



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0073546
(43) 공개일자 2022년06월03일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 3/04 (2006.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2020-0161714
(22) 출원일자 2020년11월26일
심사청구일자 없음

(71) 출원인
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
윤정빈
경기도 성남시 분당구 판교원로82번길 30, 1310동 401호
조민수
경상북도 포항시 남구 청암로 77
강다현
경상북도 포항시 남구 청암로 77
(74) 대리인
특허법인이상

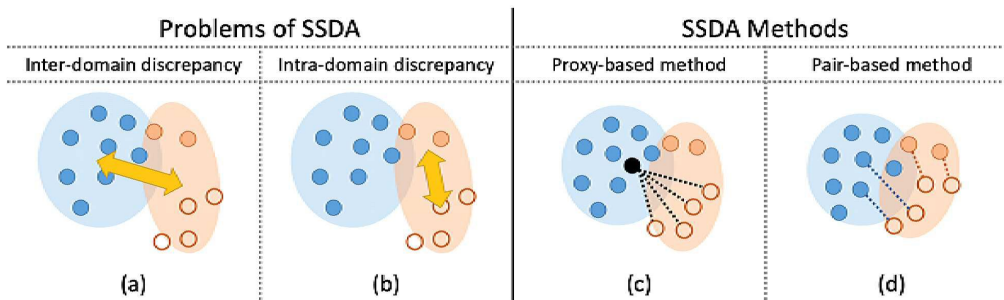
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 **페어 기반 자기 증류 방법 및 장치**

(57) 요약

본 발명은 페어 기반 자기 증류 방법을 개시한다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

| | |
|-------------|---------------------------|
| 과제고유번호 | 1711116284 |
| 과제번호 | 2017M3C4A7069369 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 한국연구재단 |
| 연구사업명 | 차세대정보컴퓨팅기술개발사업 |
| 연구과제명 | 비전 모델 기반 공간 상황 인지 원천기술 연구 |
| 기 여 율 | 1/1 |
| 과제수행기관명 | 한양대학교 |
| 연구기간 | 2020.04.01 ~ 2020.12.31 |

명세서

청구범위

청구항 1

페어 기반 자기 증류 방법.

발명의 설명

기술 분야

[0001] 본 발명은 자기 증류 방법 및 장치에 관한 것으로, 상세하게는 반지도 도메인 적응을 위한 페어 기반 자기 증류 방법 및 장치에 관한 것이다.

배경 기술

[0002] 도메인 적응은 다양한 산업 분야에 적용 가능하다. 모델이 새로운 도메인에 대한 학습을 필요로 할 때 언제나 사용할 수 있다. 대표적으로는 다양한 도로 상황을 처리해야 하는 자율 주행 분야에 응용될 수 있다. 이제까지의 도메인 적응은 비지도 도메인 적응 분야가 주를 이뤘다. 최근 새롭게 제안된 반지도 도메인 적응은 타겟 도메인에 아주 적은 레이블을 부여하여 비지도 도메인 적응에서보다 타겟 도메인에서 높은 성능을 얻을 수 있기 때문에 주목받고있는 분야이다. 즉, 아주 적은 비용으로 타겟 도메인에서 높은 성능향상을 얻을 수 있기 때문에 많은 관심을 받고 있다.

[0003] 반지도 도메인 적응 분야에서 해결해야 하는 문제는 크게 두 가지이다. 하나는 도메인 간 차이이고 다른 하나는 타겟 도메인 내 차이이다. 도메인 간 차이는 비지도 도메인에서도 해결해야 하는 문제로, 레이블이 존재하는 소스 도메인 데이터셋은 모델이 성공적으로 이미지를 분류하는 반면 타겟 도메인 데이터셋의 이미지는 모델이 잘 분류하지 못하는 문제를 일컫는다. 이러한 문제는 도메인 적응에서 해결해야 할 근본적인 문제이다. 타겟 도메인 내 차이는 반지도 도메인 적응 분야에서 타겟 도메인 데이터셋 일부의 이미지에 레이블을 부여해서 발생한 문제이다. 타겟 도메인 내에서도 모델이 레이블이 부여된 이미지는 잘 분류하는 반면, 레이블이 없는 이미지는 잘 분류하지 못하게 된다.

