



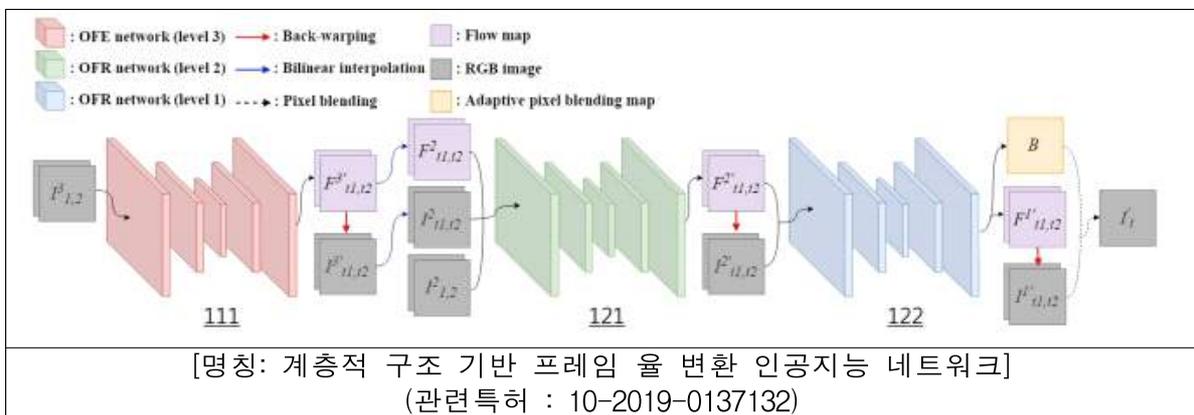
■ 기술명 : (인공지능 기반 고해상도 동영상 프레임 율 고속 변환 기술)  
 (영문 : High-speed video frame rate conversion based on AI)

산업기술분류	정보통신 - 디지털 방송 - 디지털 방송 콘텐츠 (300203)
Key-word(국문)	인공지능, 프레임 율 변환, 고속 변환, 슬로 모션
Key-word(영문)	AI, frame rate conversion, fast conversion, slow motion

■ 기술의 개요

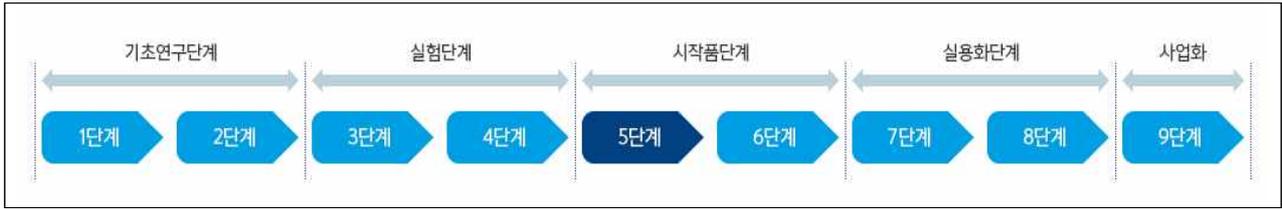
- (배경) 4K 이상의 고해상도 대화면 디스플레이가 주류로 자리 잡음. 이런 디스플레이를 위한 고품질의 영상은 높은 해상도뿐만 아니라 높은 프레임 율이 요구됨. 그러나 대부분의 기존 영상은 낮은 프레임 율을 지니고 있으므로 이를 높은 프레임 율로 변환할 수 있는 기술이 요구됨. 또한 최근 스마트폰에서 240fps의 초고속 촬영을 지원하면서 인상적인 장면을 슬로 모션으로 기록하는 경우가 많아짐. 기존의 낮은 프레임 율 영상을 슬로 모션으로 보기 위해서는 높은 프레임 율로 변환할 수 있는 기술이 요구됨
- (개요) 최근 인공 지능 기반 프레임 율 고속 기법이 등장하여 전통적인 기법에 비해 더욱 우수한 성능을 제공함. 그러나 기존 인공 지능 기법들은 고해상도 영상에서 과도한 GPU 메모리와 연산량을 요구함. 이를 극복하기 위해 계층적 인공 지능 네트워크를 개발하여 저 복잡도 및 고성능 프레임 율 변환 기술을 개발함

<기술 개요도 >





## ■ 기술의 구현수준(TRL)



## ■ 구현기술 명세 및 장점(경쟁기술과의 차별성)

### ○ 당해단계 기술명세

- 본 기술은 고해상도 영상의 프레임 을 변환을 효과적으로 수행하기 위해 다음과 같이 3단계 네트워크로 구성되어 있음.
- (1단계: 옵티컬 플로우 산출) 첫 번째 네트워크는 원영상 해상도의 1/4로 감소된 영상을 옵티컬 플로우 네트워크의 입력으로 사용하여 옵티컬 플로우를 산출함.
- (2단계: 옵티컬 플로우 정제) 네트워크는 1/4 해상도의 옵티컬 플로우를 1/2의 해상도로 증가시켜주는 옵티컬 플로우 정제 (refinement) 네트워크임
- (3단계: 초해상화) 1/2 해상도의 중간 프레임을 원본 해상도로 증가시키는 초해상도 (Super Resolution) 수행 네트워크임.

### ○ 추후단계 확보대상 기술명세

- 네트워크 양자화 기술을 통한 딥러닝 네트워크 추가 경량화
- 클라우드 기반 딥러닝 네트워크 시스템

### ○ 보유기술 장점 및 차별성(경쟁기술대비)

- 경쟁 기술 대비 더 적은 GPU 메모리를 사용하고 기존 기술 대비 최대 10배 이상의 속도 향상을 보임
- 4K와 같은 고해상도 영상에 대하여 기존 대비 더욱 높은 품질의 영상을 제공
- 조명 변화와 텍스트에 더욱 강건함.



## ■ 활용범위 및 응용분야

[프레임 올 업 컨버전 분야]
[슬로 모션 비디오 생성기]

- 프레임 올 업 컨버전 엔진 SW
- 슬로 모션 (slow motion) 비디오 생성기 (스마트폰, PC 등)
- TV 및 셋톱 박스의 프레임 올 업 컨버전 엔진

## ■ 지식재산권 현황

구분	발명의 명칭	출원번호 (출원일)	등록번호 (등록일)
특허	딥러닝 기반 동영상 프레임 올 변환 방법 및 장치	2018-0144595 (2018.11.21)	
특허	고해상도 동영상 프레임 올 고속 변환 방법 및 장치	2019-0137132 (2019.10.31)	
특허	데이터 변형을 통한 고해상도 동영상 프레임 올 고속 변환 방법 및 장치	2019-0167488 (2019.12.16.)	



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2020년07월21일  
(11) 등록번호 10-2136468  
(24) 등록일자 2020년07월15일

(51) 국제특허분류(Int. Cl.)  
HO4N 7/01 (2006.01) GO6N 3/08 (2006.01)  
(52) CPC특허분류  
HO4N 7/0127 (2013.01)  
GO6N 3/08 (2013.01)  
(21) 출원번호 10-2018-0144595  
(22) 출원일자 2018년11월21일  
심사청구일자 2019년02월12일  
(65) 공개번호 10-2020-0059627  
(43) 공개일자 2020년05월29일  
(56) 선행기술조사문헌  
KR1020150004916 A\*  
US20090279111 A1\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
전자부품연구원  
경기도 성남시 분당구 새나리로 25 (야탑동)  
(72) 발명자  
정진우  
서울특별시 송파구 송파대로 111, 파크하비오105동 1205호  
안하은  
서울특별시 성북구 성북로8다길 37, 102호  
김제우  
경기도 성남시 분당구 수내로 181, 303동 901호  
(74) 대리인  
남충우

전체 청구항 수 : 총 11 항

심사관 : 박재학

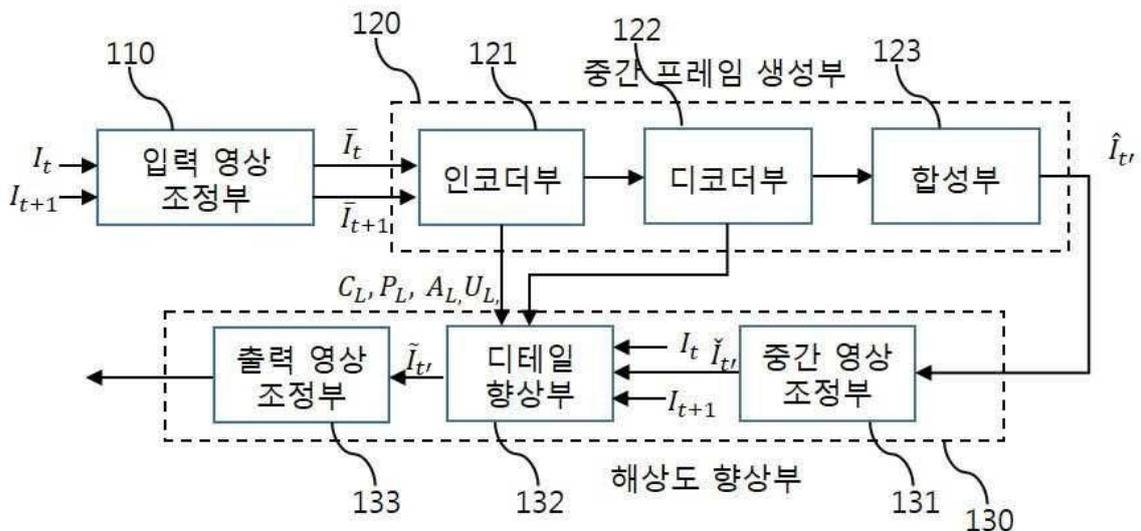
(54) 발명의 명칭 딥러닝 기반 동영상 프레임 율 변환 방법 및 장치

(57) 요약

딥러닝에 기반하여 다양한 컬러 영상 포맷에 대응하는 동시에 고품질, 고속으로 프레임 율 보간하는 동영상 프레임 율 변환 방법 및 장치가 제공된다. 본 발명의 실시예에 따른 동영상 프레임 율 변환 장치는 각 채널의 영상에 대한 해상도를 조정하는 조정부; 조정부에서 해상도가 조정된 t초와 t+1초의 연속된 두 영상으로 중간 영상을 생성하는 생성부; 및 생성부에서 생성된 중간 영상의 해상도를 향상시키는 향상부;를 포함한다.

이에 의해, 딥러닝 기반 동영상 프레임 율 변환에 있어, 딥러닝의 입력이 RGB 컬러 포맷 뿐만 아니라 YCbCr 등 다양한 컬러 포맷에 적용할 수 있게 되며, 프레임 보간 방법과 해상도 향상 방법을 동시에 이용하여 고품질, 고속의 프레임 율 변환이 가능해진다.

대표도



(52) CPC특허분류

*H04N 7/0135* (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711076193

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 차세대(UHD)방송서비스활성화기술개발

연구과제명 영상콘텐츠 초고속/초고화질 변환 기술 개발

기 여 율 1/1

주관기관 (주)픽스트리

연구기간 2018.07.01 ~ 2018.12.31

---

## 명세서

### 청구범위

#### 청구항 1

각 채널의 영상에 대한 해상도를 조정하는 조정부;  
조정부에서 해상도가 조정된  $t$ 초와  $t+1$ 초의 연속된 두 영상으로 중간 영상을 생성하는 생성부; 및  
생성부에서 생성된 중간 영상의 해상도를 향상시키는 향상부;를 포함하고,  
중간 영상은,  
 $t+0.5$ 초의 영상이며,  
조정부는,  
각 채널의 영상에 대한 해상도를 동일하게 서브 샘플링하는 것을 특징으로 하는 동영상 프레임 윌 변환 장치.

#### 청구항 2

삭제

#### 청구항 3

청구항 1에 있어서,  
생성부는,  
해상도가 조정된 영상의 특성을 계산하는 인코더부;  
계산된 특성을 영상의 해상도 크기만큼 복원하는 디코더부;  
조정부에서 해상도가 조정된 두 영상과 디코더의 출력을 이용하여 중간 영상을 생성하는 합성부;를 포함하는 것  
을 특징으로 하는 동영상 프레임 윌 변환 장치.

#### 청구항 4

청구항 1에 있어서,  
향상부는,  
생성부에서 생성된 중간 영상의 해상도를 원본 영상의 해상도로 향상시키는 것을 특징으로 하는 동영상 프레임  
윌 변환 장치.

#### 청구항 5

청구항 4에 있어서,  
향상부는,  
디테일을 개선할 중간 영상의 해상도를 조정하는 제1 조정부;  
원본 영상을 이용하여, 해상도가 조정된 중간 영상의 디테일을 개선하는 디테일 향상부;  
제1 조정부에서 해상도가 조정되지 않은 중간 영상의 해상도를 조정하는 제2 조정부;를 포함하는 것을 특징으  
로 하는 동영상 프레임 윌 변환 장치.

**청구항 6**

청구항 5에 있어서,  
디테일 향상부는,  
선별적으로 선택된 채널만을 입력받는 것을 특징으로 하는 동영상 프레임 윌 변환 장치.

**청구항 7**

청구항 6에 있어서,  
선택된 채널은,  
밝기 신호인 것을 특징으로 하는 동영상 프레임 윌 변환 장치.

**청구항 8**

청구항 5에 있어서,  
디테일 향상부는,  
생성부에서 생성되는 중간 정보를 참조하여, 중간 영상의 디테일을 개선하는 것을 특징으로 하는 동영상 프레임 윌 변환 장치.

**청구항 9**

청구항 5에 있어서,  
디테일 향상부는,  
딥러닝 네트워크를 사용하는 것을 특징으로 하는 동영상 프레임 윌 변환 장치.

**청구항 10**

각 채널의 영상에 대한 해상도를 조정하는 단계;  
해상도가 조정된  $t$ 초와  $t+1$ 초의 연속된 두 영상으로 중간 영상을 생성하는 단계; 및  
생성된 중간 영상의 해상도를 향상시키는 단계;를 포함하고,  
중간 영상은,  
 $t+0.5$ 초의 영상이며,  
조정 단계는,  
각 채널의 영상에 대한 해상도를 동일하게 서브 샘플링하는 것을 특징으로 하는 동영상 프레임 윌 변환 장치.

**청구항 11**

각 채널의 영상에 대한 해상도를 조정하는 조정부;  
조정부에서 해상도가 조정된  $t$ 초와  $t+1$ 초의 연속된 두 영상으로 생성된 중간 영상의 해상도를 향상시키는 향상부;를 포함하고,

중간 영상은,  
 $t+0.5$ 초의 영상이며,  
 조정부는,  
 각 채널의 영상에 대한 해상도를 동일하게 서브 샘플링하는 것을 특징으로 하는 동영상 프레임 율 변환 장치.

**청구항 12**

각 채널의 영상에 대한 해상도를 조정하는 단계;  
 조정부에서 해상도가 조정된  $t$ 초와  $t+1$ 초의 연속된 두 영상으로 생성된 중간 영상의 해상도를 향상시키는 단계;를 포함하고,  
 중간 영상은,  
 $t+0.5$ 초의 영상이며,  
 조정 단계는,  
 각 채널의 영상에 대한 해상도를 동일하게 서브 샘플링하는 것을 특징으로 하는 동영상 프레임 율 변환 장치.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 동영상의 프레임 율 변환 방법에 관한 것으로, 더욱 상세하게는 딥러닝 기법에 기반한 프레임 보간 기법과 해상도 증가 기법을 적용하여 고속으로 프레임 율을 증가시키는 방법 및 장치에 관한 것이다.

**배경 기술**

- [0003] 1) 동영상 프레임 율 변환 기법 개요
- [0004] 동영상은 연속된 정지 영상의 집합으로 구성된다. 비디오에서 정지 영상을 프레임이라고 부르며 단위 시간 당 프레임의 수를 동영상의 프레임 율 (frame rate)이라고 한다. 예를 들어 1초에 24장의 프레임으로 구성되면 프레임 율은 24 fps (frame per second)가 된다.
- [0005] 프레임 율은 촬영자의 의도, 영상의 포맷, 카메라의 한계 등에 의하여 결정된다. 관찰자가 영상을 연속된 화면으로 느끼기 위해서는 어느 정도 이상의 프레임 율이 필요하고 이보다 낮을 경우 움직임이 부드럽지 않아 보인다. 이 현상은 디스플레이의 크기, 조명, 시청 거리 등에 의해 달라질 수 있다.
- [0006] 이를 개선하기 위해 동영상의 프레임 율을 후처리에 의해 증가시키는 것을 동영상 프레임 율 변환이라고 한다.
- [0008] 2) 종래 기술
- [0009] 동영상 프레임 율을 증가시키는 가장 간단한 방법은 프레임을 반복하는 것이다. 예를 들어 30 fps 영상을 60 fps 영상으로 증가시킬 경우, 각 프레임마다 한 프레임을 반복하여 출력하는 것이다. 그러나 이 방법의 경우 동영상의 정보량은 동일하고 움직임에 대한 연속성은 변하지 않았으므로 관찰자가 느끼는 불편감은 동일하다.
- [0010] 이를 해결하기 위해 연속된 프레임들을 이용하여 가상의 프레임을 생성하는 기술이 개발되었다. 즉  $t$  초와  $t+1$  초 사이의 영상을 이용하여  $t+0.5$  초의 중간 영상을 새롭게 생성하며 이를 프레임 보간 (frame interpolation) 기술이라고 한다.
- [0011] 프레임 보간은 다양한 방법이 개발되었으며 일반적으로는 다음과 같은 두 단계 과정을 거친다. 첫 번째 단계는 움직임 또는 옵티컬 플로우 (optical flow)를 획득하는 단계 이며, 두 번째 단계는 움직임 정보를 바탕으로 중간 프레임을 생성 (Synthesis)하는 단계이다.
- [0012] 동영상에서 물체의 움직임이 부드럽게 보이려면 중간 영상은 물체의 움직임이 두 영상 사이의 중간에 해당되어야 한다. 따라서 물체의 움직임 정보를 가지고 있는 옵티컬 플로우를 정확하게 찾는 것이 매우 중요하다. 이에

기반한 다양한 기법들이 제안되어 왔다.

[0013] 최근 딥러닝 (Deep learning) 알고리즘이 등장하여 컴퓨터 비전, 음성 인식 등 다양한 분야에 널리 사용되고 있으며 종래에 방법에 비해 월등한 성능을 보이고 있다. 이에 발맞추어 딥러닝을 사용한 다양한 프레임 보간 기법이 등장하였다. 이 기법들은 옵티컬 플로우와 합성에 기반한 종래의 방법보다 더욱 뛰어난 보간 결과를 보여줌에 따라 최근 지속적으로 연구되고 있다.

[0015] 3) 종래 기술 문제점

[0016] 종래의 딥러닝에 기반한 기술은 입력 영상이 RGB 형식의 컬러 포맷을 갖는다고 가정한다. 그러나 대부분의 동영상은 YCbCr 형식의 컬러 포맷으로 압축되어 저장되어 있다. 또한 색차 신호인 Cb, Cr은 서브 샘플링 (sub-sampling) 되어 있어 밝기 신호인 Y와 해상도가 다르다.

[0017] 따라서 기존 방법에 일반적인 동영상을 적용하기 위해서는 색차 신호를 업샘플링 (up-sampling)하여 밝기 신호와 영상 크기를 맞추는 후, 다시 YCbCr 컬러 형식을 RGB로 변환하는 작업을 수행하여야 한다. 이런 컬러 변환 작업들은 비효율적이고 특히 영상 크기가 커질 경우 실시간 동작을 어렵게 하는 요소로 작동한다.

[0018] 기존 방법의 두 번째 문제는 네트워크 구조로 인한 문제로 4K (3840x2160) 해상도와 같이 큰 해상도에 대하여 GPU 메모리 부족으로 연산이 불가능 하거나, 아니면 매우 느린 연산 속도를 보여준다. 이와 같은 현상은 실시간 연산을 요구하는 상용 애플리케이션에는 딥러닝을 이용한 프레임 보간 방법 적용을 어렵게 한다.

### 발명의 내용

#### 해결하려는 과제

[0020] 본 발명은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 본 발명의 목적은, 딥러닝에 기반하여 다양한 컬러 영상 포맷에 대응하는 동시에 고품질, 고속으로 프레임을 보간하는 동영상 프레임 윌 변환 방법 및 장치를 제공함에 있다.

#### 과제의 해결 수단

[0022] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른, 동영상 프레임 윌 변환 장치는 각 채널의 영상에 대한 해상도를 조정하는 조정부; 조정부에서 해상도가 조정된 t초와 t+1초의 연속된 두 영상으로 중간 영상을 생성하는 생성부; 및 생성부에서 생성된 중간 영상의 해상도를 향상시키는 향상부;를 포함한다.

[0023] 조정부는, 각 채널의 영상에 대한 해상도를 동일하게 서브 샘플링하는 것일 수 있다.

[0024] 생성부는, 해상도가 조정된 영상의 특성을 계산하는 인코더부; 계산된 특성을 영상의 해상도 크기만큼 복원하는 디코더부; 조정부에서 해상도가 조정된 두 영상과 디코더의 출력을 이용하여 중간 영상을 생성하는 합성부;를 포함할 수 있다.

[0025] 향상부는, 생성부에서 생성된 중간 영상의 해상도를 원본 영상의 해상도로 향상시키는 것일 수 있다.

[0026] 향상부는, 디테일을 개선할 중간 영상의 해상도를 조정하는 제1 조정부; 원본 영상을 이용하여, 해상도가 조정된 중간 영상의 디테일을 개선하는 디테일 향상부; 제1 조정부에서 해상도가 조정되지 않은 중간 영상의 해상도를 조정하는 제2 조정부를 포함할 수 있다.

[0027] 디테일 향상부는, 선별적으로 선택된 채널만을 입력받는 것일 수 있다.

[0028] 선택된 채널은, 밝기 신호일 수 있다.

[0029] 디테일 향상부는, 생성부에서 생성되는 중간 정보를 참조하여, 중간 영상의 디테일을 개선하는 것일 수 있다.

[0030] 디테일 향상부는, 딥러닝 네트워크를 사용하는 것일 수 있다.

[0031] 한편, 본 발명의 다른 실시예에 따른, 동영상 프레임 윌 변환 장치는 각 채널의 영상에 대한 해상도를 조정하는 단계; 해상도가 조정된 t초와 t+1초의 연속된 두 영상으로 중간 영상을 생성하는 단계; 및 생성된 중간 영상의 해상도를 향상시키는 단계;를 포함한다.

[0032] 한편, 본 발명의 다른 실시예에 따른, 동영상 프레임 윌 변환 장치는 각 채널의 영상에 대한 해상도를 조정하는 조정부; 조정부에서 해상도가 조정된 t초와 t+1초의 연속된 두 영상으로 생성된 중간 영상의 해상도를 향상시키

는 향상부;를 포함한다.

[0033] 한편, 본 발명의 다른 실시예에 따른, 동영상 프레임 을 변환 장치는 각 채널의 영상에 대한 해상도를 조정하는 단계; 조정부에서 해상도가 조정된 t초와 t+1초의 연속된 두 영상으로 생성된 중간 영상의 해상도를 향상시키는 단계;를 포함한다.

