



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년05월22일
(11) 등록번호 10-2811986
(24) 등록일자 2025년05월20일

(51) 국제특허분류(Int. Cl.)
G06T 5/00 (2024.01) G06T 3/60 (2024.01)
G06T 7/11 (2017.01) G06V 10/46 (2022.01)
G06V 10/762 (2022.01) G06V 20/40 (2022.01)
(52) CPC특허분류
G06T 5/00 (2024.01)
G06T 3/608 (2024.01)
(21) 출원번호 10-2023-0166915
(22) 출원일자 2023년11월27일
심사청구일자 2023년11월28일
(56) 선행기술조사문헌
CN116469153 A*
KR1020230103790 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박형민
서울특별시 강남구 삼성로 212, 3동 1002호 (대치동, 은마아파트)
박영후
서울특별시 영등포구 영중로 119, 14층 B1418호 (영등포동8가, 리마크빌 영등포)
박래홍
서울특별시 영등포구 당산로4길 12, 112동 1603호 (문래동3가, 문래자이)
(74) 대리인
이준영

전체 청구항 수 : 총 13 항

심사관 : 김광식

(54) 발명의 명칭 영상 처리 장치 및 이의 동작방법

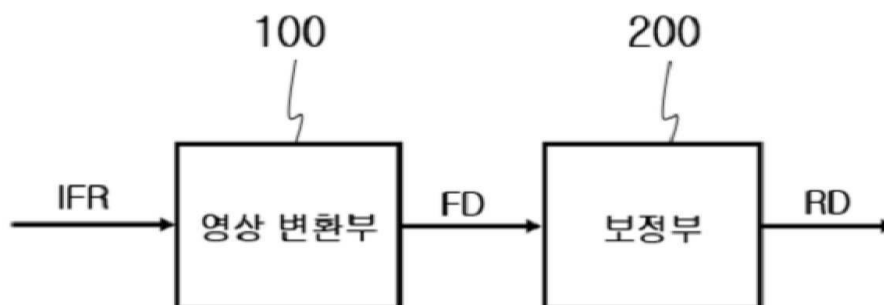
(57) 요약

본 발명의 실시예에 따른 영상 처리 장치는 영상 변환부 및 보정부를 포함할 수 있다. 영상 변환부는 복수의 입력영상들의 각각을 일정한 크기의 묶음으로 그룹화한 후 각각을 일정한 크기에 해당하는 패치 사이즈로 구분한 패치영상에 기초하여 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 제공할 수 있다. 보정부는 특징 데이터를 보정하여 보정 데이터를 제공할 수 있다.

본 발명에 따른 영상 처리 장치는 복수의 입력영상들의 각각을 일정한 크기의 묶음으로 그룹화한 후 각각을 일정한 크기에 해당하는 패치 사이즈로 구분한 패치영상에 기초하여 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 도출함으로써 발화자가 전달하고자 하는 의미를 보다 빠르고 정확하게 파악할 수 있다.

대표도 - 도1

10



(52) CPC특허분류

G06T 7/11 (2017.01)
G06V 10/46 (2023.08)
G06V 10/762 (2023.08)
G06V 20/41 (2022.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711196187
과제번호	00237117
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	차세대유망Seed기술실용화패스트트랙
연구과제명	실세계 강인음성인식을 위한 실시간 멀티채널-멀티모달 시청각음성인식 시스템 개발
기 여 율	1/1
과제수행기관명	서강대학교
연구기간	2023.04.01 ~ 2027.03.31

명세서

청구범위

청구항 1

복수의 입력영상들의 각각을 일정한 크기에 해당하는 패치 사이즈로 구분한 패치영상에 기초하여 상기 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 제공하는 영상 변환부; 및

상기 특징 데이터를 보정하여 보정 데이터를 제공하는 보정부를 포함하고,

상기 복수의 입력영상들의 각각은, 상기 패치 사이즈보다 작은 복수의 윈도우들로 구분되는 것을 특징으로 하는 영상 처리 장치.

청구항 2

제1항에 있어서,

상기 영상 변환부는,

상기 복수의 입력영상들을 그룹 사이즈의 묶음으로 그룹화한 후 각각을 상기 패치 사이즈로 구분하여 상기 패치 영상을 제공하는 영상 제공부; 및

상기 패치영상에 기초하여 상기 입술모양에 상응하는 특징 데이터를 제공하는 데이터 제공부를 포함하는 것을 특징으로 하는 영상 처리 장치.

청구항 3

제2항에 있어서,

상기 복수의 입력영상들을 상기 그룹 사이즈의 묶음으로 그룹화하는 것을 특징으로 하는 영상 처리 장치.

청구항 4

제3항에 있어서,

상기 그룹영상들의 각각은 일정한 개수에 해당하는 오버랩 개수만큼 오버랩되는 것을 특징으로 하는 영상 처리 장치.