[0004] 반지도 도메인 적응의 두 문제를 해결하기 위해 이전 연구들은 각 클래스를 대표하는 프로토타입을 구하여 레이블이 없는 타겟 도메인 샘플이 가장 가까운 클래스의 프로토타입을 수퍼비전으로 삼아 학습시키는 방법을 사용한다. 그러나 클래스의 프로토타입을 이용한 방법은 여러 레이블이 없는 타겟 도메인 샘플들이 하나의 프로토타입을 수퍼비전으로 삼기 때문에 샘플들을 다양한 수퍼비전으로 학습시킬 수 없다는 단점이 있다. 지식 증류법은 많은 파라미터를 가진 강력한 선생 모델로부터 추출한 지식을 사용하여 상대적으로 적은 파라미터를 가진 학생 모델을 가르치는 학습 방법이다. 이러한 학습 방법은 학생 모델이 주어진 데이터셋의 레이블로 학습하는 것보다 모델의 성능을 높여준다.

[0005] 여러 연구에서 지식 증류법은 모델의 안정적인 학습을 가능하게 하고 모델이 과적합되는 것을 막는다는 것을 밝혀냈다. 지식 증류법은 크게 두 가지 방법으로 수행할 수 있다. 하나는 두 개의 독립적인 선생 모델과 학생 모델을 두는 것이다. 이 방법은 선생 모델의 결과 예측 값을 학생 모델을 학습시키는 데에 사용한다. 또 다른 방법은 자기 증류법이라고 불리며 학생 모델 자신이 선생 모델의 역할을 수행하는 것이다. 즉, 학생 모델이 자신의 결과 예측 값을 스스로의 학습을 위해 사용하는 것이다.

[0006] 반지도 도메인 적응에서는 기존에 비지도 도메인 적응에서 다루었던 문제인 도메인 간 차이뿐만 아니라 타겟 도메인 내 차이 또한 해결하는 것이 중요하다. 기존 연구에서는 두 문제를 해결하기 위해 각 클래스를 대표하는 프로토타입을 생성하여 문제를 해결하였다. 구체적으로, 레이블이 없는 타겟 샘플들에 대한 학습은 가장 가까운 클래스 프로토타입을 수퍼비전으로 삼아 학습이 진행되는 것이다. 그러나 프로토타입은 하나의 클래스에 대한 모든 샘플의 대표이기 때문에 각 샘플에 존재하는 클래스의 다양한 특징들을 표현하는 것은 어렵다.

발명의 내용

해결하려는 과제

- [0007] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은 레이블이 없는 타겟 도메인 샘플에 더 다양한 수퍼비전을 주기 위하여 레이블이 존재하는 샘플 자체를 수퍼비전으로 주는 방식을 사용하여 모델이 예측한 레이블이 존재하는 샘플의 확률값을 자기 증류법을 통해 레이블이 없는 타겟 도메인 샘플의 학습 수퍼비전으로 주는 방식을 사용하는 데 있다.
- [0008] 상기와 같은 문제점을 해결하기 위한 본 발명의 다른 목적은 지식 증류법 중에서도 자기 증류법을 사용하여 반지도 도메인 적용에서 해결해야 하는 도메인 간 차이와 타겟 도메인 내 차이를 극복하는 데 있다.
- [0009] 상기와 같은 문제점을 해결하기 위한 본 발명의 또 다른 목적은 레이블이 없는 타겟 샘플들에 다양하고 풍부한 수퍼비전을 제공하기 위해 클래스의 프로토타입 대신 레이블이 있는 샘플 자체를 수퍼비전으로 삼아 모델을 학습시키는 방법을 제공하는 데 있다.

과제의 해결 수단

- [0010] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 자기 증류 방법은, 반지도 도메인 적용을 위한 페어 기반 자기 증류 단계를 포함할 수 있다.

발명의 효과

- [0011] 본 발명의 일 실시예에 따르면 반지도 도메인 적용의 도메인간 차이와 타겟 도메인 내 차이를 기존 연구들에 비해 효과적으로 줄일 수 있다.
- [0012] 또한, 본 발명의 다른 실시예에 따르면 본 발명은 다양한 수퍼비전을 제공할 수 있다.
- [0013] 또한, 본 발명의 일 실시예에 따르면 반지도 도메인 적용뿐만 아니라 비지도 도메인 적용에서도 기존 연구들보다 효과적이다.

