



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2024년03월25일  
(11) 등록번호 10-2651383  
(24) 등록일자 2024년03월21일

- (51) 국제특허분류(Int. Cl.)  
G06N 3/08 (2023.01) G06N 20/00 (2019.01)  
G06T 3/40 (2024.01) G06T 5/00 (2024.01)  
HO4N 5/14 (2023.01)
- (52) CPC특허분류  
G06N 3/08 (2023.01)  
G06N 20/00 (2021.08)
- (21) 출원번호 10-2020-0188668
- (22) 출원일자 2020년12월31일  
심사청구일자 2020년12월31일
- (65) 공개번호 10-2022-0096331
- (43) 공개일자 2022년07월07일
- (56) 선행기술조사문헌  
KR1020160036974 A  
KR1020220081869 A  
KR1020220089431 A

- (73) 특허권자  
포항공과대학교 산학협력단  
경상북도 포항시 남구 청암로 77 (지곡동)
- (72) 발명자  
이승용  
경상북도 포항시 남구 지곡로 155, 8동 803호  
조성현  
경상북도 포항시 남구 지곡로 278, 207동 102호  
(뒷면에 계속)
- (74) 대리인  
특허법인이상

전체 청구항 수 : 총 16 항

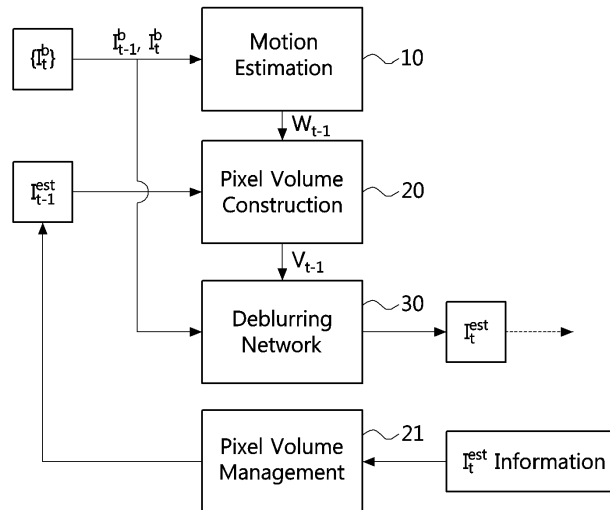
심사관 : 송근배

(54) 발명의 명칭 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법 및 이를 이용하는 장치

(57) 요약

픽셀 볼륨 기반으로 비디오 화질 개선을 수행하는 기계학습 방법 및 이를 이용하는 장치가 개시된다. 픽셀 볼륨 기반의 기계학습 방법은, 시간적으로 인접한 타겟 비디오 프레임용 현재 시간대의 입력 비디오 프레임으로 정합하기 위한 픽셀 볼륨을 생성하는 단계와, 입력 비디오 프레임과 픽셀 볼륨을 컨볼루션 신경망에 입력하여 보정 이미지를 생성하는 단계와, 보정 이미지로 계산된 컨볼루션 신경망의 손실 함수에 기초하여 기계 학습을 수행하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

*G06T 3/4046* (2024.01)

*G06T 5/73* (2024.01)

*H04N 5/144* (2013.01)

*G06T 2207/20084* (2013.01)

(72) 발명자

**이준용**

경상북도 포항시 북구 양덕로50번길 33, 101동  
1602호

**손형석**

경상북도 포항시 남구 청암로 77, 13동 106호

이 발명을 지원한 국가연구개발사업

과제고유번호

1711103034

과제번호

2015-0-00174-006

부처명

과학기술정보통신부

과제관리(전문)기관명

정보통신기획평가원

연구사업명

SW컴퓨팅산업원천기술개발

연구과제명

(SW 스타랩) 빅 비주얼 데이터 기반의 고품질 사진 메이크업 SW 개발

기 여 율

1/1

과제수행기관명

포항공과대학교 산학협력단

연구기간

2020.01.01 ~ 2020.12.31

---

## 명세서

### 청구범위

#### 청구항 1

컴퓨터 판독 가능 매체에 기록되거나 메모리(memory)에 저장된 프로그램 명령을 수행하는 컴퓨터 수단에 의해 구현되고 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법으로서,

상기 컴퓨터 수단의 픽셀 볼륨 건설부에 의해, 시간적으로 인접한 타겟 비디오 프레임을 현재 시간대의 입력 비디오 프레임으로 정합하기 위한 픽셀 볼륨을 생성하는 단계;

상기 입력 비디오 프레임과 상기 픽셀 볼륨을 상기 픽셀 볼륨 건설부에 연결된 디블러링 네트워크의 컨볼루션 신경망에 입력하여 보정 이미지를 생성하는 단계; 및

상기 컴퓨터 수단의 옵티마이저를 통해, 상기 보정 이미지로 계산된 컨볼루션 신경망의 손실 함수에 기초하여 상기 컨볼루션 신경망의 가중치를 업데이트하도록 기계 학습을 수행하는 단계;를 포함하며,

상기 픽셀 볼륨을 생성하는 단계는, 상기 타겟 비디오 프레임과 상기 입력 비디오 프레임 사이의 모션 또는 옵티컬 플로우를 예측하고, 공간 간섭성을 이용한 인접한 픽셀의 모션을 이용하여 상기 픽셀 볼륨 내의 각 픽셀 위치에 대해 상기 타겟 비디오 프레임으로부터 다수의 매칭 후보 픽셀을 생성하는 일련의 과정을 포함하는 픽셀 볼륨 기반의 기계학습 방법.

#### 청구항 2

삭제

#### 청구항 3

청구항 1에 있어서,

상기 픽셀 볼륨을 생성하는 단계는, 상기 타겟 비디오 프레임으로서 복원된 이전 시간대의 비디오 프레임인 이전 블러 프레임의 복원 결과를 사용하는, 픽셀 볼륨 기반의 기계학습 방법.

#### 청구항 4

청구항 1에 있어서,

상기 기계 학습을 수행하는 단계는, 상기 컴퓨터 수단에 의해, 상기 보정 이미지의 픽셀 값과 상기 타겟 비디오 프레임에 대응하는 선명한 교사 비디오 프레임의 픽셀 값 차이의 절대값 혹은 제곱값의 합을 최소화하도록 수행되는, 픽셀 볼륨 기반의 기계학습 방법.

#### 청구항 5

컴퓨터 판독 가능 매체에 기록되거나 메모리(memory)에 저장된 프로그램 명령을 수행하는 컴퓨터 수단에 의해 구현되고 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법으로서,

상기 컴퓨터 수단에 의해, 현재 블러 프레임을 참조하여 이전 블러 프레임의 복원 결과 프레임에 대한 모션을 추정하는 단계;

상기 컴퓨터 수단에 의해, 상기 모션의 추정 결과를 토대로 상기 현재 블러 프레임의 모든 픽셀에 대하여 상기 복원 결과 프레임 상에서 각각 매칭되는 픽셀 후보들을 모아 쌓은 픽셀 볼륨을 생성하는 단계; 및

상기 컴퓨터 수단에 의해, 상기 픽셀 볼륨에 기초하여 상기 현재 블러 프레임을 포함한 연속적인 블러 프레임들에 대해 딥러닝 기반의 디블러링을 수행하는 단계;를 포함하는 픽셀 볼륨 기반의 기계학습 방법.

#### 청구항 6

청구항 5에 있어서,

상기 픽셀 볼륨을 생성하는 단계는, 상기 현재 블러 프레임에 대응하는 레퍼런스 프레임 내 픽셀로서 상기 현재 블러 프레임 상의 특정 사이즈의 공간 윈도우 내에 위치하는 각 픽셀을 옵티컬 플로우에 따라 상기 복원 결과 프레임에 대응하는 타겟 프레임 내의 픽셀로 매칭하고 매칭된 픽셀을 매칭 픽셀로서 모으고, 축 방향 혹은 채널 방향으로 상기 각 픽셀에 대하여 쌓은 대응 매칭 픽셀들로 이루어진 픽셀 볼륨을 생성하는, 픽셀 볼륨 기반의 기계학습 방법.

