(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2015년01월09일
(11) 등록번호 10-1480256
(24) 등록일자 2014년12월31일

(51) 국제특허분류(Int. Cl.)
H04N 7/14 (2006.01) *G06T 7/00* (2006.01)
(21) 출원번호 10-2013-0155650
(22) 출원일자 2013년12월13일
심사청구일자 2013년12월13일
(56) 선행기술조사문헌
KR1020130028344 A
KR1020100129783 A

(73) 특허권자
계명대학교 산학협력단
대구광역시 달서구 달구벌대로 1095 (신당동)
(72) 발명자
성연식
대구광역시 수성구 범어동 동대구로 376 범어숲 화성파크드림S 103동 1003호
(74) 대리인
특허법인태백

전체 청구항 수 : 총 12 항 심사관 : 박금옥

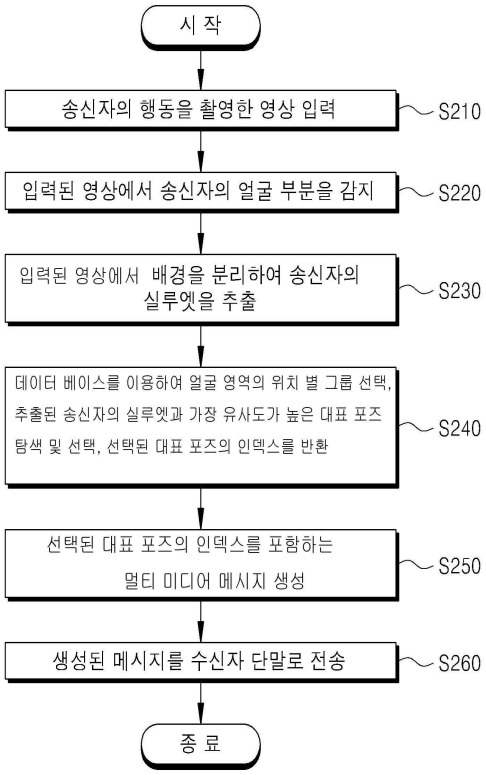
(54) 발명의 명칭 **얼굴 영역의 위치를 고려한 화상 통신 장치 및 그 방법**

(57) 요 약

본 발명은 얼굴 영역의 위치를 고려한 화상 통신 장치 및 그 방법에 관한 것이다. 본 발명에 따른 얼굴 영역의 위치를 고려한 화상 통신 장치는, 촬영장치로부터 송신자의 행동을 촬영한 영상을 입력받는 영상 입력부; 상기 입력된 영상으로부터 상기 송신자의 얼굴 영역을 검출하는 얼굴 검출부; 상기 입력된 영상으로부터 배경을 분리

(뒷면에 계속)

대 표 도 - 도2



하여 상기 송신자의 실루엣을 추출하는 실루엣 추출부; 미리 촬영된 시범 영상에서 추출된, 사람의 얼굴 영역의 위치 별로 그룹핑된 복수의 대표 포즈가 저장된 데이터 베이스를 이용하여, 상기 얼굴 검출부에 따른 송신자의 얼굴 영역의 위치에 대응되는 그룹을 선택하고, 상기 추출된 송신자의 실루엣과 가장 유사도가 높은 상기 대표 포즈를 탐색 및 선택하고, 상기 선택된 대표 포즈(이하 선택 포즈)의 인덱스를 반환하는 포즈 추정부; 상기 반환된 선택 실루엣의 인덱스를 포함하는 멀티 미디어 메시지를 생성하는 메시지 생성부; 및 상기 생성된 메시지를 수신자 단말로 전송하는 메시지 전송부를 포함한다.

