



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2019년10월04일

(11) 등록번호 10-2028705

(24) 등록일자 2019년09월27일

(51) 국제특허분류(Int. Cl.)
G06T 3/60 (2006.01) G06T 5/00 (2019.01)
G06T 7/60 (2017.01)

(52) CPC특허분류
G06T 3/60 (2013.01)
G06T 5/001 (2013.01)

(21) 출원번호 10-2017-0182797

(22) 출원일자 2017년12월28일

심사청구일자 2017년12월28일

(65) 공개번호 10-2019-0080388

(43) 공개일자 2019년07월08일

(56) 선행기술조사문헌

KR1020160061856 A*

Philipp Fischer et al., Image Orientation Estimation with Convolutional Networks, Pattern Recognition:37th German Conference, GCPR 2015, 368-378(11pages), 2015.10.31.*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

포항공과대학교 산학협력단

경상북도 포항시 남구 청암로 77 (지곡동)

(72) 발명자

홍은빈

울산광역시 중구 변영로 445, 204동 804호(복산동, 효성해링턴 플레이스 2차)

전준호

경상남도 통영시 광도면 덕포로 25

(뒷면에 계속)

(74) 대리인

김합곤, 안광석, 박영일

전체 청구항 수 : 총 8 항

심사관 : 박상철

(54) 발명의 명칭 CNN을 이용한 영상 수평 보정 방법 및 레지듀얼 네트워크 구조

(57) 요약

본 발명은 주어진 입력이 수평이 맞지 않는 영상일 때, 영상의 기울어짐 정도를 측정한 후 이를 이용해 영상을 수평에 맞게 바로 세우는 방법에 관한 것으로, 네트워크 구조의 마지막에, 1x1 크기의 특징 맵을 출력으로 갖는 제1 풀링 레이어; 2x2 크기의 특징 맵을 출력으로 갖는 제2 풀링 레이어; 및 제1 풀링 레이어에서 얻은 특징 벡터와 제2 풀링 레이어에서 얻은 특징 벡터를 결합하여 최종 결과 예측으로 사용하는 각도 예측부;를 포함하는 레

$$\Phi_R(I; \theta_R)$$

지듀얼 네트워크 구조에 따른 각도 측정 네트워크()를 학습시키기 위한 학습 데이터셋을 생성하는

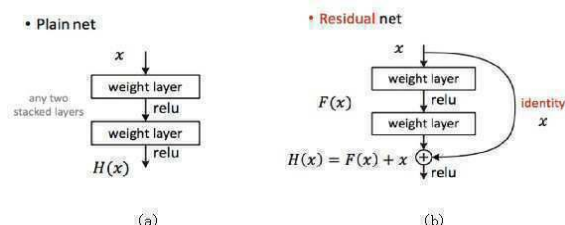
$$\Phi_R \quad \theta_R$$

단계; (b) 학습 세팅을 달리해가며 상기 각도 측정 네트워크 의 최적 파라미터 를 학습하는 단계; 및 (c)

$$\Phi_R$$

입력 영상에 대하여 상기 각도 측정 네트워크 를 이용하여 기울어진 각도를 측정하고, 측정된 기울어진 각도 만큼 반대 방향으로 영상을 회전 시키고, 빈 픽셀 영역을 크롭하는 단계;를 포함하여 구성된다.

대표도



(52) CPC특허분류

G06T 7/60 (2013.01)
G06T 2207/20081 (2013.01)
G06T 2207/20084 (2013.01)
G06T 2207/20132 (2013.01)

이승용

경상북도 포항시 남구 청암로 77 포항공과대학교
 공학4동 121호 (지곡동)

(72) 발명자

조성현

대구광역시 달성군 유가면 테크노대로 165, 213동
 802호(제일풍경채센트럴)

이 발명을 지원한 국가연구개발사업

과제고유번호 R0126-15-1078

부처명 미래창조과학부

연구관리전문기관 정보통신산업진흥원

연구사업명 SW 스타랩

연구과제명 빅 비주얼 데이터 기반의 고품질 사진 메이크업 SW 개발

기 여 율 1/1

주관기관 포항공과대학교 산학협력단

연구기간 2017.03.01 ~ 2017.12.31

공지예외적용 : 있음

명세서

청구범위

청구항 1

네트워크 구조의 마지막에 각각 연결되는, 영상 전체의 특징을 추출하는 1x1 크기의 특징 맵을 출력으로 갖는 제1 풀링 레이어와, 상기 영상을 동일한 크기로 4등분한 각 영역의 특징을 추출하는 2x2 크기의 특징 맵을 출력으로 갖는 제2 풀링 레이어; 및

제1 풀링 레이어에서 얻은 특징 벡터와 제2 풀링 레이어에서 얻은 특징 벡터를 결합하여 최종 결과 예측으로 사용하는 각도 예측부;를 포함하는 것을 특징으로 하는 레지듀얼 네트워크 구조.

청구항 2

제1항에 있어서, 상기 각도 예측부는

2개의 특징 벡터의 크기를 통일하고, 크기가 통일된 2개의 특징 벡터의 평균 평균값을 산출하여 최종 결과 예측으로 사용하는 것을 특징으로 하는 레지듀얼 네트워크 구조.

청구항 3

(a) 제1항의 레지듀얼 네트워크 구조에 따른 각도 측정 네트워크($\Phi_R(I; \theta_R)$)를 학습시키기 위한 학습 데이터셋을 생성하는 단계;

(b) 학습 세팅을 달리해가며 상기 각도 측정 네트워크 Φ_R 의 최적 파라미터 θ_R 를 학습하는 단계; 및

(c) 입력 영상에 대하여 상기 각도 측정 네트워크 Φ_R 를 이용하여 기울어진 각도를 측정하고, 측정된 기울어진 각도만큼 반대 방향으로 영상을 회전 시키고, 빈 픽셀 영역을 크롭하는 단계;를 포함함을 특징으로 하는 CNN을 이용한 영상 수평 보정 방법.

청구항 4

제3항에 있어서, 상기 (a) 단계는

바로 세워진 다수의 일반 영상에 대하여 회전된 영상과 그때의 회전된 각도를 레이블로 배정하여 학습 데이터셋을 생성하는 것을 특징으로 하는 CNN을 이용한 영상 수평 보정 방법.

청구항 5

제4항에 있어서, 학습 데이터셋 생성을 위한 영상을 회전 시키기 전에 가로, 세로 길이의 1/2만큼 대칭적 패딩(symmetric padding)을 하는 것을 특징으로 하는 CNN을 이용한 영상 수평 보정 방법.

청구항 6

제3항에 있어서, 상기 (a) 단계 이전에

상기 각도 측정 네트워크 Φ_R 를 ImageNet 영상 분류 데이터셋으로 학습시킨 후, 상기 각도 측정 네트워크 Φ_R 의 후반부에 위치한 소정 수의 레이어들의 가중치 파라미터들을 초기화 하는 것을 특징으로 하는 CNN을 이용한 영상 수평 보정 방법.