**발명의 효과**

[0035] 이상 설명한 바와 같이, 본 발명의 실시예들에 따르면, 딥러닝 기반 동영상 프레임 을 변환에 있어, 딥러닝의 입력이 RGB 컬러 포맷 뿐만 아니라 YCbCr 등 다양한 컬러 포맷에 적용할 수 있게 되며, 프레임 보간 방법과 해상도 향상 방법을 동시에 이용하여 고품질, 고속의 프레임 을 변환이 가능해진다.

**도면의 간단한 설명**

- [0037] 도 1 : 프레임 보간 기술
- 도 2 : 발명의 도면
- 도 3 : 색차 신호 서브 샘플링, Y: 밝기 신호, Cb, Cr: 색차 신호
- 도 4 : 본 발명의 대표 도면
- 도 5 : 중간 프레임 생성부와 해상도 향상부 개념도
- 도 6 : 입력 영상 조정부와 디테일 향상부의 입출력 해상도간의 관계
- 도 7 : YCbCr 4:2:0 형식에 대한 본 발명의 예시
- 도 8 : YCbCr 4:2:2 형식에 대한 본 발명의 예시 1
- 도 9 : YCbCr 4:2:2 형식에 대한 본 발명의 예시 2
- 도 10 : YCbCr 4:4:4 또는 RGB 형식에 대한 본 발명의 예시 1
- 도 11 : YCbCr 4:4:4 또는 RGB 형식에 대한 본 발명의 예시 2

**발명을 실시하기 위한 구체적인 내용**

- [0038] 이하에서는 도면을 참조하여 본 발명을 보다 상세하게 설명한다.
- [0039] 본 발명의 실시예에 따른 동영상 프레임 을 변환 장치는, 도 2에 도시된 바와 같이, 입력 영상 조정부(110), 중간 프레임 생성부(120), 해상도 향상부(130)를 포함하여 구성된다.
- [0040] 대부분의 동영상은 RGB 형식에서 YCbCr 형식으로 변환되어 사용된다. YCbCr에서 Y는 밝기 신호, Cb와 Cr은 색차 신호라고 한다. 일반적으로 대부분 동영상의 YCbCr 영상의 색차 신호는 서브 샘플링이 되어 사용된다.
- [0041] YCbCr의 대표적인 서브 샘플링 기법으로는 도 2와 같이 YCbCr4:4:4, YCbCr4:2:2, YCbCr4:2:0 방법이 있다. YCbCr4:4:4는 색차신호인 Cb 와 Cr 신호를 서브샘플링 하지 않는 영상 포맷을 의미하며 Cb, Cr의 해상도와 Y의 해상도가 같다. YCbCr 4:2:2의 경우 색차 신호는 가로 방향으로 2씩 서브 샘플링 되고, YCbCr 4:2:0에서 색차 신호는 가로, 세로 2씩 서브 샘플링되었다. 이와 같이 색차 신호와 밝기 신호의 해상도는 컬러 포맷에 의해 서로 상이하다.
- [0042] Y, Cb, Cr의 각각을 하나의 컬러 채널이라고 하자. 일반적인 딥러닝 네트워크의 입력은 채널 간의 해상도가 동일해야 각각의 컬러 채널을 동시에 처리할 수 있다.
- [0043] 그러나 위에서 언급하였듯이 컬러 포맷에 따라 컬러 채널 간 해상도가 상이하다. 컬러 채널 간 해상도를 맞추기 위해서 본 발명의 실시예에서는 입력 신호를 조정하는 방식을 제안한다.
- [0044] 도 2에서, 입력 영상 조정부(110)는 딥러닝에 적용하기 전 각 채널의 영상의 해상도를 동일하게 조절하도록 한다. 즉 채널 별로 각각 서브 샘플링하여 동일한 해상도를 갖게 하도록 한다.
- [0045] 구체적으로 설명하기 위해 입력 신호가 N개의 채널로 구성되어 있고, 각각의 채널에 대한 해상도가  $H_1 \times W_1, \dots, H_N \times W_N$ 라고 하자. 여기서 H는 영상의 세로 해상도를 W는 가로 해상도를 의미한다.

- [0046] 입력 영상 조정부(110)는 모든 채널의 가로 해상도와 세로 해상도를 서브 샘플링을 통해 동일한 해상도  $H_M \times W_M$  으로 변형시킨다.  $H_M \times W_M$  은 가장 큰 해상도를 갖는 채널의 해상도보다는 크지 않는 것을 권장한다. 왜냐하면 딥러닝 네트워크의 입력 해상도가 커질 경우 그에 비례해 연산량이 증가하기 때문이다.
- [0047] 입력 영상 조정부(110)에서는 영상의 해상도를 조절함으로써 보간 프로세서의 연산량을 줄일 수 있다. 참고로 채널 별로 별도의 네트워크로 처리할 경우 밝기와 색차신호의 움직임이 다르게 예측되어 밝기 신호와 색차 신호가 합성된 신호는 불일치 생길 수 가 있다. 이와 같이 입력 영상 조정부(110)는 채널 별로 입력된 신호의 해상도를 조절하여 네트워크에 입력으로 사용될 영상을 만든다.
- [0048] 본 발명의 실시예에서 제안하는 네트워크 구조는 도 5와 같이 크게 두 가지 단계로 이루어져 있다. 첫 번째는 중간 영상을 생성하는 네트워크로 중간 프레임 생성부로 지칭한다. 이 네트워크는 기존의 중간 영상을 생성하는 네트워크가 사용될 수 있다. 두 번째 네트워크는 생성된 중간 영상을 원본 해상도와 일치하게 조절하는 네트워크로 해상도 향상부(130)라고 지칭한다.
- [0049] 도 4를 참조하여 본 발명의 실시예에 따른 동영상 프레임 을 변환 장치에 대하여 보다 상세히 설명한다.
- [0050] 중간 프레임 생성부(120)는 t초와 t+1초의 연속된 두 프레임의 영상( $\bar{I}_t, \bar{I}_{t+1}$ )을 입력으로 받는다. 두 프레임은 입력 영상 조정부(110)를 거쳐 두 프레임 간의 모든 채널의 영상 해상도는 일치된 상태가 된다.
- [0051] 중간 프레임 생성부(120)는 인코더부(121), 디코더부(122), 합성부(123)로 구성된다.
- [0052] 인코더부(121)는 컨볼루션 레이어 (convolutional layer), 풀링 레이어 (Pooling layer), 활성화 레이어 (Activation layer) 등으로 구성되어 있고 이에 한정하지 않는다.
- [0053] 인코더부(121)의 출력은 디코더부(122)의 입력이 되고 디코더부(122)에서는 디컨볼루션 레이어 (deconvolutional layer) 및 활성화 레이어 (activation layer) 등으로 구성되어 있으며 이에 한정하지 않는다.
- [0054] 디코더부(122)의 출력과 네트워크의 입력이 합성부(123)의 입력으로 들어가 1차 중간 (보간) 영상을 생성한다. 1차 중간 영상의 해상도는 중간 프레임 생성부(120)의 입력의 해상도와 동일하다. 디코더부(122)에서는 인코더부(121)의 출력뿐만 아니라 인코더부(121)의 중간 결과도 입력으로 받을 수 있고, 이를 이용해 성능을 개선할 수 있다.
- [0055] 도 5는 인코더부(121), 디코더부(122), 합성부(123)에 대한 개념을 보여준다.
- [0056] 인코더부(121)에서는 영상의 해상도가 압축되어 영상의 특성 (feature)을 계산하고, 디코더부(122)에서는 특성을 영상의 해상도 크기만큼 복원한다. 인코더부(121)의 중간 정보는 디코더부(122)에 직접적으로 전달되고 이는 도 5에서 화살표로 표현된다.
- [0057] 또한 인코더부(121)와 디코더부(122)의 중간 정보도 디테일 향상부(132)에 전달되어 재활용되어 진다.
- [0058] 해상도 향상부(130)는 해상도가 감소된 채널 신호에 대하여 해상도 증가를 실시한다. 중간 프레임 생성부(120)에 생성된 1차 중간 프레임의 모든 채널의 해상도는  $H_M \times W_M$  이다. 원본 영상의 해상도는  $H_1 \times W_1, \dots, H_N \times W_N$  이므로, 해상도 향상부(130)에서는 원본 영상의 해상도와 일치하게 영상의 크기를 조절한다. 1차 중간 프레임의 해상도는 원본 해상도에 비해 작은 것이 권장되므로 해상도 향상부(130)는 주로 해상도 증가 작업이 수행된다.
- [0059] 해상도 향상부(130)는 중간 영상 조정부(131), 디테일 향상부(132), 출력 영상 조정부(133)로 구성되어 있고, 이 중 디테일 향상부(132)만 딥러닝 네트워크를 사용하여 많은 연산량이 필요로 한다.
- [0060] 따라서 디테일 향상부(132)에 사용되는 입력은 모든 채널이 아닌 선별적으로 선택된 채널만 사용될 수 있다. 일반적으로는 밝기 신호의 대부분의 디테일이 집중되어 있으므로 디테일 향상부(132)의 입력으로는 밝기 신호만을 사용하는 것을 권장한다.
- [0061] 중간 영상 조정부(131)에서는 디테일 향상부(132)에 사용될 입력의 해상도를 조절한다. 도 6은 입력 영상 조정부(110)와 디테일 향상부(132)의 입출력 해상도간의 관계를 보여준다. 원본 영상의 n 번째 채널의 해상도는  $H_n \times W_n$  이라고 하고 n 번째 채널을 디테일 향상부(132)에 적용하는 상태이다.

- [0062] 디테일 향상부(132)의 입출력의 해상도는 원본 해상도와 동일해야 하고, 이를 위해 입력 영상 조정부(110)에서 1차 중간 프레임의 해상도를 조절하여 영상 크기 원본 해상도와 동일하게 한다. 조절하는 방법은 bilinear, bicubic 등의 필터를 사용할 수 있고 특정 방법에 제한되지 않는다.
- [0063] 디테일 향상부(132)의 입력은 1차 중간 영상과 원본 입력 영상  $I_t$  와  $I_{t+1}$  이다. 원본 입력 영상은 비록 다른 시간의 영상들이지만 원본 해상도의 디테일들을 보존하고 있다. 따라서 디테일 향상부(132)에는 이 정보를 사용하여 1차 중간 영상의 디테일을 개선하도록 한다.
- [0064] 이외에도 해상도 향상부(130)의 입력으로 중간 프레임 생성부(120)의 중간 결과들을 입력으로 받는다. 중간 프레임 생성부(120)의 중간 과정을 디테일 향상부(132)에서는 입력으로 받는다. 이 입력은 영상의 특징 (feature)을 포함하고 있으며 이 특징은 디테일 향상부(132)에 사용됨으로써 더욱 높은 성능을 제공한다.
- [0065] 출력 영상 조정부(133)에서는 최종적으로 보간 프레임 각각의 채널 해상도가 원본의 채널 해상도와 동일하게 조절해 준다. 디테일 향상부(132)에서 해상도가 향상된 채널은 원본 영상의 해상도와 동일하겠지만, 다른 채널의 영상은 원본 해상도와 다를 수 있다. 다시 말하면 원본 해상도와 다른 보간 영상의 채널에 대해서만 출력 영상 조정부(133)에서 해상도 조절이 수행된다.
- [0066] 본 발명의 실시예에서, 네트워크의 학습(training)은 중간 프레임 생성부(120)를 먼저 학습한 후, 디테일 향상부(132)를 학습하는 것을 권장한다. 또는 위와 같이 학습을 수행한 후, 엔드 투 엔드(end-to-end)로 전체 네트워크를 한 번에 학습하는 것을 권장한다. 다른 방법으로는 중간 프레임 생성부(120)를 학습한 후, 중간 프레임과 디테일 향상부(132)를 동시에 학습하는 것을 권장한다.
- [0068] 다음은 본 발명의 다양한 구형예들을 보여준다.
- [0069] 전체 네트워크의 입력은 동영상에서 연속된 두 프레임이며, 이를  $I_t$  와  $I_{t+1}$  이라고 하자. 전체 네트워크의 출력은  $I_{t'}$  이며  $t'$  은 0과 1사이의 값이다. 즉 출력 영상은  $t$  와  $t+1$  사이의 영상이 된다. 여기서  $t$ 는 시간을 의미한다. 입력 영상이 YCbCr 컬러 포맷을 갖는다고 하면 입력 영상  $I_t$  와  $I_{t+1}$ 은  $I_t = \{Y_t, Cb_t, Cr_t\}$  와  $I_{t+1} = \{Y_{t+1}, Cb_{t+1}, Cr_{t+1}\}$ 로 표현할 수 있다.  $Y_t$ 와  $Y_{t+1}$ 은 밝기 신호를,  $Cb_t, Cr_t, Cb_{t+1}, Cr_{t+1}$  은 색차 신호를 의미한다. 입력 영상 조정부(110)의 입력은  $I_t$  와  $I_{t+1}$ 이 되며 출력은  $\bar{I}_t = \{\bar{Y}_t, \bar{Cb}_t, \bar{Cr}_t\}$  와  $\bar{I}_{t+1} = \{\bar{Y}_{t+1}, \bar{Cb}_{t+1}, \bar{Cr}_{t+1}\}$  이 된다.  $\bar{X}$  은 서브 샘플링된 영상을 의미하며  $\bar{X}$ 사이의 해상도는 모두 같아야 한다. 다시 말하면 채널 간 해상도는 같아야 한다.
- [0070] 입력 영상 조정부(110)에 대한 실시예를 YCbCr 4:2:0 컬러 형식을 갖는 동영상에 대하여 도 7과 같이 살펴보자. 밝기 신호 Y의 해상도가 H x W이면 색차 신호 Cb, Cr은 H/2 x W/2 의 해상도를 갖는다. 입력 영상 조정부(110)에서 영상 신호의 출력 해상도는 H/2 x W/2로 설정한다. 이와 같은 경우 Y의 신호만 가로, 세로 2배씩 서브 샘플링하여 H/2 x W/2로 생성한다.  $\bar{Y}_t, \bar{Cb}_t, \bar{Cr}_t$ 과  $\bar{Y}_{t+1}, \bar{Cb}_{t+1}, \bar{Cr}_{t+1}$ 은 모두 H/2 x W/2의 해상도를 갖는다. 출력 영상의 해상도는 갈 컬러 채널 중 가장 작은 해상도를 갖는 채널보다 작거나 같아야 한다. 즉, YCbCr 4:2:0영상의 경우에는 출력 해상도는 H/2 x W/2 보다 작거나 작아야 한다. 다른 실시예로 YCbCr 4:2:2 영상에 대하여 살펴보자. Y의 해상도를 H x W라고 하면 Cb, Cr의 해상도는 H x W/2이다. 그럼 입력 영상 조정부(110)의 출력 해상도는 H x W/2 이거나 그 이하여야 한다. 따라서 사용자의 설정에 따라  $\bar{Y}_t, \bar{Cb}_t, \bar{Cr}_t$ 과  $\bar{Y}_{t+1}, \bar{Cb}_{t+1}, \bar{Cr}_{t+1}$ 의 해상도는 H x W/2 가 될 수도 있고, H/2 x W/2 등도 될 수 있다.
- [0071] 중간 프레임 생성부(120)의 입력은  $\bar{I}_t = \{\bar{Y}_t, \bar{Cb}_t, \bar{Cr}_t\}$  과  $\bar{I}_{t+1} = \{\bar{Y}_{t+1}, \bar{Cb}_{t+1}, \bar{Cr}_{t+1}\}$  이다. 도 7의 YCbCr 4:2:0 포맷의 경우에는  $\bar{Cb}_{t+1}, \bar{Cr}_{t+1}$  과  $Cb_{t+1}, Cr_{t+1}$ 이 같고 t 프레임도 마찬가지다. 중간 프레임 생성부의 출력은  $\hat{I}_{t'} = \{\hat{Y}_{t'}, \hat{Cb}_{t'}, \hat{Cr}_{t'}\}$  이 된다.  $\hat{I}_{t'}$  은 t와 t+1 프레임 사이에서 보간된 중간 프레임이며  $\bar{I}_t$  와  $\hat{I}_{t'}$  의 해상도는 동일하다. 중간 프레임 생성부(120)는 인코더부, 디코더부, 합성부로 구성되나 이에 한정되지 않는다. 기존의 프레임 보간 방법을 사용하여도 무관하다. 인코더부는 컨볼루션층 레이어, 풀링 레이어, 활성화 레이어로 구성되어 있다. 아래 도 처럼 입력 영상에 대하여 컨볼루션을 수행하고 풀링 레이어를 수행하고 활성화 레이어를 수행한다. 풀링 레이어는 영상의 해상도를 감소시킨다. 일반적으로 가로, 세로 각각 2배씩 감소시킨다. 컨볼루션층

레이어의 출력을  $C_L$ , 풀링 레이어의 출력을  $P_L$ , 활성 레이어의 출력을  $A_L$ 이라고 한다. L은 레이어의 순서를 의미하며 이 각각의 레이어의 출력은 해상도 증가부의 입력이 된다. 디코더부는 인코더부에서 줄어든 해상도를 다시 증가시키면서 특징을 추출한다. 해상도 증가는 Bilinear 필터 또는 디콘볼루션 레이어 등 다양한 방법을 사용할 수 있으며 특정 방법에 제한되지 않는다. 각각의 업샘플링 레이어의 출력을  $U_L$ 이라고 정의하며 L은 레이어의 순서를 의미한다. 각각의 업샘플링 레이어의 출력  $U_L$ 은 해상도 향상부(130)의 입력으로 사용된다. 디코더부의 최종 출력  $U_F$ 와  $\bar{I}_t = \{\bar{Y}_t, \bar{Cb}_t, \bar{Cr}_t\}$ ,  $\bar{I}_{t+1} = \{\bar{Y}_{t+1}, \bar{Cb}_{t+1}, \bar{Cr}_{t+1}\}$ 은 합성부의 입력이 된다. 합성은 콘볼루션 필터, 와핑 (Warping) 등 다양한 방법이 사용될 수 있으며 본 발명에서는 특정 방법에 제한되지 않는다. 합성부의 출력은  $\hat{I}_t = \{\hat{Y}_t, \hat{Cb}_t, \hat{Cr}_t\}$ 이 되며 이를 1차 중간 프레임으로 부른다.

[0072] 1차 중간 프레임의 해상도는 원영상에 비해서 작은 해상도를 갖는다. 따라서 해상도 향상부(130)에서는 1차 중간 프레임의 해상도를 증가시켜 원영상과 같은 해상도를 갖게 하는 동작을 수행한다. 해상도 향상부(130)의 입력은 해상도 향상부(130)의 입출력은 다양하게 정의될 수 있으며 다음에서 몇 가지 실시 예에 대하여 살펴본다.

[0073] 먼저 원본 입력 영상이 YCbCr 4:2:0 컬러 포맷이며 Y의 해상도는 H x W, Cb, Cr의 해상도는 H/2 X W/2이라고 하자. 입력 영상 조정부(110)에서 Y의 해상도를 서브샘플링하여 Cb, Cr의 해상도인 H/2 x W/2 와 일치시킨다. 따라서 중간 프레임 생성부(120)의 해상도는 Y, Cb, Cr의 경우 H/2 x W/2 가 되며, Y의 경우 원영상에 비해 해상도가 줄었지만 Cb, Cr의 경우 해상도가 변하지 않았으므로 Y 영상만 해상도 향상부(130)를 사용하여 해상도를 증가시킨다. 해상도 향상부(130)는 중간 영상 조정부(131)와 디테일 향상부(132)의 두 단계로 구성된다. 중간 영상 조정부(131)는 1차 중간 프레임을 업샘플링하여 원 영상의 크기와 동일하게 만든다.

[0074] 위의 경우 중간 영상 조정부(131)의 입력은 H/2 x W/2 의 해상도를 갖는 1차 중간 프레임  $\bar{Y}_t$ 이고 출력은 H x W 해상도를 갖는 영상  $\tilde{Y}_t$ 이다. 디테일 향상부(132)의 입력은  $\tilde{Y}_t$ 와 원영상  $Y_t, Y_{t+1}$ 이다.  $\tilde{Y}_t$ 의 디테일은 고해상도 디테일을 가지고 있는  $Y_t, Y_{t+1}$ 의 협력을 받아 더욱 정교하게 복원될 수 있다. 기존의 다중 프레임 (multi-frame) 해상도 향상 방법은 여러 장의 저해상도 영상을 사용하여 고해상도 영상을 사용하였다. 그러나 본 발명에서는 저해상도 영상과 고해상도 영상을 동시에 이용하여 해상도 향상을 꾀하였다. 디테일 향상부(132)는 딥러닝 네트워크로 구성되어 있으며 출력은  $\bar{Y}_t$ 이다. 디테일 향상부(132)를 위한 딥러닝 네트워크는 컨벌루션 셔널 레이어, 풀링 레이어, 활성 레이어 등으로 구성되어 있으며, 본 논문에서는 특정 네트워크 형태로 제한하지 않는다. 디테일 향상부(132)는 중간 프레임 생성부(120)의 중간 결과물을 입력으로 받아 디테일 향상부(132)의 중간 과정에 더하거나 적층하여 (concatenation) 성능을 개선할 수 있다. 중간 생성부의 중간 결과물과 디테일 향상부(132)의 중간 결과물의 해상도는 일치해야한다.  $\bar{Y}_t$ 와 1차 중간 프레임의 색차 신호  $\bar{Cb}_t, \bar{Cr}_t$ 이 최종 결과물이 된다.

[0075] 원본 입력 영상이 YCbCr 4:2:2 컬러 포맷인 실시예에 대하여 살펴보자. Y의 해상도는 H x W, Cb, Cr의 해상도는 H x W/2 라고 하자. 도 8과 같이 입력 영상 조정부(110)에서 Y의 해상도를 서브샘플링하여 Cb, Cr의 해상도인 H x W/2 와 일치시킨다. 따라서 중간 프레임 생성부(120)의 해상도는 Y, Cb, Cr의 경우 H x W/2 가 되며, Y의 경우 원영상에 비해 해상도가 줄었지만 Cb, Cr의 경우 해상도가 변하지 않았으므로 Y 영상만 해상도 향상부(130)를 사용하여 해상도를 증가시킨다. 해상도 향상부(130)의 작동 방식은 위의 YCbCr 4:2:0 형식일 때와 동일하다. H x W의 해상도를 가지는  $\bar{Y}_t$ 와 1차 H x W/2의 해상도를 가지는 중간 프레임의 색차 신호  $\bar{Cb}_t, \bar{Cr}_t$ 이 최종 결과물이 된다.

