



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0128070
(43) 공개일자 2017년11월22일

(51) 국제특허분류(Int. Cl.)
G10H 1/00 (2006.01)

(52) CPC특허분류
G10H 1/0025 (2013.01)
G10H 2210/105 (2013.01)

(21) 출원번호 10-2017-0022493

(22) 출원일자 2017년02월20일
심사청구일자 없음

(71) 출원인
반병현

경상북도 안동시 안기1길 39, 102동403호(안기동, 안기동대원아파트)

(72) 발명자
반병현

경상북도 안동시 안기1길 39, 102동403호(안기동, 안기동대원아파트)

박주희

서울특별시 송파구 오금로36길 4-30, 1001호 (가락동, 가락동현대빌라트)

정윤의

대전광역시 유성구 대학로 291, 309호 (구성동, 한국과학기술원 진리관)

(74) 대리인

특허법인(유한)아이시스

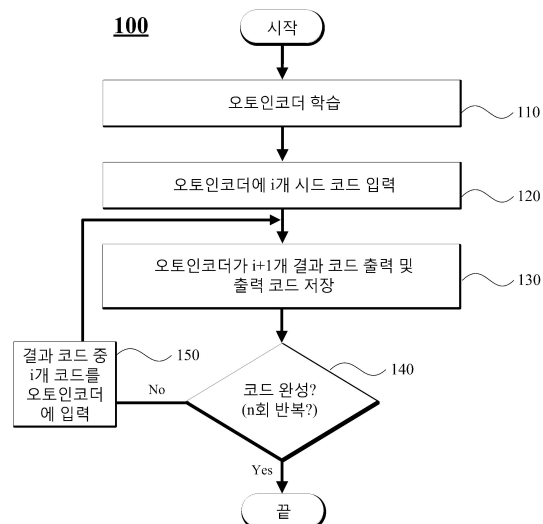
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 순환형 신경망에 기반한 작곡 방법

(57) 요약

순환형 신경망에 기반한 작곡 방법은 컴퓨터 장치가 RNN(Recurrent Neural Network)에 기반하여 학습된 오토인코더(autoencoder)에 i 개의 시드 코드를 입력하는 단계 및 상기 컴퓨터 장치가 상기 오토인코더에서 출력되는 $i+1$ 개의 코드 중 첫 번째 코드를 제외한 나머지 i 개의 코드를 다시 상기 오토인코더에 입력하여 $i+1$ 개의 코드를 출력하는 단계를 포함한다. 상기 컴퓨터 장치는 $i+1$ 개의 코드를 출력하는 단계를 n 번 반복하여 $i+n$ 개의 코드를 생성한다.

대표도 - 도6



(52) CPC특허분류

G10H 2210/111 (2013.01)

G10H 2210/341 (2013.01)

G10H 2240/121 (2013.01)

명세서

청구범위

청구항 1

컴퓨터 장치가 RNN(Recurrent Neural Network)에 기반하여 학습된 오토인코더(autoencoder)에 i 개의 시드 코드를 입력하는 단계; 및

상기 컴퓨터 장치가 상기 오토인코더에서 출력되는 $i+1$ 개의 코드 중 첫 번째 코드를 제외한 나머지 i 개의 코드를 다시 상기 오토인코더에 입력하여 $i+1$ 개의 코드를 출력하는 단계를 포함하되, 상기 컴퓨터 장치는 $i+1$ 개의 코드를 출력하는 단계를 n 번 반복하여 $i+n$ 개의 코드를 생성하는 순환형 신경망에 기반한 작곡 방법.

청구항 2

제1항에 있어서,

상기 오토인코더는 음악 코드에 대응되는 요소를 갖는 N 차원 벡터를 포함하는 샘플 데이터를 이용한 자율학습(unsupervised learning) 방식으로 학습되는 순환형 신경망에 기반한 작곡 방법.

청구항 3

제1항에 있어서,

상기 오토인코더는 RNN 구조를 갖는 인코더 및 RNN 구조를 갖는 디코더를 포함하는 순환형 신경망에 기반한 작곡 방법.