**청구항 7**

청구항 5에 있어서,

상기 디블러링을 수행하는 단계는, 상기 연속적인 블러 프레임들의 특징을 변환한 제1 변환 결과와 상기 픽셀 볼륨의 특징을 변환한 제2 변환 결과를 연결(concatenation)하여 컨볼루션 네트워크에 입력하는 것을 포함하는, 픽셀 볼륨 기반의 기계학습 방법.

**청구항 8**

청구항 7에 있어서,

상기 디블러링을 수행하는 단계는, 컨볼루션 네트워크의 디블러링 결과에 상기 현재 블러 프레임을 엘리먼트-와이즈 합(element-wise sum)하여 디블러링 결과로서 출력하는 것을 포함하는, 픽셀 볼륨 기반의 기계학습 방법.

**청구항 9**

청구항 5에 있어서,

상기 컴퓨터 수단에 의해, 상기 디블러링을 통해 얻은 디블러링 결과 프레임과 미리 준비된 선명한 교사 프레임 사이의 평균절대오차 또는 평균제곱오차로 계산된 손실 함수에 기초하여 딥러닝 네트워크의 가중치를 업데이트 하는 단계를 더 포함하는 픽셀 볼륨 기반의 기계학습 방법.

**청구항 10**

청구항 5에 있어서,

상기 컴퓨터 수단에 의해, 상기 디블러링을 통해 얻은 디블러링 결과에 기초하여 상기 모션의 추정 결과를 생성하는 단계와 상기 픽셀 볼륨을 생성하는 단계에서 사용할 상기 이전 블러 프레임의 복원 결과 프레임을 제공하는 단계를 더 포함하는 픽셀 볼륨 기반의 기계학습 방법.

**청구항 11**

비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 장치로서,

현재 블러 프레임을 참조하여 이전 블러 프레임의 복원 결과 프레임에 대한 모션을 추정하는 모션 추정부;

상기 모션의 추정 결과를 토대로 상기 현재 블러 프레임의 모든 픽셀에 대하여 상기 복원 결과 프레임 상에서 각각 매칭되는 픽셀 후보들을 모아 쌓은 픽셀 볼륨을 생성하는 픽셀 볼륨 건설부; 및

상기 픽셀 볼륨에 기초하여 상기 현재 블러 프레임을 포함한 연속적인 블러 프레임들에 대해 딥러닝 기반의 디블러링을 수행하는 디블러링 네트워크;

를 포함하는 픽셀 볼륨 기반의 기계학습 장치.

**청구항 12**

청구항 11에 있어서,

상기 픽셀 볼륨 건설부는, 상기 현재 블러 프레임에 대응하는 레퍼런스 프레임 내 픽셀로서 상기 현재 블러 프레임 상의 특정 사이즈의 공간 윈도우 내에 위치하는 각 픽셀을 옵티컬 플로우에 따라 상기 복원 결과 프레임에 대응하는 타겟 프레임 내의 픽셀로 매칭하고 매칭된 픽셀을 픽셀 후보로서 모으고, 축 방향 혹은 채널 방향으로 상기 각 픽셀에 대하여 쌓은 대응 매칭 픽셀들로 이루어진 픽셀 볼륨을 생성하는, 픽셀 볼륨 기반의 기계학습 장치.

**청구항 13**

청구항 11에 있어서,

상기 디블러링 네트워크는, 상기 연속적인 블러 프레임들의 특징을 변환한 제1 변환 결과와 상기 픽셀 볼륨의 특징을 변환한 제2 변환 결과를 연결(concatenation)하여 컨볼루션 네트워크에 입력하는 연결 네트워크를 구비하는, 픽셀 볼륨 기반의 기계학습 장치.

**청구항 14**

청구항 13에 있어서,

상기 디블러링 네트워크는, 상기 컨볼루션 네트워크의 디블러링 결과에 상기 현재 블러 프레임을 엘리먼트-와이즈 합(element-wise sum)하여 디블러링 결과로서 출력하는 출력부를 더 구비하는, 픽셀 볼륨 기반의 기계학습 장치.

**청구항 15**

청구항 11에 있어서,

상기 디블러링을 통해 얻은 디블러링 결과 프레임과 미리 준비된 선명한 교사 프레임 사이의 평균절대오차 또는 평균제곱오차로 계산된 손실 함수에 기초하여 딥러닝 네트워크의 가중치를 업데이트하는 옵티마이저를 더 포함하는 픽셀 볼륨 기반의 기계학습 장치.

**청구항 16**

청구항 11에 있어서,

상기 디블러링을 통해 얻은 디블러링 결과에 기초하여 상기 모션의 추정 결과와 상기 이전 블러 프레임의 복원 결과 프레임을 상기 픽셀 볼륨 건설부에 제공하는 픽셀 볼륨 관리부를 더 포함하는 픽셀 볼륨 기반의 기계학습 장치.

**청구항 17**

청구항 16에 있어서,

상기 픽셀 볼륨 관리부와 상기 픽셀 볼륨 건설부에 연결되고 상기 모션의 추정 결과와 상기 이전 블러 프레임의 복원 결과 프레임을 저장하는 저장부를 더 포함하는 픽셀 볼륨 기반의 기계학습 장치.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 비디오 화질 개선을 위한 기계학습 방법에 관한 것으로, 보다 상세하게는 픽셀 볼륨 기반으로 비디오 화질 개선을 수행하는 기계학습 방법 및 이를 이용하는 장치에 관한 것이다.

**배경 기술**

[0002] 머신 러닝 기반의 비디오 화질 개선에서 인접 프레임의 정보를 잘 활용하기 위하여 인접한 프레임들을 처리할 현재 프레임에 정합하여 네트워크에 함께 입력으로 넣어주는 방법들이 제안되었다.

[0003] 예를 들어, 종래 기술 중 하나(Su et al, "Deep Video Deblurring for Hand-held Cameras", CVPR 2017.)에서는 비디오 디블러링 수행 시 인접 프레임들을 현재 프레임으로 정합하여 네트워크의 입력으로 사용한다.

[0004] 또한, 종래 기술의 또 다른 하나(Sajjadi et al, "Frame-Recurrent Video Super-Resolution", CVPR 2018.)에서는 비디오 초해상화(super-resolution) 수행 시 이전 프레임을 현재 프레임으로 정합하여 네트워크의 입력으로 사용한다.

[0005] 이러한 종래 기술의 비디오 화질 개선 방법들은 모션 정합을 위해 먼저 프레임 간 모션을 예측한 뒤, 예측된 모션으로 프레임을 와핑(warping)하여 정합을 수행한다. 그러나 이러한 기존 방법들은 모션 예측의 에러를 보정할 수 있는 방법이 존재하지 않기 때문에, 모션 에러가 정합 결과에 그대로 반영되는 문제가 있다.

[0006] 이와 같이 종래 기술에서는 머신 러닝 기반의 비디오 화질 개선 시, 인접한 다중 프레임들을 정합(alignment)하여 네트워크의 입력으로 사용하는 것이 정합없이 단순히 여러 프레임을 입력하는 것보다 화질 개선 성능 향상 정도가 더 크다는 것이 알려져 있으나, 비디오가 블러 등의 여러 화질 저하 요소를 포함하는 경우, 프레임 간의 모션을 정확하게 추정하기 어려워 정합 성능이 저하되고, 결국 비디오 화질 개선 성능을 저하시키는 문제가 있다.