이와 같이 본 발명에 따르면, 사용자의 얼굴 영역의 위치를 고려하여 실제 영상에 대응하는 대표 포즈의 인덱스 값을 수신자에게 전송함으로써 가상 객체에 의해 표현할 송신자의 행동을 탐색하는 알고리즘이 간단해져서 탐색 시간을 줄일 수 있다. 또한, 행위 네트워크를 통해 가상 객체에 의한 송신자의 행동을 실제에 가깝게 구현하여 부드럽고 자연스럽게 수신자 단말에서 구현할 수 있다.

특허청구의 범위

청구항 1

촬영장치로부터 송신자의 행동을 촬영한 영상을 입력받는 영상 입력부;

상기 입력된 영상으로부터 상기 송신자의 얼굴 영역을 검출하는 얼굴 검출부;

상기 입력된 영상으로부터 배경을 분리하여 상기 송신자의 실루엣을 추출하는 실루엣 추출부;

미리 촬영된 시범 영상에서 추출된, 사람의 얼굴 영역의 위치 별로 그룹핑된 복수의 대표 포즈가 저장된 데이터 베이스를 이용하여, 상기 얼굴 검출부에 따른 송신자의 얼굴 영역의 위치에 대응되는 그룹을 선택하고, 상기 추출된 송신자의 실루엣과 가장 유사도가 높은 대표 포즈를 탐색 및 선택하고, 상기 선택된 대표 포즈(이하 선택 포즈)의 인덱스를 반환하는 포즈 추정부;

상기 반환된 선택 포즈의 인덱스를 포함하는 멀티 미디어 메시지를 생성하는 메시지 생성부; 및

상기 생성된 메시지를 수신자 단말로 전송하는 메시지 전송부를 포함하며,

상기 수신자 단말은,

상기 선택 포즈의 인덱스를 포함하는 메시지를 연속적으로 수신하고, 상기 선택 포즈 인덱스와 상기 복수의 대표 포즈가 미리 저장된 데이터 베이스를 매핑하고, 상기 인덱스에 대응하는 대표 포즈로 액션을 구성하고, 상기 액션을 상기 송신자의 행위로 추정하여 가상 객체(virtual human)의 행위로 표시하되,

상기 대표 포즈를 기초로 다음 수학식의 행위 네트워크를 이용하여 상기 대표 포즈에 대응하는 송신자의 액션을 추정하는 얼굴 영역의 위치를 고려한 화상 통신 장치:

$$a^j \in A,\ A = \{a_1, a_2, \ldots\}$$

$$a^j = < p^i, d^j, c_1^j, c_2^j, \ldots >$$

여기서, a^j 는 j번째 대표 포즈로 추정되는 j번째 액션, p^j 는 j번째 대표 포즈, d^j 는 j번째 대표 포즈의 지속 시간, c^j_n 는 상기 가상 객체가 j번째 액션인 a^j 수행시에 요구되는 제n 관절 각도를 나타낸다.

청구항 2

제 1항에 있어서,

상기 데이터 베이스는,

시작 노드인 루트 노드 겸 부모 노드에서 순차적으로 2배수로 분기되는 자식 노드 형태의 트리 구조로 구성되고, 상기 부모 및 자식 노드 각각에는 유사도 판단에 필요한 제1 임계치 및 제2 임계치를 갖는 대표 포즈가 배치되는 복수의 의사 결정 트리(decision tree)를 포함하되, 상기 각각의 의사 결정 트리(decision tree)는 동일한 패턴의 얼굴 영역의 위치를 갖는 대표 포즈를 포함하는 얼굴 영역의 위치를 고려한 화상 통신 장치.

청구항 3

제 2항에 있어서,

상기 포즈 추정부는,

상기 제1 임계치를 이용하여 제1 임계치에 해당하는 대표 포즈가 상기 추출된 실루엣에 해당되는지 여부를 판단하고,

상기 제1 임계치에 해당하는 대표 포즈가 상기 추출된 실루엣에 해당되지 않을 경우,

상기 제2 임계치를 이용하여 이하에 분기된 자식 노드에 위치하는 복수의 대표 포즈 중에서 유사도를 판단할 하

나의 대표 포즈를 지정하는 얼굴 영역의 위치를 고려한 화상 통신 장치.