청구항 5

제4항에 있어서,

상기 오버랩 개수는 상기 그룹 사이즈의 1/3 이상인 것을 특징으로 하는 영상 처리 장치.

청구항 6

제5항에 있어서,

상기 오버랩 개수는 상기 복수의 입력영상들의 수가 증가함에 따라 증가하는 것을 특징으로 하는 영상 처리 장치.

청구항 7

삭제

청구항 8

제6항에 있어서,

상기 영상 처리 장치는,

상기 복수의 입력영상들 각각에서 입술이 차지하는 입술영역을 탐지하는 입술영역 탐지부를 더 포함하는 것을 특징으로 하는 영상 처리 장치.

청구항 9

제8항에 있어서,

상기 영상 처리 장치는,

상기 복수의 입력영상들의 각각에 대한 상기 입술영역에 상응하는 상기 윈도우들의 개수를 산출하는 윈도우 산출부를 더 포함하는 것을 특징으로 하는 영상 처리 장치.

청구항 10

제9항에 있어서,

상기 영상 처리 장치는,

상기 입술영역에 상응하는 상기 윈도우들의 개수에 기초하여 상기 패치 사이즈를 결정하는 결정부를 더 포함하는 것을 특징으로 하는 영상 처리 장치.

청구항 11

제10항에 있어서,

상기 패치 사이즈는 상기 입술영역에 상응하는 상기 윈도우들의 개수들의 평균값 및 상기 평균값의 절반을 합한 합산 값으로 결정되는 것을 특징으로 하는 영상 처리 장치.

청구항 12

제11항에 있어서,

상기 데이터 제공부는,

상기 패치영상을 상기 윈도우들 각각에 해당하는 단위 윈도우만큼 쉬프트하면서 생성되는 쉬프트 영상을 제공하는 쉬프트부; 및

상기 쉬프트 영상에 기초하여 상기 특징 데이터를 제공하는 데이터 출력부를 더 포함하는 것을 특징으로 하는 영상 처리 장치.

청구항 13

제12항에 있어서,

상기 데이터 제공부의 입력들은 복수의 패치영상들로 구분되고,

상기 복수의 패치영상들의 각각은 상기 복수의 입력영상들 각각의 상기 패치영상에 상응하는 것을 특징으로 하는 영상 처리 장치.

청구항 14

복수의 입력영상들의 각각을 일정한 크기의 묶음으로 그룹화한 후 각각을 일정한 크기에 해당하는 패치 사이즈로 나누어진 패치영상에 기초하여 상기 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 제공하는 영상 변환부;

상기 특징 데이터에 상응하는 의미에 해당하는 의미정보를 제공하는 정보부를 포함하고,

상기 복수의 입력영상들의 각각은, 상기 패치 사이즈보다 작은 복수의 윈도우들로 구분되는 것을 특징으로 하는 영상 처리 시스템.

청구항 15

삭제

청구항 16

삭제

발명의 설명

기술 분야

[0001] 본 발명은 영상 처리 장치 및 이의 동작방법에 관한 것이다.

배경 기술

[0002] 발화자가 전달하는 소리의 의미를 정확하게 인식하기 위하여 음성 데이터뿐만 아니라, 영상 데이터가 사용될 수 있다. 최근, 이와 관련하여 영상 이미지를 이용하여 발화자가 전달하고자 하는 의미를 파악하기 위한 다양한 연구들이 진행되고 있다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) (한국등록특허) 제10-2602319호 (등록일자, 2023.11.16)

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 복수의 입력영상들을 일정한 크기에 해당하는 그룹 사이즈의 묶음으로 그룹화한 후 각각을 일정한 크기에 해당하는 패치 사이즈로 구분한 패치영상에 기초하여 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 도출함으로써 발화자가 전달하고자 하는 의미를 보다 빠르고 정확하게 파악할 수 있는 영상 처리 장치를 제공하는 것이다.

과제의 해결 수단

[0005] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 영상 처리 장치는 영상 변환부 및 보정부부를 포함할 수 있다. 영상 변환부는 복수의 입력영상들의 각각을 일정한 크기에 해당하는 패치 사이즈로 구분한 패치영상에 기초하여 상기 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 제공할 수 있다. 보정부는 상기 특징 데이터를 보정하여 보정 데이터를 제공할 수 있다.

[0006] 일 실시예에 있어서, 상기 영상 변환부는 영상 제공부 및 데이터 제공부를 포함할 수 있다. 영상 제공부는 상기 복수의 입력영상들을 일정한 크기에 해당하는 그룹 사이즈의 묶음으로 그룹화한 후 각각을 상기 패치 사이즈로 구분하여 상기 패치영상을 제공할 수 있다. 데이터 제공부는 상기 패치영상에 기초하여 상기 입술모양에 상응하는 특징 데이터를 제공할 수 있다.

[0007] 일 실시예에 있어서, 상기 입력영상들의 묶음 각각은 일정한 개수에 상응하는 오버랩 개수만큼 오버랩(Overlap)될 수 있다.

