



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0159101
(43) 공개일자 2024년11월05일

(51) 국제특허분류(Int. Cl.)

G06V 10/74 (2022.01) G06N 3/08 (2023.01)
G06V 10/42 (2022.01) G06V 10/75 (2022.01)
G06V 10/774 (2022.01) G06V 10/82 (2022.01)
G06V 40/10 (2022.01)

(52) CPC특허분류

G06V 10/761 (2023.08)
G06N 3/08 (2023.01)

(21) 출원번호 10-2023-0055959

(22) 출원일자 2023년04월28일

심사청구일자 없음

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

김경환

서울시 서대문구 연희로11사길 13

김성수

서울시 서대문구 증가로 150 DMC 센트럴 아이파크 112동 1104호

(74) 대리인

특허법인에스알비

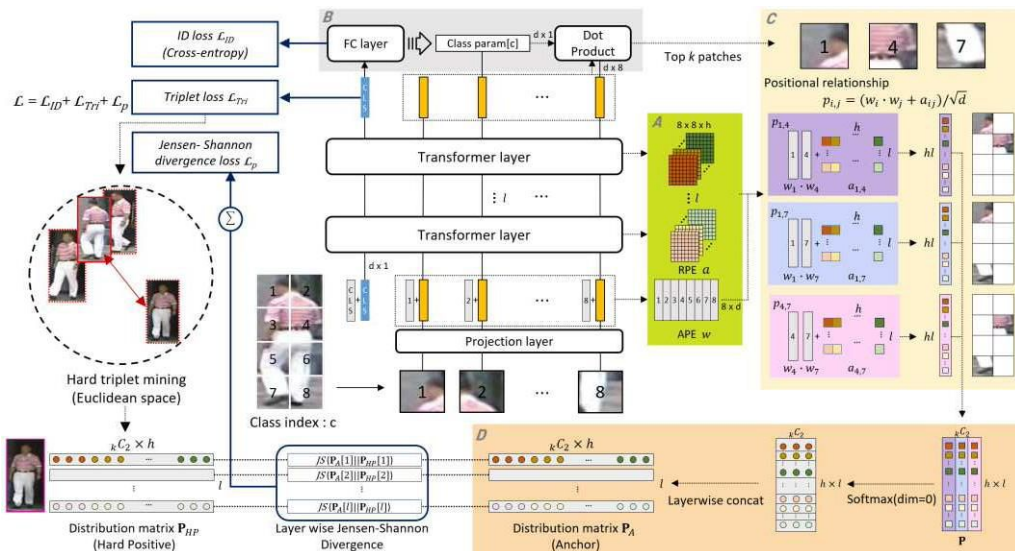
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 트랜스포머 기반 인물 재인식 성능 개선을 위한 모델 학습 방법

(57) 요약

트랜스포머 기반 인물 재인식 성능 개선을 위한 모델 학습 방법, 이를 이용한 딥뉴럴 네트워크 기반 인물 재인식 방법 및 그 장치가 개시된다. 이 방법은 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치를 활용, 인물 이미지의 특징을 해석하여 동일 인물을 구분하는 방법이다. 이 방법은 트랜스포머를 이용하여 인물 이미지를 격자 패치 단위로 나누는 단계, 주요 특징을 담은 패치를 선별하는 단계, 상기 패치의 상대적인 위치 관계를 통해 신체 구조를 분포의 형태로 모델링하는 단계, 손실 함수를 통해 인물 신체 구조의 일관성을 학습하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06V 10/42 (2023.08)

G06V 10/751 (2023.08)

G06V 10/774 (2023.08)

G06V 10/82 (2022.01)

G06V 40/10 (2022.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1615012688

과제번호 153369

부처명 국토교통부

과제관리(전문)기관명 국토교통과학기술진흥원

연구사업명 (이관)스마트 도로조명 플랫폼 개발 및 실증연구(R&D)(국토부)

연구과제명 스마트 도로조명 플랫폼 개발 및 실증연구

기 여 율 1/1

과제수행기관명 한국건설기술연구원

연구기간 2022.01.01 ~ 2023.04.30

명세서

청구범위

청구항 1

트랜스포머 기반의 인물 재인식 모델의 학습방법으로서,
 변형된 Relative positional bias(RPB)를 통해 패치간 상대적 위치관계를 반영하는 단계;
 이미지에 등장하는 인물의 신체적 특성을 대표하는 대표 패치들을 선택하는 단계;
 선별된 상기 대표 패치들로부터 신체 구조적 특성을 모델링하는 단계; 및
 모델링된 신체 구조적 특성이 같은 인물에 대해 일관된 분포를 형성하도록 손실함수를 통해 상기 모델을 학습시키는 단계;
 를 포함하는 것을 특징으로 하는 인물 재인식 모델의 학습방법.

