 분류 방법은, 특히 메탈, 힙합, 재즈, 클래식, 혼합에서 기존의 SVM 기반의 음성/음악 분류 방법보다 우수한 성능을 보이고 있음을 확인할 수 있다.

표 2

TEST	Method	Music(P_d)	Speech(P_d)	P_e
Metal	SMV	0.22	0.91	0.65
	SVM	0.90	0.92	0.10
	Proposed	0.90	0.95	0.09
Blues	SMV	0.15	0.90	0.70
	SVM	0.90	0.90	0.10
	Proposed	0.84	0.95	0.13
Hiphop	SMV	0.28	0.90	0.60
	SVM	0.66	0.90	0.29
	Proposed	0.84	0.92	0.14
Jazz	SMV	0.27	0.92	0.60
	SVM	0.35	0.90	0.54
	Proposed	0.75	0.88	0.22
Classic	SMV	0.50	0.90	0.42
	SVM	0.81	0.91	0.17
	Proposed	0.88	0.92	0.11
Mixed	SMV	0.21	0.93	0.65
	SVM	0.72	0.90	0.24
	Proposed	0.85	0.95	0.13

[0082]

[0083]

결론적으로, 본 발명의 일실시예에 따른 음성/음악 분류 방법이 기존의 SVM 기반의 음성/음악 분류 방법보다 음성/음악 분류 성능이 비등하거나, 특수한 음악에서는 더 우수한 것을 확인할 수 있다. 특히, 입력벡터에 대해 최적화된 가중치가 다양한 데이터베이스를 기반으로, 기계 학습에 기반한 MCE기법에 의해 도출될 경우, 계산량의 증가가 거의 없으면서도, 효과적으로 음성/음악 분류 성능을 향상시킬 수 있는 효과가 있다.

[0084]

이상 설명한 본 발명은 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에 의하여 다양한 변형이나 응용이 가능하며, 본 발명에 따른 기술적 사상의 범위는 아래의 특허청구범위에 의하여 정해져야 할 것이다.

도면의 간단한 설명

[0085]

도 1은 본 발명의 일실시예에 따른 SMV코덱 인코딩 부분의 전처리 과정에서 추출한 6개의 특징벡터에 대한 음성/음악 분포도.

[0086]

도 2는 본 발명의 일실시예에 따른 SMV의 특징벡터에 적용된 가중치 분포를 나타내는 도면.

[0087]

도 3은 본 발명의 일실시예에 따른 SVM의 입력벡터에 최적화된 가중치를 적용하여 SMV 코덱의 음성/음악 분류 성능을 향상시키는 방법의 흐름도.

[0088]

도 4는 기존의 SVM 기반 음성/음악 분류방법과 본 발명의 일실시예에 따른 음성/음악 분류방법의 ROC곡선을 나타낸 도면.

[0089]

<도면 중 주요 부분에 대한 부호의 설명>

[0090]

S100: SMV 코덱의 인코딩 부분의 전처리 과정에서, 적어도 하나 이상의 특징벡터를 추출하는 단계

[0091]

S200: 단계 S100에서 추출된 특징벡터 및 라그랑지안 최적화 기법을 이용하여, 음성/음악 분류 판별식을 도출하는 단계

[0092]