청구항 7

제3항 내지 제6항 중 어느 한 항의 CNN을 이용한 영상 수평 보정 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체.

청구항 8

제3항 내지 제6항 중 어느 한 항의 CNN을 이용한 영상 수평 보정 방법을 컴퓨터에서 실행시키기 위하여 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 주어진 입력이 수평이 맞지 않는 영상일 때(카메라의 시선 방향 축을 기준으로 한 회전인 롤(roll) 회전이 포함된 영상일 때), 영상의 기울어짐 정도를 측정한 후 이를 이용해 영상을 수평에 맞게 바로 세우는 방법에 관한 것이다.

배경 기술

[0002] 영상의 수평 보정 여부는 사진을 미학적으로 평가할 때 중요한 역할을 차지한다. 예를 들어 도 1의 왼쪽 영상(도 1의 (a))같이 카메라가 기울어져 찍힌 사진은 선명도, 색이 아무리 좋아도 미학적으로 좋지 않아 보인다. 이에 비해 도 1의 오른쪽 영상(도 1의 (b))은 영상의 수평이 잘 맞기 때문에 미학적으로 좋은 영상이라 할 수 있다.

[0003] 영상 수평 보정 기술은 실제 사용자가 디지털 카메라나 스마트폰으로 사진을 찍을 때 가장 필요로 하는 기술 중 하나이다. 기존의 영상 수평 보정 기술은 입력 영상 내에 존재하는 선 정보에 기반하여 기울어져 찍힌 영상을 바로 세워 주는데 이러한 방법들은 뚜렷한 선 정보가 없는 인물 사진의 경우에는 제대로 동작하지 않는다.

[0004] 본 발명은 깊은 신경망 기술을 이용하여 영상의 수평을 보정함으로써 실제 사용자들이 일상 생활 속에서 촬영한 다양한 장면의 사진들에 대해 강인하게 동작할 수 있다. 이와 같은 강인한 영상 수평 보정 기술은 다양한 어플리케이션에 유용하게 적용 될 수 있는 기술이다.

[0005] 영상의 수평 보정과 관련하여 지금까지 많은 연구가 진행되어 왔다.

[0006] (비특허문헌 1)은 영상 내의 선 정보를 분석해 이를 이용하여 영상을 바로 세워주는 방법을 제시하였다. 그러나 영상 내 존재하는 선 정보에 기반하기 때문에 인공 구조물이 많이 포함된 사진이 아닌 경우나 뚜렷한 선이 없는 사진(e.g. 사람, 자연물)의 경우에는 잘 동작하지 않는다는 한계점이 있다.

[0007] (비특허문헌 2)는 깊은 신경망 기술을 이용하여 세 가지 어려움 단계($\pm 30^\circ$, $\pm 40^\circ$, $\pm 360^\circ$ 사이의 각도)에 대해 기울어진 각도 측정을 수행하였다. 이 방법은 최근 뛰어난 성능을 보이고 있는 깊은 신경망 기술을 영상 수평 보정 기술에 적용했다는 데에 큰 의의가 있지만, 단순한 구조의 학습 네트워크를 사용하여 수평 보정 결과 영상이 $3^\circ \sim 20^\circ$ 정도의 큰 오차를 보이기 때문에 실제 실용성은 떨어진다.

선행기술문헌

비특허문헌

[0008] (비특허문헌 0001) Hyunjoon Lee, Eli Shechtman, Jue Wang, and Seungyong Lee. Automatic upright adjustment of photographs. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 877-884. IEEE, 2012.

(비특허문헌 0002) Philipp Fischer, Alexey Dosovitskiy, and Thomas Brox. Image orientation estimation with convolutional networks. In German Conference on Pattern Recognition, pages 368-378. Springer, 2015.

(비특허문헌 0003) Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770-778, 2016.

(비특허문헌 0004) Giorgos Toliás and Yannis Avrithis. Speeded-up, relaxed spatial matching. In Computer Vision (ICCV), 2011 IEEE International Conference on, pages 1653-1660. IEEE, 2011.

발명의 내용

해결하려는 과제

- [0009] 본 발명은 주어진 입력이 기울어져 찍힌 영상일 때 CNN을 기반으로 영상의 기울어짐 정도를 측정한 후 이 각도에 대해 회전시킴으로써 수평 보정이 된 결과영상을 얻을 수 있는 CNN을 이용한 영상 수평 보정 방법 및 그 방법에 사용되는 레지듀얼 네트워크 구조를 제공함을 그 목적으로 한다.

과제의 해결 수단

- [0010] 상기의 목적을 달성하기 위하여, 본 발명에 의한 레지듀얼 네트워크 구조는, 네트워크 구조의 마지막에, 1x1 크기의 특징 맵을 출력으로 갖는 제1 풀링 레이어; 2x2 크기의 특징 맵을 출력으로 갖는 제2 풀링 레이어; 및 제1 풀링 레이어에서 얻은 특징 벡터와 제2 풀링 레이어에서 얻은 특징 벡터를 결합하여 최종 결과 예측으로 사용하는 각도 예측부;를 포함하는 것을 특징으로 한다.
- [0011] 상기 레지듀얼 네트워크 구조에 있어서, 상기 각도 예측부는 2개의 특징 벡터의 크기를 통일하고, 크기가 통일된 2개의 특징 벡터의 평균 평균값을 산출하여 최종 결과 예측으로 사용하는 것을 특징으로 한다.
- [0012] 상기의 목적을 달성하기 위하여, 본 발명에 의한 CNN을 이용한 영상 수평 보정 방법은 상기 레지듀얼 네트워크 구조에 따른 각도 측정 네트워크를 학습시키기 위한 학습 데이터셋을 생성하는 단계; (b) 학습 세팅을 달리해가며 상기 각도 측정 네트워크의 최적 파라미터를 학습하는 단계; 및 (c) 입력 영상에 대하여 상기 각도 측정 네트워크를 이용하여 기울어진 각도를 측정하고, 측정된 기울어진 각도만큼 반대 방향으로 영상을 회전 시키고, 빈 픽셀 영역을 크롭하는 단계;를 포함하여 구성된다.
- [0013] 상기 CNN을 이용한 영상 수평 보정 방법에 있어서, 상기 (a) 단계는 바로 세워진 다수의 일반 영상에 대하여 회전된 영상과 그때의 회전된 각도를 레이블로 배정하여 학습 데이터셋을 생성하는 것을 특징으로 한다.
- [0014] 상기 CNN을 이용한 영상 수평 보정 방법에 있어서, 학습 데이터셋 생성을 위한 영상을 회전 시키기 전에 가로, 세로 길이의 1/2만큼 대칭적 패딩(symetric padding)을 하는 것을 특징으로 한다.
- [0015] 상기 CNN을 이용한 영상 수평 보정 방법에 있어서, 상기 (a) 단계 이전에 상기 각도 측정 네트워크를 ImageNet 영상 분류 데이터셋으로 학습시킨 후, 상기 각도 측정 네트워크를 후반부에 위치한 소정 수의 레이어들의 가중치 파라미터들을 초기화 하는 것을 특징으로 한다.
- [0016] 상기의 목적을 달성하기 위하여, 본 발명에 의한 컴퓨터로 읽을 수 있는 기록 매체는 상기 CNN을 이용한 영상 수평 보정 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한다.
- [0017] 상기의 목적을 달성하기 위하여, 본 발명에 의한 컴퓨터 프로그램은 상기 CNN을 이용한 영상 수평 보정 방법을 컴퓨터에서 실행시키기 위하여 매체에 저장된다.