[0076] YCbCr 4:2:2 컬러 포맷의 다른 실행 예시는 도 9와 같다. Y의 해상도는 H x W, Cb, Cr의 해상도는 H x W/2 라고 하자. 입력 조정부에서 Y, Cb, Cr의 해상도를 모두 H/2 x W/2로 서브 샘플링한다. 그리고 위의 YCbCr 4:2:0 과 동일한 방식으로 처리해준다. 이와 같은 과정을 거칠 경우 디테일 향상부(132)의 출력은 Y신호는 H x W의 해상도를 갖으며, Cb, Cr 신호는 H/2 x W/2 의 해상도를 갖는다. Cb, Cr의 신호가 원본 신호와 다르기 때문에 출력 영상 조정부(133)에서 Cb, Cr의 해상도를 원본 신호와 동일하게 조정하여 준다. 조정 방법은 Bicubic 필터 등이 사용될 수 있으며 특정 방법에 제한 받지 않는다. 이와 같이 할 경우 위의 방법에 비해 네트워크에서 처리되는 입력 영상 크기가 작으므로 연산량과 메모리를 절약할 수 있다는 장점이 있다. 그리고 YCbCr 4:2:0에서 사

용되었던 네트워크를 공유할 수 있다는 장점이 있다.

[0077] YCbCr 4:4:4나 RGB 입력 영상이 입력인 경우를 고려하여 보자. YCbCr 4:4:4나 RGB 컬러 형식의 동영상은 모든 채널의 해상도가 동일함으로 입력 영상 조정부(110)와 해상도 향상부(130)가 사용되지 않아도 중간 프레임이 생성될 수 있다. 그러나 연산량을 감소시키기 위해서 본 발명에서는 입력 영상 조정부(110)와 해상도 향상부(130)를 사용할 수 있다. 예를 들어 입력 영상이 RGB 형식이고 각 채널의 해상도는 H x W라고 가정하자. 입력 영상 조정부(110)에서 각 채널을 H/2 x W/2 의 크기로 서브샘플링 한 후 중간 프레임 생성부(120)의 입력으로 넣는다. 그리고 H/2 x W/2의 해상도의 1차 중간 프레임은 해상도 향상부(130)에서 H x W의 해상도로 복원된다. 전체 프로세스에서 속도의 이점은 중간 프레임 생성부(120)의 연산량이 해상도 향상부(130)의 연산량보다 높다는 데에서 온다. 따라서 연산량이 높은 중간 프레임 생성부(120)는 작은 해상도로 처리하고 이를 다시 해상도 향상부(130)에서 복원하여 연산량을 줄일 수 있도록 한다. YCbCr 4:4:4 의 경우도 유사한 방식으로 적용할 수 있다. 각 채널의 해상도가 H x W이면 입력 영상 조절부에서 H/2 x W/2로 변환한다. 그리고 중간 프레임 생성부(120)에서 H/2 x W/2의 Y, Cb, Cr 결과를 생성한다. 1차 중간 프레임은 해상도 향상부(130)에 2가지 방법으로 적용될 수 있다. 첫 번째 방법은 위에서 설명한 RGB 영상처럼 도 10과 같이 Y, Cb, Cr 모든 채널에 대하여 디테일 향상부(132)의 딥러닝 네트워크를 통과시키는 것이다. 두 번째 방법은 위에서 설명한 YCbCr 4:2:2 방식처럼 도 11과 같이 Y 영역만 딥러닝 네트워크를 통과하여 고화질 영상으로 복원하고 Cb, Cr은 출력영상 조정부에서 단순 업스케일링을 통해 복원하는 것이다. 이것은 Cb, Cr은 일반적으로 디테일 영역이 적기 때문에 단순 업스케일링을 통해서도 복원이 가능하다. 이와 같은 방법은 전체 채널 영상에 대하여 딥러닝 네트워크를 사용하는 것에 비해 더욱 적은 연산량으로 복원이 가능하다. 또한 YCbCr4:2:0 네트워크와 공통으로 사용할 수 있다는 장점이 있어 범용성이 높다는 장점이 있다.

[0078] 한편, 본 실시예에 따른 장치와 방법의 기능을 수행하게 하는 컴퓨터 프로그램을 수록한 컴퓨터로 읽을 수 있는 기록매체에도 본 발명의 기술적 사상이 적용될 수 있음은 물론이다. 또한, 본 발명의 다양한 실시예에 따른 기술적 사상은 컴퓨터로 읽을 수 있는 기록매체에 기록된 컴퓨터로 읽을 수 있는 코드 형태로 구현될 수도 있다. 컴퓨터로 읽을 수 있는 기록매체는 컴퓨터에 의해 읽을 수 있고 데이터를 저장할 수 있는 어떤 데이터 저장 장치이더라도 가능하다. 예를 들어, 컴퓨터로 읽을 수 있는 기록매체는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광디스크, 하드 디스크 드라이브, 등이 될 수 있음은 물론이다. 또한, 컴퓨터로 읽을 수 있는 기록매체에 저장된 컴퓨터로 읽을 수 있는 코드 또는 프로그램은 컴퓨터간에 연결된 네트워크를 통해 전송될 수도 있다.

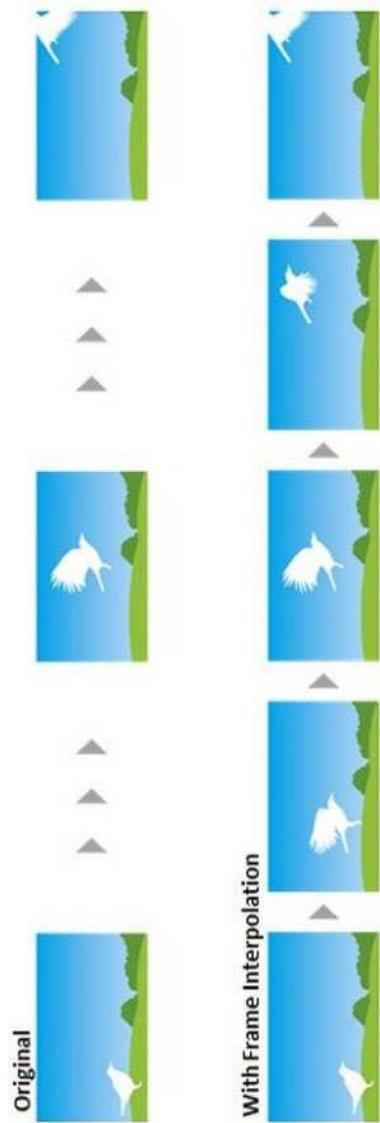
[0079] 또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특성의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어져서는 안될 것이다.

**부호의 설명**

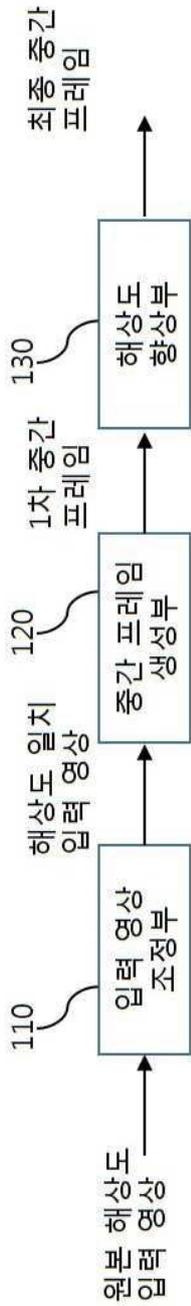
- [0081] 110 : 입력 영상 조정부
- 120 : 중간 프레임 생성부
- 121 : 인코더부
- 122 : 디코더부
- 123 : 합성부
- 130 : 해상도 향상부
- 131 : 중간 영상 조정부
- 132 : 디테일 향상부
- 133 : 출력 영상 조정부

도면

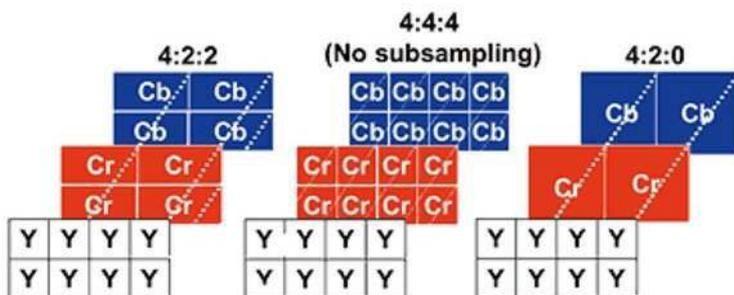
도면1



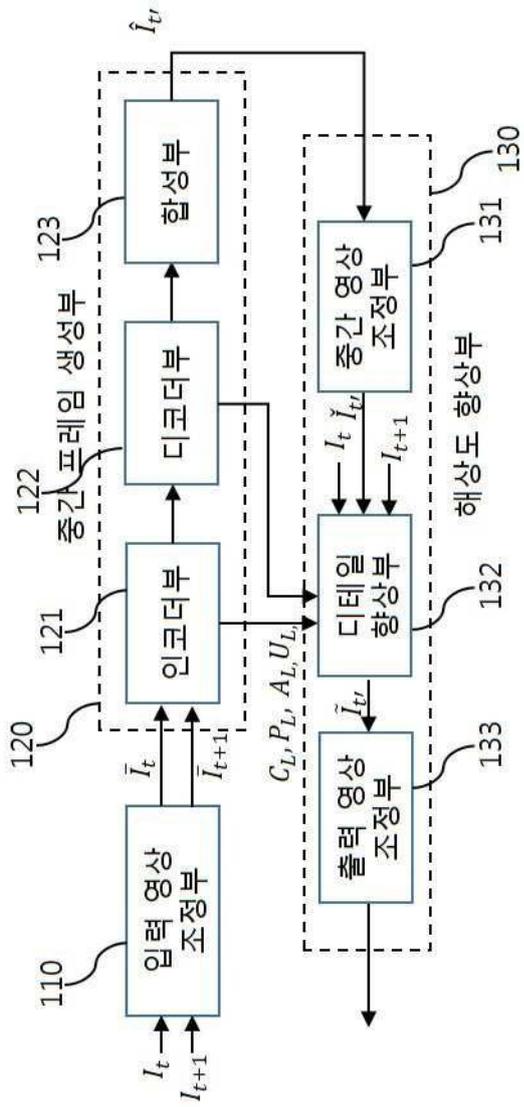
도면2



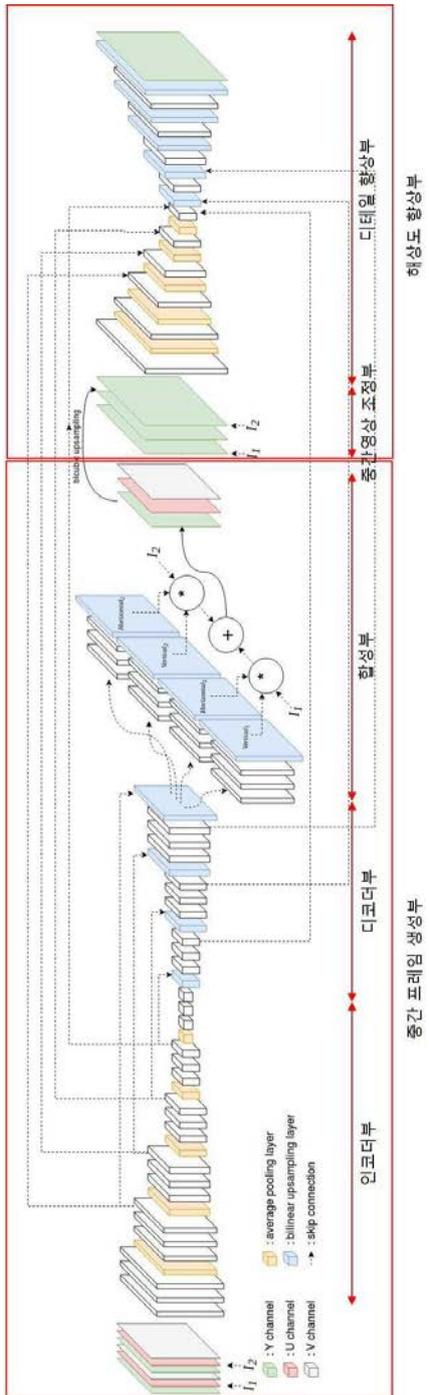
도면3



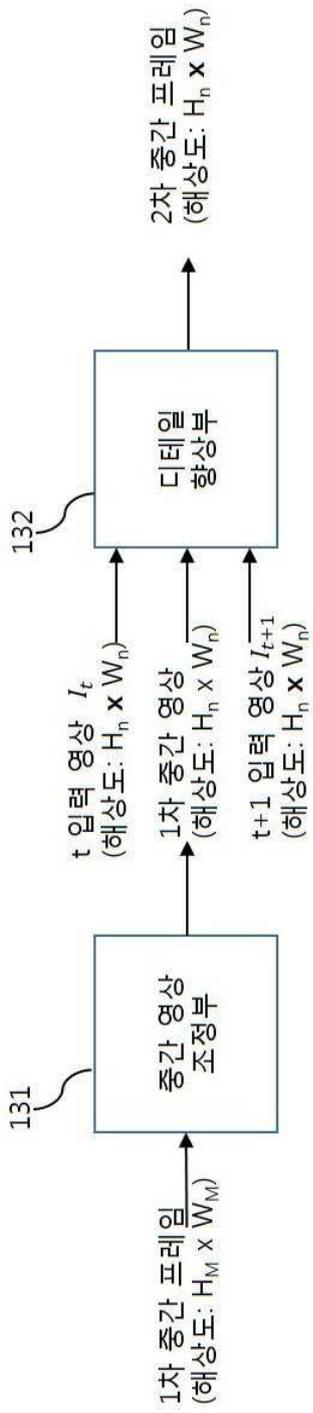
도면4



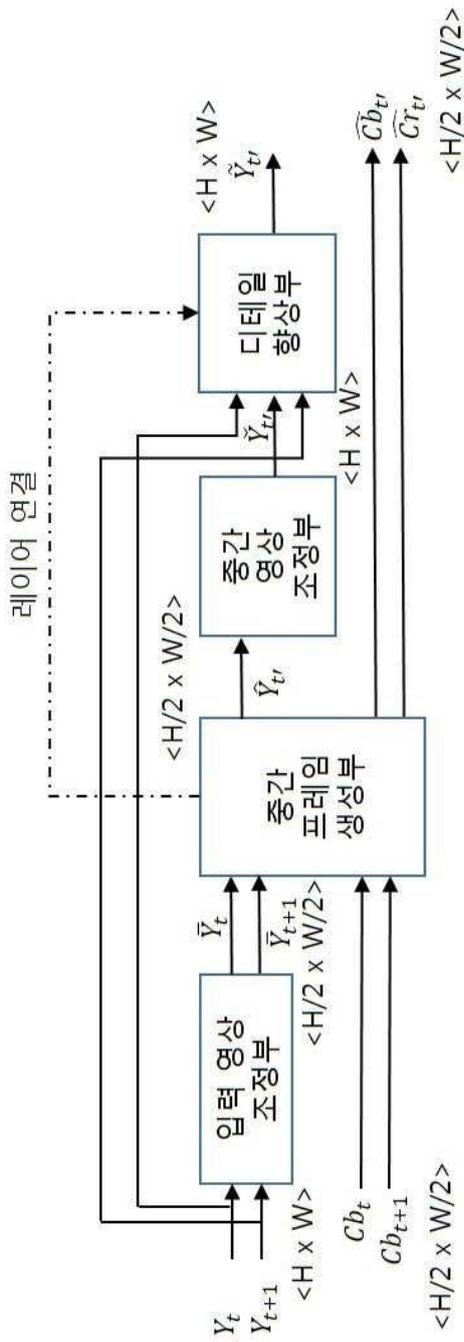
도면5



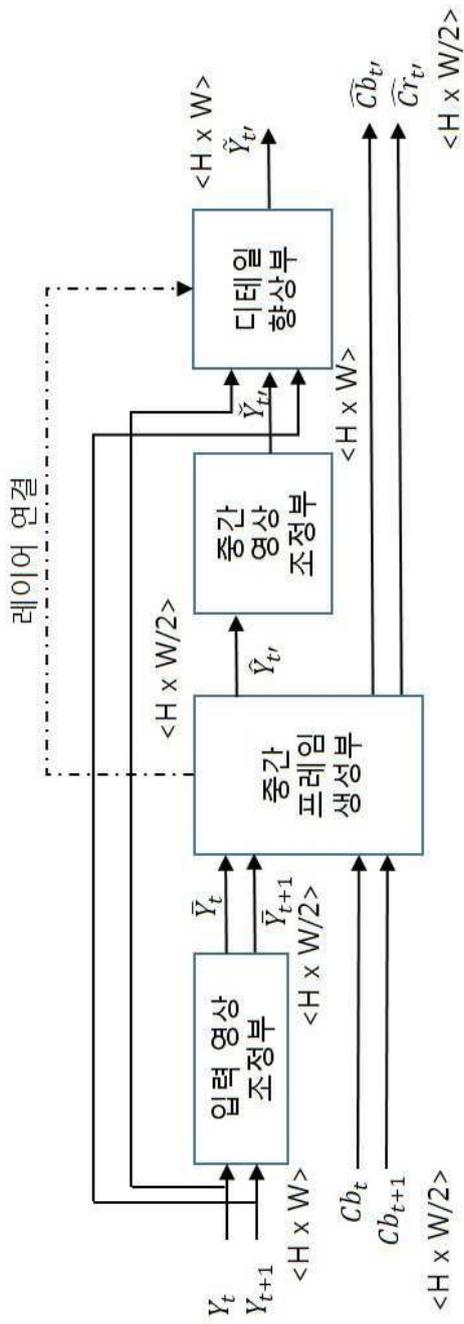
도면6



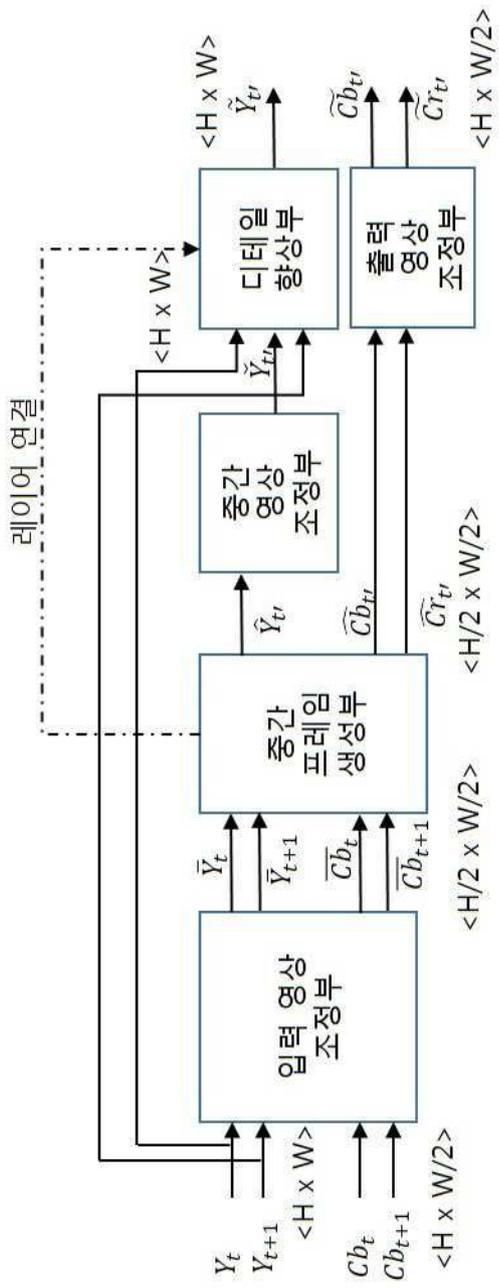
도면7



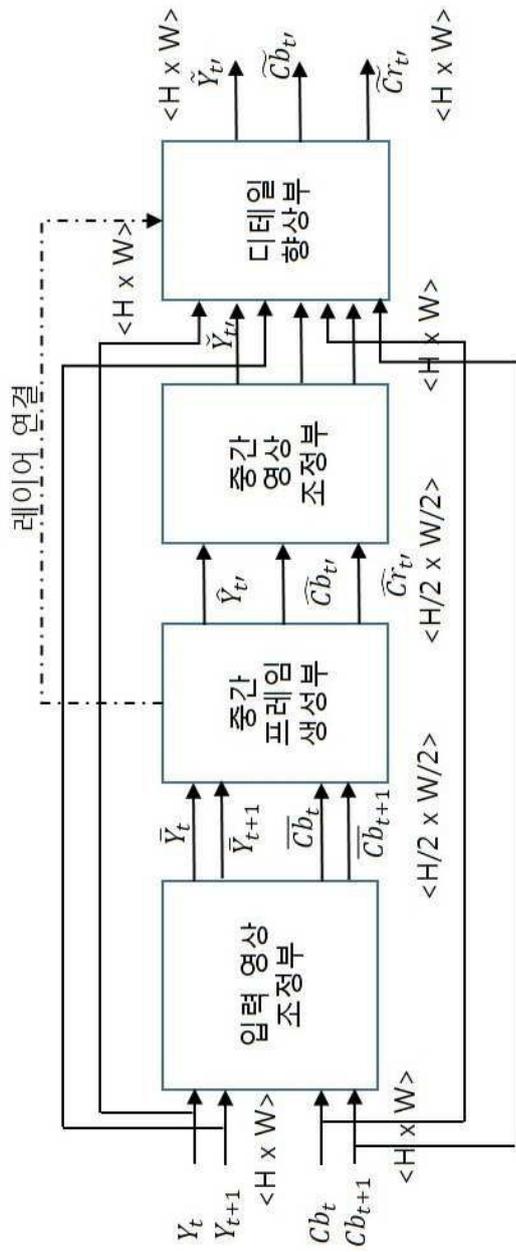
도면8



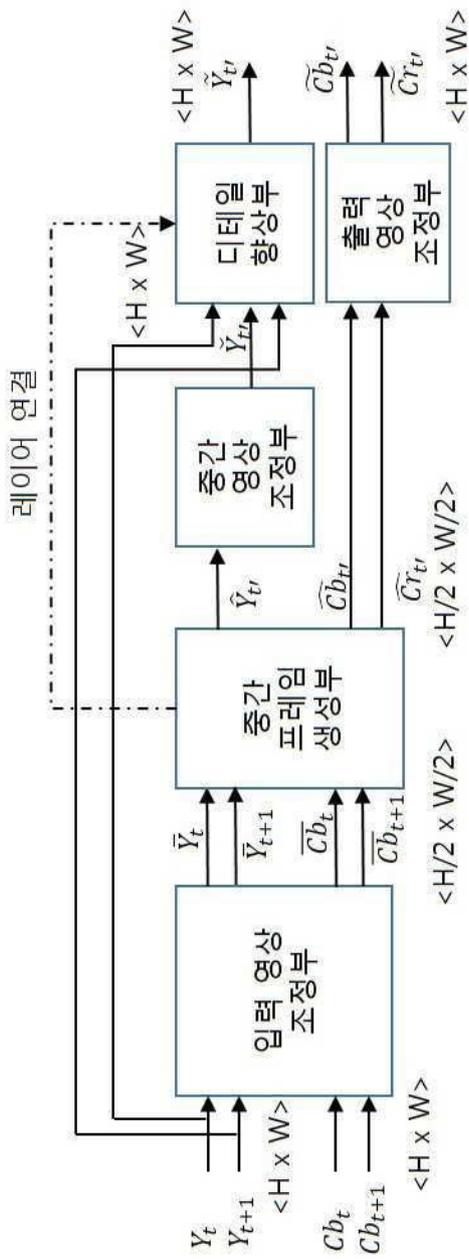
도면9



도면10



도면11





(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2021년04월20일  
(11) 등록번호 10-2242343  
(24) 등록일자 2021년04월14일