청구항 4

제1항에 있어서,

상기 컴퓨터 장치는 상기 오토인코더에 대한 학습과정에서

일정한 순서를 갖는 $k+1$ 개의 입력 데이터 중 어느 하나를 제외한 k 개의 입력 데이터를 상기 오토인코더에 입력하여 $k+1$ 개의 출력 데이터를 획득하고,

상기 k 개의 입력 데이터 및 상기 $k+1$ 개의 출력 데이터 중 마지막 데이터를 제외한 k 개의 출력 데이터의 차이와 상기 $k+1$ 개의 입력 데이터 및 상기 $k+1$ 개의 출력 데이터의 차이를 합산한 값이 감소하도록 상기 오토인코더를 구성하는 순환형 신경망에 기반한 작곡 방법.

청구항 5

제4항에 있어서,

상기 컴퓨터 장치는 음악 코드로 구성된 샘플 데이터에서 시작 위치를 하나씩 증가시키면서 상기 k 개의 입력 데이터를 선택하면서 상기 학습 과정을 반복하는 순환형 신경망에 기반한 작곡 방법.

발명의 설명

기술 분야

[0001] 이하 설명하는 기술은 딥 러닝(deep learning) 기법을 이용하여 음악을 작곡하는 기법에 관한 것이다.

배경 기술

[0002] 종래 컴퓨터 장치를 이용하여 음악을 자동으로 작곡하려는 연구가 있었다. 대부분의 종래 연구는 멜로디를 입력으로 사용하여 화음을 생성하거나, 몇 가지 변수를 고려하여 확률적 방법에 기반하여 음을 생성하였다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 한국공개특허 제2012-0060085호(2012.06.11)

발명의 내용

해결하려는 과제

[0004] 이하 설명하는 기술은 RNN(Recurrent Neural Network)에 기반한 오토인코더(autoencoder)를 이용하여 음악을 작곡하는 기법을 제공하고자 한다.

과제의 해결 수단

[0005] 순환형 신경망에 기반한 작곡 방법은 컴퓨터 장치가 RNN(Recurrent Neural Network)에 기반하여 학습된 오토인코더(autoencoder)에 i 개의 시드 코드를 입력하는 단계 및 상기 컴퓨터 장치가 상기 오토인코더에서 출력되는 $i+1$ 개의 코드 중 첫 번째 코드를 제외한 나머지 i 개의 코드를 다시 상기 오토인코더에 입력하여 $i+1$ 개의 코드를 출력하는 단계를 포함한다. 상기 컴퓨터 장치는 $i+1$ 개의 코드를 출력하는 단계를 n 번 반복하여 $i+n$ 개의 코드를 생성한다.

발명의 효과

[0006] 이하 설명하는 기술은 큰 규모의 시계열적 데이터를 높은 정확도를 유지하며 생성하여 인간이 납득할 수 있는 수준의 아름다운 음악을 작곡할 수 있다.

도면의 간단한 설명

[0007] 도 1은 오토인코더 구조에 대한 예이다.

도 2는 RNN에 기반한 오토인코더에 대한 예이다.

도 3은 오토인코더를 학습하는 예이다.

도 4는 오토인코더를 학습하는 과정에 대한 예이다.

도 5는 오토인코더를 학습하는 과정에서 이용하는 샘플 데이터의 예이다.

도 6은 Deep RNN에 기반한 작곡 방법에 대한 순서도의 예이다.

발명을 실시하기 위한 구체적인 내용

[0008] 이하 설명하는 기술은 다양한 변경을 가할 수 있고 여러 가지 실시례를 가질 수 있는 바, 특정 실시례들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 이하 설명하는 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 이하 설명하는 기술의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0009] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 이하 설명하는 기술의 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.

[0010] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함한다" 등의 용어는 실시된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.

[0011] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로

합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다. 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다.

[0012] 또, 방법 또는 동작 방법을 수행함에 있어서, 상기 방법을 이루는 각 과정들은 문맥상 명백하게 특정 순서를 기재하지 않은 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 과정들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.

[0014] 이하 설명하는 기술은 딥 러닝(deep learning) 기법을 이용하여 음악을 작곡하는 기법이다. 이하 설명하는 기술은 음의 높이를 나타내는 코드(chord)를 생성한다. 음악은 음의 높이, 하나의 음의 길이, 음의 강약과 같은 요소들이 조합된다. 다만 이하 음의 높이에 해당하는 일련의 코드(멜로디)를 생성하는 과정을 중심으로 설명한다. 이하 설명하는 기술은 딥 러닝을 위한 신경망을 이용한다. 이하 신경망을 이용하여 음악을 작곡하는 장치를 컴퓨터 장치라고 명명한다.