[0008] 일 실시예에 있어서, 상기 오버랩 개수는 상기 그룹 사이즈의 1/3 이상일 수 있다.

[0009] 일 실시예에 있어서, 상기 오버랩 개수는 상기 복수의 입력영상들의 수가 증가함에 따라 증가할 수 있다.

[0010] 일 실시예에 있어서, 상기 복수의 입력영상들의 각각은 상기 패치 사이즈보다 작은 복수의 윈도우들로 구분될 수 있다.

[0011] 일 실시예에 있어서, 상기 영상 처리 장치는 입술영역 탐지부를 더 포함할 수 있다. 입술영역 탐지부는 상기 복수의 입력영상들 각각에서 입술이 차지하는 입술영역을 탐지할 수 있다.

[0012] 일 실시예에 있어서, 상기 영상 처리 장치는 윈도우 산출부를 더 포함할 수 있다. 윈도우 산출부는 상기 복수의 입력영상들의 각각에 대한 상기 입술영역에 상응하는 상기 윈도우들의 개수를 산출할 수 있다.

- [0013] 일 실시예에 있어서, 상기 영상 처리 장치는 결정부를 더 포함할 수 있다. 결정부는 상기 입술영역에 상응하는 상기 윈도우들의 개수에 기초하여 상기 패치 사이즈를 결정할 수 있다.
- [0014] 일 실시예에 있어서, 상기 패치 사이즈는 상기 입술영역에 상응하는 상기 윈도우들의 개수들의 평균값 및 상기 평균값의 절반을 합한 합산 값으로 결정될 수 있다.
- [0015] 일 실시예에 있어서, 상기 데이터 제공부는 쉬프트부 및 데이터 출력부를 더 포함할 수 있다. 쉬프트부는 상기 패치영상을 상기 윈도우들 각각에 해당하는 단위 윈도우만큼 쉬프트하면서 생성되는 쉬프트 영상을 제공할 수 있다. 데이터 출력부는 상기 쉬프트 영상에 기초하여 상기 특징 데이터를 제공할 수 있다.
- [0016] 일 실시예에 있어서, 상기 데이터 제공부의 입력들은 복수의 입력영상들 각각의 상기 패치영상에 상응할 수 있다.
- [0017] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 영상 처리 시스템은 영상 변환부 및 정보부를 포함할 수 있다. 영상 변환부는 복수의 입력영상들을 일정한 크기에 해당하는 그룹 사이즈의 묶음으로 그룹화한 후 각각을 일정한 크기에 해당하는 패치 사이즈로 나누어진 패치영상에 기초하여 상기 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 제공할 수 있다. 정보부는 상기 특징 데이터에 상응하는 의미에 해당하는 의미정보를 제공할 수 있다.
- [0018] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 영상 처리 장치의 동작방법에서는, 영상 변환부는 복수의 입력영상들을 일정한 크기에 해당하는 그룹 사이즈의 묶음으로 그룹화한 후 각각을 일정한 크기에 해당하는 패치 사이즈로 나누어진 패치영상에 기초하여 상기 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 제공할 수 있다. 보정부는 상기 특징 데이터를 보정하여 보정 데이터를 제공할 수 있다.
- [0019] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 영상 처리 시스템의 동작방법에서는, 영상 변환부가 복수의 입력영상들을 일정한 크기에 해당하는 그룹 사이즈의 묶음으로 그룹화한 후 각각을 일정한 크기에 해당하는 패치 사이즈로 나누어진 패치영상에 기초하여 상기 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 제공할 수 있다. 정보부가 상기 특징 데이터에 상응하는 의미에 해당하는 의미정보를 제공할 수 있다.
- [0020] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

발명의 효과

- [0021] 이상과 같은 본 발명에 따르면 다음과 같은 효과가 있다.
- [0022] 본 발명에 따른 영상 처리 장치는 복수의 입력영상들을 일정한 크기에 해당하는 그룹 사이즈의 묶음으로 그룹화한 후 각각을 일정한 크기에 해당하는 패치 사이즈로 구분한 패치영상에 기초하여 복수의 입력영상들에 포함되는 입술모양에 상응하는 특징 데이터를 도출함으로써 발화자가 전달하고자 하는 의미를 보다 빠르고 정확하게 파악할 수 있다.
- [0023] 이 밖에도, 본 발명의 실시 예들을 통해 본 발명의 또 다른 특징 및 이점들이 새롭게 파악될 수도 있을 것이다.

도면의 간단한 설명

- [0024] 도 1은 본 발명의 실시예들에 따른 영상 처리 장치를 나타내는 도면이다.
- 도 2는 도 1의 영상 처리 장치에 포함되는 영상 변환부를 나타내는 도면이다.
- 도 3은 도 1의 영상 처리 장치에서 사용되는 입력영상의 일 예를 나타내는 도면이다.
- 도 4는 도 1의 영상 처리 장치에 포함되는 입술영역 탐지부를 나타내는 도면이다.
- 도 5는 도 1의 영상 처리 장치에 포함되는 윈도우 산출부를 나타내는 도면이다.
- 도 6은 도 1의 영상 처리 장치에 포함되는 입술영역 탐지부 및 윈도우 산출부를 설명하기 위한 도면이다.
- 도 7 및 8은 도 1의 영상 처리 장치에 포함되는 결정부를 설명하기 위한 도면들이다.