[0007] 다시 말해서, 딥러닝 기반 비디오 프로세싱에서 특정 프레임을 복원할 때 주변 인접 프레임을 정합하여 같이 네트워크의 입력으로 넣어주면 성능이 향상된다. 그러나 블러가 존재하는 경우 프레임 간 정확한 모션 보정이 어렵다. 일반적인 모션 보정의 경우, 모션 보정을 수행할 두 프레임들 즉 타겟 프레임(target frame)과 레퍼런스 프레임(reference frame)이 존재할 때 두 프레임들 간의 모션을 예측하고(optical flow) 이를 이용하여 타겟 프레임을 레퍼런스 프레임으로 와핑(warping)한다. 그러나 이러한 단순 와핑(warping)의 경우, 블러에 의해 생기는 모션 예측 결과의 에러가 모션 정합 결과에 그대로 반영되게 된다. 이렇게 에러를 포함한 모션 정합 결과를 네트워크에 입력으로 사용하게 되면 네트워크가 해당 정보를 활용하기 어렵다.

**발명의 내용**

**해결하려는 과제**

[0008] 본 발명은 종래 기술의 문제점을 해결하기 위해 도출된 것으로, 본 발명의 목적은 머신러닝 기반 비디오 화질 개선 모델에서 사용할 수 있는 모션 예측 에러에 강인한 모션 보정 방법을 제안하는데 있다. 즉, 옵티컬 플로우(Optical flow) 등의 모션 예측 결과의 경우 공간적으로 간섭성을 가지는(spatially coherent) 특성이 있으며, 본 발명은 이러한 공간적 간섭성을 이용하여 각 픽셀에 대한 다수의 매칭 후보(matching candidates)를 얻고, 픽셀마다 매칭 후보를 쌓은 픽셀 볼륨을 구축하고, 단순 정합(wrapping) 대신 픽셀 볼륨 형태로 주변 프레임들 정합한 입력을 네트워크에 전달함으로써, 네트워크가 정합 에러에 강인하게 비디오 프레임 복원을 수행할 수 있도록 한, 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법을 제공하는데 그 목적이 있다.

[0009] 본 발명의 다른 목적은 전술한 픽셀 볼륨 기반의 기계학습 방법을 이용하는 비디오 화질 개선용 기계학습 장치를 제공하는데 있다.

[0010] 본 발명의 또 다른 목적은 전술한 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법의 일련의 과정을 구현하는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하는데 있다.

**과제의 해결 수단**

[0011] 상기 기술적 과제를 해결하기 위한 본 발명의 일 측면에 따른 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법은, 시간적으로 인접한 타겟 비디오 프레임들 현재 시간대의 입력 비디오 프레임으로 정합하기 위한 픽셀 볼륨을 생성하는 단계; 상기 입력 비디오 프레임과 상기 픽셀 볼륨을 컨볼루션 신경망에 입력하여 보정 이미지를 생성하는 단계; 및 상기 보정 이미지로 계산된 컨볼루션 신경망의 손실 함수에 기초하여 기계 학습을 수행하는 단계를 포함한다.

[0012] 일실시예에서, 상기 픽셀 볼륨을 생성하는 단계는, 상기 타겟 비디오 프레임과 상기 입력 비디오 프레임 사이의 모션 또는 옵티컬 플로우를 예측하고, 공간 간섭성을 이용한 인접한 픽셀의 모션을 이용하여 상기 픽셀 볼륨 내의 각 픽셀 위치에 대해 상기 타겟 비디오 프레임으로부터 다수의 매칭 후보 픽셀을 생성하는 일련의 과정을 포함한다.

[0013] 일실시예에서, 상기 픽셀 볼륨을 생성하는 단계는, 상기 타겟 비디오 프레임으로서 복원된 이전 시간대의 비디오 프레임인 이전 블러 프레임의 복원 결과를 사용한다.

[0014] 일실시예에서, 상기 기계 학습을 수행하는 단계는, 상기 보정 이미지의 픽셀 값과 상기 타겟 비디오 프레임에 대응하는 선명한 교사 비디오 프레임의 픽셀 값 차이의 절대값 혹은 제곱값의 합을 최소화하도록 수행된다.

[0015] 상기 기술적 과제를 해결하기 위한 본 발명의 다른 측면에 따른 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법은, 현재 블러 프레임을 참조하여 이전 블러 프레임의 복원 결과 프레임에 대한 모션을 추정하는 단계; 상기 모션의 추정 결과를 토대로 상기 현재 블러 프레임의 모든 픽셀에 대하여 상기 복원 결과 프레임 상에서 각각 매칭되는 픽셀 후보들을 모아 쌓은 픽셀 볼륨을 생성하는 단계; 및 상기 픽셀 볼륨에 기초하여 상기 현재 블러 프레임을 포함한 연속적인 블러 프레임들에 대해 딥러닝 기반의 디블러링을 수행하는 단계를 포함한다.

- [0016] 일실시예에서, 상기 픽셀 볼륨을 생성하는 단계는, 상기 현재 블러 프레임에 대응하는 레퍼런스 프레임 내 픽셀로서 상기 현재 블러 프레임 상의 특정 사이즈의 공간 윈도우 내에 위치하는 각 픽셀을 옵티컬 플로우에 따라 상기 복원 결과 프레임에 대응하는 타겟 프레임 내의 픽셀로 매칭하고 매칭된 픽셀을 매칭 픽셀로서 모으고, 축 방향 혹은 채널 방향으로 상기 각 픽셀에 대하여 쌓은 대응 매칭 픽셀들로 이루어진 픽셀 볼륨을 생성한다.
- [0017] 일실시예에서, 상기 디블러링을 수행하는 단계는, 상기 연속적인 블러 프레임들의 특징을 변환한 제1 변환 결과와 상기 픽셀 볼륨의 특징을 변환한 제2 변환 결과를 연결(concatenation)하여 컨볼루션 네트워크에 입력하는 것을 포함한다.
- [0018] 일실시예에서, 상기 디블러링을 수행하는 단계는, 컨볼루션 네트워크의 디블러링 결과에 상기 현재 블러 프레임을 엘리먼트-와이즈 합(element-wise sum)하여 디블러링 결과로서 출력하는 것을 포함한다.
- [0019] 일실시예에서, 픽셀 볼륨 기반의 기계학습 방법은, 상기 디블러링을 통해 얻은 디블러링 결과 프레임과 미리 준비된 선명한 교사 프레임 사이의 평균절대오차 또는 평균제곱오차로 계산된 손실 함수에 기초하여 딥러닝 네트워크의 가중치를 업데이트하는 단계를 더 포함할 수 있다.
- [0020] 일실시예에서, 픽셀 볼륨 기반의 기계학습 방법은, 상기 디블러링을 통해 얻은 디블러링 결과에 기초하여 상기 모션의 추정 결과를 생성하는 단계와 상기 픽셀 볼륨을 생성하는 단계에서 사용할 상기 이전 블러 프레임의 복원 결과 프레임을 제공하는 단계를 더 포함할 수 있다.
- [0021] 상기 기술적 과제를 해결하기 위한 본 발명의 또 다른 측면에 따른 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 장치는, 현재 블러 프레임을 참조하여 이전 블러 프레임의 복원 결과 프레임에 대한 모션을 추정하는 모션 추정부; 상기 모션의 추정 결과를 토대로 상기 현재 블러 프레임의 모든 픽셀에 대하여 상기 복원 결과 프레임 상에서 각각 매칭되는 픽셀 후보들을 모아 쌓은 픽셀 볼륨을 생성하는 픽셀 볼륨 건설부; 및 상기 픽셀 볼륨에 기초하여 상기 현재 블러 프레임을 포함한 연속적인 블러 프레임들에 대해 딥러닝 기반의 디블러링을 수행하는 디블러링 네트워크를 포함한다.
- [0022] 일실시예에서, 상기 픽셀 볼륨 건설부는, 상기 현재 블러 프레임에 대응하는 레퍼런스 프레임 내 픽셀로서 상기 현재 블러 프레임 상의 특정 사이즈의 공간 윈도우 내에 위치하는 각 픽셀을 옵티컬 플로우에 따라 상기 복원 결과 프레임에 대응하는 타겟 프레임 내의 픽셀로 매칭하고 매칭된 픽셀을 픽셀 후보로서 모으고, 축 방향 혹은 채널 방향으로 상기 각 픽셀에 대하여 쌓은 대응 매칭 픽셀들로 이루어진 픽셀 볼륨을 생성한다.
- [0023] 일실시예에서, 상기 디블러링 네트워크는, 상기 연속적인 블러 프레임들의 특징을 변환한 제1 변환 결과와 상기 픽셀 볼륨의 특징을 변환한 제2 변환 결과를 연결(concatenation)하여 컨볼루션 네트워크에 입력하는 연결 네트워크를 구비한다.
- [0024] 일실시예에서, 상기 디블러링 네트워크는, 상기 컨볼루션 네트워크의 디블러링 결과에 상기 현재 블러 프레임을 엘리먼트-와이즈 합(element-wise sum)하여 디블러링 결과로서 출력하는 출력부를 더 구비한다.
- [0025] 일실시예에서, 픽셀 볼륨 기반의 기계학습 장치는, 상기 디블러링을 통해 얻은 디블러링 결과 프레임과 미리 준비된 선명한 교사 프레임 사이의 평균절대오차 또는 평균제곱오차로 계산된 손실 함수에 기초하여 딥러닝 네트워크의 가중치를 업데이트하는 옵티마이저를 더 포함할 수 있다.
- [0026] 일실시예에서, 픽셀 볼륨 기반의 기계학습 장치는, 상기 디블러링을 통해 얻은 디블러링 결과에 기초하여 상기 모션의 추정 결과와 상기 이전 블러 프레임의 복원 결과 프레임을 상기 픽셀 볼륨 건설부에 제공하는 픽셀 볼륨 관리부를 더 포함할 수 있다.
- [0027] 일실시예에서, 픽셀 볼륨 기반의 기계학습 장치는, 상기 픽셀 볼륨 관리부와 상기 픽셀 볼륨 건설부에 연결되고 상기 모션의 추정 결과와 상기 이전 블러 프레임의 복원 결과 프레임을 저장하는 저장부를 더 포함할 수 있다.
- [0028] 상기 기술적 과제를 해결하기 위한 본 발명의 또 다른 측면에 따른 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 장치는, 진술한 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법의 단계들을 구현한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 포함할 수 있다.