청구항 4

제 2항에 있어서,

상기 제1 임계치는,

해당 대표 포즈 및 비교되는 사용자 실루엣 간의 유사도가 제1 임계치 보다 클 경우, 상기 대표 포즈가 선택 포즈가 되도록 설정되고,

상기 제2 임계치는,

상기 부모 노드에서 분기된 이하의 자식 노드들이 양분되는 중앙값이어서, 상기 유사도가 제2 임계치 보다 클 경우에는 상기 분기된 자식 노드에 위치한 대표 포즈 중에서 해당 제1 임계치가 큰 쪽의 유사도를 판단하도록 설정되는 얼굴 영역의 위치를 고려한 화상 통신 장치.

청구항 5

제 1항에 있어서,

상기 송신자로부터 음성 또는 텍스트 중 적어도 하나를 입력받는 음성 입력부 또는 텍스트 입력부를 더 포함하고,

상기 생성된 메시지는 상기 입력된 음성 또는 텍스트 중 적어도 하나를 더 포함하여 상기 수신자 단말로 전송되는 얼굴 영역의 위치를 고려한 화상 통신 장치.

청구항 6

삭제

청구항 7

삭제

청구항 8

제 1항에 있어서,

상기 수신자 단말은,

송신자의 액션을 포함하는 액션 노드를 다음 수학식의 행위 네트워크를 이용하여 추정하는 얼굴 영역의 위치를 고려한 화상 통신 장치:

$$m^k = < a^j, s^k, g^k, o_1^k, o_2^k, \ldots, q_1^k, q_2^k >$$

여기서, m^k 는 a^j를 포함하는 액션 노드, s^k 및 g^k 는 액션 노드(m^k)와 연결된 시작 포즈 및 목표 포즈의 인덱스, o_x^k 는 액션노드(m^k)가 x 번째로 연결된 다른 액션 노드, q_x^k 는 o_x^k 로의 전이 확률을 각각 나타낸다.

청구항 9

화상 통신 장치를 이용한 얼굴 영역의 위치를 고려한 화상 통신 방법에 있어서,

촬영장치로부터 송신자의 행동을 촬영한 영상을 입력받는 단계;

상기 입력된 영상으로부터 상기 송신자의 얼굴 영역을 검출하는 단계;

상기 입력된 영상으로부터 배경을 분리하여 상기 송신자의 실루엣을 추출하는 단계;

미리 촬영된 시범 영상에서 추출된, 사람의 얼굴 영역의 위치 별로 그룹핑된 복수의 대표 포즈가 저장된 데이터베이스를 이용하여, 상기 얼굴 검출부에 따른 송신자의 얼굴 영역의 위치에 대응되는 그룹을 선택하고, 상기 추출된 송신자의 실루엣과 가장 유사도가 높은 상기 대표 포즈를 탐색 및 선택하고, 상기 선택된 대표 포즈(이하

선택 포즈)의 인덱스를 반환하는 단계;

상기 반환된 선택 포즈의 인덱스를 포함하는 멀티 미디어 메시지를 생성하는 단계; 및

상기 생성된 메시지를 수신자 단말로 전송하는 단계를 포함하며,

상기 수신자 단말은,

상기 선택 포즈의 인덱스를 포함하는 메시지를 연속적으로 수신하고, 상기 선택 포즈의 인덱스와 상기 복수의 대표 포즈가 미리 저장된 데이터 베이스를 매핑하고, 상기 인덱스에 대응하는 대표 포즈로 액션을 구성하고, 상기 액션을 상기 송신자의 행위로 추정하여 가상 객체(virtual human)의 행위로 표시하되,