도 9 내지 12는 도 1의 영상 처리 장치에 포함되는 데이터 제공부의 동작을 설명하기 위한 도면들이다.

도 13는 도 1의 영상 처리 장치에 포함되는 데이터 제공부의 입력들을 설명하기 위한 도면들이다.

도 14는 도 1의 영상 처리 장치에 포함되는 오버랩이 있는 그룹화를 설명하기 위한 도면들이다.

도 15는 본 발명의 실시예들에 따른 영상 처리 시스템을 나타내는 도면이다.

도 16은 본 발명의 실시예들에 따른 영상 처리 장치의 동작방법을 나타내는 순서도이다.

도 17은 본 발명의 실시예들에 따른 영상 처리 시스템의 동작방법을 나타내는 순서도이다.

발명을 실시하기 위한 구체적인 내용

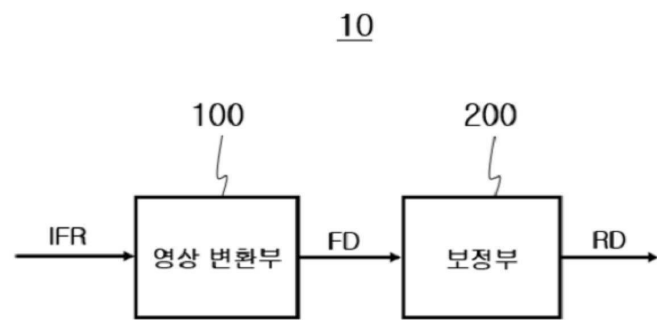
- [0025] 본 명세서에서 각 도면의 구성 요소들에 참조번호를 부가함에 있어서 동일한 구성 요소들에 한해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 번호를 가지도록 하고 있음에 유의하여야 한다.
- [0026] 한편, 본 명세서에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0027] 단수의 표현은 문맥상 명백하게 다르게 정의하지 않는 한, 복수의 표현을 포함하는 것으로 이해되어야 하는 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다.
- [0028] "포함하다" 또는 "가지다" 등의 용어는 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0029] 이하, 첨부되는 도면을 참고하여 상기 문제점을 해결하기 위해 고안된 본 발명의 바람직한 실시예들에 대해 상세히 설명한다.
- [0030] 도 1은 본 발명의 실시예들에 따른 영상 처리 장치를 나타내는 도면이고, 도 2는 도 1의 영상 처리 장치에 포함되는 영상 변환부를 나타내는 도면이고, 도 3은 도 1의 영상 처리 장치에서 사용되는 입력영상의 일 예를 나타내는 도면이다.
- [0031] 도 1 내지 3을 참조하면, 본 발명의 실시예에 따른 영상 처리 장치(10)는 영상 변환부(100) 및 보정부(200)를 포함할 수 있다. 영상 변환부(100)는 복수의 입력영상들(IFR)의 각각을 일정한 크기에 해당하는 패치 사이즈(PS)로 구분한 패치영상(PI)에 기초하여 복수의 입력영상들(IFR)에 포함되는 입술모양에 상응하는 특징 데이터(FD)를 제공할 수 있다.
- [0032] 일 실시예에 있어서, 영상 변환부(100)는 영상 제공부(110) 및 데이터 제공부(120)를 포함할 수 있다. 영상 제공부(110)는 복수의 입력영상들(IFR)을 그룹 사이즈(GS)로 그룹화한 후 그룹영상들(GFR)의 각각을 패치 사이즈(PS)로 구분하여 패치영상(PI)을 제공할 수 있다.
- [0033] 영상 제공부(110)는 복수의 입력영상들(IFR)을 일정한 그룹 사이즈(GS)로 그룹화하여 그룹영상(GFR)들을 제공할 수 있다. 이 경우, 입력영상(IFR)들에 비해 총 발화길이가 줄어들어 보다 빠른 데이터 처리를 할 수 있다. 복수의 그룹영상(GFR)들은 깊이별 분리 합성곱 신경망 (Depthwise separable convolutional neural network)을 사용하여 그룹화된 각각의 입력영상(IFR)들을 개별적인 필터로 처리할 수 있다. 여기서 합성곱 연산은 그래픽카드의 병렬연산으로 수행할 수 있다.
- [0034] 다른 일 실시예에 있어서, 복수의 입력영상들(IFR) 혹은 그룹영상들(GFR) 각각은 패치 사이즈(PS)보다 작은 복수의 윈도우(WI)들로 구분될 수 있다. 예를 들어, 복수의 입력영상들(IFR)은 제1 입력영상(IFR1) 내지 제N 입력영상(IFRN)을 포함할 수 있다. 제1 입력영상(IFR1)은 복수의 윈도우(WI)들로 구분될 수 있다. 복수의 윈도우(WI)들은 제1 윈도우 내지 제36 윈도우를 포함할 수 있고, 패치 사이즈(PS)는 정사각형 형태로 구성되는 4개의 윈도우들의 크기일 수 있다. 이 경우, 제1 입력영상(IFR1)을 4개의 윈도우(WI)로 구성되는 패치 사이즈(PS)로 구분하여 9개의 패치영상(PI)들을 구현할 수 있다.
- [0035] 데이터 제공부(120)는 패치영상(PI)에 기초하여 입술모양에 상응하는 특징 데이터(FD)를 제공할 수 있다. 여기서 데이터 제공부(120)는 스윈 트랜스포머(Swin transformer)를 포함할 수 있고, 데이터 제공부(120)는 스윈 트랜스포머를 이용하여 패치영상(PI)으로부터 특징 데이터(FD)를 제공할 수 있다.
- [0036] 보정부(200)는 특징 데이터(FD)를 보정하여 보정 데이터(RD)를 제공할 수 있다. 예를 들어, 여기서 사용되는 보정부(200)는 컨포머(Conformer)를 포함할 수 있고, 컨포머는 음성처리에서 주로 사용되고 있으나, 본 발명에 따