청구항 2

제 1 항에 있어서,
 상기 패치간 상대적 위치관계를 반영하는 단계는,
 트랜스포머 모델의 토큰 간의 상대적 위치 정보를 모델링하는 RPB가 클래스 토큰이 배제되는 형태로 변형되어
 상기 트랜스포머 모델의 모든 계층에 더해지는 것을 특징으로 하는 인물 재인식 모델의 학습 방법.

청구항 3

제 1 항에 있어서,
 상기 대표 패치들을 선택하는 단계는,
 출력 클래스 매개변수와 출력 패치 토큰 간의 내적을 통해 얻은 유사도가 높은 패치 토큰을 선택하는 것을 특징
 으로 하는 인물 재인식 모델의 학습 방법.

청구항 4

제 1 항에 있어서,
 상기 신체 구조적 특성을 모델링 하는 단계는,
 선별된 패치 쌍들의 위치 임베딩을 통해 패치 쌍들의 위치관계를 추출하고, 추출된 상기 위치관계에 관한 정보
 를 통합하고 정규화 하여 인물의 신체 구조적 특성을 나타내는 분포행렬로 모델링하는 것을 특징으로 하는 인물
 재인식 모델의 학습 방법.

청구항 5

제 1 항에 있어서,
 상기 손실함수는,
 상기 분포행렬이 동일 인물 이미지에 대해 같아지도록 유도하는 Jensen-Shannon Divergence(JSD) 손실함수와 인
 물 재인식 모델에서 주로 사용되는 교차 엔트로피 손실과 삼중항 손실을 더한 값을 사용하는 것을 특징으로 하

는 인물 재인식 모델의 학습방법.

발명의 설명

기술 분야

- [0001] 본 발명은 트랜스포머 기반 인물 재인식 네트워크 모델의 학습 방법에 관한 것으로, 특히 기존 네트워크에 변형된 relative positional embedding을 적용하여 인물의 신체 구조적 특성을 설계하고, 네트워크가 인간의 외형적 특징에 더해 신체부위의 구조적 특징을 추가적으로 학습하도록 제안된 손실함수를 이용한 트랜스포머 기반 인물 재인식 네트워크 모델의 학습 방법에 관한 것이다.

배경 기술

- [0003] 인물 재인식(Person Re-Identification)은 하나의 카메라에 국한되지 않고 여러 카메라에 등장하는 사람을 인식하고, 동일한 사람인지 식별하는 기술이다. 이 기술에 사용되는 네트워크 모델은 다양한 샘플 이미지를 이용하여 각 인물 이미지로부터 분별 가능한 특징을 추출하여 각 인물을 구별할 수 있도록 학습된다.
- [0004] 이 작업에서 같은 인물 이미지라 하더라도 카메라의 시점에 따른 인상착의의 변화가 발생하거나 신체 일부가 가려지는 등의 제약 상황이 발생하기 때문에 인물 재인식 연구는 신체의 일부를 포함하는 이미지의 지역적인 부분에서 자세한 특징을 추출하고, 각 신체 부위 간의 비교를 통해 모델이 더 나은 특징을 추출하는 방법을 학습할 수 있도록 유도한다.
- [0005] Vision Transformer(ViT)는 입력 이미지를 일렬로 배치된 일련의 이미지 패치로 나누고 이를 트랜스포머 인코더 네트워크의 연속된 입력으로 제공한다. ViT는 Multi-head Self-attention(MHSA) 모듈과 위치 별 feed-forward 네트워크를 사용하여 패치 토큰에서 특징을 추출하는데, 패치 단위로 자른 이미지의 위치정보를 보존하기 위해 위치 임베딩(Positional embedding)을 추가한다. 최근에는 ViT 기반의 인물 재인식 연구에서 지역 패치를 활용하여 CNN기반의 모델에 비해 좋은 결과를 보여주고 있다.
- [0006] TransReID는 전체 이미지에서 무작위로 선택된 패치 토큰만을 다루는 Jigsaw 패치 모듈이라는 추가 브랜치를 transformer 모듈에 도입했다. PAT는 이미지의 여러 부분에서의 정보를 통합하기 위해 지역 패치 토큰을 사용했다. 또한 일부 연구는 자세 추정 모듈(Pose estimation module)을 도입하여 자세정보를 활용하고 몸 관절을 포함하는 지역 패치들을 활용했다.
- [0007] 이전의 기술들은 ViT기반 인물 재인식 모델이 사람을 더 잘 구별하도록 했지만 트랜스포머 아키텍처에서 추가 모듈 또는 수정이 필요하여 다양한 트랜스포머에 적용하기 어렵고, 여전히 각 지역 패치에 반영된 인물의 외형적 특징에만 집중하고 있다.