[0015] 이하 설명하는 기술은 신경망 중 소위 오토인코더(autoencoder)라고 불리는 신경망을 사용한다. 도 1은 오토인코더 구조에 대한 예이다. 오토인코더는 신경망 중 자율학습(Unsupervised learning)을 사용하여 학습하는 모델을 의미한다.

[0016] 오토인코더는 입력 계층(input layer)과 은닉 계층(hidden layer)을 포함하는 인코더(encoder) 및 은닉 계층과 출력 계층(output layer)을 포함하는 디코더(decoder)로 구성된다.

[0017] 인코더와 디코더에 학습된 모델은 학습 과정에서 계속 데이터를 생성해 내고, 학습이 끝난 이후에는 인코더의 출력품 형식의 수학적 함수 분포(distribution)를 디코더에 넣어 주면 일정한 출력물을 생성할 수 있다. 일반적으로 오토인코더는 FNN(Feedforward Neural Networks)를 이용하여 학습된다. 이 경우 오토인코더는 입력된 데이터에 대한 결과물을 한 번에 출력한다.

[0019] 이하 설명하는 기술에서 오토인코더는 RNN을 이용하여 학습한다. RNN은 신경망 학습 과정에서 시계열적으로 이전 또는/및 이후 데이터를 활용하여 학습을 수행한다. RNN은 시간 스텝 t에서의 출력값이 이전 시간 스텝 및/또는 이후의 시간 스텝에서 들어오는 입력값에도 영향을 받을 수 있다는 아이디어에 기반한다. 예를 들어, 영어 문제에서 빈칸에 가장 알맞는 단어를 채우기 위해서는 빈칸보다 앞쪽 문장들도 봐야겠지만, 빈칸 이후의 단어들도 문맥을 파악하는데 도움이 될 수 있다. 네트워크 구조는 몇 가지 형태가 가능하다. 예컨대, 양방향(birectional) RNN은 두 개의 RNN을 포함하고, 출력값은 두 개의 RNN의 은닉 계층에 의존하여 결정된다. 도 2는 RNN에 기반한 오토인코더에 대한 예이다. 도 2는 Deep RNN의 유형 중 하나의 예이다. Deep RNN(심층 순환형 신경망)은 양방향 RNN과 유사하지만, 매 시간 스텝마다 복수의 계층이 존재한다. 이하 Deep RNN을 이용하여 음악을 작곡한다고 가정한다.

[0020] 먼저 오토인코더를 학습하는 과정에 대해 설명한다. 도 3은 오토인코더를 학습하는 예이다. 컴퓨터 장치는 샘플 데이터인 코드 샘플을 이용하여 오토인코더를 학습한다. 컴퓨터 장치가 일정한 코드 샘플을 오토인코더에 입력하면 일정한 결과물인 출력 코드가 출력된다. 컴퓨터 장치는 입력된 코드와 출력된 코드를 비교하면서 일정한 방향으로 학습을 수행한다.

[0021] 인코더는 입력값으로 N차원 텐서(tensor)를 받는다. 이 텐서는 N차원 행벡터를 연결한 것이다. 텐서는 예시 열차원의 크기는 각 배치별 트레이닝 데이터의 크기와 일치한다. N차원 행벡터는 One-Hot vector로 구성되며, 벡터의 요소들은 C, Dm, G7등의 음악적 코드에 사상된다.

[0022] 도 4는 오토인코더를 학습하는 과정에 대한 예이다. 도 4는 12개의 연속된 코드를 갖는 코드 샘플을 예로 도시한다. 도 4에서 숫자로 표시한 부분은 코드 샘플의 위치를 의미한다. 컴퓨터 장치는 코드 샘플 중 일정한 길이의 샘플을 기준으로 학습을 수행한다. 도 4는 5개 길이의 샘플 코드를 이용한 예를 도시한다. 도 4(a)는 5개의 샘플 코드를 오토인코더에 입력하여 6개의 코드를 출력하는 예이다. 도 4(a)는 코드 샘플 중 첫 번째 샘플인 입력 샘플을 입력하여 출력 코드를 출력한 예이다. 입력 샘플은 [0,1,2,3,4]이고, 이에 대한 출력 코드는 [a,b,c,d,e,f]이다.