른 영상처리 분야에도 적용될 수 있다.

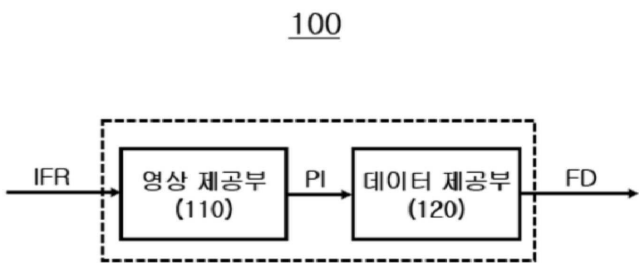
- [0037] 본 발명에 따른 영상 처리 장치(10)는 복수의 입력영상들(IFR)의 각각을 일정한 크기에 해당하는 패치 사이즈(PS)로 구분한 패치영상(PI)에 기초하여 복수의 입력영상들(IFR)에 포함되는 입술모양에 상응하는 특징 데이터(FD)를 도출함으로써 발화자가 전달하고자 하는 의미를 보다 정확하게 파악할 수 있다.
- [0038] 도 4는 도 1의 영상 처리 장치에 포함되는 입술영역 탐지부를 나타내는 도면이고, 도 5는 도 1의 영상 처리 장치에 포함되는 윈도우 산출부를 나타내는 도면이고, 도 6은 도 1의 영상 처리 장치에 포함되는 입술영역 탐지부 및 윈도우 산출부를 설명하기 위한 도면이고, 도 7 및 8은 도 1의 영상 처리 장치에 포함되는 결정부를 설명하기 위한 도면들이다.
- [0039] 도 1 내지 8을 참조하면, 일 실시예에 있어서, 영상 처리 장치(10)는 입술영역 탐지부(300)를 더 포함할 수 있다. 입술영역 탐지부(300)는 복수의 입력영상들(IFR) 각각에서 입술이 차지하는 입술영역(LR)을 탐지할 수 있다. 예를 들어, 복수의 입력영상들(IFR)은 제1 입력영상(IFR1) 내지 제16 입력영상(IFR16)을 포함할 수 있다. 이 경우, 입술영역 탐지부(300)는 제1 입력영상(IFR1) 내지 제16 입력영상(IFR16) 각각에 포함되는 사람의 입술부분을 탐지하여 사람의 입술을 포함하는 입술영역(LR)을 제공할 수 있다.
- [0040] 일 실시예에 있어서, 영상 처리 장치(10)는 윈도우 산출부(400)를 더 포함할 수 있다. 윈도우 산출부(400)는 복수의 입력영상들(IFR)의 각각에 대한 입술영역(LR)에 상응하는 윈도우(WI)들의 개수(WN)를 산출할 수 있다. 예를 들어, 복수의 입력영상들(IFR) 중 제1 입력영상(IFR1)에 포함되는 입술영역(LR)은 도 6과 같이 도시될 수 있다. 여기서, 제1 입력영상(IFR1)에 포함되는 입술영역(LR)을 윈도우(WI)의 개수(WN)로 산출하면 8일 수 있다. 이와 동일한 방식으로 윈도우 산출부(400)는 제2 입력영상 내지 제16 입력영상(IFR16)에 포함되는 입술영역(LR)에 상응하는 윈도우(WI)의 개수(WN)를 산출할 수 있다.
- [0041] 일 실시예에 있어서, 영상 처리 장치(10)는 결정부(500)를 더 포함할 수 있다. 결정부(500)는 입술영역(LR)에 상응하는 윈도우(WI)들의 개수(WN)에 기초하여 패치 사이즈(PS)를 결정할 수 있다.