도면의 간단한 설명

- [0014] 도 1은 반지도 도메인 적용의 예시도이다.
 도 2는 본 발명의 일 실시예에 따른 페어기반 자기 증류 방법의 예시도이다.

발명을 실시하기 위한 구체적인 내용

- [0015] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.
- [0016] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는 데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0017] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.
- [0018] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

- [0019] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0021] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.
- [0022] 본 발명은 컴퓨터 비전 및 기계 학습 기술을 이용한 이미지 인식에 관한 것으로서, 서로 다른 두 도메인에 속하는 이미지의 인식에 관한 것이다.
- [0023] 도메인 적응의 목표는 학습 데이터가 풍부한 소스 도메인에서 얻은 정보를 활용하여 학습 데이터가 부족한 새로운 타겟 도메인에서 모델이 잘 동작하도록 하는 것이다. 많은 연구에서 도메인 적응을 비지도 도메인 적응의 세팅에서 풀고자 한다.
- [0024] 비지도 도메인 적응의 세팅에서 소스 도메인 데이터셋은 이미지와 레이블로 구성되고 타겟 도메인 데이터셋은 이미지만으로 구성될 수 있다.
- [0025] 본 발명에서 다루는 반지도 도메인 적응에서는 타겟 도메인 데이터셋에 일부의 레이블만 주어질 수 있다. 모델이 새로운 도메인에 적응해야 하는 목표를 공통적으로 갖고 있기 때문에 일반적인 비지도 도메인 적응 기술에 쓰이는 방법론을 다수 차용할 수 있다.
- [0026] 구체적으로, 도메인 차이를 줄이기 위하여 각 도메인을 대표하는 프로토타입을 구하여 두 프로토타입들의 차이를 줄이는 방식이 있다. 하지만 프로토타입을 이용하여 도메인 차이를 줄이는 기술은 모델이 학습할 때 다양한 수퍼비전을 고려하기 어렵다.
- [0027] 상기의 목적을 달성하기 위하여 본 발명은 반지도 도메인 적응을 위해 샘플을 수퍼비전으로 삼는 페어기반 자기 증류법을 제안할 수 있다. 먼저, 본 발명은 반지도 도메인 적응은 세가지 데이터셋을 사용할 수 있다.
- [0028] 첫 번째로 소스 도메인 데이터셋은 소스 도메인의 이미지와 이에 대한 레이블로 구성될 수 있다. 두 번째로 레이블이 있는 타겟 도메인 데이터셋은 타겟 도메인의 이미지와 이에 대한 레이블로 구성될 수 있다. 해당 데이터셋은 다른 데이터셋들에 비해 상대적으로 크기가 작을 수 있다. 세 번째로 레이블이 없는 타겟 도메인 데이터셋은 타겟 도메인 이미지만으로 구성될 수 있다.
- [0029] 모든 데이터셋은 같은 종류와 개수의 클래스에 대해 다룬다. 모델은 자기 증류법을 적용하기 전 레이블이 존재하는 데이터셋만으로 미리 학습을 시킬 수 있다. 모델의 성능이 수렴한 이후에는 자기 증류법을 이용하여 도메인 간, 타겟 도메인 내의 차이를 줄일 수 있다.
- [0031] 본 연구에서 제안하는 페어기반 자기 증류법은 두 단계를 번갈아 반복하며 모델의 성능을 높일 수 있다. 첫 번째 단계는 학생 데이터셋 생성 단계이고 두 번째 단계는 샘플들 간 자기 증류법 단계일 수 있다.
- [0032] 첫 번째 단계인 학생 데이터셋 생성 단계는 레이블이 없는 타겟 데이터셋의 모든 이미지에 레이블을 부여하는 단계일 수 있다. 이를 위하여 모델이 예측한 결과값들 중 가장 높은 값을 가지는 클래스를 레이블로 삼는 방법인 수도 레이블 방법을 사용할 수 있다. 더욱 정확한 수도 레이블을 얻기 위해 첫 번째로 높은 예측값과 두 번째로 높은 예측값 사이의 차이가 일정 마진값 이상일 경우에만 해당 샘플에 수도 레이블을 부여할 수 있다. 이와 같은 방식으로 단순하게 수도 레이블을 부여하는 것 보다 더 정확하게 샘플들에게 레이블을 부여하여 다음 단계인 샘플들 간 자기 증류법에서 더욱 정확한 페어를 만들 수 있다.
- [0033] 두 번째 단계인 샘플들 간 자기 증류법 단계에서는 선생 샘플의 예측값을 학생 샘플의 수퍼비전으로 주어 모델을 학습시킬 수 있다. 선생 샘플은 레이블이 존재하는 소스 도메인 데이터셋과 타겟 도메인 데이터셋에서 뽑고 학생 샘플은 학생 데이터셋 생성 단계에서 수도 레이블이 부여된 타겟 데이터셋에서 뽑을 수 있다. 이 때 페어는 같은 클래스의 레이블을 가지는 선생 샘플과 학생 샘플로 이루어질 수 있다. 선생 샘플이 소스 도메인 데이터셋으로부터 선택되면 자기 증류법으로 도메인 간 차이를 줄일 수 있고 선생 샘플이 레이블이 존재하는 타겟 도메인 데이터셋으로부터 선택되면 자기 증류법으로 타겟 도메인 내 차이를 줄일 수 있다.