(51) 국제특허분류(Int. Cl.)  
H04N 7/01 (2006.01) H04N 21/2343 (2011.01)  
H04N 21/4402 (2011.01)  
(52) CPC특허분류  
H04N 7/0127 (2013.01)  
H04N 21/234381 (2013.01)  
(21) 출원번호 10-2019-0137132  
(22) 출원일자 2019년10월31일  
심사청구일자 2020년02월20일  
(56) 선행기술조사문헌  
A Fast 4K Video Frame Interpolation Using a Hybrid Task-Based Convolutional Neural Network.\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
한국전자기술연구원  
경기도 성남시 분당구 새나리로 25 (야탑동)  
(72) 발명자  
정진우  
서울특별시 송파구 송파대로 111, 파크하비오105동 1205호  
안하은  
서울특별시 성북구 성북로8다길 37, 102호  
김제우  
경기도 성남시 분당구 수내로 181, 303동 901호  
(74) 대리인  
남충우

전체 청구항 수 : 총 12 항

심사관 : 박재학

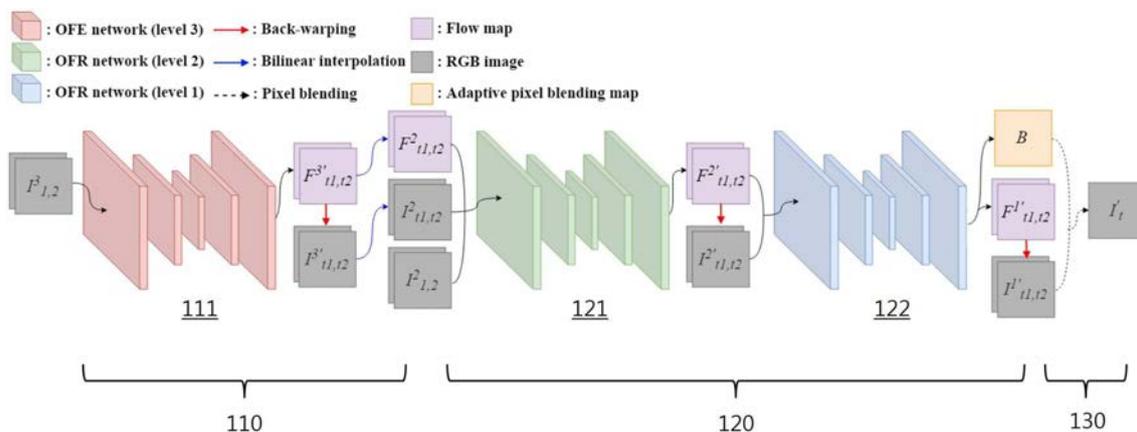
(54) 발명의 명칭 **고해상도 동영상 프레임 울 고속 변환 방법 및 장치**

(57) 요약

고해상도 비디오 영상에 대하여 고품질, 고속으로 프레임 보간을 수행하기 위한 방안으로, 입력 고해상도 영상을 저해상도 영상으로 변환하여 고속으로 광학 흐름 지도를 생성하고 이를 원본 고해상도로 복원하여 고해상도 영상을 고속으로 보간하는 동영상 프레임 울 고속 변환 방법 및 장치가 제공된다. 본 발명의 실시예에 따른, 동영상

(뒷면에 계속)

대표도 - 도4



프레임 을 변환 방법은 시간적으로 연속된 고해상도의 프레임들로부터 생성한 저해상도의 프레임들들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 제1 생성단계; 저해상도의 광학 흐름 지도들의 해상도를 단계적으로 높이면서, 고해상도의 중간 프레임들을 생성하는 제2 생성단계; 생성된 고해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임들을 생성하는 제3 생성단계:를 포함한다.

이에 의해, 입력 고해상도 영상을 저해상도 영상으로 변환하여 고속으로 광학 흐름 지도를 생성하고 이를 원본 고해상도로 복원하여 고해상도 영상을 보간함으로써, 4K와 같은 고해상도 비디오 영상에 대하여 실시간성을 요구하는 시스템 환경에서도 고품질, 고속으로 프레임 보간을 수행할 수 있게 된다.

(52) CPC특허분류

**H04N 21/440281** (2013.01)

**H04N 7/0137** (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711081116
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대(UHD)방송서비스활성화기술개발(R&D)
연구과제명	영상콘텐츠 초고속/초고화질 변환 기술 개발
기 여 율	1/1
과제수행기관명	(주)픽스트리
연구기간	2019.01.01 ~ 2019.12.31

공지예외적용 : 있음

## 명세서

### 청구범위

#### 청구항 1

시간적으로 연속된 고해상도의 프레임들로부터 생성한 저해상도의 프레임들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 제1 생성단계;

저해상도의 광학 흐름 지도들의 해상도를 단계적으로 높이면서, 고해상도의 중간 프레임들을 생성하는 제2 생성단계;

생성된 고해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임을 생성하는 제3 생성단계:를 포함하는 것을 특징으로 하는 동영상 프레임 을 변환 방법.

#### 청구항 2

청구항 1에 있어서,

제1 생성단계는,

입력되는 시간적으로 연속된 프레임들의 해상도인 제1 해상도 보다 낮은 제3 해상도의 프레임들로 광학 흐름을 예측하여, 제3 해상도의 광학 흐름 지도들을 생성하는 단계;

제3 해상도의 광학 흐름 지도들을 이용하여, 제3 해상도의 중간 프레임들을 생성하는 단계;

제3 해상도의 중간 프레임들과 광학 흐름 지도들로, 제3 해상도 보다 높은 제2 해상도의 중간 프레임들과 광학 흐름 지도들을 복원하는 단계;를 포함하고,

제2 생성단계는,

제2 해상도의 중간 프레임들과 광학 흐름 지도들을 이용하여, 제2 해상도 보다 높은 제1 해상도의 중간 프레임들을 생성하는 제2 생성단계;를 포함하는 것을 특징으로 하는 동영상 프레임 을 변환 방법.

#### 청구항 3

청구항 2에 있어서,

제2 생성단계는,

복원된 제2 해상도의 중간 프레임들과 광학 흐름 지도들로 광학 흐름을 예측하여, 제2 해상도의 광학 흐름 지도들을 생성하는 단계;

제2 해상도의 광학 흐름 지도들을 이용하여, 제2 해상도의 중간 프레임들을 생성하는 단계;

생성된 제2 해상도의 중간 프레임들과 광학 흐름 지도들로 광학 흐름을 예측하여, 제1 해상도의 광학 흐름 지도들을 생성하는 단계;

제1 해상도의 광학 흐름 지도들을 이용하여, 제1 해상도의 중간 프레임들을 생성하는 단계;를 포함하고,

제3 생성단계는,

생성된 제1 해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임을 생성하는 것을 특징으로 하는 동영상 프레임 을 변환 방법.

#### 청구항 4

청구항 3에 있어서,

제2 해상도의 광학 흐름 지도 생성단계는,

복원된 제2 해상도의 중간 프레임들과 광학 흐름 지도들에 제1 해상도의 프레임들을 제2 해상도로 다운 샘플링 한 프레임들을 추가로 이용하여, 광학 흐름을 예측하는 것을 특징으로 하는 동영상 프레임 윌 변환 방법.

#### 청구항 5

청구항 3에 있어서,

복원 단계는,

선형 보간 방법을 이용하여, 제3 해상도의 중간 프레임들과 광학 흐름 지도들로, 제2 해상도의 중간 프레임들과 광학 흐름 지도들을 복원하는 것을 특징으로 하는 동영상 프레임 윌 변환 방법.

#### 청구항 6

청구항 3에 있어서,

제3 해상도의 광학 흐름 지도 생성단계, 제2 해상도의 광학 흐름 지도 생성단계 및 제1 해상도의 광학 흐름 지도 생성단계는,

훈련된 인공지능 모델을 이용하여, 제3 해상도의 광학 흐름 지도, 제2 해상도의 광학 흐름 지도 및 제1 해상도의 광학 흐름 지도를 생성하는 것을 특징으로 하는 동영상 프레임 윌 변환 방법.

#### 청구항 7

청구항 6에 있어서,

제3 해상도의 광학 흐름 지도 생성단계는,

제3 해상도의 프레임들과 제3 해상도의 광학 흐름 지도들을 bak-warping 연산하여 생성하고,

제2 해상도의 광학 흐름 지도 생성단계는,

제2 해상도의 프레임들과 제2 해상도의 광학 흐름 지도들을 bak-warping 연산하여 생성하며,

제1 해상도의 광학 흐름 지도 생성단계는,

제1 해상도의 프레임들과 제1 해상도의 광학 흐름 지도들을 bak-warping 연산하여 생성하는 것을 특징으로 하는 동영상 프레임 윌 변환 방법.

#### 청구항 8

청구항 6에 있어서,

제3 생성단계는,

$$B = \frac{1}{1 + e^{-k_1 k_2}}$$

위의 블렌딩 파라미터 B로 제1 해상도의 중간 프레임들을 블렌딩 연산하고,

$k_1$ 은 sigmoid 함수의 기울기를 조절하며,

$K_2$ 는 sigmoid 함수의 입력으로 주어지는 것을 특징으로 하는 동영상 프레임 율 변환 방법.

**청구항 9**

청구항 6에 있어서,

인공지능 모델의 훈련은,

원본 입력 영상과 보간 영상을 다시 원본 영상으로 back-warping한 영상과의 차이를 계산하는 함수, 생성된 광학 흐름 지도들의 전과 후의 차이를 계산하는 함수, adversarial 손실 함수를 이용하는 것을 특징으로 하는 동영상 프레임 율 변환 방법.

**청구항 10**

시간적으로 연속된 고해상도의 프레임들로부터 생성한 저해상도의 프레임들들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 예측부;

저해상도의 광학 흐름 지도들의 해상도를 단계적으로 높이면서, 고해상도의 중간 프레임들을 생성하는 향상부;

생성된 고해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임을 생성하는 생성부:를 포함하는 것을 특징으로 하는 동영상 프레임 율 변환 장치.

**청구항 11**

저해상도의 프레임들들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 제1 생성단계;

저해상도의 광학 흐름 지도들을 이용하여, 고해상도의 중간 프레임들을 생성하는 제2 생성단계; 및

생성된 고해상도의 중간 프레임들을 이용하여, 최종 보간 프레임을 생성하는 제3 생성단계:를 포함하는 것을 특징으로 하는 동영상 프레임 율 변환 방법.

**청구항 12**

시간적으로 연속된 고해상도의 프레임들로부터 생성한 저해상도의 프레임들들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 제1 생성단계;

저해상도의 광학 흐름 지도들의 해상도를 단계적으로 높이면서, 고해상도의 중간 프레임들을 생성하는 제2 생성단계;

생성된 고해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임을 생성하는 제3 생성단계:를 포함하는 것을 특징으로 하는 동영상 프레임 율 변환 방법을 수행할 수 있는 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록매체.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 동영상의 프레임 율 변환 기술에 관한 것으로, 더욱 상세하게는 딥러닝 기법에 기반한 광학 흐름 추정 지도를 저해상도에서 예측하고 이를 고해상도로 복원하여 고해상도 동영상의 프레임을 고속으로 보간하는 동영상 프레임 율 고속 변환 방법 및 장치에 관한 것이다.

**배경 기술**

- [0003] 1. 동영상 프레임 율 변환 기법 개요
- [0004] 동영상은 연속된 정지 영상의 집합으로 구성된다. 비디오에서 정지 영상을 프레임이라고 부르며 단위 시간 당 프레임의 수를 동영상의 프레임 율 (frame rate)이라고 한다. 예를 들어 1초에 24장의 프레임으로 구성되면 프레임 율은 24 fps (frame per second)가 된다. 프레임 율은 촬영자의 의도, 영상의 포맷, 카메라의 한계 등에 의하여 결정된다.
- [0005] 관찰자가 영상을 연속된 화면으로 느끼기 위해서는 어느 정도 이상의 프레임 율이 필요하고 이보다 낮을 경우 움직임이 부드럽지 않아 보인다. 이 현상은 디스플레이의 크기, 조명, 시청 거리 등에 의해 달라질 수 있다. 이를 개선하기 위해 동영상의 프레임 율을 후처리에 의해 증가시키는 것을 동영상 프레임 율 변환이라고 한다.
- [0007] 2. 종래 기술
- [0008] 동영상 프레임 율을 증가시키는 가장 간단한 방법은 프레임을 반복하는 것이다. 예를 들어 30 fps 영상을 60 fps 영상으로 증가시킬 경우, 각 프레임마다 한 프레임을 반복하여 출력하는 것이다.
- [0009] 그러나 이 방법의 경우 동영상의 정보량은 동일하고 움직임에 대한 연속성은 변하지 않았으므로 관찰자가 느끼는 불편감은 동일하다. 이를 해결하기 위해 연속된 프레임들을 이용하여 가상의 프레임을 생성하는 기술이 개발되었다. 즉 t 초와 t+1 초 사이의 영상을 이용하여 t+0.5 초의 중간 영상을 새롭게 생성하며 이를 프레임 보간 (frame interpolation) 기술(도 1 참조)이라고 한다.
- [0010] 프레임 보간은 다양한 방법이 개발되었으며 일반적으로는 다음과 같은 두 단계 과정을 거친다. 첫 번째 단계는 움직임 또는 광학 흐름 지도를 획득하는 단계이며 두 번째 단계는 움직임 정보를 바탕으로 중간 프레임을 생성 (Synthesis)하는 단계이다.
- [0011] 동영상에서 물체의 움직임이 부드럽게 보이려면 중간 영상은 물체의 움직임이 두 영상 사이의 중간에 해당되어야 한다. 따라서 물체의 움직임 정보를 가지고 있는 광학 흐름 지도를 정확하게 찾는 것이 매우 중요하다. 이에 기반한 다양한 기법들이 제안되어 왔다.
- [0012] 최근 딥러닝 (Deep learning) 알고리즘이 등장하여 컴퓨터 비전, 음성 인식 등 다양한 분야에 널리 사용되고 있으며 종래에 방법에 비해 월등한 성능을 보이고 있다. 이에 발맞추어 딥러닝을 사용한 다양한 프레임 보간 기법이 등장하였다. 이 기법들은 딥러닝을 이용하여 고품질의 광학 흐름 지도를 예측하여 종래의 방법보다 더욱 뛰어난 보간 결과를 보여주고 있다.
- [0014] 3. 종래 기술 문제점
- [0015] 기존 방법의 문제는 네트워크 구조로 인한 문제로 4K (3840x2160) 해상도와 같이 큰 해상도에 대하여 GPU 메모리 부족으로 연산이 불가능 하거나, 매우 느린 연산 속도를 보여준다. 이와 같은 현상은 실시간 연산을 요구하는 상용 애플리케이션에 딥러닝을 이용한 프레임 보간 방법 적용을 어렵게 한다.
- [0016] 또한, 고해상도 영상은 저해상도 영상들에 비하여 일반적으로 큰 움직임을 가진다. 기존 방법들은 이러한 큰 움직임에 대하여 저품질의 광학 흐름 지도를 생성하는 경향이 있으며, 이는 보간된 영상의 품질 저하를 야기하는 문제를 가진다.

**발명의 내용**

**해결하려는 과제**

- [0018] 본 발명은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 본 발명의 목적은, 고해상도 비디오 영상에 대하여 고품질, 고속으로 프레임 보간을 수행하기 위한 방안으로, 입력 고해상도 영상을 저해상도 영상으로 변환하여 고속으로 광학 흐름 지도를 생성하고 이를 원본 고해상도로 복원하여 고해상도 영상을 고속으로 보간하는 동영상 프레임 율 고속 변환 방법 및 장치를 제공함에 있다.

**과제의 해결 수단**

- [0020] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른, 동영상 프레임 율 변환 방법은 시간적으로 연속된 고해상도의 프레임들로부터 생성한 저해상도의 프레임들들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 제1 생성단계; 저해상도의 광학 흐름 지도들의 해상도를 단계적으로 높이면서, 고해상도의 중간 프레임들을 생성하는 제2 생성단계; 생성된 고해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임을 생성하

는 제3 생성단계:를 포함한다.

[0021] 제1 생성단계는, 입력되는 시간적으로 연속된 프레임들의 해상도인 제1 해상도 보다 낮은 제3 해상도의 프레임들로 광학 흐름을 예측하여, 제3 해상도의 광학 흐름 지도들을 생성하는 단계; 제3 해상도의 광학 흐름 지도들을 이용하여, 제3 해상도의 중간 프레임들을 생성하는 단계; 제3 해상도의 중간 프레임들과 광학 흐름 지도들로, 제3 해상도 보다 높은 제2 해상도의 중간 프레임들과 광학 흐름 지도들을 복원하는 단계:를 포함하고, 제2 생성단계는, 제2 해상도의 중간 프레임들과 광학 흐름 지도들을 이용하여, 제2 해상도 보다 높은 제1 해상도의 중간 프레임들을 생성하는 제2 생성단계:를 포함할 수 있다.

[0022] 제2 생성단계는, 복원된 제2 해상도의 중간 프레임들과 광학 흐름 지도들로 광학 흐름을 예측하여, 제2 해상도의 광학 흐름 지도들을 생성하는 단계; 제2 해상도의 광학 흐름 지도들을 이용하여, 제2 해상도의 중간 프레임들을 생성하는 단계; 생성된 제2 해상도의 중간 프레임들과 광학 흐름 지도들로 광학 흐름을 예측하여, 제1 해상도의 광학 흐름 지도들을 생성하는 단계; 제1 해상도의 광학 흐름 지도들을 이용하여, 제1 해상도의 중간 프레임들을 생성하는 단계:를 포함하고, 제3 생성단계는, 생성된 제1 해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임을 생성하는 것일 수 있다.

[0023] 제2 해상도의 광학 흐름 지도 생성단계는, 복원된 제2 해상도의 중간 프레임들과 광학 흐름 지도들에 제1 해상도의 프레임들을 제2 해상도로 다운 샘플링한 프레임들을 추가로 이용하여, 광학 흐름을 예측하는 것일 수 있다.

[0024] 복원 단계는, 선형 보간 방법을 이용하여, 제3 해상도의 중간 프레임들과 광학 흐름 지도들로, 제2 해상도의 중간 프레임들과 광학 흐름 지도들을 복원하는 것일 수 있다.

[0025] 제3 해상도의 광학 흐름 지도 생성단계, 제2 해상도의 광학 흐름 지도 생성단계 및 제1 해상도의 광학 흐름 지도 생성단계는, 훈련된 인공지능 모델을 이용하여, 제3 해상도의 광학 흐름 지도, 제2 해상도의 광학 흐름 지도 및 제1 해상도의 광학 흐름 지도를 생성하는 것일 수 있다.

[0026] 제3 해상도의 광학 흐름 지도 생성단계는, 제3 해상도의 프레임들과 제3 해상도의 광학 흐름 지도들을 bak-warping 연산하여 생성하고, 제2 해상도의 광학 흐름 지도 생성단계는, 제2 해상도의 프레임들과 제2 해상도의 광학 흐름 지도들을 bak-warping 연산하여 생성하며, 제1 해상도의 광학 흐름 지도 생성단계는, 제1 해상도의 프레임들과 제1 해상도의 광학 흐름 지도들을 bak-warping 연산하여 생성하는 것일 수 있다.

[0027] 제3 생성단계는,

$$B = \frac{1}{1 + e^{-k_1 k_2}}$$

[0028] 위의 블렌딩 파라미터 B로 제1 해상도의 중간 프레임들을 블렌딩 연산하고,  
 [0029] k<sub>1</sub>은 sigmoid 함수의 기울기를 조절하며,

[0030] k<sub>2</sub>는 sigmoid 함수의 입력으로 주어지는 것일 수 있다.

[0032] 인공지능 모델의 훈련은, 원본 입력 영상과 보간 영상을 다시 원본 영상으로 back-warping한 영상과의 차이를 계산하는 함수, 생성된 광학 흐름 지도들의 전과 후의 차이를 계산하는 함수, adversarial 손실 함수를 이용하는 것일 수 있다.

[0033] 본 발명의 다른 측면에 따르면, 시간적으로 연속된 고해상도의 프레임들로부터 생성한 저해상도의 프레임들들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 예측부; 저해상도의 광학 흐름 지도들의 해상도를 단계적으로 높이면서, 고해상도의 중간 프레임들을 생성하는 향상부; 생성된 고해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임을 생성하는 생성부:를 포함하는 것을 특징으로 하는 동영상 프레임 윌 변환 장치가 제공된다.

[0034] 본 발명의 또다른 측면에 따르면, 저해상도의 프레임들들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 제1 생성단계; 저해상도의 광학 흐름 지도들을 이용하여, 고해상도의 중간 프레임들을 생성하는 제2 생성단계; 및 생성된 고해상도의 중간 프레임들을 이용하여, 최종 보간 프레임을 생성하는 제3 생성단계:를 포함하는 것을 특징으로 하는 동영상 프레임 윌 변환 방법이 제공된다.

[0035] 본 발명의 또다른 측면에 따르면, 시간적으로 연속된 고해상도의 프레임들로부터 생성한 저해상도의 프레임들들로 광학 흐름을 예측하여, 저해상도의 광학 흐름 지도들을 생성하는 제1 생성단계; 저해상도의 광학 흐름 지도들의 해상도를 단계적으로 높이면서, 고해상도의 중간 프레임들을 생성하는 제2 생성단계; 생성된 고해상도의 중간 프레임들을 블렌딩하여, 최종 보간 프레임을 생성하는 제3 생성단계:를 포함하는 것을 특징으로 하는 동영상 프레임 윌 변환 방법을 수행할 수 있는 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록매체가 제공된다.

**발명의 효과**

[0037] 이상 설명한 바와 같이, 본 발명의 실시예들에 따르면, 입력 고해상도 영상을 저해상도 영상으로 변환하여 고속으로 광학 흐름 지도를 생성하고 이를 원본 고해상도로 복원하여 고해상도 영상을 보간함으로써, 4K와 같은 고해상도 비디오 영상에 대하여 실시간성을 요구하는 시스템 환경에서도 고품질, 고속으로 프레임 보간을 수행할 수 있게 된다.