상기 대표 포즈를 기초로 다음 수학식의 행위 네트워크를 이용하여 상기 대표 포즈에 대응하는 송신자의 액션을 추정하는 얼굴 영역의 위치를 고려한 화상 통신 방법:

$$a^j \in A,\ A = \{a_1, a_2, \ldots\}$$

$$a^j = < p^i, d^j, c_1^j, c_2^j, \ldots >$$

여기서, a^j 는 j번째 대표 포즈로 추정되는 j번째 액션, p^j 는 j번째 대표 포즈, d^j 는 j번째 대표 포즈의 지속 시간, c^j_n 는 상기 가상 객체가 j번째 액션인 a^j 수행시에 요구되는 제n 관절 각도를 나타낸다.

청구항 10

제 9항에 있어서,

상기 데이터 베이스는,

시작 노드인 루트 노드 겸 부모 노드에서 순차적으로 2배수로 분기되는 자식 노드 형태의 트리 구조로 구성되고, 상기 부모 및 자식 노드 각각에는 유사도 판단에 필요한 제1 임계치 및 제2 임계치를 갖는 대표 포즈가 배치되는 복수의 의사 결정 트리(decision tree)를 포함하되, 상기 각각의 의사 결정 트리(decision tree)는 동일한 패턴의 얼굴 영역의 위치를 갖는 대표 포즈를 포함하는 얼굴 영역의 위치를 고려한 화상 통신 방법.

청구항 11

제 10항에 있어서,

상기 대표 포즈의 인덱스를 반환하는 단계는,

상기 제1 임계치를 이용하여 제1 임계치에 해당하는 대표 포즈가 상기 추출된 실루엣에 해당되는지 여부를 판단하고,

상기 제1 임계치에 해당하는 대표 포즈가 상기 추출된 실루엣에 해당되지 않을 경우,

상기 제2 임계치를 이용하여 이하에 분기된 자식 노드에 위치하는 복수의 대표 포즈 중에서 유사도를 판단할 하나의 대표 포즈를 지정하는 얼굴 영역의 위치를 고려한 화상 통신 방법.

청구항 12

제 10항에 있어서,

상기 제1 임계치는,

해당 대표 포즈 및 비교되는 사용자 실루엣 간의 유사도가 제1 임계치 보다 클 경우, 상기 대표 포즈가 선택 포즈가 되도록 설정되고,

상기 제2 임계치는,

상기 부모 노드에서 분기된 이하의 자식 노드들이 양분되는 중앙값이어서, 상기 유사도가 제2 임계치 보다 클 경우에는 상기 분기된 자식 노드에 위치한 대표 포즈 중에서 해당 제1 임계치가 큰 쪽의 유사도를 판단하도록

설정되는 얼굴 영역의 위치를 고려한 화상 통신 방법.

청구항 13

제 9항에 있어서,

상기 송신자로부터 음성 또는 텍스트 중 적어도 하나를 입력받는 단계를 더 포함하고,

상기 생성된 메시지는 상기 입력된 음성 또는 텍스트 중 적어도 하나를 더 포함하여 상기 수신자 단말로 전송되는 얼굴 영역의 위치를 고려한 화상 통신 방법.

청구항 14

삭제

청구항 15

삭제

청구항 16

제 9항에 있어서,

상기 수신자 단말은,

송신자의 액션을 포함하는 액션 노드를 다음 수학식의 행위 네트워크를 이용하여 추정하는 얼굴 영역의 위치를 고려한 화상 통신 방법:

$$m^k = < a^j, s^k, g^k, o_1^k, o_2^k, \ldots, q_1^k, q_2^k >$$

여기서, m^k 는 a^j를 포함하는 액션 노드, s^k 및 g^k 는 액션 노드(m^k)와 연결된 시작 포즈 및 목표 포즈의 인덱스, o_x^k 는 액션노드(m^k)가 x 번째로 연결된 다른 액션 노드, q_x^k 는 o_x^k 로의 전이 확률을 각각 나타낸다.