발명의 내용

해결하려는 과제

- [0009] 본 발명은 인물 재인식(Person Re-Identification)이 인물 이미지만을 다룬다는 점과, Transformer 기반 모델이 이미지를 패치단위로 구분한 후, 각 영역에 포함된 패치에서 특징을 뽑아 통합하는 구조적 특징을 지닌다는 것을 활용하여, 인물의 외형적 특징에 더해 인물의 신체 구조적 특징을 반영하는 재인식 모델의 학습 방식을 제공하는 것을 목적으로 한다.

과제의 해결 수단

- [0011] 상기 목적을 달성하기 위하여 본 발명은, 트랜스포머 기반의 인물 재인식 모델의 학습방법으로서, 변형된 Relative positional bias(RPB)를 통해 패치간 상대적 위치관계를 반영하는 단계, 이미지에 등장하는 인물의 신체적 특성을 대표하는 대표 패치들을 선택하는 단계, 선별된 상기 대표 패치들로부터 신체 구조적 특성을 모델

링하는 단계 및 모델링 된 신체 구조적 특성이 같은 인물에 대해 일관된 분포를 형성하도록 손실함수를 통해 상기 모델을 학습시키는 단계를 포함하는 것을 특징으로 한다.

[0012] 바람직하게는, 상기 패치간 상대적 위치관계를 반영하는 단계는, 트랜스포머 모델의 토큰 간의 상대적 위치 정보를 모델링하는 RPB가 클래스 토큰이 배제되는 형태로 변형되어 상기 트랜스포머 모델의 모든 계층에 더해질 수 있다.

[0013] 바람직하게는, 상기 대표 패치들을 선택하는 단계는, 출력 클래스 매개변수와 출력 패치 토큰 간의 내적을 통해 얻은 유사도가 높은 패치 토큰을 선택할 수 있다.

[0014] 바람직하게는, 상기 신체 구조적 특성을 모델링 하는 단계는, 선별된 패치 쌍들의 위치 임베딩을 통해 패치 쌍들의 위치관계를 추출하고, 추출된 상기 위치관계에 관한 정보를 통합하고 정규화 하여 인물의 신체 구조적 특성을 나타내는 분포행렬로 모델링할 수 있다.

[0015] 바람직하게는, 상기 손실함수는, 상기 분포행렬이 같은 인물 이미지에 대해 같아지도록 유도하는 Jensen-Shannon Divergence(JSD) 손실함수와 인물 재인식 모델에서 주로 사용되는 교차 엔트로피 손실과 삼중항 손실을 더한 값을 사용할 수 있다.

발명의 효과

[0017] 본 발명에 따르면, RPB는 Transformer 계층에 쉽게 적용할 수 있어 추가적인 무거운 모듈 없이 인물 재인식 모델이 시각적 특징 뿐만 아니라 위상적 및 위치적 특징을 추가적으로 학습하여 기존 모델에 비해 높은 정확도를 얻을 수 있다는 이점이 있다.

도면의 간단한 설명

[0019] 도 1은 본 발명의 실시예에 따른 인물 재인식 모델의 학습 파이프라인 예시도를 나타낸다.

도 2는 본 발명의 실시예에 따른 변형된 RPB를 적용하는 방법을 시각화한 것을 나타낸다.

도 3은 본 발명의 실시예에 따른 대표 패치를 선택하는 과정의 예시도를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0020] 이하, 첨부된 도면들에 기재된 내용들을 참조하여 본 발명을 상세히 설명한다. 다만, 본 발명이 예시적 실시 예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일 참조부호는 실질적으로 동일한 기능을 수행하는 부재를 나타낸다.