발명의 효과

- [0019] 본 발명에 의하면, 영상 수평 보정을 위한 영상의 기울어짐의 정도를 측정함에 있어서, 선분과 같이 기울어짐을 판단하기 위해 미리 정의된 특징이 존재하지 않는 영상이나, 영상의 기울어짐을 판단할 수 있는 정보가 특정 영역에만 존재하는 영상의 경우에도 그 정확성을 훨씬 높일 수 있게 된다.

도면의 간단한 설명

- [0020] 도 1은 수평이 맞지 않는 영상과 수평이 잘 맞는 영상의 예를 보여주는 도면이다.
- 도 2는 CNN을 이용한 영상 수평 보정 방법의 전체적인 구조를 보여주는 도면이다.
- 도 3은 영상 수평 보정 방법이 수행되는 과정을 보여주는 도면이다.
- 도 4는 2개의 레이어를 갖는 일반적인 CNN 구조(a)와 2개의 레이어에 대해 skip connection 구조가 추가된 레지듀얼 네트워크 구조(b)를 보여주는 도면이다.
- 도 5는 34 레이어로 구성된 일반 CNN 구조(a)와 residual network 구조(b)를 보여주는 도면이다.
- 도 6은 본 발명에 의한 기울어진 각도 측정을 위한 레지듀얼 네트워크 구조를 도시한 것이다.
- 도 7은 대칭적 패딩을 이용하여 회전된 영상을 만드는 과정을 보여주는 도면이다.

도 8은 한 영상에 대해 7가지 임의의 각도로 회전시킨 결과를 보여주는 도면이다.

도 9는 영상 내 주요 선 존재 여부를 판별하는 과정을 보여주는 도면이다.

도 10은 본 발명에 의한 CNN을 이용한 영상 수평 보정 결과를 보여주는 도면이다.

도 11은 기존 상용 기술 중 하나인 Adobe Lightroom 과 본 발명에 의한 방법의 결과를 비교하여 보여주는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0021] 이하에서, 첨부된 도면을 참조하면서 본 발명의 바람직한 실시예에 대하여 상세히 설명하기로 한다.

[0023] <알고리즘 개요>

[0024] 본 발명에서 제안하는 영상 수평 보정 기술의 전체 과정은 도 2와 같다. 기울어진 영상 I 에 대한 각도 측정 네트워크를 $\Phi_R(I; \theta_R)$ 라 정의하면, 네트워크의 학습은 크게 세 단계로 구성된다.

[0025] 우선 학습할 네트워크 Φ_R 의 구조를 구성한다. 네트워크 구조는 비특허문헌 3에서 제안한 레지듀얼 네트워크(residual network)를 기반으로 하여 두 갈래로 나뉘어진 풀링(pooling) 레이어를 추가하여 새롭게 구성한다.

[0026] 그 후 네트워크 Φ_R 을 학습시키기 위해 학습 데이터셋을 생성하는 것이 두 번째 단계인데, 바로 세워져 있다고 판단한 대규모 영상들을 임의의 각도로 회전시킨 후 회전된 영상과 그 때 회전시킨 각도를 레이블로 하여 학습 데이터셋을 구성한다.

[0027] 마지막으로 네트워크와 학습 데이터셋이 준비되면 학습 세팅을 달리해가며 네트워크 Φ_R 의 최적 파라미터 θ_R 를 학습한다. 입력 영상이 네트워크의 입력으로 들어왔을 때 수행되는 과정은 도 3에서 표시된 과정을 거치게 된다.

[0029] <네트워크 구조>

[0030] 본 발명에서 사용한 네트워크 구조는 물체 인식에 뛰어난 효과를 보이는 레지듀얼 네트워크(residual network)에 기반하고 있다.

[0031] 기존의 CNN(Convolutional Neural Network) 구조들은 역전파(backpropagation) 기법을 이용하여 학습을 진행하는데 CNN 구조가 깊을수록 전 레이어의 그라디언트(gradient) 정도에 따라서 다음 레이어의 그라디언트 값이 지수적으로 작아지는 현상이 발생할 수 있다. 따라서 출력 레이어 근처의 그라디언트들은 값이 존재하지만, 입력 레이어에서는 그 값이 0에 가까워지는 문제가 발생해 학습이 더 이상 진행되지 않는 문제가 발생한다(gradient vanishing problem).

[0032] 레지듀얼 네트워크(residual network)는 이러한 gradient vanishing 문제를 해결하여 깊은 네트워크 구조의 효과적인 학습을 가능하게 한 구조이다. 이 네트워크에서는 skip connection이라는 구조를 이용하여 입력 값을 몇 개의 레이어에 통과시킨 출력 값(도 4의 $F(x)$)과 입력 값(도 4의 identity)의 합을 계산하여 결과 값을 얻을 수 있도록 구성한다.

[0033] 도 4의 (a)가 기존의 일반적인 네트워크 구조인데, 이러한 구조일 때 두 레이어를 거친 후 매핑된 결과는 $H(x)$ 로 표현할 수 있다. 도 4의 (b)는 레지듀얼 네트워크(residual network) 구조인데, 2개의 레이어를 건너 뛴

skip connection을 활용한 것이 특징이다. 여기서 $F(x) + x$ 로 표현할 수 있다. 이러한 구성은 그라디언트를 계산하는데 있어 그 값이 너무 작아지지 않게 해주기 때문에 기존 CNN에서 발생되던 gradient vanishing 문제를 해결할 수 있다. 따라서 보다 많은 레이어의 모델을 잘 학습시킬 수 있고 적은 수의 레이어를 갖는 모

델로도 높은 성능을 얻을 수 있다.

[0034] 본 발명에서는 도 5의 (b)에 도시된 바와 같은 34개의 레이어로 구성된 레지듀얼 네트워크(residual network)를 사용한다. 일반적으로 레지듀얼 네트워크(residual network)는 입력 영상에 존재하는 물체 인식에 사용되기 때문에 입력 영상 전체의 정보를 한번에 이용하며, 이를 위한 구조가 도 5의 네트워크 구조의 마지막에 존재하는 하나의 애버리지 풀링(average pooling) 레이어다.

$$\Phi_R$$

[0035] 도 6은 본 발명에서 입력 영상의 기울어진 각도를 측정하는 네트워크 Φ_R 의 구조를 보여준다. 이를 위해 사용한 네트워크 구조는 기존의 레지듀얼 네트워크(residual network) 구조를 기반으로 한다.