[0035] 도 1은 반지도 도메인 적응의 예시도이다.

[0036] 도 1은 반지도 도메인 적응의 두 가지 문제인 도메인 간, 타겟 도메인 내 차이와 두 가지 접근법인 프로토타입, 페어 기반 법을 나타낸 그림일 수 있다. 도 1을 참조하면, 먼저 도메인 간 차이는 소스 도메인과 타겟 도메인 사이 샘플 분포의 차이를 나타낸다. 타겟 도메인 내 차이는 타겟 도메인 내 샘플 분포의 차이를 나타낸다. 반지도 도메인 적응의 두 가지 접근법들 중 하나인 프로토타입 기반 접근법은 레이블이 없는 타겟 도메인 샘플들에 슈퍼비전으로 프로토타입을 제공할 수 있다. 페어 기반 접근법은 레이블이 없는 타겟 도메인 샘플들에 슈퍼비전으로 다양한 레이블이 존재하는 샘플들의 예측값을 제공할 수 있다.

[0038] 도 2는 본 발명의 일 실시예에 따른 페어기반 자기 증류 방법의 예시도이다.

$$f f \quad g g$$

[0039] 도 2를 참조하면, 특징을 추출하는 모델 와 분류기 는 레이블이 존재하는 데이터셋들로 미리 학습을 마친 상태일 수 있다.

[0040] 여기서, 페어기반 자기 증류법은 학생 데이터셋 생성 단계와 샘플들 간 자기 증류법, 두 단계로 이루어질 수 있다. 첫 번째 단계에서는 레이블이 존재하지 않는 타겟 데이터셋에 수도 레이블로 레이블을 생성할 수 있다. 두 번째 단계에서는 선생 샘플과 학생 샘플을 짝 지어 자기 증류법을 사용할 수 있다.

수학식 1

$$\mathcal{D}_s = \{(\mathbf{x}_s^{(i)}, y_s^{(i)})\}_{i=1}^N, \quad \mathcal{D}_{tl} = \{(\mathbf{x}_{tl}^{(j)}, y_{tl}^{(j)})\}_{j=1}^{N_{tl}}, \quad \mathcal{D}_{tu} = \{(\mathbf{x}_{tu}^{(k)})\}_{k=1}^{N_{tu}}, \quad \mathcal{D}_{ts} = \{(\mathbf{x}'_{ts}, \hat{y}_{ts})\}_{i=1}^{N_{ts}}$$

[0041]

[0042] 수학식 1은 데이터셋의 종류로서 $\mathcal{D}_s = \{(\mathbf{x}_s^{(i)}, y_s^{(i)})\}_{i=1}^N$, $\mathcal{D}_{tl} = \{(\mathbf{x}_{tl}^{(j)}, y_{tl}^{(j)})\}_{j=1}^{N_{tl}}$, $\mathcal{D}_{tu} = \{(\mathbf{x}_{tu}^{(k)})\}_{k=1}^{N_{tu}}$, $\mathcal{D}_{ts} = \{(\mathbf{x}'_{ts}, \hat{y}_{ts})\}_{i=1}^{N_{ts}}$ 는 각각 소스 도메인 데이터셋, 레이블이 있는 타겟 도메인 데이터셋, 레이블이 없는 타겟 도메인 데이터셋, 수도 레이블이 부여된 타겟 도메인 데이터셋을 의미할 수 있다.