명 세 서

기 술 분 야

[0001] 본 발명은 화상 통신 장치 및 그 방법에 관한 것으로서, 더욱 상세하게는 얼굴 영역의 위치를 고려하여 의사 결정 트리에 의한 행동 인지 기법을 사용하는 화상 통신 장치 및 그 방법에 관한 것이다.

배 경 기 술

[0002] 기술의 진보에 따라 사용자 단말기들은 상시로 네트워크에 연결되어 데이터 송수신이 가능해졌으며, 이에 따라 네트워크를 통한 커뮤니케이션 기술 또한 더불어 발전하게 되었다. 이러한 네트워크 커뮤니케이션 수단으로 오늘날 널리 활용되고 있는 통신 서비스는 인스턴트 메신저(Instant Messenger) 내지 화상 통신 시스템이다. 인스턴트 메신저란 인터넷을 통해 쪽지, 파일, 자료들을 실시간 전송할 수 있는 서비스로 채팅이나 전화와 마찬가지로 실시간으로 의사소통이 가능하다는 특징을 갖는다. 동일한 메신저 프로그램을 설치한 사용자들은 별도의 웹사이트와의 연결 없이도 단말기 대 단말기로 직접 통신기능이 가능하다. 별도의 인터넷이나 PC 통신 서비스에 가입하지 않고도 메신저 프로그램 설치만으로 상호 간의 대화가 가능하다.

[0003] 이러한 메신저 프로그램을 비롯하여 다양한 통신 서비스들은 최근 음성이나 화상 등의 데이터를 첨부하여 송수신하는 수준으로 진화하기에 이르렀으나, 다수의 사람이 동시에 화상을 통해서 데이터를 주고 받을 때 발생되는 데이터량이 상대적으로 크다는 점이 문제점으로 지적되었다. 뿐만 아니라, 네트워크의 물리적인 한계 내지 제약조건들로 인해 통신 시스템을 통해 실시간으로 다수의 영상을 처리하기 어려운 문제가 발생할 수 있다. 따라서, 통신 시스템을 통해 영상과 같이 대용량 데이터를 주고 받음에 있어서 송수신 데이터의 크기를 줄이기 위한 기술적 수단이 요구된다.

[0004] 본 발명의 배경이 되는 기술은 대한민국 공개특허공보 제2013-0028344호(2013.03.19)에 기재되어 있다.

발명의 내용

해결하려는 과제

[0005] 본 발명이 해결하고자 하는 기술적 과제는 화상 통신 시스템에 있어서 대용량 데이터의 전송을 줄여 네트워크의 트래픽 문제를 해결하기 위하여 실제 영상 전송을 대체할 수 있도록 가상 객체를 이용한 송신자의 예측 행동을 실제의 송신자의 행동에 가깝게 표현하는 것이다.

과제의 해결 수단

[0006] 상기한 바와 같은 목적을 달성하기 위한 본 발명의 하나의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 통신 장치는, 촬영장치로부터 송신자의 행동을 촬영한 영상을 입력 받는 영상 입력부; 상기 입력된 영상으로부터 상기 송신자의 얼굴 영역을 검출하는 얼굴 검출부; 상기 입력된 영상으로부터 배경을 분리하여 상기 송신자의 실루엣을 추출하는 실루엣 추출부; 미리 촬영된 시범 영상에서 추출된, 사람의 얼굴 영역의 위치 별로 그룹핑된 복수의 대표 포즈가 저장된 데이터 베이스를 이용하여, 상기 얼굴 검출부에 따른 송신자의 얼굴 영역의 위치에 대응되는 그룹을 선택하고, 상기 추출된 송신자의 실루엣과 가장 유사도가 높은 상기 대표 포즈를 탐색 및 선택하고, 상기 선택된 대표 포즈(이하 선택 포즈)의 인덱스를 반환하는 포즈 추정부; 상기 반환된 선택 실루엣의 인덱스를 포함하는 멀티 미디어 메시지를 생성하는 메시지 생성부; 및 상기 생성된 메시지를 수신자 단말로 전송하는 메시지 전송부를 포함한다.