[0036] 그런데, 기존의 레지듀얼 네트워크(residual network) 구조에서 가장 마지막 풀링(pooling) 레이어는 최종 특징 맵을 1x1로 만들어 주는 역할을 하는데, 본 발명에서는 이 레이어를 두 개의 풀링(pooling) 레이어로 대체하여 특징 맵을 각각 1x1, 2x2로 만든 후 이를 결합한 벡터를 최종 결과 예측에 사용한다. 이를 통해 입력 영상의 다중 스케일 특징을 고려할 수 있게 된다.

[0037] 두 풀링(pooling) 레이어는 각각 1x1, 2x2 크기의 특징 맵을 출력으로 갖는데, 1x1 특징 맵은 영상 전체의 특징을 추출하고, 2x2 특징 맵은 각 셀(cell)이 영상을 동일한 크기로 4등분한 각 영역의 특징을 추출한다. 두 풀링(pooling) 레이어로부터 얻은 특징 벡터는 그 크기가 서로 다르기 때문에 선형 레이어를 통해 256 크기로 통일한 뒤, 두 벡터의 평균값을 취한다. 마지막으로 선형 레이어를 통해 영상의 최종 기울어짐 각도를 예측하도록 한다. 예측 정확도를 측정하는 로스(loss) 함수로는 L1 loss(absolute difference)를 사용한다.

[0038] 위와 같이 두 개의 (pooling) 레이어를 통해 추출한 다중 스케일의 특징을 이용할 경우, 1x1 특징 맵으로부터 얻은 정보는 영상 전체 기울어짐의 구조적인 정보를 표현하고, 2x2 특징 맵의 각 셀(cell)은 영상의 서로 다른 영역에 대한 특징 정보를 표현하는데, 이후의 fully connected 레이어를 통해 서로 다른 영역 간의 위치 관계가 분석되어 영상 기울어짐 측정에 도움이 될 수 있다.

[0039] 예를 들어 사람의 얼굴 영상의 경우 얼굴 전체 외곽선의 기울어짐이 영상의 기울어짐에 대한 특징 정보가 될 수도 있지만, 양 눈 사이의 관계, 혹은 코와 입이 이루는 상대적인 위치 관계에 대한 정보가 영상의 기울어짐 측정에 단서가 될 수 있는 것이다. 이와 같이 다중 스케일 특징을 고려하는 경우, 선분과 같이 기울어짐을 판단하기 위해 미리 정의된 특징이 존재하지 않는 다양한 영상의 경우에도 많은 데이터를 통해 영상 영역간의 위치 관계를 학습함으로써 정확한 영상 기울어짐을 측정할 수 있다.

[0040] 더불어 2x2로 만들어진 특징 맵은 영상의 상하좌우 각 영역의 정보를 독립적으로 판단하기 때문에 영상의 기울어짐을 판단할 수 있는 정보가 특정 영역에만 존재할 경우에도 해당 정보가 존재하지 않는 영역에 의해 방해받지 않으며, 영상 내 다양한 크기의 특징들을 동시에 학습할 수 있기 때문에 보다 강인하게 최종 기울어진 각도 측정 성능을 높일 수 있다.

[0042] <네트워크 학습>

$$\Phi_R$$

[0043] 영상 기울어짐 측정 네트워크 Φ_R 를 학습하기 위해 우선 학습 데이터셋을 생성한다. 학습 데이터셋은 이미 바로 세워져 있는 영상들을 임의의 각도로 회전시킨 뒤 이때의 각도를 레이블로 사용하는 방식으로 생성한다.

[0045] <데이터셋 생성>

$$\Phi_R$$

[0046] 영상 기울어짐 측정 네트워크 Φ_R 를 학습하기 위한 영상 데이터셋을 만드는 과정은 도 7과 같다. 기울어지지 않고 바로 세워져 있는 영상에 대해 $-20^\circ \sim +20^\circ$ 사이의 각도 7개를 무작위로 선택하여 그 각도만큼 회전시킨다. 이 때, 영상을 회전시키면 빈 픽셀 영역이 발생하게 되는데 이 부분을 채우기 위해 영상을 회전시키기 전에 가로, 세로 길이의 1/2만큼 대칭적 패딩(symmetrical padding)을 한다. 그 후 영상을 특정 각도만큼 회전시키고, 다시 원래 크기만큼 크롭하면 빈 픽셀 영역이 없는 사각형 꼴의 회전된 영상을 얻을 수 있다.

[0047] 최종 학습 데이터셋은 대칭적 패딩을 이용하여 얻은 회전된 영상과 그때의 회전된 각도를 레이블로 배정하여 구성한다. 한 영상에 대해 임의의 7가지 각도로 회전하여 얻은 데이터셋 예제는 도 8에서 보여진다. 회전되지 않은 원본 영상도 레이블을 0으로 하여 학습 데이터셋에 포함시킨다.

[0048] 기울어진 영상 데이터셋을 생성하기 위한 원본 영상 데이터셋은 비특허문헌 4에서 소개된 World Cities Dataset 이다. Flickr 웹사이트로부터 40개의 주요 도시들의 지리학적 쿼리(geographic query)를 이용하여 수집한 220 만장의 영상들로 구성되어 있는데, 본 발명에서는 이 중 일부인 21,994장을 학습 데이터셋, 1,000장은 검증 데이터셋으로 사용하였다. World Cities Dataset의 영상들 중에는 기울어져 찍힌 영상들도 소수 포함되어 있으나, 대부분은 바로 세워져 있기 때문에 학습 시 전체적인 경향에 큰 영향이 없을 것이라 가정하였고, 실제로 실험 결과를 통해 동작이 잘 됨을 확인할 수 있었다. 검증 데이터셋은 직접 눈으로 확인하여 기울어져 있는 영상들은 제외시켰다(865장).

[0049] 본 발명에서는 선분과 같이 기울어짐을 판단하는데 도움이 되는 정보가 부족한 영상에 대해서도 강건하게 동작 하는 기울어짐 보정을 목표로 하였기 때문에 학습 데이터셋 내에 주요 선이 존재하는 영상(건물, 나무 등) 외에도 주요 선이 존재하지 않는 영상(인물, 자연 풍경 등)도 충분히 포함되도록 각 그룹의 비율을 조정하였다. 이를 위해 영상에서 찾은 선분을 클러스터링 한 뒤 선분의 길이의 합이 가장 긴 클러스터의 중심을 주요 선분으로 검출했을 때, 만약 검출된 주요 선이 영상의 높이, 너비 중 짧은 쪽의 1/3 보다 짧다면 이 영상에서 주요 선은 존재하지 않는다고 가정한다(도 9 참조). 이를 통해 최종 학습 데이터셋을 주요 선이 존재하는 영상과 주요 선이 존재하지 않는 영상을 각각 2:1 비율로 구성하였다.