**발명의 효과**

- [0029] 진술한 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법 및 이를 이용하는 장치를 채용하는 경우에는 머신러닝 기반 비디오 화질 개선 모델에서 사용할 수 있는 모션 추정 에러에 강인한 모션 보정 방법을 제공할

수 있다.

[0030] 본 발명에 의하면, 옵티컬 플로우(Optical flow) 등의 모션 추정 결과에 공간적 간섭성을 가지는(spatially coherent) 특성을 이용하여 현재 블러 프레임에 대응하는 레퍼런스 프레임의 각 픽셀에 대한 다수의 매칭 후보(matching candidates)를 현재 블러 프레임의 이전 블러 프레임에 대한 복원 결과 프레임에서 얻고, 각 픽셀마다 매칭 후보를 쌓은 픽셀 볼륨을 구축하고, 픽셀 볼륨 형태로 주변 프레임을 정합한 입력을 신경망에 전달함으로써, 신경망이 정합 에러에 강인하게 비디오 프레임을 복원할 수 있다.

[0031] 또한, 본 발명에 의하면, 정합 에러에 강인하게 복원된 비디오 프레임의 디블러링 결과와 선명한 교사 프레임을 이용하여 계산한 손실 함수로 네트워크의 가중치를 업데이트함으로써 모션 보정 모델을 학습하는 네트워크를 효과적으로 학습시킬 수 있다.

**도면의 간단한 설명**

[0032] 도 1은 본 발명의 실시예에 따른 픽셀 볼륨 기반의 비디오 디블러링 프레임워크에 대한 개략적인 블록도이다.

도 2는 도 1의 비디오 디블러링 프레임워크에 채용할 수 있는 픽셀 볼륨 건설 과정을 도식화하여 나타낸 도면이다.

도 3은 본 발명의 다른 실시예에 따른 픽셀 볼륨 기반의 비디오 디블러링 프레임워크에 대한 블록도이다.

도 4는 도 3의 비디오 디블러링 프레임워크에 의한 픽셀 볼륨 효과를 나타낸 예시도이다.

도 5는 본 발명의 또 다른 실시예에 따른 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법을 나타낸 흐름도이다.

**발명을 실시하기 위한 구체적인 내용**

[0033] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 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.

[0034] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 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.

[0035] 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.

[0036] 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.

[0037] 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 가진 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0038] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.



- [0039] 도 1은 본 발명의 실시예에 따른 픽셀 볼륨 기반의 비디오 디블러링 프레임워크에 대한 개략적인 블록도이다.
- [0040] 도 1을 참조하면, 본 실시예에 따른 픽셀 볼륨 기반의 비디오 디블러링 프레임워크(100)는, 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법을 이용하는 장치의 일종으로서, 모션 추정부(Motion Estimation, 10), 픽셀 볼륨 건설부(Pixel Volume Construction, 20), 디블러링 네트워크(Deblurring Network, 30) 및 픽셀 볼륨 관리부(Pixel Volume Management, 21)를 구비한다. 비디오 디블러링 프레임워크(100)는 와핑(wrapping)을 사용하여 단순하게 모션을 정합 및 보정한 프레임을 네트워크에 넣는 기존 방식 대신에 픽셀 볼륨 건설부(20)에서 생성한 픽셀 볼륨을 디블러링 네트워크(30)의 입력으로 사용한다. 이것은 디블러링 네트워크(30)가 모션 정합된 주변 프레임의 정보를 더 효과적으로 활용하도록 하여 고품질의 비디오 화질 개선을 수행할 수 있도록 한다.
- [0041] 좀더 구체적으로 설명하면, 모션 추정부(10)는 이전 블러 프레임( $I_{t-1}^b$ )의 복원 결과( $I_{t-1}^{est}$ )와 현재 블러 프레임( $I_t^b$ ) 사이의 모션( $W_{t-1}$ )을 추정한다. 여기서 모션은 모션 추정 결과에 대응된다. 모션 추정부(10)는 이전 블러 프레임( $I_{t-1}^b$ )의 복원 결과( $I_{t-1}^{est}$ )를 사용함으로써 모션 추정의 정확도와 신뢰도를 높인다.
- [0042] 픽셀 볼륨 건설부(20)는 모션 추정부(10)로부터의 모션 추정 결과( $W_{t-1}$ )를 이용해 이전 프레임의 복원 결과( $I_{t-1}^{est}$ )에서 각 픽셀의 픽셀 후보들을 모아 축 방향 또는 채널 방향으로 쌓은 픽셀 볼륨( $V_{t-1}$ )을 생성한다. 픽셀 볼륨 건설부(20)는 이전 프레임의 복원 결과( $I_{t-1}^{est}$ )에서 각 픽셀의 픽셀 후보들을 모아 픽셀 볼륨( $V_{t-1}$ )을 생성함으로써 비디오 디블러링을 위하여 인접한 선명한 프레임을 정합할 때 기존의 단순 정합에 비해 선명한 정보를 더욱 중점적으로 활용할 수 있다.
- [0043] 디블러링 네트워크(30)는 연속된 블러 프레임들 $\{I_{t-1}^b, I_t^b, I_{t+1}^b\}$ 과 픽셀 볼륨( $V_{t-1}$ )을 입력으로 받아 현재 블러 프레임( $I_t^b$ )의 복원 결과( $I_t^{est}$ )를 추정한다. 디블러링 네트워크(30)에서는 모션 추정 방법으로써 임의의 옵티컬 플로우(optical flow) 방법을 채용할 수 있다.
- [0044] 픽셀 볼륨 관리부(21)는 현재 블러 프레임에 대한 복원 결과 정보( $I_t^{est}$  Information)에 기초하여 픽셀 볼륨 건설부(20)에 이전 블러 프레임의 복원 결과( $I_{t-1}^{est}$ )가 입력되도록 동작한다. 픽셀 볼륨 관리부(21)는 블러 프레임의 복원 결과를 저장하는 메모리나 저장유닛에 연결되어 픽셀 볼륨 건설부(20)에 이전 블러 프레임의 복원 결과( $I_{t-1}^{est}$ )를 공급하도록 구현될 수 있다.
- [0045] 전술한 비디오 디블러링 프레임워크(100)의 작동 원리를 설명하면 다음과 같다.
- [0046] 디블러링 네트워크(30)에서 한 장의 블러 프레임( $I_t^b$ )을 복원할 때, 연속된 세 장의 블러 프레임들  $\{I_{t-1}^b, I_t^b, I_{t+1}^b\}$ 과 이전 프레임의 복원 결과( $I_{t-1}^{est}$ )를 사용한다. 여기서, 세 장의 입력 블러 프레임들  $\{I_{t-1}^b, I_t^b, I_{t+1}^b\}$ 은 기재된 순서대로 이전 프레임, 현재 프레임 및 이후 프레임으로 각각 지칭되거나 이전 블러 프레임, 현재 블러 프레임 및 이후 프레임으로 각각 지칭될 수 있다.
- [0047] 이 때, 픽셀 볼륨을 이용한 모션 보정은 이전 프레임의 복원 결과에만 적용한다. 그 이유로서, 비디오는 프레임을 순차적으로 처리하기 때문에 이전 프레임의 복원 결과가 이미 계산되어 있고, 이러한 복원 결과는 현재 블러 프레임보다 선명한 정보를 가질 확률이 크기 때문이다. 따라서 이미 복원을 거친 이전 프레임의 복원 결과의 좀더 선명한 정보를 중점적으로 활용하기 위하여 모션 보정은 이전 프레임의 복원 결과( $I_{t-1}^{est}$ )에만 적용한다. 이것은 이전 블러 프레임과 현재 블러 프레임 사이의 모션과 이전 블러 프레임의 복원 결과와 현재 블러 프레임 사이의 모션은 동일하다고 가정하는 것에 대응될 수 있다.
- [0048] 다시 말해서, 본 실시예의 픽셀 볼륨 기반의 모션 보정을 위하여 픽셀 볼륨 건설부(20)는 모션 추정부(10)로부터의 이전 프레임의 복원 결과( $I_{t-1}^{est}$ )와 현재 블러 프레임( $I_t^b$ ) 사이의 모션 추정 결과( $W_{t-1}$ )를 이용해 이전 프레임의 복원 결과( $I_{t-1}^{est}$ )를 통해 얻은 픽셀 후보들로 픽셀 볼륨( $V_{t-1}$ )을 생성하고, 디블러링 네트워크(30)는 연속된 입력 블러 프레임들 $\{I_{t-1}^b, I_t^b, I_{t+1}^b\}$ 과 픽셀 볼륨( $V_{t-1}$ )을 입력으로 받아 현재 블러 프레임( $I_t^b$ )의 복원 결과( $I_t^{est}$ )를 효과적으로 추정할 수 있다.
- [0049] 본 실시예에서 제공하는 픽셀 볼륨 건설(Pixel Volume Construction) 방법은 위치적으로 인접한 픽셀들은 비슷