[0023] 컴퓨터 장치는 입력 샘플과 디코더가 출력한 출력 코드 중 마지막 코드를 제외한 5개의 코드 [a,b,c,d,e]를 비

교한다. 입력 샘플 $[0,1,2,3,4]$ 와 출력 코드 중 5개인 $[a,b,c,d,e]$ 의 차이를 기본 손실(basal loss)로 정의한다. 도 4(b)는 기본 손실을 판단하는 과정에 대한 예이다.

[0024] 컴퓨터 장치는 코드 샘플에서 입력 샘플과 연속된 하나의 샘플 $[0,1,2,3,4,5]$ 과 출력 코드 $[a,b,c,d,e,f]$ 를 비교한다. 입력 샘플 +1과 출력 코드의 차이를 예측 손실(prediction loss)로 정의한다. 도 4(c)는 예측 손실을 판단하는 과정에 대한 예이다.

[0025] 도 4의 내용을 일반화하여 다시 설명한다. 컴퓨터 장치는 코드 샘플 중 일부인 k 개의 입력 샘플을 오토인코더에 입력하여, $k+1$ 개의 출력 코드를 획득한다. 컴퓨터 장치는 k 개의 입력 샘플과 출력 코드 중 k 개를 비교하여 기본 손실을 결정한다. 컴퓨터 장치는 입력 샘플과 코드 샘플에서 입력 샘플에 인접한 하나의 샘플($k+1$ 개)와 출력 코드($k+1$)을 비교하여 예측 손실을 결정한다. 컴퓨터 장치는 기본 손실과 예측 손실을 합산한 값이 줄어드는 방향으로 오토인코더를 학습한다. 이와 같은 과정을 반복하면서 컴퓨터 장치는 학습된 오토인코더를 마련한다. k 는 음악의 종류나 시스템의 성능에 따라 달라질 수 있다.

[0026] 도 5는 오토인코더를 학습하는 과정에서 이용하는 샘플 데이터의 예이다. 도 5는 샘플 데이터에 대해 반복적으로 학습을 수행하는 과정에 대한 예이다. 컴퓨터 장치는 입력 샘플 $[0,1,2,3,4]$ 를 이용하여 오토인코더를 학습하고, 이후 입력 샘플의 시작 위치를 하나 변경(증가)하여 입력 샘플 $[1,2,3,4,5]$ 를 이용하여 오토인코더를 학습한다. 이와 같이 컴퓨터장치는 입력 샘플의 위치를 변경하면서 전체 샘플 데이터를 이용하여 오토인코더를 학습한다.

[0027] 악적 코드는 작곡하려는 곡의 종류에 따라 최소 8개, 최대 120개의 vocabulary로 거의 모든 기성음악을 표현할 수 있다. 마이너한 장르의 음악을 포괄하도록 예외적인 코드들까지 고려하더라도 Vocabulary가 200개 내외면 충분하다.

[0029] 도 6은 RNN에 기반한 작곡 방법(100)에 대한 순서도의 예이다. 컴퓨터 장치는 전술한 과정에 따라 학습된 오토인코더를 준비한다(110). 컴퓨터 장치는 오토인코더에 i 개의 시드 코드를 입력한다(120). 시드 코드는 음악을 작곡하기 위해 이용하는 시작 코드에 해당한다. 컴퓨터 장치는 오토인코더의 디코더에 시드 코드를 입력할 수 있다.