[0051] <영상 기울어짐 측정 네트워크 학습>

[0052] 학습 네트워크는 ImageNet 영상 분류 데이터셋에 대해 미리 학습된 레지듀얼 네트워크(residual network) 모델을 사용하였다. ImageNet 분류 문제는 대상의 기울어짐에 대해 강건하게 분류할 수 있도록 네트워크가 학습된다. 따라서 미세한 기울어짐 정보를 구분할 수 있어야 하는 본 알고리즘의 목적과는 반대된다고 할 수 있다. 이를 보완하기 위해 네트워크의 후반부 레이어들(9개의 residual block)의 가중치 파라미터들을 초기화 한 후, 앞에서 설명한 학습 데이터셋으로 다시 학습하였다. 이를 통해 네트워크의 초반부 레이어들에서는 ImageNet 영상 분류 데이터셋으로부터 학습된 영상의 저수준 특징을 추출하고, 후반부 레이어들에서는 기울어짐을 측정할 수 있는 구조적 정보나 영역 간의 상대적 위치 관계 등이 새롭게 학습되도록 유도하였다.

[0054] <결과 영상 및 정량적 평가>

[0055] 본 발명은 CNN을 이용한 영상 수평 보정 방법을 제안한다. 학습 과정을 통해 생성한 네트워크 Φ_R 은 입력 영상 I 의 기울어진 각도를 측정하기 위해 사용된다. 이렇게 얻은 각도와 반대 방향으로 입력 영상 I 를 회전시키면 기울어짐이 보정된 결과 영상을 생성할 수 있다. 도 10은 다양한 입력 영상을 본 발명의 프레임워크에 적용시킨 결과이다. 본 방법의 결과는 영상 내 직선이 많은 경우뿐만 아니라 주요 선이 존재하지 않는 복잡한 장면에 대해서도 잘 동작하는 것을 확인할 수 있다. 또한 배경이 어두운 영상에 대해서도 강인하게 동작한다.

[0056] 기울어짐 측정 네트워크 Φ_R 의 정확도를 정량적으로 평가하고, 영상의 기울어진 정도에 따라 측정 정확도가 어떻게 변하는지 평가하기 위해 865장의 다양한 영상으로 이루어진 검증 데이터셋을 다섯 가지 각도(3° , 5° , 10° , 15° , 20°)에 대해 (+, -) 방향으로 회전한 후, 각 기울어진 각도에 대해 정확도를 측정하였다. 결과는 표 1에서 보여진다.

표 1

GT	3°	5°	10°	15°	20°	average
error	1.2268	0.6524	0.4596	0.4659	0.1532	0.7915

[0057]

[0058] 표 1에서의 오차는 본 방법을 통해 측정된 각도와 GT(ground truth) 각도의 차이 값의 평균이다. 실험 결과, 다섯 가지 경우 모두 1° 내외의 오차를 보일 정도로 높은 정확도를 갖는 것을 확인 할 수 있다.

[0060] <네트워크 구조에 따른 성능 비교>

$$\Phi_R$$

[0061] 본 발명에서 제안한 기울어짐 측정 네트워크 Φ_R 의 구조는 영상의 다중 스케일 특징을 동시에 고려하기 때문에 단일 스케일 특징만을 고려한 네트워크 구조에 비해 높은 성능을 얻을 수 있다. 이를 확인하기 위해 단일 스케일 특징만을 고려하는 네트워크를 동일한 조건에서 학습한 후 다중 스케일 네트워크와의 성능 비교를 수행하였다. 단일 스케일 네트워크 구조는 도 6에서 보여지는 다중 스케일 네트워크 구조에서 2개의 pooling 레이어를 1개의 pooling 레이어로 대체하여 512x1x1의 특징맵을 생성한 후 선형 레이어를 통해 512 크기의 특징을 거쳐 최종 기울어짐 각도를 측정하도록 구성하였다. 이를 본 알고리즘과 동일한 데이터셋으로 동일한 epoch 만큼 학습하고, 결과를 비교하였다. 전체 검증 데이터셋에 대해 단일 스케일, 다중 스케일 네트워크 각각을 이용하여 기울어짐 각도를 측정한 뒤 GT(ground truth) 각도와 평균 오차를 측정, 학습 epoch가 진행됨에 따라 두 네트워크의 성능 변화를 비교하였다.

[0062] 비교 결과는 [표 2]와 같다.

표 2

	Epoch 4	Epoch 8	Epoch 12	Epoch 16
Single	1.7836	1.4095	1.0359	0.8730
Multi	1.8081	1.2921	0.9585	0.7915

[0063]

[0064] 비교 결과, 실험 초반 epoch에서 단일 스케일 네트워크가 더 높은 정확도를 보이는 현상이 관찰되었는데, 이는 다중 스케일 네트워크의 경우 단일 스케일 네트워크에 비해 학습해야 할 특징의 종류, 파라미터의 개수가 더 많기 때문이다. 그러나 학습이 계속 진행됨에 따라 다중 스케일 네트워크의 파라미터 학습이 충분히 이루어지고 결과적으로 더 높은 정확도를 보이는 것을 확인할 수 있다.

[0066] <데이터셋 구성에 따른 결과 비교>

[0067] 딥러닝 기반의 영상 분석 알고리즘은 네트워크 학습 시 사용하는 데이터셋을 어떻게 구성하느냐에 따라 학습 결

$$\Phi_R$$

과가 달라질 수 있다. 본 장에서는 학습 데이터셋을 세 가지 경우로 다르게 구성하여 네트워크 Φ_R 를 학습했을 때 검증 데이터셋에 대한 기울임 측정 정확도가 어떻게 차이 나는지 비교한다. 이 실험에 사용한 네트워크 구조는 앞에서 설명한 다중 스케일 네트워크이다.

[0068] 첫 번째는 주요 선이 존재하는 영상들(건물, 나무 등이 존재하는 영상들) 만으로 학습 데이터셋을 구성한 경우이다. 학습 데이터셋 구성 시, 도 9를 통해 설명한 영상 내 주요 선 존재 여부 판별 알고리즘을 이용하여 주요

$$\Phi_R$$

선이 존재하는 영상들로만 구성된 데이터셋으로 네트워크 Φ_R 를 학습하였다. 이렇게 학습된 모델은 선분 등의 구조적 특징을 갖고 있는 영상들에 대해서는 잘 동작할 것으로 예상할 수 있다.

[0069] 두 번째는 주요 선이 존재하지 않는 영상들(인물 중심 또는 자연 풍경 배경의 영상들) 만으로 데이터셋을 구성한 경우이다. 이 역시 주요 선 존재 여부 판별 알고리즘을 사용해 주요 선이 존재하지 않는 영상들로만 구성된

$$\Phi_R$$

학습 데이터셋을 생성하여 네트워크 Φ_R 를 학습하였는데, 이 모델은 선이 존재하는 영상들에 대해서는 잘 작동하지 않을 것으로 예상할 수 있다.

$$\Phi_R$$

[0070] 마지막으로 주요 선이 존재하는 영상과 주요 선이 존재하지 않는 영상을 각각 2:1로 구성하여 네트워크 Φ_R 를 학습하였다.