한 모션을 갖는다는 특성을 이용하여 다수의 모션 보정 후보인 매칭 후보(matching candidates)를 생성하여 픽셀 볼륨 형태로 만든다. 픽셀 볼륨은 다수의 모션 보정 후보를 갖고 있기 때문에 기존 모션 추정이 틀렸더라도 후보 중에 정확한 모션이 있다면 에러 보정이 가능하다.

[0050] 다시 말해서, 모션이 정확할 가능성이 큰 곳에서는 다수의 모션 보정 후보들이 같은 픽셀 값을 가진다. 이러한 특성으로 인해 와핑(warping) 기반의 단순 모션 정합과 달리 모션 보정 후보들의 픽셀값 분포를 통해 해당 모션 보정을 얼마나 신뢰할 수 있는지 추정할 수 있다. 이러한 과정은 인위적으로 설정하는 것이 아니라 네트워크에 픽셀 볼륨을 입력함으로써 네트워크가 알아서 모션 보정 후보 픽셀들의 정보들을 활용하여 에러를 보정하도록 학습된다. 따라서 와핑을 사용하여 단순히 모션을 정합 및 보정한 프레임에 네트워크에 입력하는 기존 방식과 달리 픽셀 볼륨을 건설하여 넣는 경우, 네트워크가 모션 정합된 주변 프레임의 정보를 더 효과적으로 활용하여 고품질의 비디오 화질 개선을 수행할 수 있다.

[0051] 도 2는 도 1의 비디오 디블러링 프레임워크에 채용할 수 있는 픽셀 볼륨 건설 과정을 도식화하여 나타낸 도면이다.

[0052] 도 2를 참조하면, 본 실시예에 따른 디블러링 프레임워크에 채용할 수 있는 픽셀 볼륨 건설부는 현재 블러 프레임( $I_t^b$ )의 현재 픽셀  $(x, y)$ 와 이전 블러 프레임( $I_{t-1}^b$ )의 복원 결과인 이전 복원 프레임( $I_{t-1}^{est}$ )의 모션 보정 후보 픽셀  $(x', y')$ 을 매칭시킨다. 모션 보정 후보 픽셀은 간략히 매칭 픽셀, 매칭 후보 또는 픽셀 후보라고도 한다. 이 때, 현재 픽셀  $(x, y)$ 와 매칭 픽셀  $(x', y')$ 은 모션 추정에서 정확한(accurate) 픽셀 값을 갖거나 정확하지 않은(inaccurate) 픽셀 값을 가질 수 있다.

[0053] 여기에서, 픽셀 볼륨 건설부의 입력은 타겟(target) 프레임에 대응하는 이전 프레임의 복원 결과( $I_{t-1}^{est}$ )와 레퍼런스(reference) 프레임에 대응하는 현재 블러 프레임( $I_t^b$ ) 사이의 모션 추정 결과( $W_{t-1}$ )이다. 이와 같이, 현재 블러 프레임을 참조하여 이전 블러 프레임의 복원 결과 프레임에 대한 모션 추정 결과를 생성할 수 있다.

[0054] 픽셀 볼륨 건설부에서 여러 개의 매칭 후보(matching candidate)를 얻는 과정을 구체적으로 살펴보면 다음과 같다.

[0055] 레퍼런스 프레임 내 픽셀  $(x, y)$ 에 대하여 가로 세로 사이즈  $k \times k$ 의 공간 윈도우(spatial window)를 고려할 수 있다. 본 실시예에서는  $k=5$ 를 예시하지만 이에 한정되지는 아니한다.

[0056] 공간 윈도우에서  $\Delta_x, \Delta_y \in \{-\lfloor \frac{k}{2} \rfloor, \dots, +\lfloor \frac{k}{2} \rfloor\}$  일 때, 윈도우 내의 각 픽셀  $(x + \Delta_x, y + \Delta_y)$ 은 모션 추정부의 모션 추정 결과인 옵티컬 플로우( $W_{t-1}$ )에 따라서 이전 프레임의 복원 결과 프레임에 대응하는 타겟 프레임 내의 매칭 픽셀  $(x', y')$ 을 갖는다. 즉, 도 2에 나타난 바와 같이 인접한 매칭 픽셀들은 비슷한 모션을 갖는다는 공간 간섭성(spatial coherency)에 따라서 타겟 프레임의 픽셀  $(x' - \Delta_x, y' - \Delta_y)$ 은 레퍼런스 프레임의 특정 픽셀  $(x, y)$ 에 매칭하는 픽셀 후보(pixel candidates)가 된다.