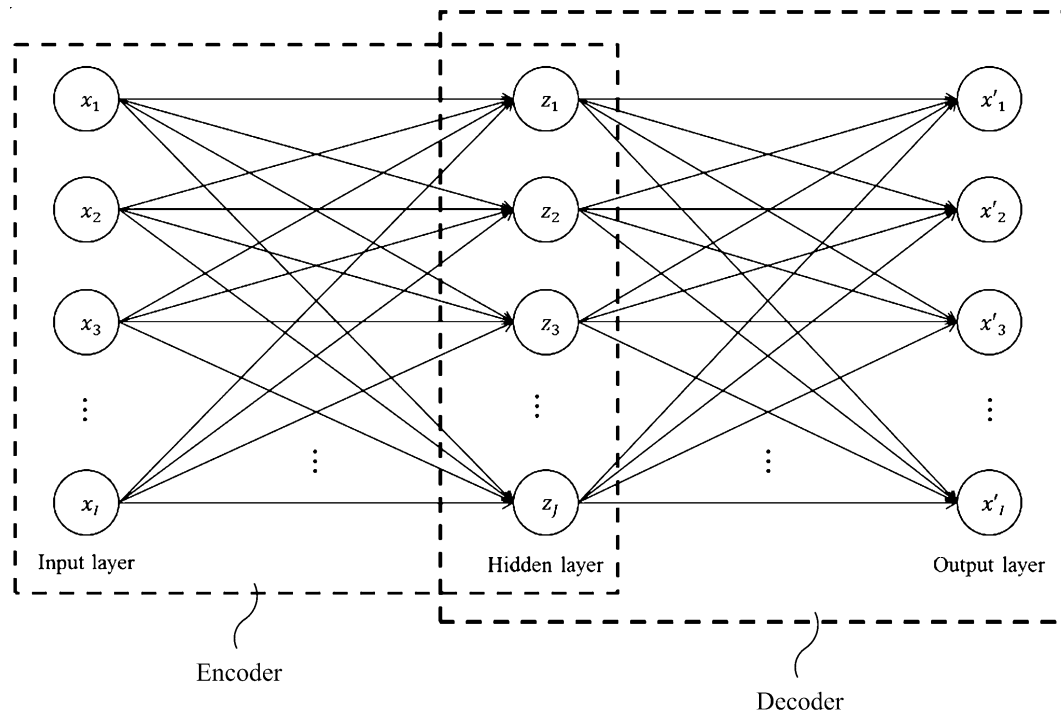
[0030] 오토인코더가 $i+1$ 개의 음악 코드를 출력한다(130). 오토인코더가 출력하는 코드를 결과 코드라고 명명한다. 컴퓨터 장치는 오토인코더가 출력하는 결과 코드를 별도로 저장한다(130). 컴퓨터 장치는 코드 작곡이 완료되었는지 확인한다(140). 코드의 길이는 사전에 설정할 수 있다. 설명의 편의를 위해 오토인코더가 n 개의 추가 코드를 생성한다고 가정한다. 아직 코드 생성이 완성되지 않았다면, 컴퓨터 장치는 결과 코드 중 i 개의 코드를 다시 오토인코더에 입력한다(150). 여기서 입력되는 i 개의 코드는 결과 코드($i+1$) 중 가장 앞에 있는 코드를 제외한 나머지 코드에 해당한다. 컴퓨터 장치가 결과 코드를 획득하는 과정을 n 번 반복하면, 최종적으로 $i + n$ 개의 길이를 갖는 코드가 완성된다.

[0031] 컴퓨터 장치가 n 번의 과정에서 생성되는 n 개의 출력 코드를 일정하게 조합하여 최종 코드를 생성할 수도 있다. 또는 정확도가 높은 경우라면 컴퓨터 장치는 최초 i 개의 코드에서 추가되는 마지막 코드만을 추가하는 방식으로 최종 코드를 생성할 수도 있다.