[0021] 본 발명의 목적 및 효과는 하기의 설명에 의해서 자연스럽게 이해되거나 보다 분명해질 수 있으며, 하기의 기재만으로 본 발명의 목적 및 효과가 제한되는 것은 아니다. 또한, 본 발명을 설명함에 있어서 본 발명과 관련된 공지 기술에 대한 구체적인 설명이, 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다.

[0022] 도면을 참조하여 본 발명의 실시예에 대하여 설명한다.

[0023] 도 1은 본 발명의 실시예에 따른 인물 재인식 모델의 학습 파이프라인 예시도를 나타낸다.

[0024] 도 1을 참조하면, 도 1은 인물 이미지가 입력되었을 때, 이미지를 분할한 패치 조각에서 인물을 대표할 수 있는 대표 패치들을 선별하는 단계와 선별된 패치들 사이의 위치관계를 추출하여 신체 구조적 특성을 분포 행렬의 형태로 모델링하는 단계, 모델링된 분포행렬과 트랜스포머 모델의 출력인 클래스 토큰을 이용하여 ID 손실(교차 엔트로피), 삼중항 손실, JSD 손실을 합쳐 손실함수를 생성하는 단계를 도식화한다.

[0025] 트랜스포머 기반 인물 재인식 네트워크는 인물 이미지를 일정 크기의 패치로 나누고, 패치 단위 이미지에 초기 위치 정보를 도입하기 위해 절대 위치 임베딩(Absolute positional embedding, APE)을 각 패치 이미지에 더하고 초기화된 클래스 토큰과 함께 입력 받는다.

[0026] 트랜스포머 기반 인물 재인식 네트워크를 입력 패치 토큰들이 통과하는 단계에서 패치 토큰들 사이의 위치 정보

를 유지하기 위해 트랜스포머의 매 계층에서 self-attention 과정 중에 변형된 RPB를 더해준다.

[0027] 도 2는 본 발명의 실시예에 따른 변형된 RPB를 적용하는 방법을 시각화한 것을 나타낸다. 도 2를 참조하면, 도 2는 Transformer 계층의 각 multi-attention head에 적용되는 self-attention score를 계산하는데 있어 RPB를 적용하는 방법을 간략하게 시각화한 도면이다.

[0028] 도 2의 C, Q, K는 각각 클래스 토큰, 쿼리 패치 토큰, 키 패치 토큰을 의미한다. 도 2의 (a)는 self-attention 과정에서 쿼리(query)와 키(key)사이의 내적(Dot product)을 나타내고, (b)는 변형된 RPB 매개변수와 이를 더해 쿼리와 키 패치 토큰 사이에 위치관계를 반영하는 방식을 보여준다. 이때, 일반적인 RPB와 다르게 클래스 토큰과 패치 토큰 사이의 상대적 위치관계는 배제한다.

[0029] 도 3은 본 발명의 실시예에 따른 대표 패치를 선택하는 과정의 예시도를 나타낸다. 도 3을 참조하면, 도 3은 도 1의 B영역을 자세하게 나타낸 예시도로 트랜스포머의 최종 출력을 이용하여 인물 이미지를 대표할 수 있는 패치 토큰을 선택하는 과정을 나타낸다.

[0030] 이미지의 전역적인 특성을 나타내는 클래스 토큰과 내적을 통해 최종 class probabilities를 출력하는데 사용되는 class parameter는 학습과정에서 동일한 클래스 이미지 간에 공유하는 특징을 나타내도록 지도 학습된다.

[0031] 따라서 이미지의 전역적인 특징을 나타내는 클래스 토큰과 패치 토큰사이의 유사도는 class parameter와 패치 토큰 사이의 내적을 통해 얻을 수 있고, class parameter와 유사도가 높은 패치를 선택한다.

[0032] 도 1의 C영역은 앞서 선별된 대표 패치들로부터 패치간 위치관계를 추출하는 단계를 나타낸다.

[0033] 선별된 패치들 사이의 위치관계는 self-attention 매커니즘의 기본공식에서 추출되며, 위치정보를 독립적으로 추출하기위해 [수학식 1]과 같이 정의된다.

[0034] [수학식 1]

$$p_{i,j} = \frac{w_i \cdot w_j + a_{i,j}}{\sqrt{d}}$$

[0035]

[0036] [수학식 1]을 참조하면, w_i, w_j 는 i 번째 패치와 j 번째 패치의 절대 위치 임베딩으로 각 패치의 절대적인 위치를, $a_{i,j}$ 는 패치 쌍 (i,j) 사이의 RPB로 상대적인 위치를 의미한다. d 는 패치 임베딩의 차원을 나타내고, 단일 패치 쌍 (i,j) 의 위치관계는 $p_{i,j}$ 로 나타낸다.