**도면의 간단한 설명**

- [0039] 도 1 : 프레임 보간 기술
- 도 2 : 본 발명의 실시예에 따른 동영상 프레임 윌 변환 장치의 블록도
- 도 3 : 피라미드 형태의 영상 표현 방법
- 도 4 : 광학 흐름 예측부, 광학 흐름 해상도 향상부 및 중간 프레임 생성부의 상세 구조

**발명을 실시하기 위한 구체적인 내용**

- [0040] 이하에서는 도면을 참조하여 본 발명을 보다 상세하게 설명한다.
- [0041] 도 2는 본 발명의 일 실시예에 따른 동영상 프레임 윌 변환 장치의 블록도이다. 본 발명의 실시예에 따른 동영상 프레임 윌 변환 장치는, 도시된 바와 같이, 광학 흐름 예측부(110), 광학 흐름 해상도 향상부(120) 및 최종 중간 프레임 생성부(130)를 포함하여 구성된다.
- [0042] 광학 흐름 예측부(110)는 입력으로 주어지는 시간적으로 연속된 프레임들의 양방향 광학 흐름을 저해상도에서 예측한다.
- [0043] 기존의 방법에서는 광학 흐름 예측을 원본 해상도에서 수행하기 때문에, 고해상도 영상을 입력으로 받는 경우 많은 양의 메모리를 요구하며, 제한된 하드웨어 환경을 가지는 시스템에서는 프레임 보간을 수행 할 수 없으며, 프레임 보간을 수행한다 하더라도, 매우 느린 동작 속도를 가지는 문제를 가졌었다. 또한 고해상도 영상의 경우 영상내의 객체들의 움직임이 매우 큰 편이며, 기존의 방법들은 이러한 경우에 blur나 ghost 열화를 발생시키는 문제가 있었다.
- [0044] 하지만, 본 발명의 실시예에서는 광학 흐름 예측을 저해상도에서 수행함으로써 상기 문제들을 효과적으로 해결한다.
- [0045] 도 3은 본 발명의 실시예에서 사용가능한 피라미드 형태의 영상 표현 방법들을 보여준다. 도 3에서 Level 1 해상도는 원본 입력 영상의 해상도와 동일한 해상도를 가지며, level 2의 영상들은 level 1의 해상도를 2분의 1크기로 downsample한 해상도를 가진다. 마지막으로 level 3 해상도는 level 1 해상도의 4분의 1크기를 가진다.
- [0046] 광학 흐름 예측부(110)는 level 3의 해상도에서 광학 흐름 지도를 생성하고, 광학 흐름 해상도 향상부(120)에서는 level 3 해상도의 광학 흐름 지도를 level 2 해상도와 level 1 해상도로 순차적으로 복원한다.
- [0047] 광학 흐름 지도의 해상도 향상을 단계적으로 수행하는 것은 광학 흐름이 급작스럽게 변화하지 않도록 하여, 최종 보간된 프레임이 자연스러운 움직임을 갖도록 하기 위함이다.
- [0048] 최종적으로 최종 중간 프레임 생성부(130)에서는, 복원된 고해상도 광학 흐름 지도를 이용하여 최종 중간 프레임을 생성하여 보간한다.
- [0049] 도 4는 광학 흐름 예측부(110), 광학 흐름 해상도 향상부(120) 및 최종 중간 프레임 생성부(130)의 상세 구조를 도식화하여 보여준다.
- [0050] 일반적으로 고품질의 광학 흐름 지도를 생성하기 위해서는 합성곱 신경망에서 충분한 크기의 receptive field가 요구된다. 하지만 receptive field가 커질수록 합성곱 신경망의 알고리즘 계산 복잡도가 증가하게 되는 문제를

가진다. 기존의 방법들에서는 원본 영상의 해상도와 동일한 크기를 가지는 광학 흐름 지도를 생성하기 때문에 4K 프레임과 같은 고해상도 영상에 대해서 느린 동작 속도를 가지는 문제를 가진다.

[0051] 이에 본 발명의 실시예에서는 원본 입력 영상의 해상도를 level 3 해상도로 줄여 광학 흐름을 예측한다. 이를 통하여 비교적 작은 크기의 receptive field를 통해서도 영상 내의 객체의 움직임을 쉽게 다룰 수 있고, 동시에 계산 연산에 필요한 메모리와 동작시간을 크게 감소시킬 수 있다.

[0052] 광학 흐름 예측부(110)는 합성곱 신경망(111)를 포함하고 있다. 합성곱 신경망(111)에서는 원본 입력 해상도의 4분의 1의 크기를 갖는 level 3 해상도의 원본 입력 영상을 이용하여 광학 흐름 지도  $F_{t1}^{3'}$ 와  $F_{t2}^{3'}$ 를 생성한다.

그리고, 광학 흐름 지도  $F_{t1}^{3'}$ 와  $F_{t2}^{3'}$ 를 이용하여, 광학 흐름 예측부(110)는 level 3 해상도의 중간 프레임  $I_{t1}^{3'}$

와  $I_{t2}^{3'}$ 를 생성한다.  $I_{t1}^{3'}$ 와  $I_{t2}^{3'}$ 는 다음의 수학적 식 (1)로 구할 수 있다.

$$I_{t1}^{3'} = b_w(I_1^3, F_{t1}^{3'}), \quad I_{t2}^{3'} = b_w(I_2^3, F_{t2}^{3'}) \quad (1)$$

[0053]

여기서  $b_w(\cdot, \cdot)$ 는 bak-warping 연산을 의미한다.

[0054]

[0055] 다음, 광학 흐름 예측부(110)는 선형 보간 방법을 이용하여, 중간 프레임  $I_{t1}^{3'}$ 과  $I_{t2}^{3'}$  그리고 광학 흐름 지도  $F_{t1}^{3'}$ 과  $F_{t2}^{3'}$ 로부터 원본 입력 해상도의 2분의 1크기를 갖는 level 2 해상도의  $I_{t1}^{2'}$ 과  $I_{t2}^{2'}$  그리고 level 2 해상도의  $F_{t1}^{2'}$ 과  $F_{t2}^{2'}$ 를 복원한다.

[0056] 복원된 영상들은 광학 흐름 해상도 향상부(120)의 입력으로 주어진다. 이 때, 2분의 1크기로 downsample 된 level 2 해상도의 원본 입력 영상을 추가적으로 사용하여 광학 흐름 지도의 해상도 향상시 품질 향상에 도움을 줄 수 있다.

[0057] 광학 흐름 해상도 향상부(120)는 2개의 합성곱 신경망을 포함하고 있다. 첫 번째 합성곱 신경망(121)에서는  $X_1 = \{I_1^2, I_2^2, F_{t1}^{2'}, F_{t2}^{2'}, I_{t1}^{2'}, I_{t2}^{2'}\}$ ,  $X_1 \in R^{2 \times \frac{H}{2} \times \frac{W}{2}}$ 를 입력으로 받아, level 2 해상도의 광학 흐름 지도  $F_{t1}^{2'}$ 과  $F_{t2}^{2'}$ 를 생성한다. 여기서 H와 W는 원본 입력 영상의 세로와 가로 크기를 의미한다.

[0058] 그리고 광학 흐름 해상도 향상부(120)는 수식 (1)과 유사하게 광학 흐름 지도  $F_{t1}^{2'}$ 과  $F_{t2}^{2'}$ 를 이용하여 level 2 해상도의 중간 프레임  $I_{t1}^{2'}$ 과  $I_{t2}^{2'}$ 를 생성한다.  $I_{t1}^{2'}$ 과  $I_{t2}^{2'}$ 는 다음의 수학적 식 (2)로 구할 수 있다.

$$I_{t1}^{2'} = b_w(I_1^2, F_{t1}^{2'}), \quad I_{t2}^{2'} = b_w(I_2^2, F_{t2}^{2'}) \quad (2)$$

[0059]

[0060] 다음 광학 흐름 해상도 향상부(120)의 두 번째 합성곱 신경망(122)에서는  $X_2 = \{F_{t1}^{2'}, F_{t2}^{2'}, I_{t1}^{2'}, I_{t2}^{2'}\}$ 를 입력으로 받아 원본 입력 영상의 해상도인 level 1 해상도를 가지는 광학 흐름 지도  $F_{t1}^{1'}$ 과  $F_{t2}^{1'}$ 를 복원한다. 즉,  $F_{t1}^{1'}$ 과  $F_{t2}^{1'}$ 는 원본 해상도인 level 1 해상도로 복원된 최종 광학 흐름 지도를 의미한다.

[0061] 그리고 광학 흐름 해상도 향상부(120)는 원본 입력 영상의 해상도인 level 1 해상도를 가지는 중간 프레임  $I_{t1}^{1'}$ 과  $I_{t2}^{1'}$ 을 다음 식 (3)을 통하여 구한다.

[0062] 
$$I_{t1}^{1'} = b_w(I_1^1, F_{t1}^{1'}) \quad , \quad I_{t2}^{1'} = b_w(I_2^1, F_{t2}^{1'}) \quad (3)$$

[0063] 마지막으로, 최종 중간 프레임 생성부(130)는 최종 중간 프레임  $I_{t1}^{1'}$  과  $I_{t2}^{1'}$ 를 블렌딩 (blending) 하여 보간할 최종 중간 프레임을 구한다. 최종 중간 프레임  $I_t^1$  을 다음 식 (4)를 통하여 구한다.

[0064] 
$$I_t^1 = B \odot I_{t1}^{1'} + (1 - B) \odot I_{t2}^{1'} \quad (4)$$

[0065] 여기서  $\odot$ 는 element-wise 곱셈 연산을 의미하며 B는 입력 영상의 blending parameter를 의미한다. 예를 들어, pixel p가 두 개의 입력 영상  $I_1$  과  $I_2$ 에 모두 존재한다면 B는 0.5로 설정되며, 만약 p가  $I_1$ 에는 존재하지만  $I_2$ 에는 존재하지 않는다면 B=1로 설정된다.

[0066] Blending parameter를 통한 중간 프레임 생성 방법은 가려짐 영역들에 의하여 발생하는 문제를 효과적으로 해결하나, 복잡한 구조의 변화나, 너무 큰 움직임을 가지는 객체들에 대해서는 저조한 성능을 보인다. 특히 이러한 경우에는 B의 값이 0.5로 설정되는 경우가 많아, 최종적으로 생성된 중간프레임이 blur한 결과를 보이는 경우가 많다.

[0067] 본 발명의 실시예에서는 이를 해결하기 위하여, sigmoid 함수를 이용하여 적응적으로 학습을 통해 blending parameter를 생성하는 방법을 이용한다. 적응적 blending parameter B는 다음을 통하여 구한다.

[0068] 
$$B = \frac{1}{1 + e^{-k_1 k_2}} \quad (5)$$

[0069] 여기서,  $k_1$ 과  $k_2$ 는 두 번째 광학 흐름 해상도 향상부(120)에서 출력된다. 첫 번째 parameter  $k_1$ 은 sigmoid 함수의 기울기를 조절하며, 두 번째 parameter  $k_2$ 는 sigmoid 함수의 입력으로 주어진다. Sigmoid 함수의 기울기 조절을 통하여 blur와 ghost 문제를 효과적으로 해결할 수 있다.

[0070] 한편, 본 발명의 실시예에서는 딥러닝 기반의 광학 흐름 예측부(110)와 광학 흐름 해상도 향상부(120)를 훈련하기 위하여 multi-scale smoothness 손실 함수, consistency 손실 함수와 adversarial 손실 함수를 새롭게 제안한다.

[0071] Consistency 손실 함수는 원본 입력 영상과 보간 영상을 다시 원본 영상으로 back-warping한 영상과의 차이를 계산하는 함수이다. 이를 통하여 광학 흐름 해상도 향상부(120)가 고 품질의 고해상도 광학 흐름 지도를 복원하고 보간 영상이 blur 열화되는 것을 방지 할 수 있다. 제안하는 consistency 손실 함수  $l_c$ 는 다음 식을 통하여 계산된다.

[0072] 
$$l_c = \| f_b(I_t, F_{1t}^{1'}) - I_1^1 \|_1 + \| f_b(I_t, F_{2t}^{1'}) - I_2^1 \|_1 \quad (6)$$

[0073] 한편, Consistency 손실 함수를 이용하여 blur 열화 문제를 방지 함에도 불구하고, 최종 보간 영상이 급작스러운 광학 흐름 변화로 인하여 over-smoothed되는 경향을 보일 수 있다.

[0074] 이를 해결하기 위하여 본 발명의 실시예에서는 mulit-scale smoothness 손실 함수를 제안한다. Multi-scale smoothness 손실함수는 광학 흐름 해상도 향상부(120)의 결과와 이전 단계의 광학 흐름 해상도 향상부(120)와 광학 흐름 예측부(110)의 결과의 차이를 계산하는 함수이다.

[0075] Multi-scale smoothness 손실함수는 regularization 역할을 수행하며, 훈련된 광학 흐름 해상도 향상부(120)가 안정적으로 고해상도의 광학 흐름 지도를 복원하도록 한다. Multi-scale smoothness 손실함수  $l_s$ 는 다음 식을 통하여 계산 된다.

$$l_s = \sum_{k=1}^2 \| f_u(2 \odot F_{t1}^{k+1'}) - F_{t1}^k \|_1 \quad (7)$$

[0076]

[0077] 여기서  $f_u$ 는 양방향 선형 보간 연산을 의미한다. 마지막으로 자연스러운 보간 영상을 생성하기 위하여 adversarial 손실 함수를 제안한다.

[0078]

Adversarial 손실 함수  $l_a$ 는 다음 식을 통하여 계산된다.

[0079]

$$l_a = \min_G \max_D E_{I_t \sim p_{data}(I_t)} [\log D(I_t)] + E_{(I_1, I_2) \sim \pi(I_1, I_2)} [\log(1 - D(G(I_1, I_2)))] \quad (8)$$

[0080]

지금까지, 고해상도 동영상 프레임 율 고속 변환 방법 및 장치에 대해 바람직한 실시예를 들어 상세히 설명하였다.

[0081]

종래 방법은 고해상도 영상을 입력시, 원본 해상도와 동일한 해상도를 가지는 광학 흐름 지도를 생성하여 프레임 보간을 수행함으로써, 많은 양의 메모리를 요구하며, 매우 느린 보간 속도를 갖는 문제점을 보였다.

[0082]

하지만, 본 발명의 실시예에서는 입력 고해상도 영상을 저해상도 영상으로 변환하여 고속으로 광학 흐름 지도를 생성하고 이를 원본 고해상도로 복원하는 형태로 동작하여, 4K 프레임과 같은 고해상도 영상을 고속으로 보간이 가능하다.

[0083]

본 발명의 실시예에 따른 고해상도 동영상 프레임 율 고속 변환 장치를 구현할 수 있는 하드웨어 구조에 대해, 이하에서 도 5를 참조하여 상세히 설명한다.

[0084]

도 5는 본 발명의 다른 실시예에 따른 고해상도 동영상 프레임 율 고속 변환 장치로 기능할 수 있는 영상 시스템의 하드웨어 구조를 나타낸 블록도이다. 본 발명의 실시예에 따른 영상 시스템은, 도 5에 도시된 바와 같이, 입력부(210), 프로세서(220), 출력부(230) 및 저장부(240)를 포함한다.

[0085]

입력부(210)는 외부 저장매체, 외부 기기, 통신망 등을 통해 영상 데이터를 입력받는 수단이고, 프로세서(220)는 입력된 영상에 대해 프레임 율 변환을 수행하기 위한 CPU들과 GPU들의 집합이다.

[0086]

동영상 프레임 율을 변환함에 있어 프로세서(220)는 전술한 실시예에서 제시한 방법을 이용한다. 저장부(240)는 프로세서(220)가 프레임 율 변환을 수행함에 있어 필요한 저장공간을 제공하는 내부 저장매체이다.

[0087]

출력부(230)는 프로세서(220)에서 프레임 율이 변환된 영상을 외부 저장매체, 외부 기기, 통신망 등으로 출력한다.

[0088]

한편, 본 실시예에 따른 장치와 방법의 기능을 수행하게 하는 컴퓨터 프로그램을 수록한 컴퓨터로 읽을 수 있는 기록매체에도 본 발명의 기술적 사상이 적용될 수 있음은 물론이다. 또한, 본 발명의 다양한 실시예에 따른 기술적 사상은 컴퓨터로 읽을 수 있는 기록매체에 기록된 컴퓨터로 읽을 수 있는 코드 형태로 구현될 수도 있다. 컴퓨터로 읽을 수 있는 기록매체는 컴퓨터에 의해 읽을 수 있고 데이터를 저장할 수 있는 어떤 데이터 저장 장치이더라도 가능하다. 예를 들어, 컴퓨터로 읽을 수 있는 기록매체는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광디스크, 하드 디스크 드라이브, 등이 될 수 있음은 물론이다. 또한, 컴퓨터로 읽을 수 있는 기록매체에 저장된 컴퓨터로 읽을 수 있는 코드 또는 프로그램은 컴퓨터간에 연결된 네트워크를 통해 전송될 수도 있다.

[0089]

또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어져서는 안될 것이다.

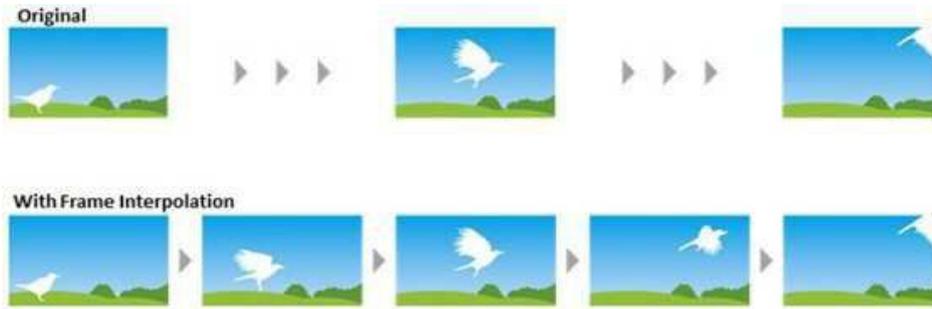
**부호의 설명**

[0091]

- 110 : 광학 흐름 예측부
- 120 : 광학 흐름 해상도 향상부
- 130 : 최종 중간 프레임 생성부

도면

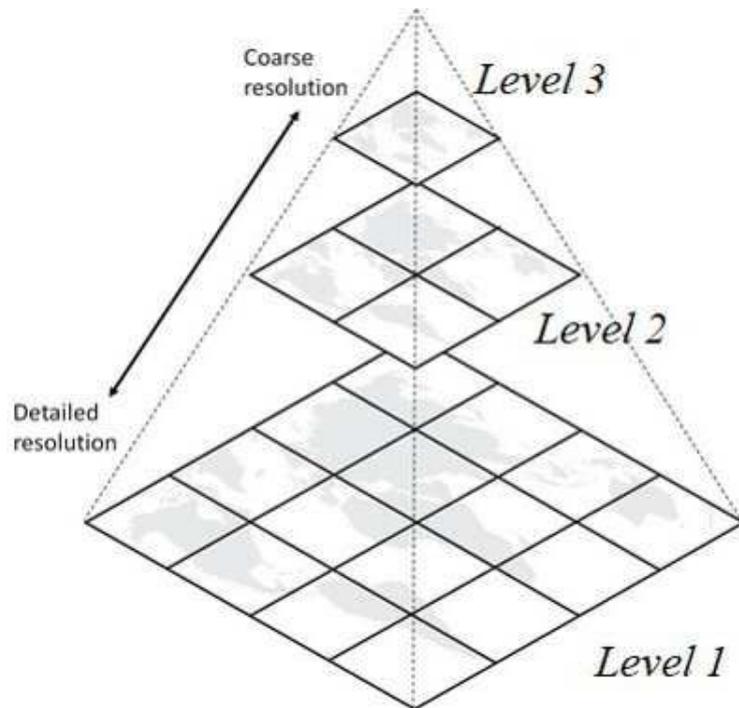
도면1



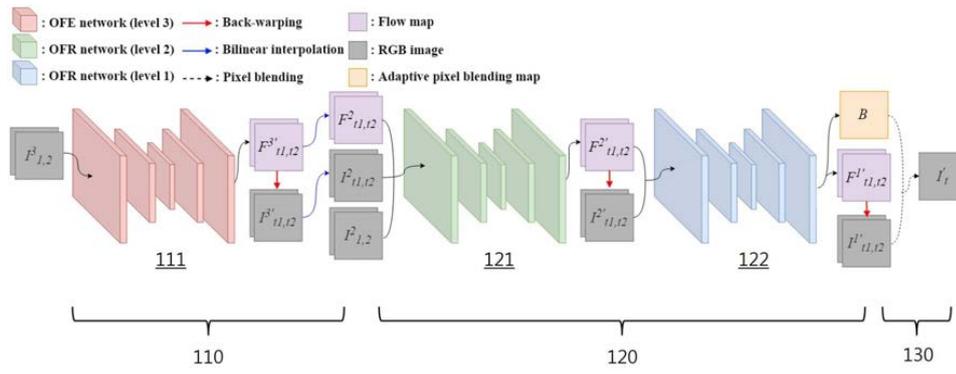
도면2



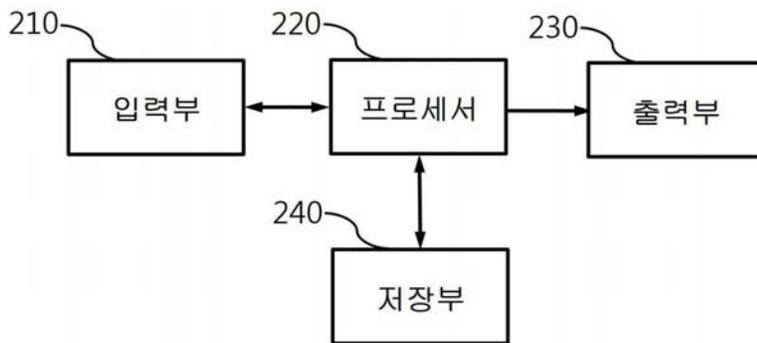
도면3



도면4



도면5





(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2021년04월20일  
(11) 등록번호 10-2242334  
(24) 등록일자 2021년04월14일

(51) 국제특허분류(Int. Cl.)  
HO4N 7/01 (2006.01) HO4N 5/262 (2006.01)  
(52) CPC특허분류  
HO4N 7/0135 (2013.01)  
HO4N 5/2628 (2013.01)  
(21) 출원번호 10-2019-0167488  
(22) 출원일자 2019년12월16일  
심사청구일자 2020년02월21일  
(56) 선행기술조사문헌  
US20070097260 A1  
US20190138889 A1  
US20190139205 A1

(73) 특허권자  
한국전자기술연구원  
경기도 성남시 분당구 새나리로 25 (야탑동)  
(72) 발명자  
정진우  
서울특별시 송파구 송파대로 111, 파크하비오105동 1205호  
안하은  
서울특별시 성북구 성북로8다길 37, 102호  
김계우  
경기도 성남시 분당구 수내로 181, 303동 901호  
(74) 대리인  
남충우

전체 청구항 수 : 총 11 항

심사관 : 박재학

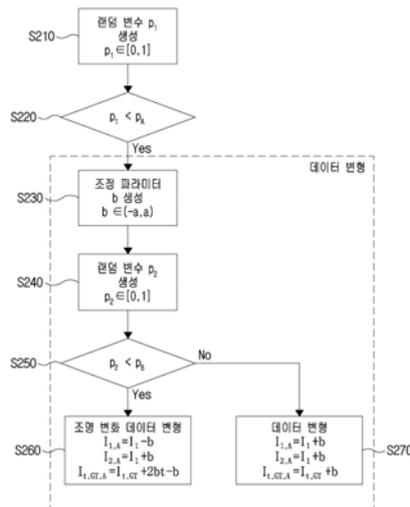
(54) 발명의 명칭 데이터 변형을 통한 고해상도 동영상 프레임 율 고속 변환 방법 및 장치

(57) 요약

고해상도 비디오 영상에 대하여 고품질, 고속으로 프레임 보간을 수행하는 딥러닝에 기반 프레임 율 고속 변환 방법 및 장치가 제공된다. 본 발명의 실시예에 따른 데이터 변형 방법은 프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들을 입력받는 단계; 입력된 입력 영상들의 영상값을 변경하여, 변형된 영상들을 생성하는 제1 생성단계; 입력된 입력 영상들의 중간 프레임의 영상값을 변경하여, 변형된 영상들의 중간 프레임을 생성하는 제2 생성단계;를 포함한다.