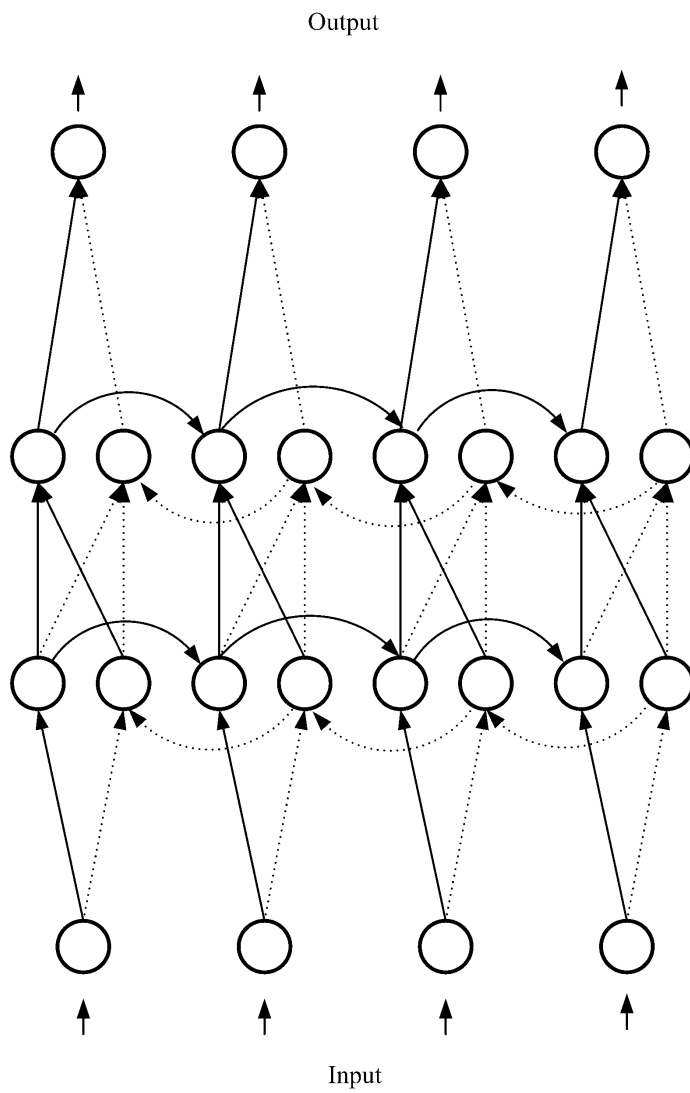
[0032] 본 실시례 및 본 명세서에 첨부된 도면은 전술한 기술에 포함되는 기술적 사상의 일부를 명확하게 나타내고 있는 것에 불과하며, 전술한 기술의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시례는 모두 전술한 기술의 권리범위에 포함되는 것이 자명하다고 할 것이다.