[0037] 이때, RPE에서 $a_{i,j}$ 는 모든 attention head에 적용되므로, 각 패치 쌍마다 $h \times l$ (헤드수 $\times$ 레이어 수)크기의 parameter가 사용된다.

[0038] 도 1의 D영역은 앞서 추출된 패치 쌍들 사이의 위치관계를 하나의 2차원 행렬로 통합하고 정규화하는 단계를 나타낸다.

[0039] 도 1의 실시예에 따라 k 개의 대표 패치가 선택되면 kC_2 개의 패치쌍이 존재한다. 이미지에 내재된 위상 및 위치 관계를 표현하기 위해 kC_2 개의 패치 쌍의 $p_{i,j}$ 를 하나의 2차원 행렬로 연결(Concatenation)한다. 그 후 행 별로 softmax함수를 사용하여 각 패치쌍의 관계의 크기를 정규화하여 하기 [수학식 2]로 표현되는 분포행렬을 얻는다.

[0040] [수학식 2]

$$P = \text{softmax}(\text{concat}(p_{1,2}, p_{1,3}, \dots, p_{(k-1)k}))$$

[0041]

[0042] 앞서 수행되는 정규화 단계는 2차원 행렬을 확률분포 행렬로 변환하고, 각 transformer 계층에 존재하는 attention patten에 맞춰 개편한다.

[0043] 정규화되고 개편된 분포행렬 P 는 대표 패치들을 기반으로 몸의 부위들의 위상 및 위치적 특성을 나타낸다.

[0044] 인물 이미지에 등장하는 신체부위 간의 위치관계만을 고려한 분포행렬 P 는 각 인물의 독특한 특성을 나타낸다. 따라서 하드 트리플렛 마이닝을 통해 얻어진 앵커와 하드-포지티브 샘플의 분포 행렬 P_A 와 P_{BP} 를 유사하게 만

들기 위해 두 확률분포 사이의 Jensen-Shannon divergence(JSD)손실 L_p 를 하기 [수학식 3]과 같이 정의한다.

[0045] [수학식 3]

$$L_p = \sum_{i=1}^I JS(P_A[i] || P_{HP}[i])$$

[0047] [수학식 3]을 참조하면, I 은 총 계층의 수, i 는 각 계층의 index를 의미한다.

[0048] JSD 손실은 두 개의 확률분포 와 가 주어졌을 때 하기 [수학식 4]와 같이 정의되고, 여기서 KL은 Kullback-Leibler divergence(KLD)를 나타낸다.

[0049] [수학식 4]

$$JS(p||q) = \frac{1}{2}KL\left(p||\frac{(p+q)}{2}\right) + \frac{1}{2}KL\left(q||\frac{(p+q)}{2}\right)$$

[0051] 도 1에 표현된 전체 모델은 분포행렬 P에 대한 Jensen-Shannon divergence(JSD) 손실과 인물 재인식 모델에서 주로 사용되는 ID 손실 및 삼중항 손실의 합으로 구성된 손실함수로 학습되고, 이를 수학식으로 나타내면 하기 [수학식 5]와 같다.

[0052] [수학식 5]