[0043] 집합 내 \mathbf{x} 는 이미지를 의미할 수 있고, y 는 이에 대응하는 레이블을 의미할 수 있다. N 은 샘플의 개수를 의미할 수 있다.

수학식 2

$$\mathcal{L}_{lab} = \mathbb{E}_{(\mathbf{x}, y) \in \mathcal{D}_s \cup \mathcal{D}_{tl}} [-\log p(y|\mathbf{x})].$$

[0044]

[0045] 수학식 2는 레이블이 존재하는 소스 도메인과 타겟 도메인의 데이터셋만을 사용하여 모델을 cross-entropy loss 로 학습시키는 수식을 의미할 수 있다.

수학식 3

$$\mathcal{L}_{pair} = \mathbb{E}_{(\mathbf{x}, y) \in \mathcal{D}_s \cup \mathcal{D}_{tl}, (\mathbf{x}', \hat{y}) \in \mathcal{D}_{ts}} [[\hat{y} = y] D_{KL}(p(y|\mathbf{x}) \| p(\hat{y}|\mathbf{x}'))],$$

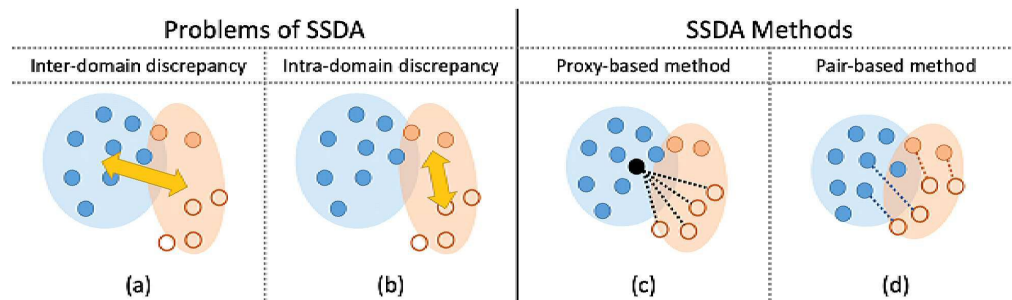
[0046]

[0047] 수학식 3은 레이블이 존재하는 데이터셋과 수도 레이블이 부여된 타겟 데이터셋으로부터 만든 페어 간 KL-divergence를 계산하여 학습시키는 수식을 의미할 수 있다.

- [0049] 본 발명의 실시예에 따른 방법의 동작은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 프로그램 또는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽혀질 수 있는 정보가 저장되는 모든 종류의 기록장치를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산 방식으로 컴퓨터로 읽을 수 있는 프로그램 또는 코드가 저장되고 실행될 수 있다.
- [0050] 또한, 컴퓨터가 읽을 수 있는 기록매체는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 프로그램 명령은 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.
- [0051] 본 발명의 일부 측면들은 장치의 문맥에서 설명되었으나, 그것은 상응하는 방법에 따른 설명 또한 나타낼 수 있고, 여기서 블록 또는 장치는 방법 단계 또는 방법 단계의 특징에 상응한다. 유사하게, 방법의 문맥에서 설명된 측면들은 또한 상응하는 블록 또는 아이템 또는 상응하는 장치의 특징으로 나타낼 수 있다. 방법 단계들의 몇몇 또는 전부는 예를 들어, 마이크로프로세서, 프로그램 가능한 컴퓨터 또는 전자 회로와 같은 하드웨어 장치에 의해(또는 이용하여) 수행될 수 있다. 몇몇의 실시예에서, 가장 중요한 방법 단계들의 하나 이상은 이와 같은 장치에 의해 수행될 수 있다.
- [0052] 실시예들에서, 프로그램 가능한 로직 장치(예를 들어, 필드 프로그래머블 게이트 어레이)가 여기서 설명된 방법들의 기능의 일부 또는 전부를 수행하기 위해 사용될 수 있다. 실시예들에서, 필드 프로그래머블 게이트 어레이는 여기서 설명된 방법들 중 하나를 수행하기 위한 마이크로프로세서와 함께 작동할 수 있다. 일반적으로, 방법들은 어떤 하드웨어 장치에 의해 수행되는 것이 바람직하다.
- [0053] 이상 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

도면1



도면2