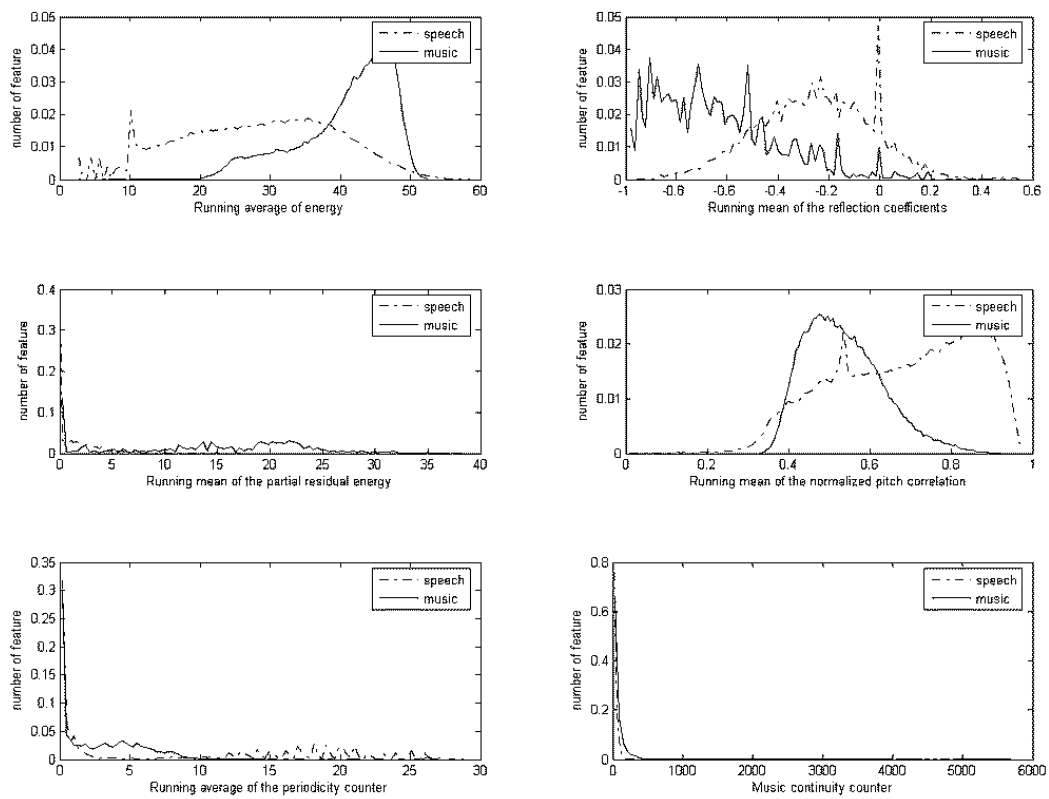
S300: GPD 기반의 MCE 훈련에 기초하여, 단계 S100에서 추출된 특징벡터에 대한 가중치를 획득하는 단계

[0093]

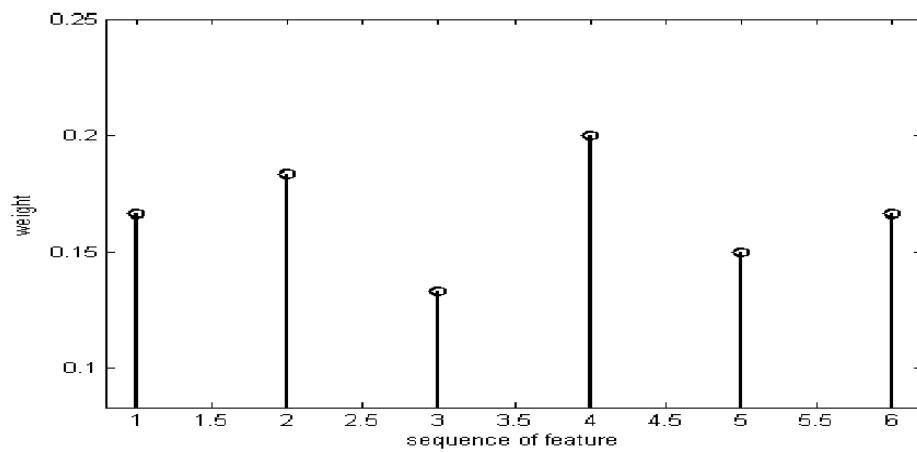
S400: 단계 S300에서 획득된 가중치를, 단계 S200에서 도출된 음성/음악 분류 판별식에 적용하여, SMV에 입력된 음성신호를 분류하는 단계

도면

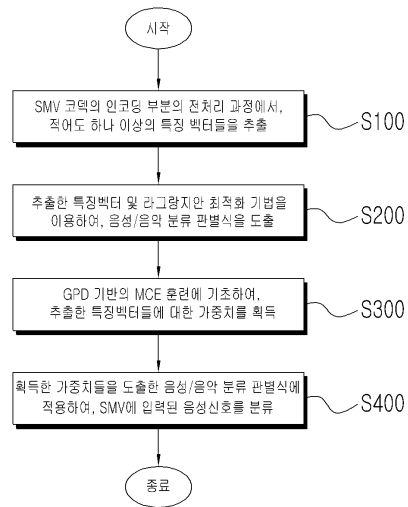
도면1



도면2



도면3



도면4