- [0042] 일 실시예에 있어서, 패치 사이즈(PS)는 입술영역(LR)에 상응하는 윈도우(WI)들의 개수(WN)들의 평균값 및 평균값의 절반을 합한 합산 값으로 결정될 수 있다. 예를 들어, 제1 입력영상(IFR1) 내지 제16 입력영상(IFR16)에 포함되는 입술영역(LR)에 상응하는 윈도우(WI) 개수(WN)의 평균값은 8일 수 있다. 이 경우, 패치 사이즈(PS)는 윈도우(WI)들의 개수(WN)들의 평균값 및 평균값의 절반을 합한 $8+4=12$ 일 수 있다. 이와 같은 방식은 패치 사이즈(PS)를 결정하는 하나의 방식이고, 이외에도 입술영역(LR)의 제1 방향(D1)으로의 길이 및 제2 방향(D2)으로의 길이 중 큰 값에 해당하는 제1 방향(D1)으로의 길이를 기준으로 패치 사이즈(PS)를 정할 수도 있다.
- [0043] 도 9 내지 12는 도 1의 영상 처리 장치에 포함되는 데이터 제공부의 동작을 설명하기 위한 도면들이고, 도 13 및 14는 도 1의 영상 처리 장치에 포함되는 데이터 제공부의 입력들을 설명하기 위한 도면들이다.
- [0044] 도 1 내지 14를 참조하면, 일 실시예에 있어서, 데이터 제공부(120)는 쉬프트부(121) 및 데이터 출력부(122)를 더 포함할 수 있다. 쉬프트부(121)는 패치영상(PI)을 윈도우(WI)들 각각에 해당하는 단위 윈도우(WI)만큼 쉬프트하면서 생성되는 쉬프트 영상(SI)을 제공할 수 있다. 예를 들어, 입력영상으로부터 제1 패치영상(PI1)을 도출하기 위해서 제1 패치(PC1)가 사용될 수 있다. 쉬프트부(121)는 제1 패치(PC1)를 이용하여 단위 윈도우(WI)만큼씩 제1 방향(D1)으로 이동하면서 쉬프트 영상(SI)을 제공할 수 있다. 도 11에 도시되는 바와 같이, 쉬프트 영상(SI)은 제1 패치(PC1)를 기준으로 제1 방향(D1)과 반대방향에 배치되는 제5 영상을 제1 패치(PC1)를 기준으로 제1 방향(D1)에 배치하고, 제1 패치(PC1)를 기준으로 제3 방향과 반대방향에 배치되는 제3 영상(3)을 제1 패치(PC1)를 기준으로 제3 방향에 배치하고, 이와 동일한 방식으로 제1 영상(1), 제2 영상(2), 제4 영상(4)을 배치함으로써 구현될 수 있다.
- [0045] 데이터 출력부(122)는 쉬프트 영상(SI)에 기초하여 특징 데이터(FD)를 제공할 수 있다. 여기서, 데이터 출력부(122)의 동작은 스윈 트랜스포머의 동작일 수 있다.
- [0046] 일 실시예에 있어서, 데이터 제공부(120)의 입력들은 복수의 입력영상들(IFR) 각각의 패치영상(PI)에 상응할 수 있다.
- [0047] 다른 일 실시예에 있어서, 복수의 입력영상들(IFR) 들은 일정한 개수에 해당하는 그룹 사이즈(GS) 만큼 그룹화될 수 있다. 복수의 그룹영상(GFR)들은 일정한 개수에 해당하는 오버랩 개수만큼 오버랩(Overlap)될 수 있다. 그룹 사이즈(GS)는 발화 음성의 길이에 기초하여 결정할 수 있다.
- [0048] 예를 들어, 복수의 입력영상들(IFR)은 제1 입력영상(IFR1) 내지 제16 입력영상(IFR16)을 포함할 수 있다. 이 경