이에 의해, 4K 프레임과 같은 고해상도 영상을 고속으로 보간이 가능하며, Fade-in/out, 조명 변화, 줌 영상에 강인하게 학습 데이터 변형을 수행하여 이런 다양한 영상에서도 정확한 광학 흐름 지도를 생성해 낼 수 있게 하여, 고해상도 영상에 대한 고속의 강인한 프레임 보간이 가능해진다.

대표도 - 도12



(52) CPC특허분류  
*H04N 7/0127* (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711081116
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대(UHD)방송서비스활성화기술개발(R&D)
연구과제명	영상콘텐츠 초고속/초고화질 변환 기술 개발
기 여 율	1/1
과제수행기관명	(주)픽스트리
연구기간	2019.01.01 ~ 2019.12.31

---

## 명세서

### 청구범위

#### 청구항 1

프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들을 입력받는 단계;

입력된 입력 영상들의 영상값을 변경하여, 변형된 영상들을 생성하는 제1 생성단계;

입력된 입력 영상들의 중간 프레임의 영상값을 변경하여, 변형된 영상들의 중간 프레임을 생성하는 제2 생성단계;

입력 영상들과 입력 영상들의 중간 프레임의 기준 패치를 결정하는 단계;

입력 영상들 중 제1 입력 영상의 기준 패치를 축소하는 단계;

입력 영상들 중 제2 입력 영상의 기준 패치를 확대하는 단계;

축소된 제1 입력 영상의 기준 패치를 원래 크기로 조정하는 제1 조정단계;

확대된 제2 입력 영상의 기준 패치를 원래 크기로 조정하는 제2 조정단계;를 포함하는 것을 특징으로 하는 데이터 변형 방법.

#### 청구항 2

청구항 1에 있어서,

제1 생성 단계는,

입력 영상들 중 제1 입력 영상에 대해서는 조정 파라미터를 감산하여 제1 입력 영상을 변형하고,

입력 영상들 중 제2 입력 영상에 대해서는 조정 파라미터를 가산하여 제2 입력 영상을 변경하며,

제2 생성 단계는,

중간 프레임의 시간에 따라 가변하는 조정 파라미터를 가산하여 중간 프레임을 변형하는 것을 특징으로 하는 데이터 변형 방법.

#### 청구항 3

청구항 2에 있어서,

제1 생성 단계는,

입력 영상들 중 제1 입력 영상에 대해 조정 파라미터를 가산하여 제1 입력 영상을 변형하고,

입력 영상들 중 제2 입력 영상에 대해 조정 파라미터를 가산하여 제2 입력 영상을 변형하며,

제2 생성 단계는,

중간 프레임에 대해 조정 파라미터를 가산하여 중간 프레임을 변형하는 것을 특징으로 하는 데이터 변형 방법.

#### 청구항 4

청구항 2 또는 청구항 3에 있어서,

조정 파라미터는,  
랜덤한 값으로 생성되는 것을 특징으로 하는 데이터 변형 방법.

#### 청구항 5

청구항 4에 있어서,  
조정 파라미터는,  
밝기, 감마 함수, 컨트라스트, 휴(hue), 채도(saturation) 중 어느 하나를 조정하기 위한 파라미터인 것을 특징으로 하는 데이터 변형 방법.

#### 청구항 6

삭제

#### 청구항 7

청구항 1에 있어서,  
결정 단계는,  
기준 패치의 위치와 크기를 결정하는 것을 특징으로 하는 데이터 변형 방법.

#### 청구항 8

청구항 1에 있어서,  
줌 파라미터를 랜덤하게 생성하는 단계;를 더 포함하고,  
축소 단계는,  
제1 입력 영상의 기준 패치를 줌 파라미터를 이용하여 축소하며,  
확대 단계는,  
제2 입력 영상의 기준 패치를 줌 파라미터를 이용하여 확대하는 것을 특징으로 하는 데이터 변형 방법.

#### 청구항 9

청구항 1에 있어서,  
제1 조정 단계 및 제2 조정 단계는,  
확대된 제2 입력 영상의 기준 패치가 제2 입력 영상을 벗어나지 않은 것으로 판단된 경우에 수행되는 것을 특징으로 하는 데이터 변형 방법.

#### 청구항 10

프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들을 입력받는 입력부;  
입력된 입력 영상들의 영상값을 변경하여 변형된 영상들을 생성하고, 입력된 입력 영상들의 중간 프레임의 영상값을 변경하여 변형된 영상들의 중간 프레임을 생성하는 프로세서;를 포함하고,  
프로세서는,

입력 영상들과 입력 영상들의 중간 프레임의 기준 패치를 결정하고,  
 입력 영상들 중 제1 입력 영상의 기준 패치를 축소하며,  
 입력 영상들 중 제2 입력 영상의 기준 패치를 확대하고,  
 축소된 제1 입력 영상의 기준 패치를 원래 크기로 조정하며,  
 확대된 제2 입력 영상의 기준 패치를 원래 크기로 조정하는 것을 특징으로 하는 데이터 변형 시스템.

**청구항 11**

프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들을 입력받는 단계;  
 입력된 입력 영상들의 영상값을 변경하여, 변형된 영상들을 생성하는 제1 생성단계;  
 입력된 입력 영상들의 중간 프레임의 영상값을 변경하여, 변형된 영상들의 중간 프레임을 생성하는 제2 생성단계;  
 제1 생성단계에서 생성된 변형된 영상들과 제2 생성단계에서 생성된 변형된 영상들의 중간 프레임으로 인공지능 모델을 학습시키는 단계;  
 입력 영상들과 입력 영상들의 중간 프레임의 기준 패치를 결정하는 단계;  
 입력 영상들 중 제1 입력 영상의 기준 패치를 축소하는 단계;  
 입력 영상들 중 제2 입력 영상의 기준 패치를 확대하는 단계;  
 축소된 제1 입력 영상의 기준 패치를 원래 크기로 조정하는 제1 조정단계;  
 확대된 제2 입력 영상의 기준 패치를 원래 크기로 조정하는 제2 조정단계;  
 제1 조정단계에서 조정된 제1 입력 영상의 기준 패치, 제2 조정단계에서 조정된 제2 입력 영상의 기준 패치 및 결정단계에서 결정된 중간 프레임의 기준 패치로 인공지능 모델을 학습시키는 단계;를 포함하는 것을 특징으로 하는 인공지능 모델 학습 방법.

**청구항 12**

프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들의 영상값을 변경하여 변형된 영상들을 생성하고, 입력된 입력 영상들의 중간 프레임의 영상값을 변경하여 변형된 영상들의 중간 프레임을 생성하는 데이터 변형 시스템; 및  
 프로세서에 의해 생성된 변형된 영상들과 변형된 영상들의 중간 프레임으로 인공지능 모델을 학습시키는 고속 프레임 보간 시스템;을 포함하고,  
 데이터 변형 시스템은,  
 입력 영상들과 입력 영상들의 중간 프레임의 기준 패치를 결정하고,  
 입력 영상들 중 제1 입력 영상의 기준 패치를 축소하며,  
 입력 영상들 중 제2 입력 영상의 기준 패치를 확대하고,  
 축소된 제1 입력 영상의 기준 패치를 원래 크기로 조정하며,  
 확대된 제2 입력 영상의 기준 패치를 원래 크기로 조정하고,  
 고속 프레임 보간 시스템은,  
 조정된 제1 입력 영상의 기준 패치, 조정된 제2 입력 영상의 기준 패치 및 결정된 중간 프레임의 기준 패치로 인공지능 모델을 학습시키는 것을 특징으로 하는 인공지능 모델 학습 시스템.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 고해상도 동영상의 프레임 율 변환 기술에 관한 것으로, 더욱 상세하게는 딥러닝 기법에 기반한 광학 흐름 추정 지도를 저해상도에서 예측하고 이를 고해상도로 복원하여 고해상도 동영상의 프레임 율을 고속으로 보간하는 방법에 관한 것이다.

**배경 기술**

[0003] 1) 동영상 프레임 율 변환 기법 개요

[0004] 동영상은 연속된 정지 영상의 집합으로 구성된다. 비디오에서 정지 영상을 프레임이라고 부르며 단위 시간 당 프레임의 수를 동영상의 프레임 율(frame rate)이라고 한다. 예를 들어 1초에 24장의 프레임으로 구성되면 프레임 율은 24 fps(frame per second)가 된다. 프레임 율은 촬영자의 의도, 영상의 포맷, 카메라의 한계 등에 의하여 결정된다. 관찰자가 영상을 연속된 화면으로 느끼기 위해서는 어느 정도 이상의 프레임 율이 필요하고 이보다 낮을 경우 움직임이 부드럽지 않아 보인다. 이 현상은 디스플레이의 크기, 조명, 시청 거리 등에 의해 달라질 수 있다. 이를 개선하기 위해 동영상의 프레임 율을 후처리에 의해 증가시키는 것을 동영상 프레임 율 변환이라고 한다.

[0006] 2) 종래 기술

[0007] 동영상 프레임 율을 증가시키는 가장 간단한 방법은 프레임을 반복하는 것이다. 예를 들어 30fps 영상을 60 fps 영상으로 증가시킬 경우, 각 프레임마다 한 프레임을 반복하여 출력하는 것이다. 그러나 이 방법의 경우 동영상의 정보량은 동일하고 움직임에 대한 연속성은 변하지 않았으므로 관찰자가 느끼는 불편감은 동일하다. 이를 해결하기 위해 연속된 프레임들을 이용하여 가상의 프레임을 생성하는 기술이 개발되었다. 즉 t 초와 t+1 초 사이의 영상을 이용하여 t+0.5 초의 중간 영상을 새롭게 생성하며 이를 프레임 보간(frame interpolation) 기술이라고 한다.

[0008] 프레임 보간은 다양한 방법이 개발되었으며 일반적으로는 다음과 같은 두 단계 과정을 거친다. 첫 번째 단계는 움직임 또는 광학 흐름 지도를 획득하는 단계이며 두 번째 단계는 움직임 정보를 바탕으로 중간 프레임을 생성(Synthesis) 하는 단계이다. 동영상에서 물체의 움직임이 부드럽게 보이려면 중간 영상은 물체의 움직임이 두 영상 사이의 중간에 해당되어야 한다. 따라서 물체의 움직임 정보를 가지고 있는 광학 흐름 지도를 정확하게 찾는 것이 매우 중요하다. 이에 기반한 다양한 기법들이 제안되어 왔다.

[0009] 최근 딥러닝(Deep learning) 알고리즘이 등장하여 컴퓨터 비전, 음성 인식 등 다양한 분야에서 사용되고 있으며 종래에 방법에 비해 월등한 성능을 보이고 있다. 이에 발맞추어 딥러닝을 사용한 다양한 프레임 보간 기법이 등장하였다. 이 기법들은 딥러닝을 이용하여 고품질의 광학 흐름 지도를 예측하여 종래의 방법보다 더욱 뛰어난 보간 결과를 보여줌에 따라 최근 지속적으로 연구되고 있다.

[0011] 3) 종래 기술 문제점

[0012] 기존 방법의 문제는 네트워크 구조로 인한 문제로 4K(3840x2160) 해상도와 같이 큰 해상도에 대하여 GPU 메모리 부족으로 연산이 불가능 하거나, 매우 느린 연산 속도를 보여준다. 이와 같은 현상은 실시간 연산을 요구하는 상용 애플리케이션에 딥러닝을 이용한 프레임 보간 방법 적용을 어렵게한다. 또한, 고해상도 영상은 저해상도 영상들에 비하여 일반적으로 큰 움직임을 가진다. 기존 방법들은 이러한 큰 움직임에 대하여 저품질의 광학 흐름 지도를 생성하는 경향이 있으며, 이는 보간된 영상의 품질 저하를 야기하는 문제를 가진다.

[0013] 또한 기존 방법의 문제는 Fade-in/out, 조명 변화, 줌 영상을 고려하지 않아 이런 종류의 영상에 있어서는 광학 흐름 지도가 정확히 생성되지 않는 문제점이 발생한다. 따라서 이런 영상에서는 프레임 보간 성능이 급격히 저하되는 단점이 존재한다.

**발명의 내용**

**해결하려는 과제**

[0015] 본 발명은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 본 발명의 목적은, 고해상도 비디오 영상에 대하여 고품질, 고속으로 프레임 보간을 수행하는 딥러닝에 기반 프레임 윌 고속 변환 방법 및 장치를 제공함에 있다.

**과제의 해결 수단**

[0017] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른, 데이터 변형 방법은 프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들을 입력받는 단계; 입력된 입력 영상들의 영상값을 변경하여, 변형된 영상들을 생성하는 제1 생성단계; 입력된 입력 영상들의 중간 프레임의 영상값을 변경하여, 변형된 영상들의 중간 프레임을 생성하는 제2 생성단계;를 포함한다.

[0018] 제1 생성 단계는, 입력 영상들 중 제1 입력 영상에 대해서는 조정 파라미터를 감소하여 제1 입력 영상을 변형하고, 입력 영상들 중 제2 입력 영상에 대해서는 조정 파라미터를 가산하여 제2 입력 영상을 변형하며, 제2 생성 단계는, 중간 프레임의 시간에 따라 가변하는 조정 파라미터를 가산하여 중간 프레임을 변형하는 것일 수 있다.

[0019] 제1 생성 단계는, 입력 영상들 중 제1 입력 영상에 대해 조정 파라미터를 가산하여 제1 입력 영상을 변형하고, 입력 영상들 중 제2 입력 영상에 대해 조정 파라미터를 가산하여 제2 입력 영상을 변형하며, 제2 생성 단계는, 중간 프레임에 대해 조정 파라미터를 가산하여 중간 프레임을 변형하는 것일 수 있다.

[0020] 조정 파라미터는, 랜덤한 값으로 생성되는 것일 수 있다.

[0021] 조정 파라미터는, 밝기, 감마 함수, 컨트라스트, 휴(hue), 채채레이션(saturation) 중 어느 하나를 조정하기 위한 파라미터일 수 있다.

[0022] 본 발명에 따른 데이터 변형 방법은 입력 영상들과 입력 영상들의 중간 프레임의 기준 패치를 결정하는 단계; 입력 영상들 중 제1 입력 영상의 기준 패치를 축소하는 단계; 입력 영상들 중 제2 입력 영상의 기준 패치를 확대하는 단계; 축소된 제1 입력 영상의 기준 패치를 원래 크기로 조정하는 제1 조정단계; 확대된 제2 입력 영상의 기준 패치를 원래 크기로 조정하는 제2 조정단계;를 더 포함할 수 있다.

[0023] 결정 단계는, 기준 패치의 위치와 크기를 결정하는 것일 수 있다.

[0024] 본 발명에 따른 데이터 변형 방법은 줌 파라미터를 랜덤하게 생성하는 단계;를 더 포함하고, 축소 단계는, 제1 입력 영상의 기준 패치를 줌 파라미터를 이용하여 축소하며, 확대 단계는, 제2 입력 영상의 기준 패치를 줌 파라미터를 이용하여 확대하는 것일 수 있다.

[0025] 제1 조정 단계 및 제2 조정 단계는, 확대된 제2 입력 영상의 기준 패치가 제2 입력 영상을 벗어나지 않은 것으로 판단된 경우에 수행되는 것일 수 있다.

[0026] 본 발명의 다른 측면에 따르면, 프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들을 입력받는 입력부; 입력된 입력 영상들의 영상값을 변경하여 변형된 영상들을 생성하고, 입력된 입력 영상들의 중간 프레임의 영상값을 변경하여 변형된 영상들의 중간 프레임을 생성하는 프로세서;를 포함하는 것을 특징으로 하는 데이터 변형 시스템이 제공된다.

[0027] 본 발명의 또다른 측면에 따르면, 프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들을 입력받는 단계; 입력된 입력 영상들의 영상값을 변경하여, 변형된 영상들을 생성하는 제1 생성단계; 입력된 입력 영상들의 중간 프레임의 영상값을 변경하여, 변형된 영상들의 중간 프레임을 생성하는 제2 생성단계; 제1 생성단계에서 생성된 변형된 영상들과 제2 생성단계에서 생성된 변형된 영상들의 중간 프레임으로 인공지능 모델을 학습시키는 단계;를 포함하는 것을 특징으로 하는 인공지능 모델 학습 방법이 제공된다.

[0028] 본 발명의 또다른 측면에 따르면, 프레임 보간을 위한 중간 프레임을 생성하는 인공지능 모델의 학습에 이용될 연속하는 입력 영상들의 영상값을 변경하여 변형된 영상들을 생성하고, 입력된 입력 영상들의 중간 프레임의 영상값을 변경하여 변형된 영상들의 중간 프레임을 생성하는 데이터 변형 시스템; 및 프로세서에 의해 생성된 변형된 영상들과 변형된 영상들의 중간 프레임으로 인공지능 모델을 학습시키는 고속 프레임 보간 시스템;을 포함하는 것을 특징으로 하는 인공지능 모델 학습 시스템이 제공된다.

**발명의 효과**

[0030] 이상 설명한 바와 같이, 본 발명의 실시예들에 따르면, 4K 프레임과 같은 고해상도 영상을 고속으로 보간이 가

능하며, Fade-in/out, 조명 변화, 줌 영상에 강인하게 학습 데이터 변형을 수행하여 이런 다양한 영상에서도 정확한 광학 흐름 지도를 생성해 낼 수 있게 하여, 고해상도 영상에 대한 고속의 강인한 프레임 보간이 가능해진다.

**도면의 간단한 설명**

[0032]

- 도 1은 프레임 보간 기술,
- 도 2는 고속 프레임 보간 시스템의 블록도,
- 도 3은 피라미드 형태의 영상 표현 방법,
- 도 4은 저해상도 광학 흐름 예측부의 상세 블록도,
- 도 5는 광학 흐름 해상도 향상부의 상세 블록도,
- 도 6은 중간 프레임 시간 간격,
- 도 7은 중간 프레임 생성부의 상세 블록도,
- 도 8은 중간 프레임 해상도 향상부의 상세 블록도,
- 도 9는 학습 데이터 구성,
- 도 10은 데이터 변형된 학습 데이터 구성,
- 도 11은 밝기 조정 파라미터 결정 방법,
- 도 12는 학습을 위한 샘플 별 데이터 변형 과정,
- 도 13은 줌 데이터 변형 패치 생성 개념도,
- 도 14는 줌 데이터 변형 패치 생성 과정,
- 도 15는 줌 데이터 변형 패치 생성 순서도,
- 도 16은 데이터 변형 시스템의 블록도이다.

**발명을 실시하기 위한 구체적인 내용**

[0033]

이하에서는 도면을 참조하여 본 발명을 보다 상세하게 설명한다.

[0034]

본 발명의 실시예에서는, 첫 번째로 딥러닝 네트워크를 이용하여 고해상도 영상에 대해 고속 프레임 보간을 수행하는 방법을 제시하고, 두 번째로 조명 변화, 페이드 인/아웃, 줌 영상에 강인한 프레임 보간 기법으로 이는 학습시 데이터를 변형(argumentation) 하는 방법을 제시한다.

[0036]

1. 고속 프레임 보간

[0037]

본 발명의 일 실시예에 따른 고속 프레임 보간 시스템은, 도 2에 도시된 바와 같이, 광학 흐름 예측부(110), 저해상도 광학 흐름 해상도 향상부(120), 중간 프레임 생성부(130), 중간 프레임 해상도 향상부(140)를 포함하여 구성된다.

[0038]

저해상도 광학 흐름 예측부(110)에서는 입력으로 주어진 시간적으로 연속된 프레임들의 양방향 광학 흐름을 저해상도에서 예측한다. 기존의 방법에서는 광학 흐름 예측을 원본 해상도에서 수행하기 때문에, 고해상도 영상을 입력으로 받는 경우 많은 양의 메모리를 요구하며, 제한된 하드웨어 환경을 가지는 시스템에서는 프레임 보간을 수행 할 수 없으며, 프레임 보간을 수행한다 하더라도, 매우 느린 동작 속도를 가지는 문제를 가진다. 또한 고해상도 영상의 경우 영상내의 객체들의 움직임이 매우 큰 편이며, 기존의 방법들은 이러한 경우에 blur나 ghost 열화를 발생시키는 문제를 갖는다. 본 발명의 실시예에서는 광학 흐름 예측을 저해상도에서 수행함으로써 상기 문제들을 효과적으로 해결한다.

[0039]

도 3은 본 발명의 실시예에서 사용하는 피라미드 형태의 영상 표현 방법을 보여준다. Level 1 해상도는 원본 입력 영상의 해상도와 동일한 해상도를 가지며, Level 2의 영상들은 Level 1의 해상도를 2분의 1크기로 downsample한 해상도를 가진다. 마지막으로 Level 3 해상도는 Level 1 해상도의 4분의 1크기를 가진다. 즉 Level 1의 원본 입력 해상도를  $W(\text{너비}) \times H(\text{높이})$  라고 하면, Level2의 해상도는  $W/2 \times H/2$ , Level 3의 해상도

는  $W/4 \times H/4$  가 된다.