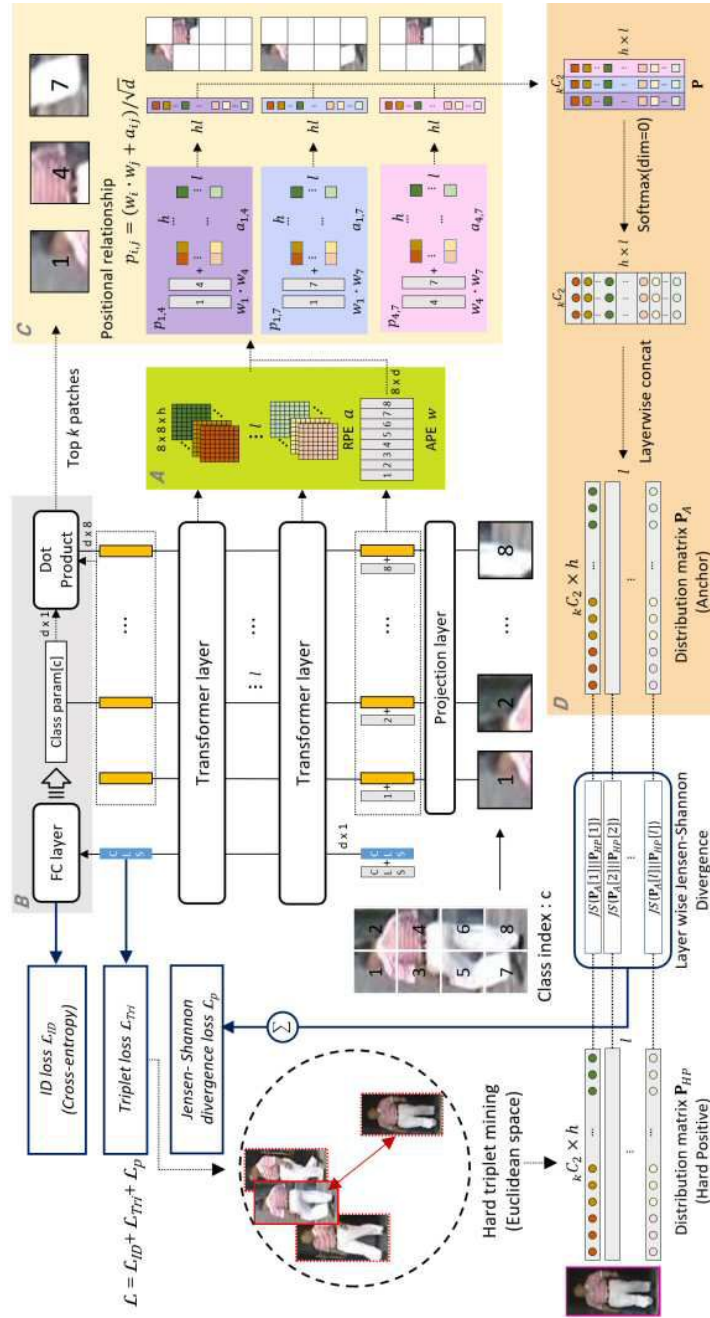
$$L = L_{ID} + L_{Tri} + L_p$$

[0054] [수학식 5]를 참조하면, L 은 제안하는 인물 재인식 모델에 사용되는 최종 손실 함수이다. L_{ID} 는 ID 할당에 대한 교차 엔트로피 손실이다. L_{Tri} 는 배치 하드 마이닝을 사용한 삼중항 손실이다. L_p 는 본 발명에서 제안하는 분포행렬 P 에 대한 JSD 손실이다.

[0055] 이상에서 대표적인 실시예를 통하여 본 발명을 상세하게 설명하였으나, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리 범위는 설명한 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위뿐만 아니라 특허청구범위와 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태에 의하여 정해져야 한다.

도면

도면1



도면2

	CLS	1	2	3
CLS	$C^Q \cdot C^K$	$C^Q \cdot 1^K$	$C^Q \cdot 2^K$	$C^Q \cdot 3^K$
1	$1^Q \cdot C^K$	$1^Q \cdot 1^K$	$1^Q \cdot 2^K$	$1^Q \cdot 3^K$
2	$2^Q \cdot C^K$	$2^Q \cdot 1^K$	$2^Q \cdot 2^K$	$2^Q \cdot 3^K$
3	$3^Q \cdot C^K$	$3^Q \cdot 1^K$	$3^Q \cdot 2^K$	$3^Q \cdot 3^K$

(a)

	1	2	3
1	a_{11}	a_{12}	a_{13}
2	a_{21}	a_{22}	a_{23}
3	a_{31}	a_{32}	a_{33}

(b)

	CLS	1	2	3
CLS	$C^Q \cdot C^K$	$C^Q \cdot 1^K$	$C^Q \cdot 2^K$	$C^Q \cdot 3^K$
1	$1^Q \cdot C^K$	$1^Q \cdot 1^K + a_{11}$	$1^Q \cdot 2^K + a_{12}$	$1^Q \cdot 3^K + a_{13}$
2	$2^Q \cdot C^K$	$2^Q \cdot 1^K + a_{21}$	$2^Q \cdot 2^K + a_{22}$	$2^Q \cdot 3^K + a_{23}$
3	$3^Q \cdot C^K$	$3^Q \cdot 1^K + a_{31}$	$3^Q \cdot 2^K + a_{32}$	$3^Q \cdot 3^K + a_{33}$

도면3