[0007] 또한, 상기 데이터 베이스는, 시작 노드인 루트 노드 겸 부모 노드에서 순차적으로 2배수로 분기되는 자식 노드 형태의 트리 구조로 구성되고, 상기 부모 및 자식 노드 각각에는 유사도 판단에 필요한 제1 임계치 및 제2 임계치를 갖는 대표 포즈가 배치되는 복수의 의사 결정 트리(decision tree)를 포함하되, 상기 각각의 의사 결정 트리(decision tree)는 동일한 패턴의 얼굴 영역의 위치를 갖는 대표 포즈를 포함할 수 있다.

[0008] 또한, 상기 포즈 추정부는, 상기 제1 임계치를 이용하여 제1 임계치에 해당하는 대표 포즈가 상기 선택 실루엣에 해당되는지 여부를 판단하고, 상기 제1임계치에 해당하는 대표 포즈가 선택 실루엣에 해당되지 않을 경우, 상기 제2 임계치를 이용하여 이하에 분기된 자식 노드에 위치하는 복수의 대표 포즈 중에서 유사도를 판단할 하나의 대표 포즈를 정할 수 있다.

[0009] 또한, 상기 제1 임계치는, 해당 대표 포즈 및 비교되는 사용자 실루엣 간의 유사도가 제1 임계치 보다 클 경우, 상기 대표 포즈가 선택 포즈가 되도록 설정되고, 상기 제2 임계치는, 상기 부모 노드에서 분기된 이하의 자식 노드들이 양분되는 중앙값이어서, 상기 유사도가 제2 임계치 보다 클 경우에는 상기 분기된 자식 노드에 위치한 대표 포즈 중에서 해당 제1 임계치가 큰 쪽의 유사도를 판단하도록 설정될 수 있다.

[0010] 또한, 상기 송신자로부터 음성 또는 텍스트 중 적어도 하나를 입력받는 음성 입력부 또는 텍스트 입력부를 더 포함하고, 상기 생성된 메시지는 상기 입력된 음성 또는 텍스트 중 적어도 하나를 더 포함하여 상기 수신자 단말로 전송될 수 있다.

[0011] 또한, 상기 수신자 단말은, 상기 선택 포즈의 인덱스를 포함하는 메시지를 연속적으로 수신하고, 상기 선택 포즈 인덱스와 상기 복수의 대표 포즈가 미리 저장된 데이터 베이스를 매핑하고, 상기 인덱스에 대응하는 대표 포즈로 액션을 구성하고, 상기 액션을 상기 송신자의 행위로 추정하여 가상 객체(virtual human)의 행위로 표시할 수 있다.

[0012] 또한, 상기 수신자 단말은, 상기 대표 포즈를 기초로 다음 수학식의 행위 네트워크(behavior network)를 이용하여 상기 대표 포즈에 대응하는 송신자의 액션을 추정할 수 있다.

$$a^j \in A,\ A = \{a_1, a_2, \ldots\}$$

$$a^j = < p^i, d^j, c_1^j, c_2^j, \ldots >$$

[0013]

[0014] 여기서, a^j 는 j번째 대표 포즈로 추정되는 j번째 액션, p^j 는 j번째 대표 포즈, d^j 는 j번째 대표 포즈의 지속 시간, c^j_n 는 상기 가상 객체가 j번째 액션인 a^j 수행시에 요구되는 제n 관절 각도를 나타낸다.

[0015] 또한, 상기 수신자 단말은, 송신자의 액션을 포함하는 액션 노드를 다음 수학식의 행위 네트워크(behavior network)를 이용하여 추정할 수 있다.