- [0040] 저해상도 광학 흐름 예측부(110)에서는 Level 3의 해상도에서 광학 흐름 지도를 생성하고, 광학 흐름 해상도 향상부(120)에서는 Level 3 해상도의 광학 흐름 지도를 Level 2 해상도로 복원한다. 중간 프레임 생성부(130)에서는 Level 2의 해상도를 갖는 광학 흐름과 Level 2 해상도의 입력 영상을 이용하여 Level 2 해상도의 중간 프레임을 생성한다. 중간 프레임 해상도 향상부(140)에서는 Level 2 해상도의 중간 프레임을 Level 1의 원해상도로 복원한다.
- [0041] 저해상도 광학 흐름 예측부(110)에서는 입력 두 프레임 사이의 광학 흐름 지도를 산출한다. 일반적으로 고품질의 광학 흐름 지도를 생성하기 위해서는 합성곱 신경망에서 충분한 크기의 receptive field가 요구된다. 하지만 receptive field가 커질수록 합성곱 신경망의 알고리즘 계산 복잡도가 증가하게 되는 문제를 가진다. 기존의 방법들에서는 원본 영상의 해상도와 동일한 크기를 가지는 광학 흐름 지도를 생성하기 때문에 4K 프레임과 같은 고해상도 영상에 대해서 느린 동작 속도를 가지는 문제를 가진다.
- [0042] 본 발명의 실시예에서는 도 4에 도시된 바와 같이, 해상도 감소부(111)가 원본 입력 영상의 해상도를 Level 3 해상도로 줄인 후에, 광학 흐름 예측부(112)가 광학 흐름을 예측한다.
- [0043] 해상도 감소부(111)에서는 Level 1 해상도의 원본 입력 영상을 Level 3 해상도로 줄인다. 줄이는 방법은 bilinear 또는 bicubic 필터 등을 사용할 수 있으며 특정 방법에 한정되지 않는다. 도 3에서  $I_1^{L1}$ ,  $I_2^{L1}$ 의 L1은 Level 1의 해상도를,  $I_1^{L3}$ ,  $I_2^{L3}$ 의 L3는 Level 3의 해상도를 각각 나타낸다.
- [0044] 광학 흐름 예측부(112)에서는 Level 3 저해상도로 영성으로 입력 두 프레임 사이의 광학 흐름 지도를 산출한다. 이를 통하여 비교적 작은 크기의 receptive field를 통해서도 영상 내의 객체의 움직임을 쉽게 다룰 수 있고, 동시에 계산 연산에 필요한 메모리와 동작시간을 크게 감소시킬 수 있다.
- [0045] 광학 흐름 예측부(112)에서는 원본 입력 해상도의 4분의 1의 크기(즉 Level 1의 크기)를 갖는 광학 흐름 지도  $F_{1 \rightarrow 2}^{L3}$ 와  $F_{2 \rightarrow 1}^{L3}$ 를 생성한다. 광학 흐름을 생성하는 방법은 다양한 딥러닝 네트워크 및 기존 방법들이 사용될 수 있으며 특정한 방법에 한정되지 않는다.
- [0046] 광학 흐름 해상도 향상부(120)는 Level 3 광학 흐름 지도를 Level 2의 광학 흐름 지도로 해상도를 개선시키는 장치로 도 5와 같이 구성된다.
- [0047] 해상도 증가부(121)은 Level 3 해상도의 광학 흐름 지도를 Level 2의 해상도의 광학 흐름 지도로 증가시킨다. 증가 방법은 bilinear 또는 bicubic 필터 등을 사용할 수 있으며 특정 방법에 한정되지 않는다.
- [0048] 광학 흐름 재조정부(123)는 입력 시간 t에 대한 광학 흐름 지도로 재조정한다. 도 6과 같이 t는  $I_1$ ,  $I_2$  사이에 생성할 중간 프레임의 시간을 의미한다.  $F_{1 \rightarrow 2}^{L3}$ 의 의미는  $I_1^{L3}$ 에서  $I_2^{L3}$ 로 진행할 때의 광학 흐름 지도를  $F_{2 \rightarrow 1}^{L3}$ 은 그 역을 나타낸다. 도 5의  $\hat{F}_{t \rightarrow 1}^{L2}$ 는 임의의 t에서 보간된 프레임인  $\hat{I}_t^{L2}$ 에서  $I_1^{L2}$ 으로 진행되는 광학 흐름 지도로 볼 수 있고  $\hat{F}_{t \rightarrow 2}^{L2}$ 도 마찬가지이다.  $\hat{F}_{t \rightarrow 1}^{L2}$ 와  $\hat{F}_{t \rightarrow 2}^{L2}$ 는 다음과 같이 산출 될 수 있다.
- $$\begin{aligned} \hat{F}_{t \rightarrow 1}^{L2} &= -(1-t)t \times F_{1 \rightarrow 2}^{L2} + t^2 \times F_{2 \rightarrow 1}^{L2} \\ \hat{F}_{t \rightarrow 2}^{L2} &= (1-t)^2 \times F_{1 \rightarrow 2}^{L2} - t(1-t) \times F_{2 \rightarrow 1}^{L2} \end{aligned}$$
- [0049] (1)
- [0050] 해상도 감소부(112)에서는 Level 1 해상도를 갖는 입력 원영상을 Level 2의 해상도로 감소시켜  $I_1^{L2}$ ,  $I_2^{L2}$ 를 생성한다. 증가 방법은 bilinear 또는 bicubic 필터 등을 사용할 수 있으며 특정 방법에 한정되지 않는다.

[0051] 와핑부 1(124)에서는 광학 흐름 지도  $\hat{F}_{t \rightarrow 1}^{L2}$  와  $\hat{F}_{t \rightarrow 2}^{L2}$  와 Level 2의 입력 영상  $I_1^{L2}$ ,  $I_2^{L2}$  를 이용하여 임시 중간 프레임  $\hat{I}_{t \rightarrow 1}^{L2}$  와  $\hat{I}_{t \rightarrow 2}^{L2}$  를 생성한다.  $\hat{I}_{t \rightarrow 1}^{L2}$  와  $\hat{I}_{t \rightarrow 2}^{L2}$  는 식 (2)와 같이 구할 수 있으며, 여기서  $b_w(\cdot, \cdot)$  은 back-warping 연산을 의미한다.

$$\begin{aligned} \hat{I}_{t \rightarrow 1}^{L2} &= b_w(I_1^{L2}, \hat{F}_{t \rightarrow 1}^{L2}) \\ \hat{I}_{t \rightarrow 2}^{L2} &= b_w(I_2^{L2}, \hat{F}_{t \rightarrow 2}^{L2}) \end{aligned} \quad (2)$$

[0053] 광학 흐름 향상부(125)에서는 와핑부1(124)의 출력  $\hat{I}_{t \rightarrow 1}^{L2}$ ,  $\hat{I}_{t \rightarrow 2}^{L2}$ , 광학 흐름 재조정부(123)의 출력  $\hat{F}_{t \rightarrow 1}^{L2}$ ,  $\hat{F}_{t \rightarrow 2}^{L2}$ , 해상도 감소부2(112)의 출력  $I_1^{L2}$ ,  $I_2^{L2}$  을 이용하여 광학 흐름을 개선한다. 즉, 위의 입력들을  $X^{L2} = \{I_1^{L2}, I_2^{L2}, \hat{F}_{t \rightarrow 1}^{L2}, \hat{F}_{t \rightarrow 2}^{L2}, \hat{I}_{t \rightarrow 1}^{L2}, \hat{I}_{t \rightarrow 2}^{L2}\}$  로 묶을 수 있고 광학 흐름 향상부(125)의 입력이 된다. 광학 흐름 향상부(125)는 합성곱 신경망으로 구성되며 개선된 광학 흐름 지도  $F_{t \rightarrow 1}^{L2}$ ,  $F_{t \rightarrow 2}^{L2}$  를 생성한다.

[0054] 중간 프레임 생성부(130)는 광학 흐름 해상도 향상부(120)의 출력인  $F_{t \rightarrow 1}^{L2}$ ,  $F_{t \rightarrow 2}^{L2}$  를 이용하여 Level 2 해상도의 중간 프레임을 생성한다. 도 7은 중간 프레임 생성부(130)의 상세 구성을 나타낸다. 도시된 바와 같이, 중간 프레임 생성부(130)는 와핑부2(131)와 중간 프레임 합성부(132)로 구성된다.

[0055] 와핑부2(131)에서는 광학 흐름 지도  $F_{t \rightarrow 1}^{L2}$ ,  $F_{t \rightarrow 2}^{L2}$  와 Level 2의 입력 영상  $I_1^{L2}$ ,  $I_2^{L2}$  를 이용하여 임시 중간 프레임  $I_{t \rightarrow 1}^{L2}$  와  $I_{t \rightarrow 2}^{L2}$  를 생성한다. 이는 식 (3)과 같이 구할 수 있다.

$$\begin{aligned} I_{t \rightarrow 1}^{L2} &= b_w(I_1^{L2}, F_{t \rightarrow 1}^{L2}) \\ I_{t \rightarrow 2}^{L2} &= b_w(I_2^{L2}, F_{t \rightarrow 2}^{L2}) \end{aligned} \quad (3)$$

[0057] 중간 프레임 합성부(132)에서는 생성된 임시 중간 프레임을 이용하여 Level 2 해상도의 최종 중간 프레임을 다음과 같이 생성한다.

$$\hat{I}_t^{L2} = B \odot I_{t \rightarrow 1}^{L2} + (1 - B) \odot I_{t \rightarrow 2}^{L2}$$

[0059] 여기서  $\odot$ 는 element-wise 곱셈 연산을 의미하며 B는 입력 영상의 blending parameter를 의미한다. 예를 들어, pixel의 p가 두 개의 입력 영상  $I_1$  과  $I_2$  에 모두 존재한다면 B는 0.5로 설정되며, 만약 p가  $I_1$  에는 존재하지 만  $I_2$  에 존재하지 않는다면 B=1 로 설정된다. B는 학습을 통해서 선택 될 수도 있고 고정된 값으로 결정될 수 있다.

[0060] 중간 프레임 해상도 향상부(140)는 Level 2 해상도의 최종 중간 프레임을  $\hat{I}_t^{L2}$  를 원영상 해상도인 Level1 해상도로 복원한다. 도 8에 도시된 바와 같이, 중간 프레임 해상도 향상부(140)는 해상도 증가부2(141과 142), 와핑부3(143), 해상도 개선부(144)로 구성된다.

[0061] 해상도 증가부2(141과 142)에서는  $\hat{I}_t^{L2}$  와  $F_{t \rightarrow 1}^{L2}$ ,  $F_{t \rightarrow 2}^{L2}$  를 Level 2 해상도에서 Level 1 해상도로 증가시켜

$\hat{I}_t^{L1}$  와  $F_{t \rightarrow 1}^{L1}$ ,  $F_{t \rightarrow 2}^{L1}$  를 생성한다. 증가 방법은 bilinear 또는 bicubic 필터 등을 사용할 수 있으며 특정 방법에 한정되지 않는다.

[0062] 해상도 개선부(144)에서는 고해상도 영상을 입력으로 넣기 위해 원영상 정보를 입력으로 사용한다. 원영상을 바로 넣을 경우  $I_1^{L1}$ ,  $\hat{I}_t^{L1}$ ,  $I_2^{L1}$  의 입력 간에 시간 차이가 존재하여 개선 효율이 저하된다. 이를 방지하기 위해  $I_1^{L1}$  과  $I_2^{L1}$  는 광학 흐름 지도를 이용하여 와핑부(143)을 거쳐 t시간의 프레임을 각각 생성하도록 한다. 이는 식 (4)와 같이 표현할 수 있다.

$$\begin{aligned}
 I_{t \rightarrow 1}^{L1} &= b_w(I_1^{L1}, F_{t \rightarrow 1}^{L1}) \\
 I_{t \rightarrow 2}^{L1} &= b_w(I_2^{L1}, F_{t \rightarrow 2}^{L1})
 \end{aligned}
 \tag{4}$$

[0064] 최종적으로 해상도 개선부(144)는  $I_{t \rightarrow 1}^{L1}$ ,  $\hat{I}_t^{L1}$ ,  $I_{t \rightarrow 2}^{L1}$  를 묶어  $X^{L1} = \{ I_{t \rightarrow 1}^{L1}, \hat{I}_t^{L1}, I_{t \rightarrow 2}^{L1} \}$  을 입력으로 하는 합성곱 신경망으로 구성된다. 해상도 개선부(144)의 출력은 Level 1의 해상도를 갖는  $\tilde{I}_t^{L1}$  이 전체 네트워크의 최종 출력이 된다.

[0065] 위에서 제시한 방법으로 중간 프레임을 생성할 경우 고해상도 영상의 빠른 움직임을 고속으로 정확하게 찾을 수 있어 정확한 프레임 보간이 가능하다.

[0067] 2. 데이터 변형

[0068] 전술하였듯이 일반적인 영상은 빈번하게 플래쉬, 조명 변화, 페이드 인/아웃이 발생하게 된다. 이런 영상에 있어서는 광학 흐름 지도가 정확하게 산출되기 어렵다. 이런 원인이 발생하는 주된 이유는 학습시 데이터에 조명 변화가 발생하는 샘플이 많이 포함되지 않아서이다. 조명 변화가 있는 샘플을 많이 포함하기 위해서는 이런 데이터를 많이 수집해야 하는데 이는 많은 비용과 노력이 필요하다. 본 발명의 실시예에서는 학습 시 데이터 변형을 통해 이를 해결하도록 한다.

[0069] 일반적으로 딥러닝을 이용한 프레임 보간 학습은, 도 9와 같이 두 개의 입력 영상  $I_1$ ,  $I_2$  과 한개의 Ground Truth(GT) 영상  $I_{t,GT}$  으로 구성된다. 딥러닝 네트워크에서 두 개의 입력 영상은 도 2에서의 입력 영상이 되고, 출력 영상은  $\tilde{I}_t$  가 된다. 학습 시  $I_{t,GT}$  와  $\tilde{I}_t$  의 차이가 최소가 되게 하도록 네트워크를 갱신하게 된다.

[0070] 도 9에서 보듯이 대부분의 영상에서는 학습 데이터 간 즉,  $I_1$ ,  $I_2$ ,  $I_{t,GT}$  간에 조명 변화가 일어나지 않는다. 조명 변화가 있는 학습 데이터 샘플이 부족하므로 조명 변화가 발생할 경우 영상이 정확히 보간되지 않게 된다.

[0071] 이 문제를 해결하기 위해 도 10과 같이 데이터 변형(argumentation)을 통해 강제로 조명 변화가 발생하는 학습 데이터를 생성시킨다. 밝기 조절을 위해 밝기 및 감마(gamma) 함수 등이 사용될 수 있다. 밝기 조절의 강도는 밝기 조정 파라미터 b에 의해 결정되어 진다. 이와 같을 경우 다음과 같이 데이터의 변형을 이루도록 한다.

$$\begin{aligned}
 I_{1,A} &= I_1 - b \\
 I_{2,A} &= I_2 + b
 \end{aligned}
 \tag{5}$$

[0073] GT 영상은 t의 위치에 따라 데이터 변형을 수행하도록 하며 밝기 변형은 선형적으로 변한다고 가정한다. 이는 도 11과 같은 그래프로 표현될 수 있다. X축은 GT가 위치한 시간을 Y축은 밝기 조정 파라미터 값을 나타낸다. t의 범위는  $I_1$  과  $I_2$  사이로 제한되므로 0과 1을 포함하지 않는 0과 1 사이의 값이다. 예를 들어 t=0.5 이면 GT의 밝기 조정 파라미터는 0이 된다. GT의 밝기 조정 파라미터는 식 (6)과 같이 일반화 할 수 있다. 밝기 조정

파라미터의 산출은 항상 선형식일 필요는 없으며 다양한 방식으로 계산이 가능하다.

$$I_{t,GT,A} = I_{t,GT} + 2 \times b \times t - b \tag{6}$$

영상의 신호 범위를 0과 1사이로 정규화 했다고 가정하면, b의 범위는 -1과 1사이로 매 배치(batch)마다 랜덤하게 선택된다. 모든 학습 샘플에서 데이터 변형을 수행하지 않고 일정 확률 P 만큼 랜덤하게 데이터 변형을 수행하게 된다. 즉 데이터 변형 횟수와 강도는 모두 랜덤하게 선택되도록 한다.

또한 데이터 변형을 입력 영상에 동일하게 하는 방법을 적용할 수 있다. 이는 특정 밝기 영상의 샘플이 부족할 때 성능이 떨어지는 점을 보완할 수 있다. 이럴 경우 데이터 변형은 식 (7)과 같이 이루어진다.

$$\begin{aligned} I_{1,A} &= I_1 + b \\ I_{2,A} &= I_2 + b \\ I_{t,GT,A} &= I_{t,GT} + b \end{aligned} \tag{7}$$

도 12는 위의 매 학습 샘플마다 이루어지는 데이터 변형 과정을 보여준다. 0과 1사이의 값을 갖는 랜덤 변수  $p_1$  을 생성시킨다(S210). 랜덤 변수  $p_1$  이  $P_A$  보다 클 경우에만(S220-Yes), 데이터 변형을 수행하도록 한다(박스 점선 부분).  $P_A$  는 0과 1사이의 값으로 데이터 변형이 적용되는 빈도를 결정한다. 만약 데이터 변형을 하도록 결정되면, 조정 파라미터 b를 생성한다(S230). 조정 파라미터 b는 데이터 변형 강도를 결정하게 된다. 조정 파라미터 범위 a는 영상의 범위 데이터 변형 종류 등에 따라 다양하게 결정될 수 있다. 그 다음 조명 변화 데이터 변환을 할지를 랜덤 변수  $p_2$  를 사용하여 결정하고(S240), 조명 변화 학습 변형을 할 경우에는(S250-Yes), 식 (5,6)을 사용하여 데이터 변형을 수행하고(S260), 그렇지 않을 경우에는(S250-No), 식 (7)을 사용하여 데이터 변형을 수행한다(S270).

위와 같은 방법은 감마 함수, 컨트라스트, 휴(hue), 채도(saturation) 등 다양한 데이터 변형 함수에 대하여 적용이 가능하다.

다른 실시 예로 줌(zoom) 영상을 효과적으로 대응하기 위한 데이터 변형 방법을 제시한다. 줌 영상은 시간이 진행함에 따라 물체의 크기가 변하게 된다. 이런 영상에 대하여 효율적으로 보간하기 위하여, 도 13와 같이 줌을 고려한 데이터 변형 패치를 생성하도록 한다.

일반적으로 학습을 진행 할 시 영상의 일부분을 패치로 사용하도록 한다. 예를 들어 영상의 크기를  $W \times H$  로 하고 패치의 크기를  $w_0 \times h_0$  라고 하자. 입력 영상과 GT 영상의 패치의 위치는 동일하다. 도 13에서와 같이 한 개의 입력은 크기를 감소시키고 다른 한 개의 입력은 크기를 증가시키면서 줌 영상에 대응하는 패치를 생성할 수 있다.

도 14와 도 15는 구체적인 줌 데이터 생성 패치 과정을 보여준다. 위의 데이터 변형 과정과 마찬가지로 0과 1사이의 값을 갖는 랜덤 변수  $p_1$  을 생성시킨다(S310). 랜덤 변수  $p_1$  가  $P_A$  보다 클 경우에만(S320-Yes) 줌 데이터 변형을 수행하도록 한다.

줌 데이터 변형을 수행하기로 결정되었다면 줌 파라미터 s를 랜덤하게 설정한다(S330). 줌 파라미터 s는 줌 배율의 강도를 의미하며 1이면 원래의 패치 크기와 같게 된다. 줌 파라미터가 너무 클 경우 패치 크기가 영상 크기를 벗어날 수 있는 경우가 많이 발생하므로 1과 1.2 사이의 값을 권장한다.

다음으로는 도 14과 같이 기준 패치의 위치  $(x_{0L}, y_{0L}, x_{0R}, y_{0R})$  를 결정한다(S340). 패치의 크기는  $w_0 \times h_0$  이다. 기준 패치의 위치, 패치의 크기 및 줌 파라미터를 이용하여 입력 영상  $I_1, I_2$  에서 사용될 위치를 식

(8~11)과 같이 결정한다. 식(8)과 식(10)에서  $\lfloor X \rfloor_{Even}$  는  $X$  를 넘지 않는 가장 큰 짝수를 의미한다.  $I_1$  영상에서 사용될 패치의 크기는 식 (8)과 같이 산출된다. 재계산된 패치 크기인  $w_1 \times h_1$  를 기반으로  $I_1$  의 패치 영역을 식 (9)와 같이 계산한다. 이는 도 14의  $I_1$  영상의 빨간 박스 영역을 의미한다.  $I_2$  영상에서 사용될 패치의 크기는 식 (10)과 같이 산출된다. 재계산된 패치 크기인  $w_2 \times h_2$  를 기반으로  $I_2$  의 패치 영역을 식 (11)와 같이 계산한다. 이는 도 14의  $I_2$  영상의 빨간 박스 영역을 의미한다.  $I_1$  과  $I_2$  의 위치는 랜덤하게 바뀔 수 있다.

$$\begin{aligned} w_1 &= \lfloor w_0 \times s \rfloor_{Even} \\ h_1 &= \lfloor h_0 \times s \rfloor_{Even} \end{aligned} \tag{8}$$

$$\begin{aligned} x_{1L} &= x_{0L} - \frac{w_1 - w_0}{2} \\ x_{1R} &= x_{0R} + \frac{w_1 - w_0}{2} \\ y_{1L} &= y_{0L} - \frac{h_1 - h_0}{2} \\ y_{1R} &= y_{0R} + \frac{h_1 - h_0}{2} \end{aligned} \tag{9}$$

$$\begin{aligned} w_2 &= \left\lfloor \frac{w_0}{s} \right\rfloor_{Even} \\ h_2 &= \left\lfloor \frac{h_0}{s} \right\rfloor_{Even} \end{aligned} \tag{10}$$

$$\begin{aligned} x_{2L} &= x_{0L} + \frac{w_0 - w_2}{2} \\ x_{2R} &= x_{0R} - \frac{w_0 - w_2}{2} \\ y_{2L} &= y_{0L} + \frac{h_0 - h_2}{2} \\ y_{2R} &= y_{0R} - \frac{h_0 - h_2}{2} \end{aligned} \tag{11}$$

만약 식(9)에서 계산된 영역이 영상의 범위를 벗어난다면 줌 데이터 변형 과정을 수행하지 않는다. 만약 경계 조건을 만족한다면(S360-Yes), 식(9)에서 계산된 좌표를 이용하여  $w_1 \times h_1$  의 크기를 갖는  $I_{1,patch}$  를 생성한다(S370).  $I_{1,patch}$  는  $w_0 \times h_0$  로 해상도를 조정하여 줌 데이터 변형된 패치인  $I_{1,patch}^Z$  를 생성한다. 마찬가지로 식 (10)에서 계산된 좌표를 이용하여  $w_2 \times h_2$  의 크기를 갖는  $I_{2,patch}$  를 생성하고  $w_0 \times h_0$  (또 해상도를 조정하여 줌 데이터 변형된 패치인  $I_{2,patch}^Z$  ) 를 생성한다(S340). 최종적으로  $w_0 \times h_0$  를 갖는  $I_{1,patch}^Z$  ,  $I_{1,GT,patch}$  ,  $I_{2,patch}^Z$  를 이용하여 학습을 수행하도록 한다.

지금까지 설명한 데이터 변형을 수행할 수 있는 시스템에 대해, 이하에서 16을 참조하여 상세히 설명한다. 도 16은 본 발명의 다른 실시예에 따른 데이터 변형 시스템의 블록도이다.