표 3

	Line test set	No-Line test set
Train with line images	1.0538	1.4014
Train with no-line images	1.9953	2.0435
Train with both images	0.7261	0.8789

[0071]

[0072]

실험 결과, 표 3에서 볼 수 있듯이 주요 선이 존재하는 영상과 존재하지 않는 영상을 모두 포함시킨 원본 데이터 셋이 가장 정확도가 높은 것으로 나타났다. 이는 영상으로부터 얻을 수 있는 선분 정보만 이용하기보다 선분이 없는 일반적인 영상에서 얻을 수 있는 다양한 피사체의 특징으로부터 얻는 추가적인 정보를 활용해 기울어짐을 측정할 때, 전체 검증 데이터 셋에 대한 정확도가 높아진다는 것을 의미한다. 또한 주요 선분이 존재하거나 하지 않는 학습 데이터 셋으로 학습한 두 네트워크 모두 선분이 존재하는 검증 데이터 셋에 대한 정확도가 선분이 존재하지 않는 데이터 셋에 비해 높았는데, 이는 단일 정보로 사용될 때 기울어진 선분 정보가 일반적인 피사체의 기울어짐 정보보다 영상의 기울어짐을 측정하기에 유리하다는 것을 뜻한다.

[0074]

<기존 상용 기술과의 비교>

[0075]

마지막으로 본 발명에서 제안한 방법과 기존의 상용 사진 기울어짐 보정 기술의 결과를 비교하였다. 본 방법과 비교한 기존 상용 기술은 Adobe Lightroom CC 2015에 내장된 기능이다.

[0076]

비교 결과는 도 11에서 보여진다. 첫 번째 예제 영상은 영상 내에 다수의 건물들이 위치하고 있기 때문에 선 검출이 용이하다. 그 결과 본 발명의 방법뿐만 아니라 기존 기술의 수행 결과 모두 기울어짐 보정이 잘 되는 것을 확인 할 수 있다. 반면 두 번째 예제 영상은 배경의 수평선을 중심으로 회전시켜야 할 지 주요 물체에 맞춰 회전시켜야 할 지 모호하다. 본 방법은 배경의 수평선을 기준으로 보정을 하여 GT 영상과 비슷한 결과를 얻은 반면 기존 기술은 결과가 좋지 않다. 나머지 예제 영상들의 경우 영상 내 주요 선이 존재하지 않거나 검출하기 어려운 영상으로, 기존 기술의 경우 제대로 동작하지 않는 반면 본 방법은 사진의 피사체인 인물 또는 동물이 똑바로 세워지도록 기울어짐 보정을 잘 수행하는 것을 확인할 수 있다.

[0078]

한편, 상술한 본 발명의 실시예는 개인용 컴퓨터를 포함한 범용 컴퓨터에서 사용되는 매체에 기록될 수 있다. 상기 매체는 마그네틱 기록매체(예를 들면, 롬, 플로피 디스크, 하드 디스크 등), 광학적 판독매체(예를 들면, 씨디롬, 디브이디 등) 및 전기적 기록매체(예를 들면, 플래쉬 메모리, 메모리 스틱 등)와 같은 기록매체를 포함한다.

[0079]

이제까지 본 발명에 대하여 그 바람직한 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예는 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

도면1

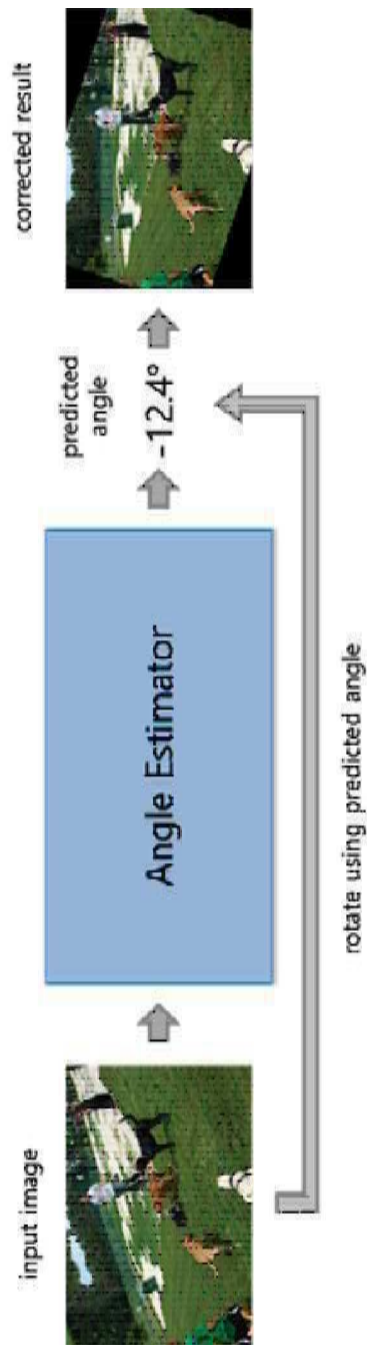


(a)

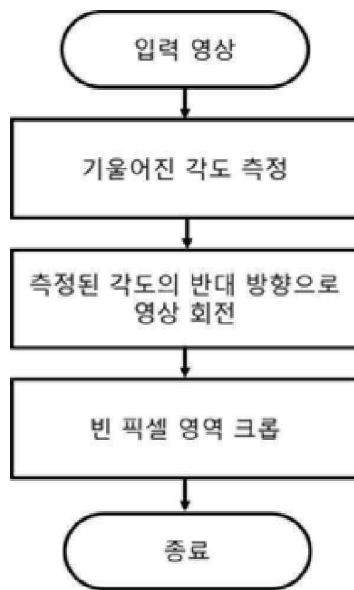


(b)