도면

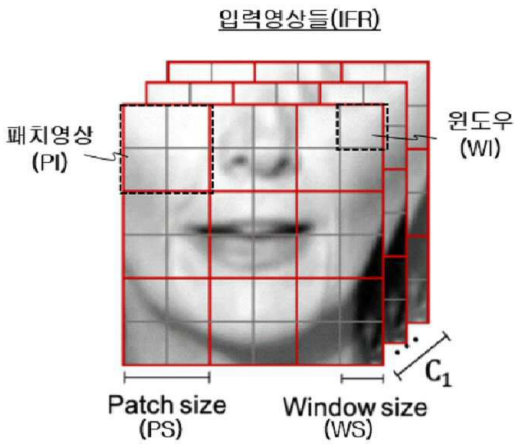
도면1



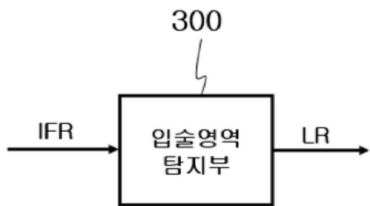
도면2



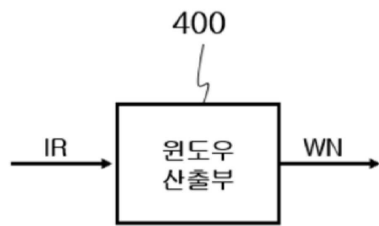
도면3



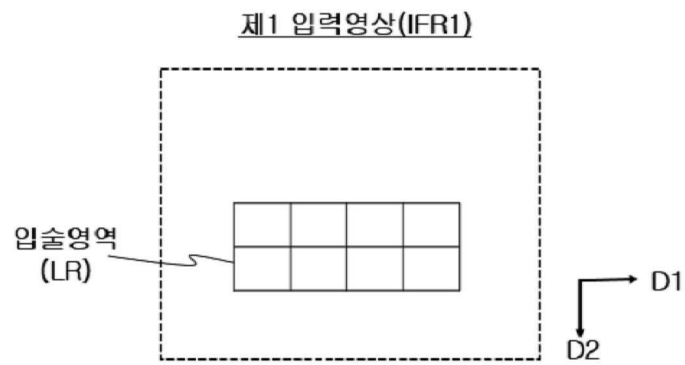
도면4



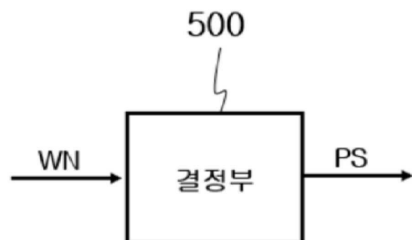
도면5



도면6



도면7

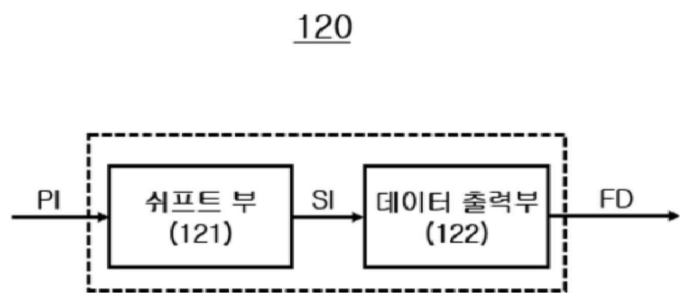


도면8

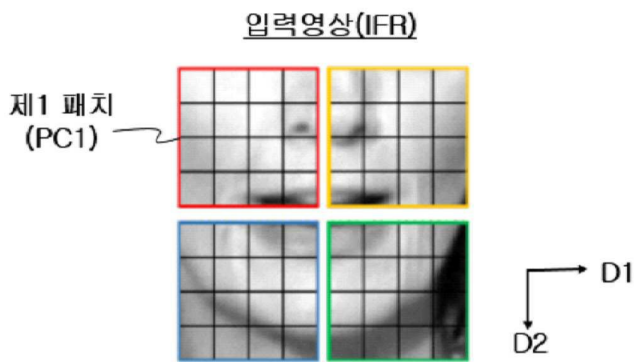
패치 사이즈(PS)