도면

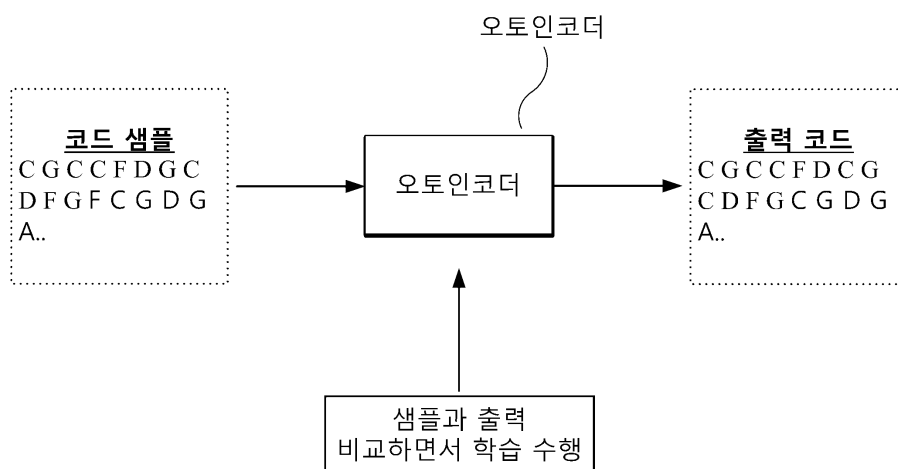
도면1



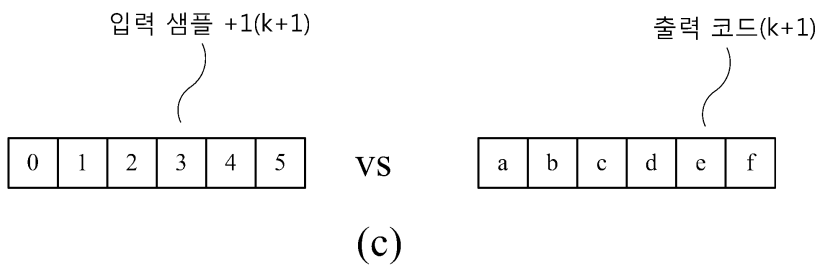
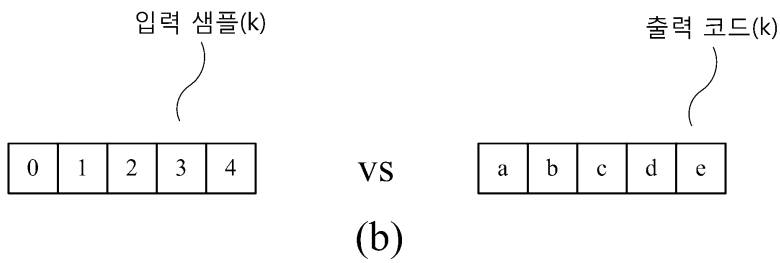
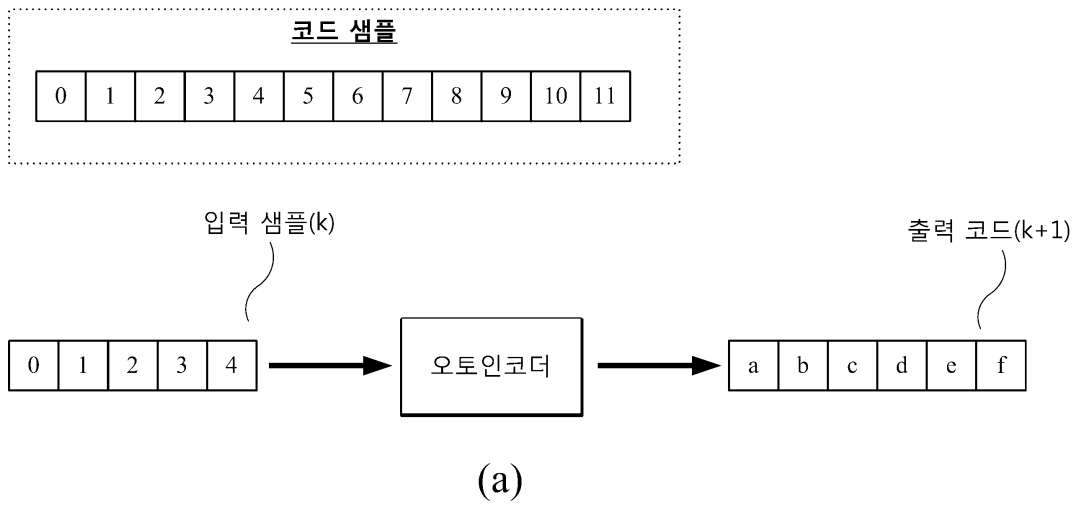
도면2



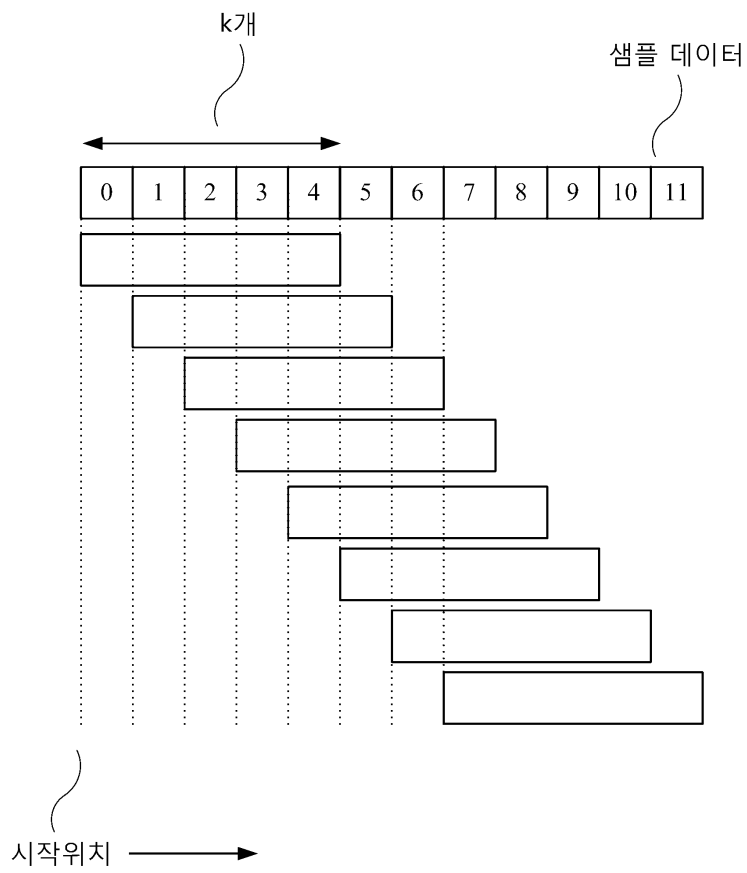
도면3



도면4



도면5



도면6