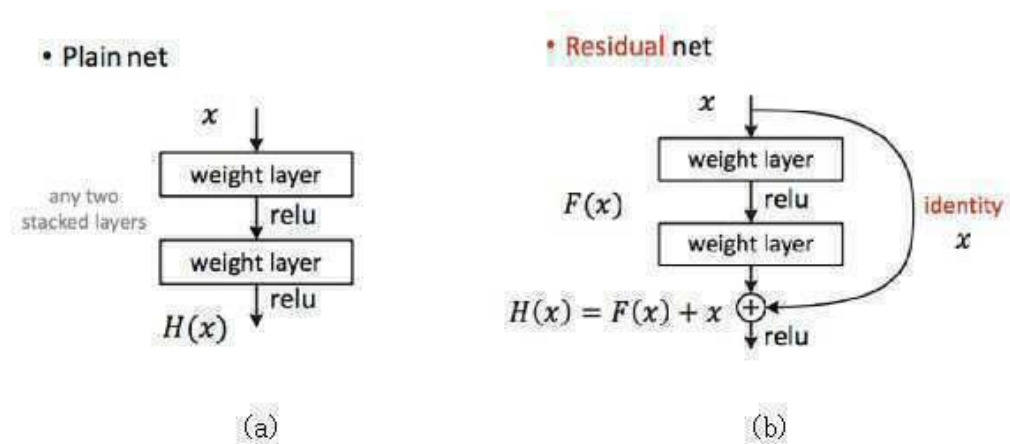
도면2



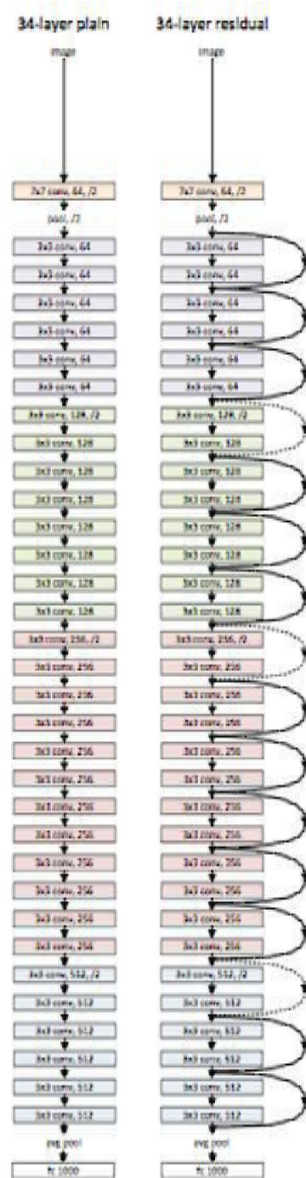
도면3



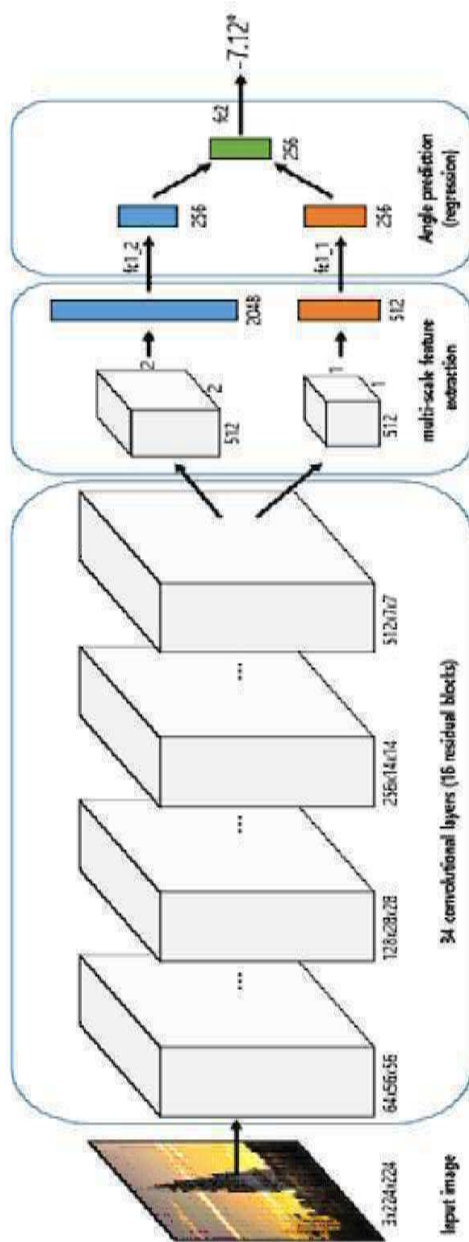
도면4



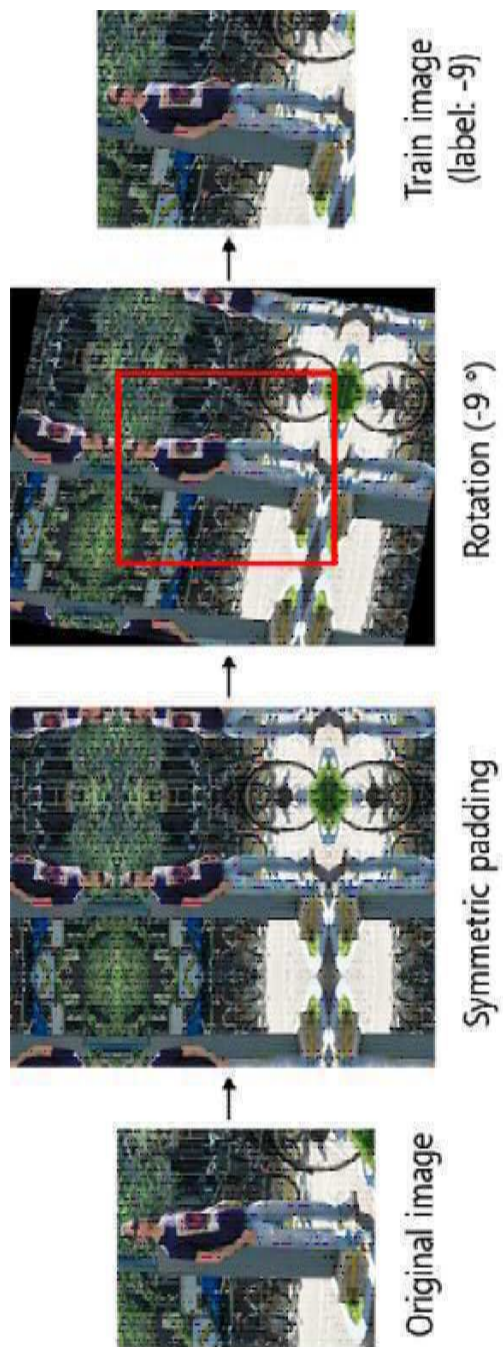
도면5



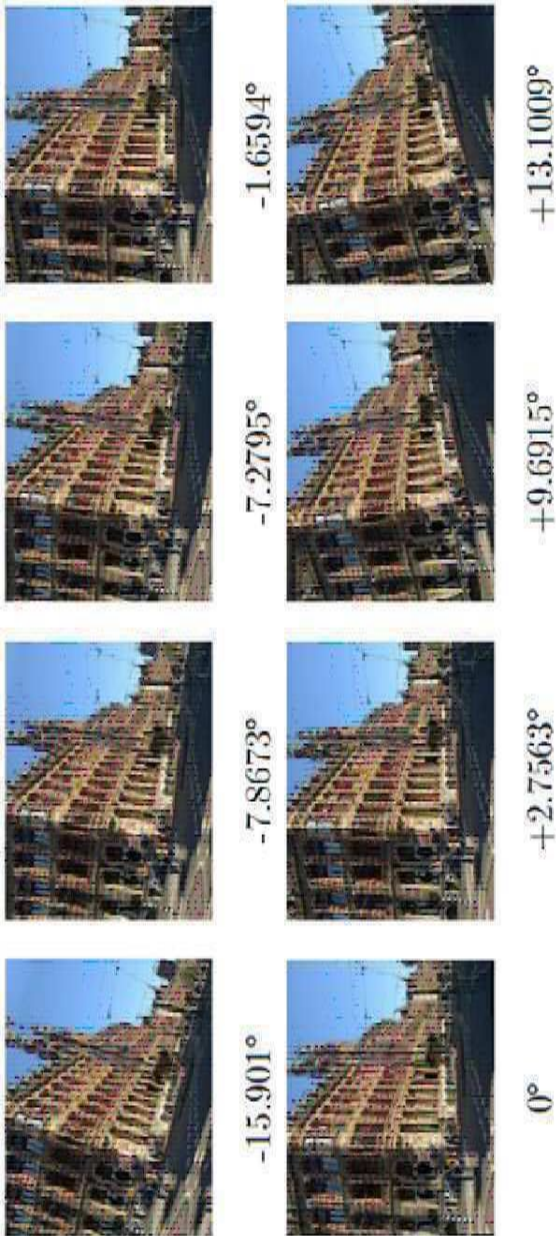