[0057] 위와 같은 픽셀 후보의 선택 방식에 따라 레퍼런스 프레임의  $k \times k$  공간 윈도우(spatial window)의 각 픽셀  $(x, y)$ 에 대하여  $k^2$ 개의 픽셀 후보를 모을 수 있다. 레퍼런스 프레임의 모든 픽셀에 대하여 픽셀 후보 선택 과정을 수행하면, 결과적으로 특정 사이즈(HxW)의 이미지에 채널 방향으로 다수의 픽셀들이 쌓인 데이터 구조(HxWxC)의 픽셀 볼륨을 구축할 수 있다. 이 때, C개의 픽셀들은 채널 방향에서의 픽셀 후보들의 적층 개수이고,  $(\Delta_x, \Delta_y) = (0, 0)$ 에 대응하는 채널은 단순 정합(naive warping) 결과와 같다.

[0058] 위의 과정을 통해 얻은 다수의 픽셀 후보는 모션 에러에 강인한 모션 정합을 위한 추가 정보를 디블러링 네트워크에 제공한다. 픽셀 후보들은 랜덤으로 수집된 것이 아니라 모션 추정 결과가 공간 간섭성을 가진다는 특성을 이용하여 수집되기 때문에 유의미한 후보 값이다. 즉, 주변 픽셀의 모션 추정이 정확하다면 해당 픽셀 후보의 픽셀 값도 정확한 픽셀 값이다. 따라서 어떤 픽셀에 해당하는 모션이 잘못되었다고 하더라도 윈도우 내에 정확한 모션이 하나 이상 존재한다면 해당 픽셀은 정확한 픽셀 값을 가진 픽셀 후보를 포함하게 되고, 그에 의해 디블러링 네트워크는 정확한 모션 보정을 수행할 수 있게 된다.

[0059] 참고로, 모션 추정이 정확할 가능성이 클 때, 대부분의 픽셀 후보들은 같은 값이나 유사한 값을 지닌다. 반대로 모션 추정이 부정확한 영역에서는 픽셀 후보들이 상대적으로 차이가 큰 서로 다른 값을 지닌다. 실제로 일례의

기존 기술(Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. 2017. Deep Video Deblurring for Hand-held Cameras. In Proc. CVPR)에서 제시하는 비디오 디블러링 테스트 셋 중 100 프레임에 대하여 실험을 했을 때, 이웃 프레임들의 스택 내에서 95.1 퍼센트의 픽셀들이 가장 자주 나타나는(majority) 픽셀 값을 지녔다. 즉,  $k^2$  개의 매칭 후보 중 절반 이상이 같은 픽셀 값을 지녔다. 이러한 다수(majority) 픽셀 값을 선명한 교사(ground truth)와 비교할 때 92%의 정확도를 보여주었다. 이러한 특성을 통해 네트워크는 매칭 후보의 분포를 모션 정합의 신뢰도로 사용할 수 있다. 대부분의 매칭 후보들이 같은 값을 지닐 때, 네트워크는 해당 값을 적극적으로 활용할 수 있다. 반면에 매칭 후보들 사이에 가장 자주 나타나는(majority) 특성이 없는 경우, 네트워크는 그 부분에 예러가 있다고 판단하여 이를 보정할 다른 정보를 활용할 수 있다. 이러한 다수(majority) 정보 활용은 후보가 존재하지 않는 단순 와핑(wrapping)에서는 구현할 수 없는 성질이다.

- [0060] 이에 본 실시예에서는 픽셀 볼륨이 제공할 수 있는 추가 정보를 활용할 수 있도록 픽셀 볼륨을 컨볼루션 네트워크의 직접적인 입력으로 사용한다. 그 경우, 컨볼루션 네트워크 혹은 이를 포함하는 딥러닝 네트워크는 픽셀 볼륨을 최대한 활용하는 형태로 학습될 수 있다. 픽셀 볼륨을 사용하는 딥러닝 네트워크는 모션 정합에 존재하는 예러를 보정하여 고품질의 프레임 모션 보정 결과를 도출할 수 있다.
- [0061] 도 3은 본 발명의 다른 실시예에 따른 픽셀 볼륨 기반의 비디오 디블러링 프레임워크에 채용할 수 있는 디블러닝 네트워크를 설명하기 위한 블록도이다.
- [0062] 본 실시예에 따른 비디오 디블러링 프레임워크에 채용할 수 있는 디블러닝 네트워크는 딥러닝 네트워크의 학습을 위해 위에서 언급한 상기 일례의 기존 기술에서 사용한 디블러링 데이터셋이나 이와 유사한 구성의 데이터셋을 이용할 수 있다. 즉, 데이터셋은 카메라 셰이크 모션 블러(camera shake motion blur)에 의한 블러를 가진 동영상 프레임과 그에 대응하는 선명한 동영상 프레임을 포함하도록 구성될 수 있다.
- [0063] 도 3을 참조하면, 본 실시예에 따른 디블러닝 네트워크는 제1 특징 변환부(31), 제2 특징 변환부(32), 연결 네트워크(33), 딥러닝 네트워크(35) 및 디블러닝 결과(deblurring result)를 출력하는 출력부(39)를 포함한다.
- [0064] 제1 특징 변환부(31)는 제1 입력부(31a)를 통해 연속적인 블러 프레임들(Consecutive blurry frames) 예컨대, 세 장의 블러 프레임들  $\{I_{t-1}^b, I_t^b, I_{t+1}^b\}$  을 입력받는다. 세 장의 연속적인 블러 프레임들 중 그 중간에 위치하는 현재 블러 프레임( $I_t^b$ )은 딥러닝 결과로 출력되는 출력 프레임에 대해 레퍼런스 프레임으로 사용된다. 제1 특징 변환부(31)는 특정 사이즈의 필터 또는 커널을 포함한다.
- [0065] 제2 특징 변환부(32)는 픽셀 볼륨 공급부(22)로부터 제2 입력부(32a)를 통해 이전 프레임( $I_{t-1}^b$ )의 복원에 사용된 픽셀 볼륨( $V_{t-1}$ )을 입력받는다. 이 픽셀 볼륨( $V_{t-1}$ )은 모션 추정부에서 상기의 픽셀 볼륨( $V_{t-1}$ )을 기준으로 하는 현재 블러 프레임과 이전 블러 프레임의 복원 결과의 모션 추정을 통해 얻은 모션 추정 결과( $W_{t-1}$ )에 기초하여 이전 블러 프레임( $I_{t-1}^b$ )의 복원 결과( $I_{t-1}^{est}$ )로부터 얻은 픽셀 볼륨이다. 제2 특징 변환부(32)는 특정 사이즈의 필터 또는 커널을 포함한다.
- [0066] 여기에서, 픽셀 볼륨 공급부(22)는 픽셀 볼륨 건설부와 픽셀 볼륨 관리부를 포함하고, 현재 블러 프레임의 복원 결과( $I_t^{est}$ ) 정보에 기초하여 제어장치나 모션 보정 장치의 제어유닛(50)에 이전 블러 프레임의 모션 추정 결과( $W_{t-1}$ )와 이전 블러 프레임의 복원 결과( $I_{t-1}^{est}$ )를 입력받도록 구현될 수 있다. 모션 보정 장치는 본 실시예의 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법을 이용하는 장치 중 하나일 수 있다.
- [0067] 연결 네트워크(33)는 제1 특징 변환부(31)의 제1 변환 결과와 제2 특징 변환부(32)의 제2 변환 결과를 연결(concatenation)하여 딥러닝 네트워크(35)에 입력한다.
- [0068] 딥러닝 네트워크(35)는 예를 들어 제1 컨볼루션(convolution) 레이어, 제1 풀링(pooling) 레이어, 제2 컨볼루션 레이어, 제2 풀링 레이어, 복수의 잔차블럭(residual blocks, 37), 제1 디컨볼루션 레이어 및 제2 디컨볼루션 레이어를 구비할 수 있다. 컨볼루션 레이어나 디컨볼루션 레이어는 필터를 1픽셀씩이나 2픽셀씩 움직이는 스트라이드(stride) 값을 가질 수 있다. 컨볼루션 레이어와 디컨볼루션 레이어 사이에는 잔차 블럭(37)을 경유하지 않는 스킵 커넥션이 구비될 수 있다. 잔차 블럭(37)은 12개의 잔차 블록을 포함하나 이에 한정되지는 않는다.
- [0069] 출력부(39)는 딥러닝 네트워크(35)의 출력에 현재 블러 프레임( $I_t^b$ )을 추가한 디블러닝 결과( $I_t^{est}$ )를 출력한다.