[0016] $$m^k = < a^j, s^k, g^k, o_1^k, o_2^k, \ldots, q_1^k, q_2^k >$$

[0017] 여기서, m^k 는 a^j를 포함하는 액션 노드, s^k 및 g^k 는 액션 노드(m^k)와 연결된 시작 포즈 및 목표 포즈의 인덱스, o_x^k 는 액션노드(m^k)가 x 번째로 연결된 다른 액션 노드, q_x^k 는 o_x^k 로의 전이 확률을 각각 나타낸다.

[0018] 본 발명의 하나의 실시예에 따른 화상 통신 장치를 이용한 얼굴 영역의 위치를 고려한 화상 통신 방법은, 촬영 장치로부터 송신자의 행동을 촬영한 영상을 입력받는 단계; 상기 입력된 영상으로부터 상기 송신자의 얼굴 영역을 검출하는 단계; 상기 입력된 영상으로부터 배경을 분리하여 상기 송신자의 실루엣을 추출하는 단계; 미리 촬영된 시범 영상에서 추출된, 사람의 얼굴 영역의 위치 별로 그룹핑된 복수의 대표 포즈가 저장된 데이터 베이스를 이용하여, 상기 얼굴 검출부에 따른 송신자의 얼굴 영역의 위치에 대응되는 그룹을 선택하고, 상기 추출된 송신자의 실루엣과 가장 유사도가 높은 상기 대표 포즈를 탐색 및 선택하고, 상기 선택된 대표 포즈(이하 선택 포즈)의 인덱스를 반환하는 단계; 상기 반환된 선택 실루엣의 인덱스를 포함하는 멀티 미디어 메시지를 생성하는 단계; 및 상기 생성된 메시지를 수신자 단말로 전송하는 단계를 포함한다.

발명의 효과

[0019] 본 발명인 얼굴 영역의 위치를 고려한 화상 통신 장치 및 그 방법에 따르면, 사용자의 얼굴 영역의 위치를 고려하여 실제 영상에 대응하는 대표 포즈의 인덱스 값을 수신자에게 전송함으로써 가상 객체에 의해 표현할 송신자의 행동을 탐색하는 알고리즘이 간단해져서 탐색 시간을 줄일 수 있다. 또한, 행위 네트워크(behavior network)를 통해 가상 객체에 의한 송신자의 행동을 실제에 가깝게 구현하여 부드럽고 자연스럽게 수신자 단말에서 구현할 수 있다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 통신 장치의 구성도이다.

도 2는 본 발명의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 방법의 순서도이다.

도 3은 본 발명의 실시예에 따른 도 2의 S230 단계에서 실루엣을 추출하는 과정을 도시한 도면이다.

도 4는 본 발명의 실시예에 따른 복수의 후보 포즈에서 복수의 대표 포즈를 선택하는데 사용되는 Maximin 알고리즘의 예시 도면이다.

도 5는 선택된 대표 포즈를 이용하여 의사 결정 트리를 구성하는 방법을 설명하기 위한 도면이다.

도 6는 도 5에서 배치된 제2행의 대표 포즈 중에서 첫 번째 및 두 번 째로 배치되는 대표 포즈를 설명하기 위한 도면이다.

도 7은 도 5에서 제3행에 배치될 자식 노드의 대표 포즈를 설명하기 위한 도면이다.

도 8은 의사 결정 트리를 이용하여 대표 포즈의 인덱스를 추정하는 과정을 설명하기 위한 예시 도면이다.

도 9은 본 발명의 실시예에 따른 행위 네트워크(behavior network)를 이용하여 액션 노드를 선택하는 과정을 설명하기 위한 예시 도면이다.

발명을 실시하기 위한 구체적인 내용

[0021] 이하에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0022] 본 발명의 실시예들을 설명하기에 앞서, 우선 본 발명의 실시예들이 구현되는 통신 서비스(메신저 서비스 또는 실시간 영상 채팅 서비스)의 특징에 대해 설명한다.