CASE1	8개
CASE2	12개

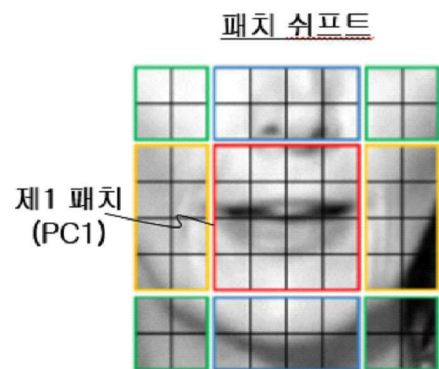
도면9



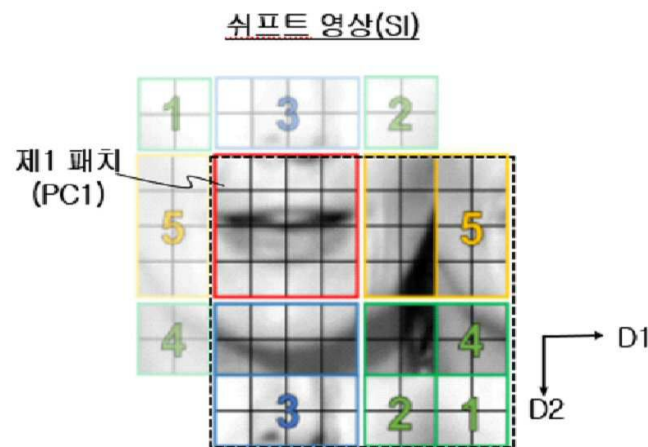
도면10



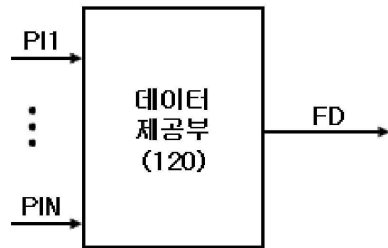
도면11



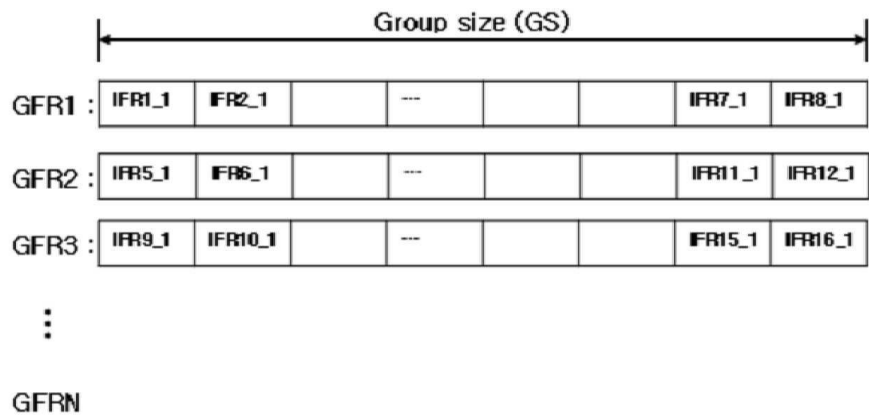
도면12



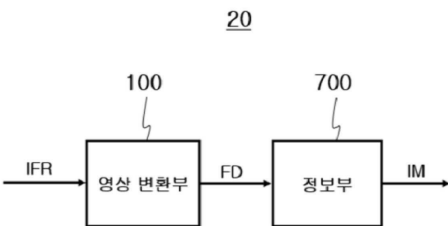
도면13



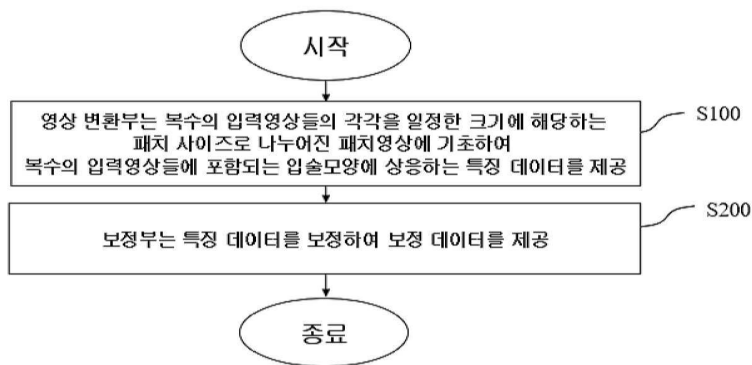
도면14



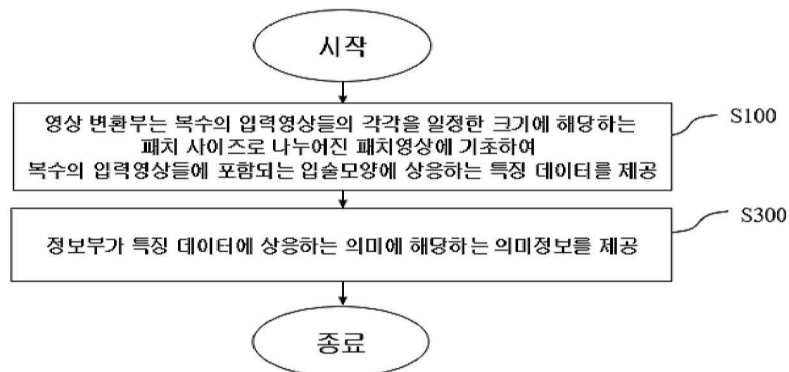
도면15



도면16



도면17



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 2

【변경전】

제1항에 있어서,

상기 영상 변환부는,

상기 복수의 입력영상들을 상기 그룹 사이즈의 묶음으로 그룹화한 후 각각을 상기 패치 사이즈로 구분하여 상기 패치영상을 제공하는 영상 제공부; 및

상기 패치영상에 기초하여 상기 입술모양에 상응하는 특징 데이터를 제공하는 데이터 제공부를 포함하는 것을 특징으로 하는 영상 처리 장치.

【변경후】

제1항에 있어서,

상기 영상 변환부는,

상기 복수의 입력영상들을 그룹 사이즈의 묶음으로 그룹화한 후 각각을 상기 패치 사이즈로 구분하여 상기 패치영상을 제공하는 영상 제공부; 및

상기 패치영상에 기초하여 상기 입술모양에 상응하는 특징 데이터를 제공하는 데이터 제공부를 포함하는 것을 특징으로 하는 영상 처리 장치.