출력부(39)는 딥러닝 네트워크(35)의 출력에 현재 블러 프레임( $I_t^b$ )을 축 기준으로 합 연산을 수행하거나 차원에  
서 같은 자리를 가진 요소들끼지만 계산하는 엘리먼트-와이즈 합(element-wise sum)을 위한 수단이나 이러한 수  
단에 상응하는 기능을 수행하는 구성부를 구비할 수 있다.

[0070] 출력부(39)에서 생성되는 복원 결과 즉, 디블러링 결과(deblurring result,  $I_t^{est}$ )나 이와 관련된 디블러링 결과  
정보 예컨대 디블러링 결과에 대응하는 이전 블러 프레임의 디블러링 결과를 저장하는 메모리의 저장 위치 정보  
등은 본 실시예의 모션 보정 장치에 의해 픽셀 볼륨 공급부(22)에 전달될 수 있다.

[0071] 위의 과정을 통해 네트워크가 복원 대상인 현재 블러 프레임( $I_t^b$ )의 복원 결과에 해당하는 디블러링 결과 또는 디  
블러링 프레임( $I_t^{est}$ )을 생성하면, 본 실시예에 따른 모션 보정 장치는 생성된 복원 결과와 선명한 교사 프레임  
( $I_t^{GT}$ ) 사이의 평균절대오차(mean absolute error, MAE) 혹은 평균제곱오차(mean squared error, MSE)를 계산하  
여 손실 함수로 사용할 수 있다. 이러한 손실 함수를 사용하면, 모션 보정 장치는 아담 옵티마이저(Adam  
optimizer)를 통해 그라디언트 역전파로 네트워크의 가중치(weights)를 업데이트하면서 네트워크를 학습시킬 수  
있다.

[0072] 도 4는 도 3의 비디오 디블러링 프레임워크에 의한 픽셀 볼륨 효과를 나타낸 예시도이다.

[0073] 도 4를 참조하면, 본 실시예에 따른 비디오 디블러링 프레임워크는 연속되는 흐린한-즉, 블러를 가진- 비디오  
프레임들에 대하여 픽셀마다 매칭 후보를 쌓은 픽셀 볼륨을 디블러링 네트워크의 입력으로 사용함으로써 개선된  
디블러링 효과를 얻는다. 여기서, 매칭 후보는 순차적으로 처리되는 비디오 프레임들의 복원 대상 프레임에 대  
한 이전 프레임의 복원 결과로 얻은 복원 결과 프레임을 사용한다.

[0074] 다시 말해서, 본 실시예에서는 입력 영상(a)과 입력 영상의 처리 직전에 복원 처리된 이전 영상의 단순 정합(b)  
대신에 입력 영상과 입력 영상의 이전 영상의 복원 결과 간의 모션 추정 결과에 기초하여 입력 영상의 적어도  
일부 픽셀들 각각에 대해 복수의 매칭 후보들을 채널 방향 등으로 모음으로써, 채널 방향 등의 일방향으로 쌓은  
매칭 후보들의 데이터 구조 즉, 픽셀 볼륨(d)을 구축할 수 있다. 구축된 픽셀 볼륨(d)은 디블러링 네트워크의  
입력으로 사용된다.

[0075] 픽셀 볼륨 기반의 네트워크 결과(e)는 비교예의 단순 정합을 이용한 네트워크 결과(c)에 비해 의미 있게 향상된  
영상 복원 성능을 나타내는 것을 확인할 수 있다.

[0076] 이와 같이 본 실시예에서는 머신러닝 기반 비디오 화질 개선 모델에서 사용할 수 있는 모션 추정 에러에 강인한  
모션 보정 방법을 제공한다. 즉, 옵티컬플로우(optical flow) 등의 모션 추정 결과의 경우 공간 간섭성을 가지  
는(spatially coherent) 특성을 고려하여 픽셀마다 다수의 매칭 후보들(matching candidates)을 쌓은 픽셀 볼  
륨을 구축하고, 구축된 픽셀 볼륨을 이용하여 주변 프레임을 정합함으로써 정합 에러에 강인하게 비디오 프레임  
을 복원하는, 픽셀 볼륨 기반의 비디오 화질 개선용 기계학습 방법을 제공할 수 있다.

[0077] 도 5는 본 발명의 또 다른 실시예에 따른 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법을 나타낸  
흐름도이다.

[0078] 도 5를 참조하면, 본 실시예에 따른 픽셀 볼륨 기반의 기계학습 방법(이하 간략히 '모션 보정 방법'이라 한다)  
은 먼저 현재 블러 프레임을 참조하여 이전 블러 프레임의 복원 결과 프레임에 대한 모션 추정 결과를 생성한다  
(S51).

[0079] 다음, 모션 추정 결과를 토대로 현재 블러 프레임의 모든 픽셀에 대하여 복원 결과 프레임 상에서 매칭하는 픽  
셀 후보들을 모아 쌓은 픽셀 볼륨을 생성한다(S52).

[0080] 전술한 두 단계들(S51 및 S52)은 시간적으로 인접한 타겟 비디오 프레임을 현재 시간대의 입력 비디오 프레임으  
로 정합하기 위한 픽셀 볼륨을 생성하는 단계에 대응될 수 있다. 여기서, 타겟 비디오 프레임은 이전 블러 프레  
임의 복원 결과 프레임에 대응되고, 입력 비디오 프레임은 현재 블러 프레임에 대응된다. 즉, 타겟 비디오 프레  
임으로는 복원된 이전 시간대의 비디오 프레임인 이전 블러 프레임의 복원 결과를 사용한다.

[0081] 위와 같은 경우, 픽셀 볼륨을 생성하는 단계는, 타겟 비디오 프레임과 입력 비디오 프레임 사이의 모션 또는 옵  
티컬 플로우를 예측하고, 공간 간섭성을 이용한 인접한 픽셀의 모션을 이용하여 픽셀 볼륨 내의 각 픽셀 위치에  
대해 타겟 비디오 프레임으로부터 다수의 매칭 후보 픽셀을 생성하는 과정을 포함할 수 있다.

[0082] 다음, 픽셀 볼륨에 기초하여 현재 블러 프레임을 포함한 연속적인 블러 프레임들에 대해 딥러닝 기반의 디블러

링을 수행한다(S53). 즉, 본 단계(S53)는 입력 비디오 프레임과 픽셀 볼륨을 컨볼루션 신경망에 입력하여 보정 이미지를 생성하는 단계에 대응될 수 있다.

[0083] 다음, 디블러링 결과 프레임과 선명한 교사 프레임 사이의 평균절대오차(MAE) 또는 평균제곱오차(MSE)로 계산된 손실 함수에 기초하여 딥러닝 네트워크의 가중치를 업데이트한다(S54). 즉, 본 단계(S54)는 보정 이미지로 계산된 컨볼루션 신경망의 손실 함수에 기초하여 기계 학습을 수행하는 단계에 대응될 수 있다. 다시 말해서, 기계 학습을 수행하는 단계는, 보정 이미지의 픽셀 값과 타겟 비디오 프레임에 대응하는 선명한 교사 비디오 프레임의 픽셀 값 차이의 절대값 혹은 제곱값의 합을 최소화하도록 수행될 수 있다.