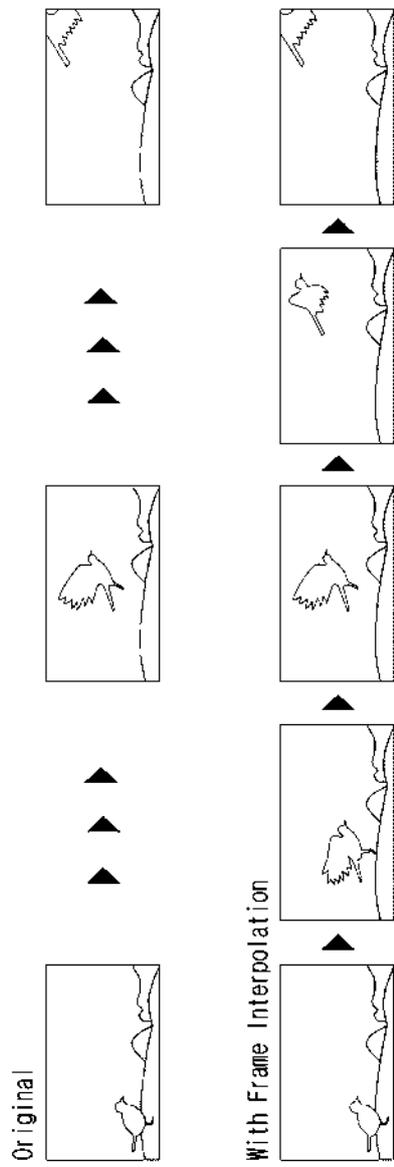
- [0091] 본 발명의 실시예에 따른 데이터 변형 시스템은, 도 16에 도시된 바와 같이, 입력부(210), 프로세서(220), 출력부(230) 및 저장부(240)를 포함하는 컴퓨팅 시스템으로 구현 가능하다.
- [0092] 입력부(210)는 외부 저장매체, 외부 기기, 통신망 등을 통해 영상들을 입력받는 수단이고, 프로세서(220)는 입력된 영상 데이터들을 변형하여 학습 영상들을 증가시킨다.
- [0093] 영상 데이터를 변형함에 있어, 프로세서(220)는 전술한 실시예에서 제시한 방법을 이용한다. 저장부(240)는 프로세서(220)가 영상 데이터들을 변형함에 있어 필요한 저장공간을 제공하는 내부 저장매체이다.
- [0094] 출력부(230)는 프로세서(220)에서 추가된 학습 영상들을 외부 저장매체, 외부 기기, 통신망 등으로 출력한다.
- [0096] 3. 변형예
- [0097] 지금까지, 딥러닝 네트워크를 이용하여 고해상도 영상에 대해 고속 프레임 보간을 수행하는 방법과 조명 변화, 페이드 인/아웃, 줌 영상에 강인한 프레임 보간 기법으로 이는 학습시 데이터를 변형하는 방법에 대해, 바람직한 실시예들을 들어 상세히 설명하였다.
- [0098] 고해상도 영상을 입력시, 원본 해상도와 동일한 해상도를 가지는 광학 흐름 지도를 생성하여 프레임 보간을 수행하여 많은 양의 메모리를 요구하며, 매우 느린 보간 속도를 갖는 종래 방법과 달리, 본 발명의 실시예에서는 입력 고해상도 영상을 저해상도 영상으로 변환하여 고속으로 광학 흐름 지도를 생성하고 이를 원본 고해상도로 복원하여, 4K 프레임과 같은 고해상도 영상을 고속으로 보간이 가능하게 하였다.
- [0099] 또한, 본 발명의 실시예에서는, Fade-in/out, 조명 변화, 줌 영상에 강인하게 학습 데이터 변형을 수행하여 이런 다양한 영상에서도 정확한 광학 흐름 지도를 생성해 낼 수 있게 하였다.
- [0100] 한편, 본 실시예에 따른 장치와 방법의 기능을 수행하게 하는 컴퓨터 프로그램을 수록한 컴퓨터로 읽을 수 있는 기록매체에도 본 발명의 기술적 사상이 적용될 수 있음은 물론이다. 또한, 본 발명의 다양한 실시예에 따른 기술적 사상은 컴퓨터로 읽을 수 있는 기록매체에 기록된 컴퓨터로 읽을 수 있는 코드 형태로 구현될 수도 있다. 컴퓨터로 읽을 수 있는 기록매체는 컴퓨터에 의해 읽을 수 있고 데이터를 저장할 수 있는 어떤 데이터 저장 장치이더라도 가능하다. 예를 들어, 컴퓨터로 읽을 수 있는 기록매체는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광디스크, 하드 디스크 드라이브, 등이 될 수 있음은 물론이다. 또한, 컴퓨터로 읽을 수 있는 기록매체에 저장된 컴퓨터로 읽을 수 있는 코드 또는 프로그램은 컴퓨터간에 연결된 네트워크를 통해 전송될 수도 있다.
- [0101] 또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어져서는 안될 것이다.

**부호의 설명**

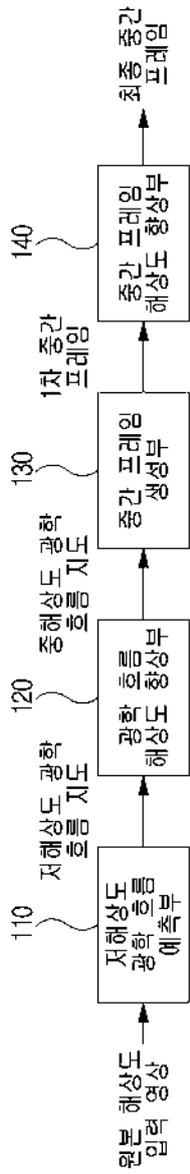
- [0103] 110 : 광학 흐름 예측부
- 120 : 저해상도 광학 흐름 해상도 향상부
- 130 : 중간 프레임 생성부
- 140 : 중간 프레임 해상도 향상부

도면

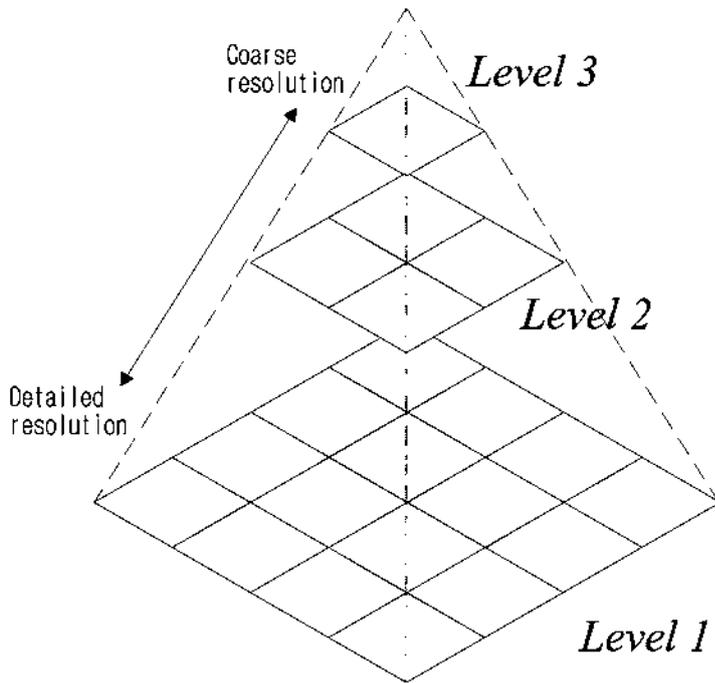
도면1



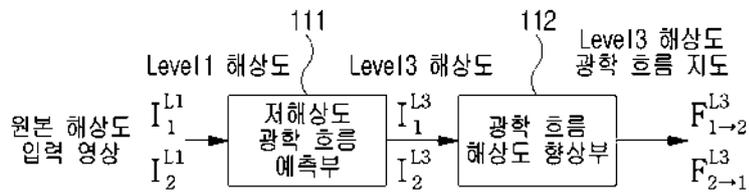
도면2



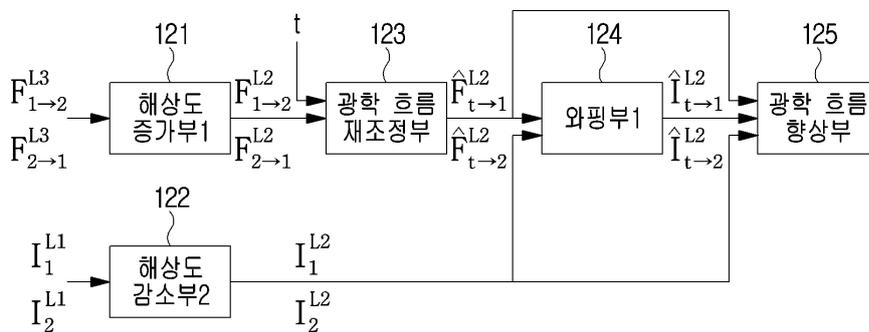
도면3



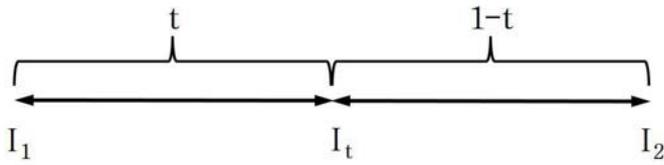
도면4



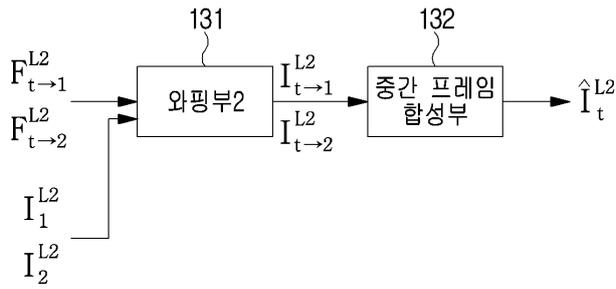
도면5



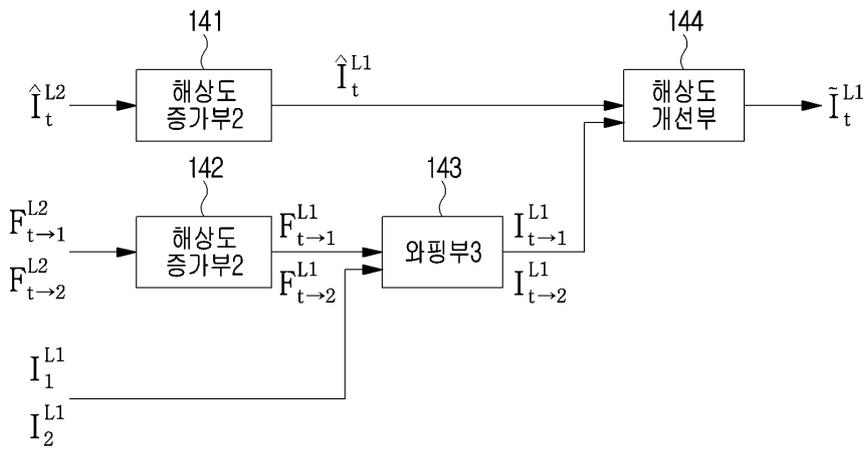
도면6



도면7



도면8



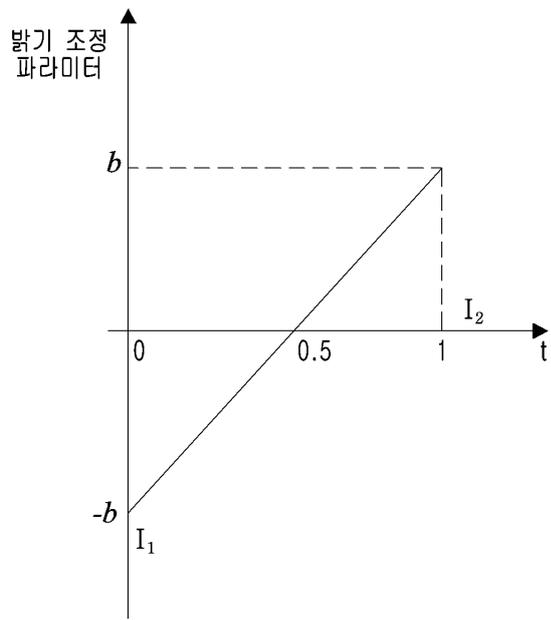
도면9



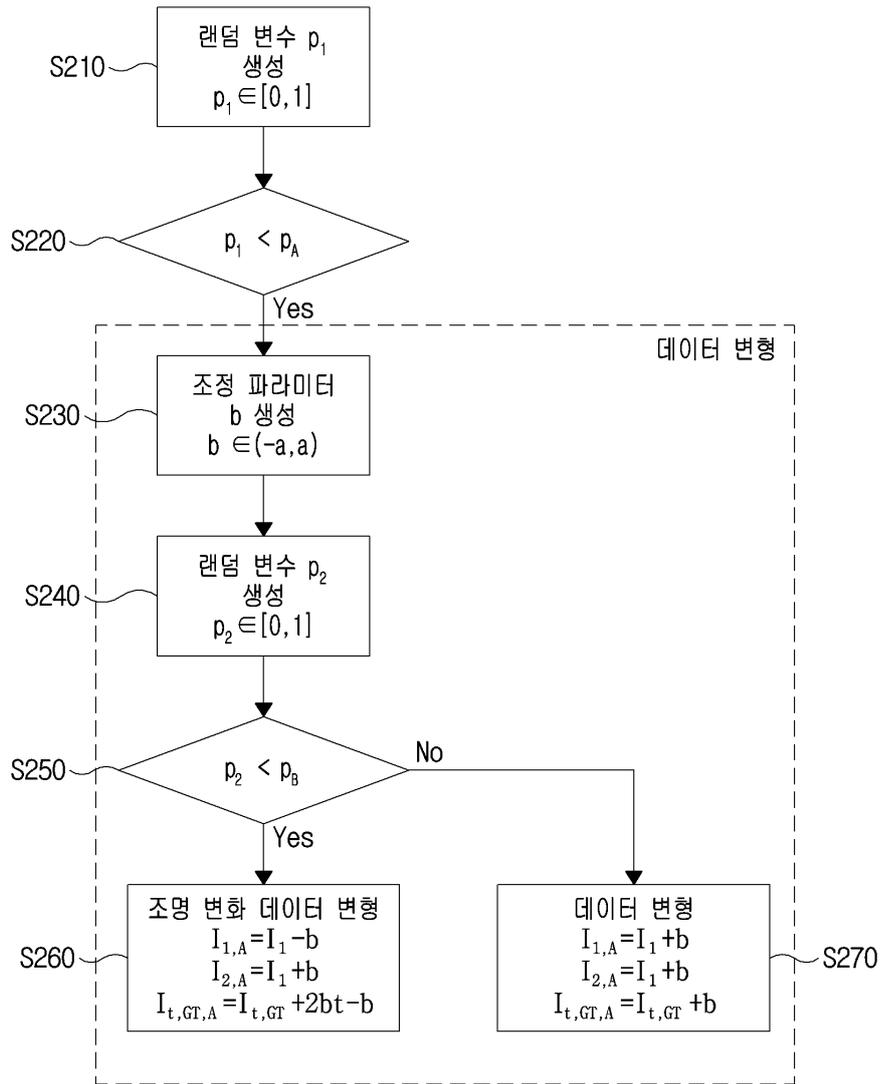
도면10



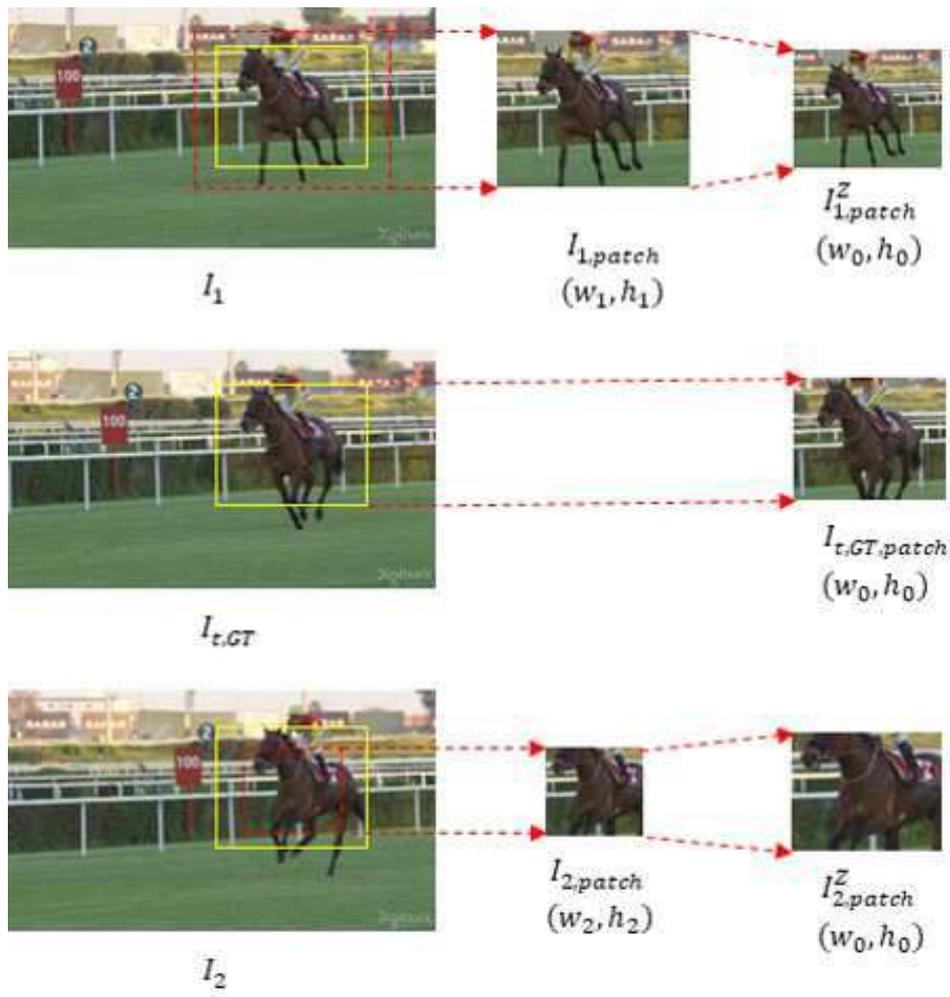
도면11



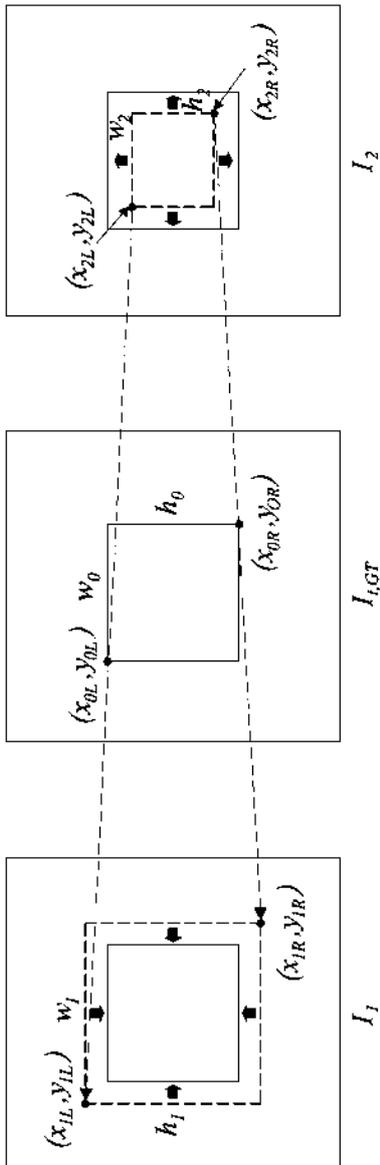
도면12



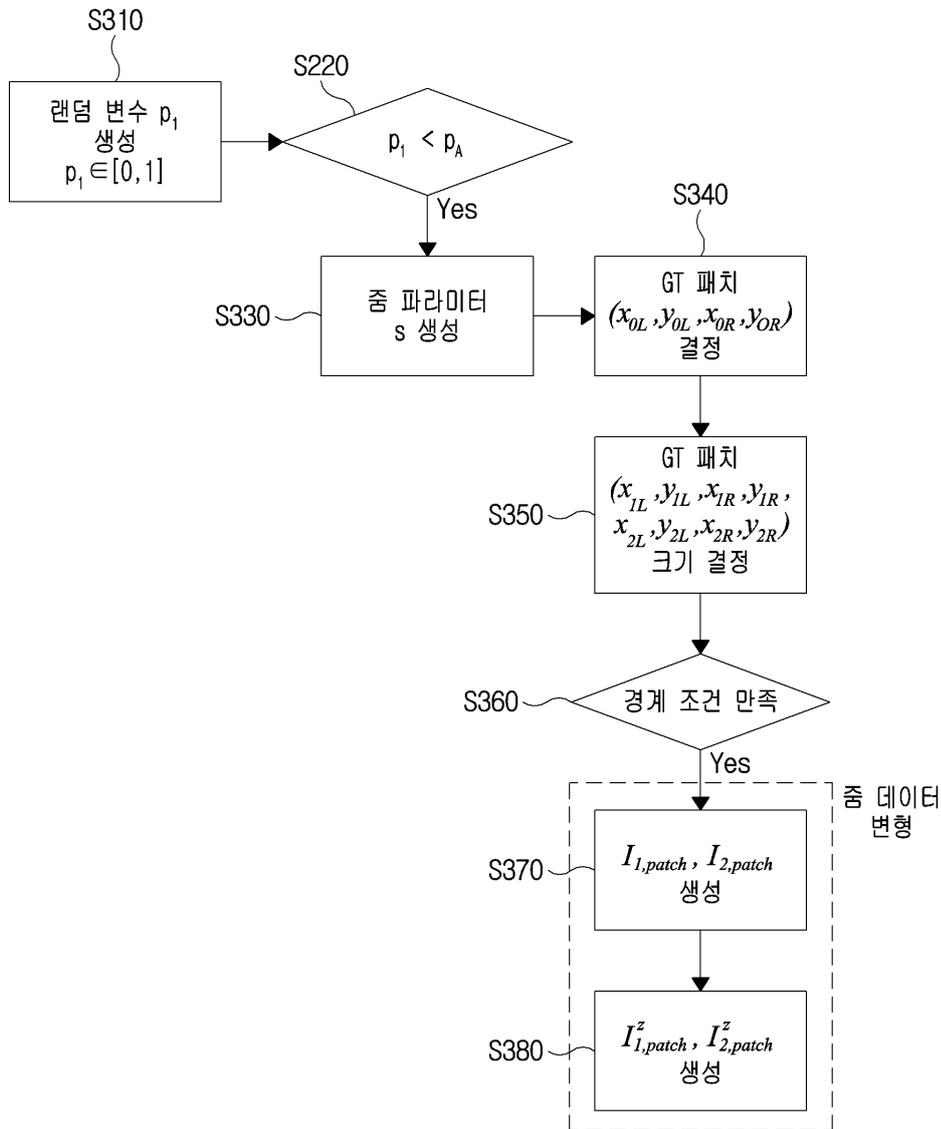
도면13



도면14



도면15



도면16