도면6



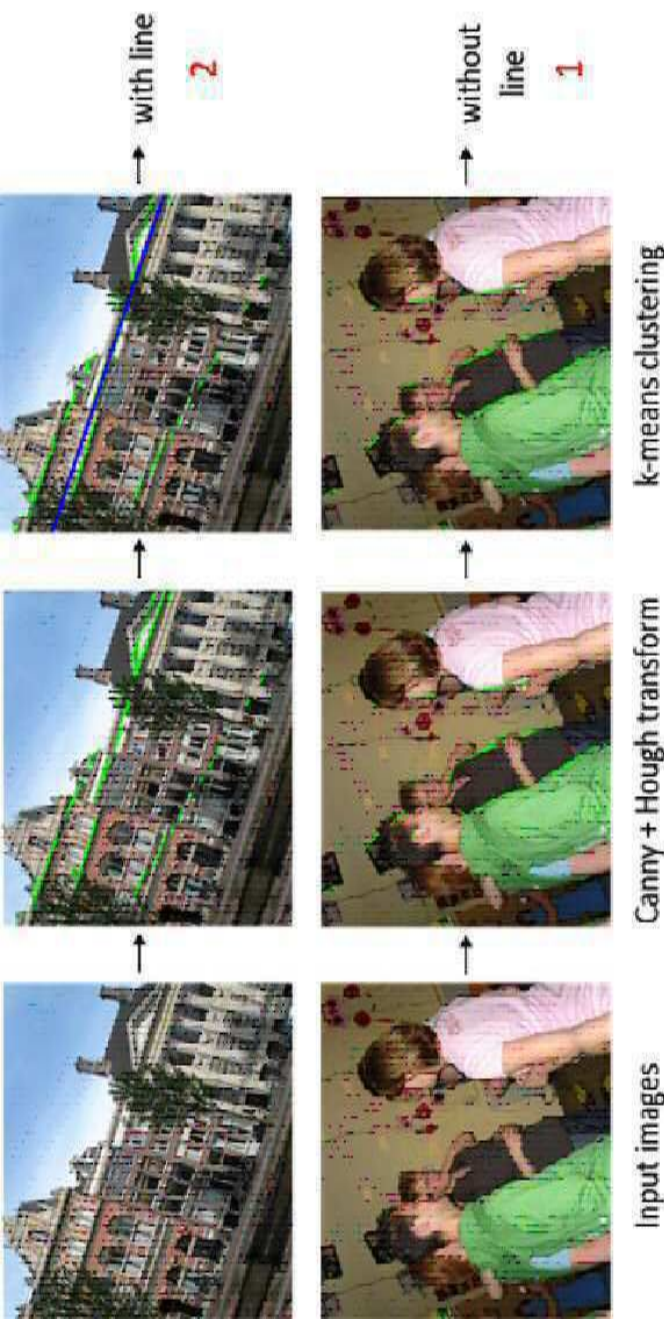
도면7



도면8



도면9



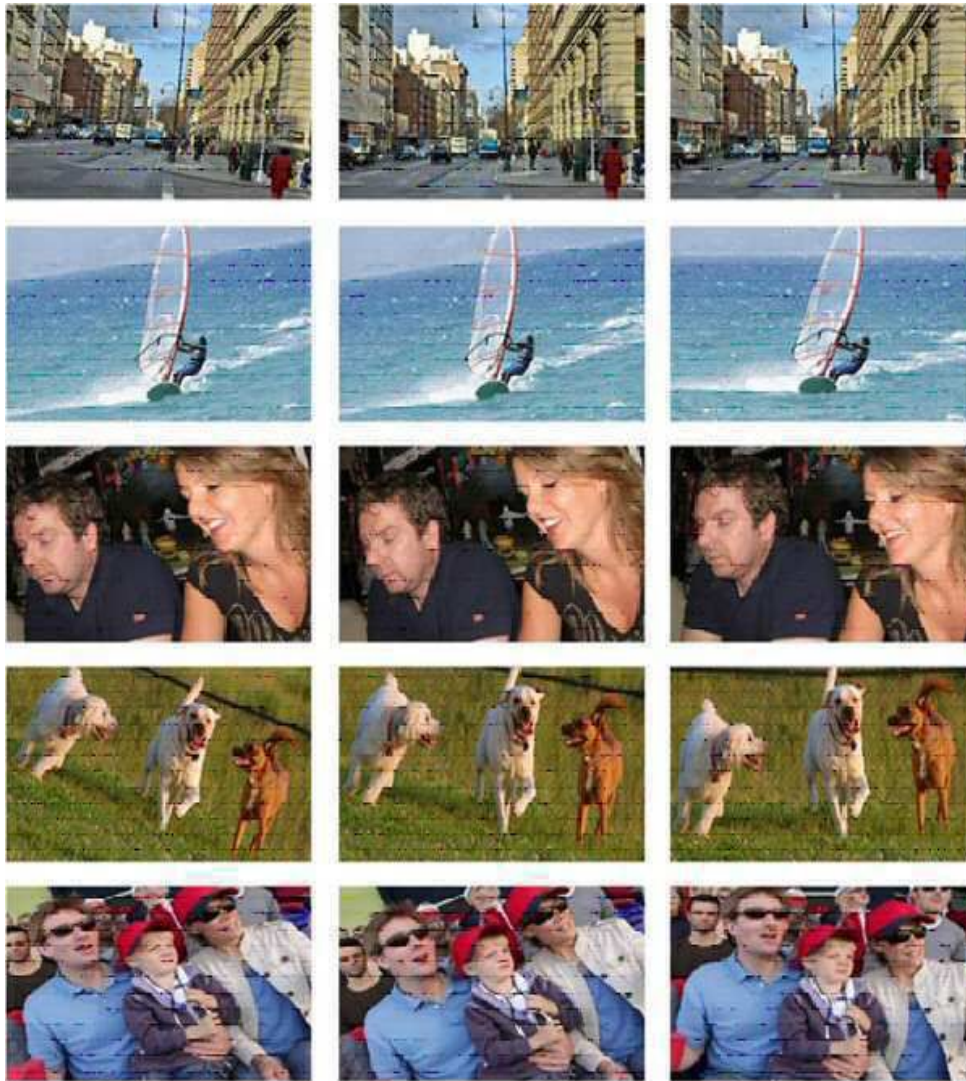
도면10



(a) 입력 영상

(b) 영상 수평 보정 결과

도면11



(a) 입력 영상

(b) Lightroom의 결과

(c) 본 방법의 결과