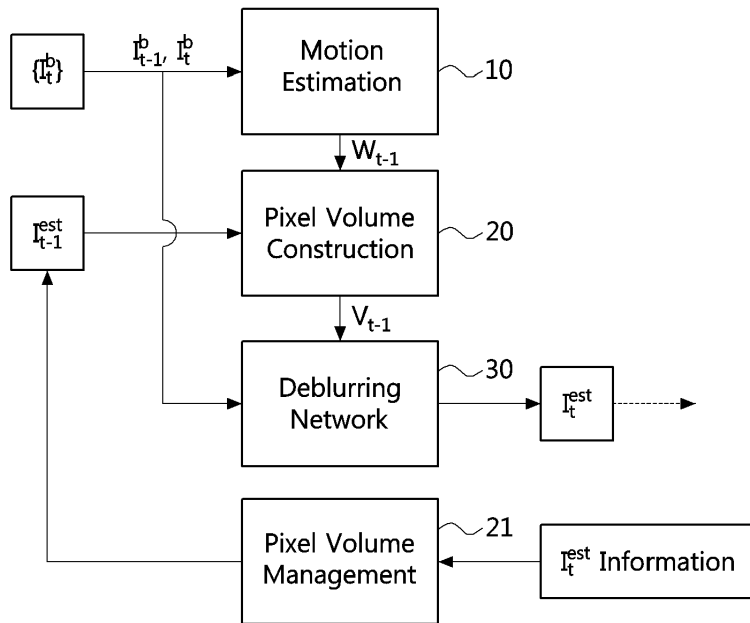
[0084] 한편, 전술한 실시예들을 통해 설명한 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법은 다양한 컴퓨터 수단을 통해 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위해 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용가능한 것일 수 있다.

[0085] 컴퓨터 판독 가능 매체의 예에는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함한다. 상술한 하드웨어 장치는 본 실시예에 따른 비디오 화질 개선을 위한 픽셀 볼륨 기반의 기계학습 방법의 일련의 동작을 수행하기 위해 적어도 하나의 소프트웨어 모듈로 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

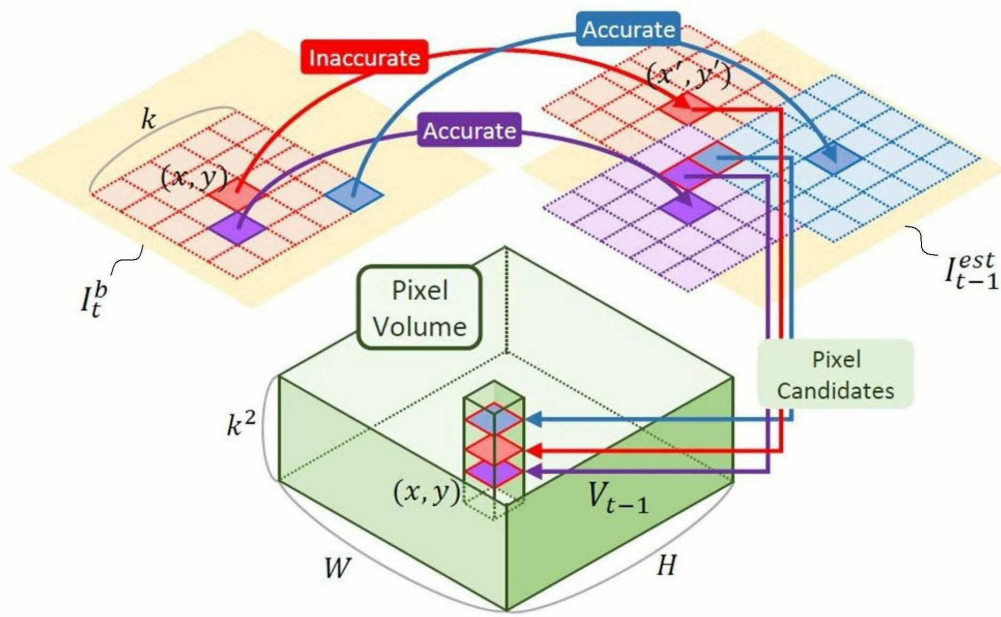
[0086] 이상 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

**도면**

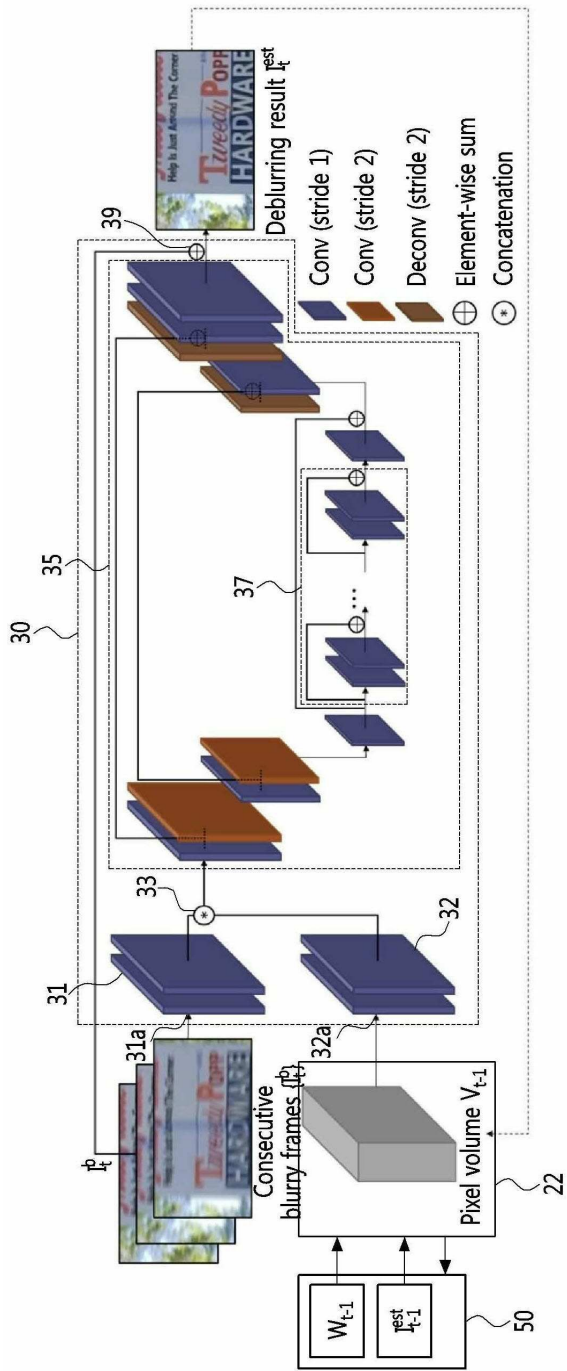
**도면1**



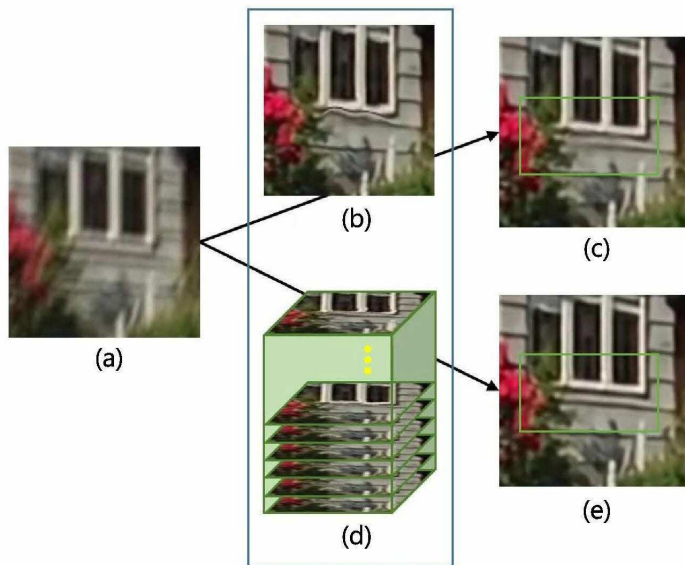
도면2



도면3



도면4



도면5