[0023] 영상을 전달하는 통신 서비스에 참여하는 사용자들 간에 영상 정보를 직접 송수신할 경우 실시간 영상 정보의 특징으로 인해 많은 네트워크 트래픽을 발생시킨다. 따라서, 이하에서 제안되는 본 발명의 다양한 실시예들은 직접 이러한 영상 정보를 전송하지 않고, 단지 영상 정보를 나타낼 수 있는 코드(code)(이하에서는 영상의 특정 포즈, 단편적인 행동, 일련의 행위 등을 식별하기 위한 인덱스라고 명명한다) 정보만을 수신자에게 전달한다. 그러면, 수신자는 이러한 인덱스만을 수신하여 송신자의 행동이 어떠한 것인지 재현하게 된다.

[0024] 이를 위해 본 발명의 실시예들은 사전에 시범(demonstration) 영상을 통해 인간의 행동을 모방하기 위한 대표 포즈를 자동으로 생성하고, 데이터베이스를 구축한다. 이 때, 각각의 대표 포즈와 이에 부여된 고유 식별자(포즈 인덱스를 의미한다)가 매칭되어 저장된다.

[0025] 이제, 이러한 대표 포즈들이 저장된 데이터 베이스를 이용하여 카메라를 통해 실시간으로 입력되는 사용자의 영상에 대응하는 대표 포즈의 인덱스 값을 결정할 수 있고, 결정된 인덱스 값은 수신자에게 전송된다. 인덱스 값을 수신한 수신자는 송신자와 동일한 데이터 베이스를 가지고 있어야 하며, 이로부터 영상을 재생할 수 있다. 이 때, 영상의 재생은 단지 일 순간의 포즈만을 재생하는 것이 아니라, 시간적 순서와 단일 포즈의 지속 시간을 고려하여 행동을 구성한다. 그리고 이를 기반으로 일련의 다양한 행위를 구성하여 재생할 수 있다.

[0026] 이하에서 첨부된 도면을 참조하여 본 발명의 실시예들 및 그 구성요소들을 각각 구체적으로 설명한다.

[0027] 먼저 도 1을 통하여 본 발명의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 통신 장치에 대해서 설명한다.

[0028] 도 1은 본 발명의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 통신 장치의 구성도이다.

[0029] 도 1에 도시된 바와 같이, 본 발명의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 통신 장치(100)는, 영상 입력부(110), 얼굴 검출부(120), 실루엣 추출부(130), 포즈 추정부(140), 데이터 베이스(150), 메시지 생성부(160) 및 메시지 전송부(170)을 포함하며, 텍스트 입력부(180) 및 음성 입력부(190)를 더 포함할 수 있다.

[0030] 영상 입력부(110)는 촬영장치(미도시)로부터 송신자의 행동을 촬영한 영상을 입력 받으며, 얼굴 검출부(120)는 영상 입력부(110)로 입력된 영상으로부터 송신자의 얼굴 영역을 검출한다.

[0031] 실루엣 추출부(130)는 입력된 영상으로부터 배경을 분리하여 송신자의 실루엣을 추출한다. 포즈 추정부(140)는 미리 촬영된 시범 영상에서 추출된, 사람의 얼굴 영역의 위치 별로 그룹핑된 복수의 대표 포즈가 저장된 데이터 베이스(150)를 이용하여, 상기 얼굴 검출부에 따른 송신자의 얼굴 영역의 위치에 대응되는 그룹을 선택하고, 상기 추출된 송신자의 실루엣과 가장 유사도가 높은 대표 포즈를 탐색 및 선택하고, 상기 선택된 대표 포즈(이하 선택 포즈)의 인덱스를 반환한다.

[0032] 메시지 생성부(160)는 반환된 선택 포즈의 인덱스를 포함하는 멀티 미디어 메시지를 생성하며, 메시지 전송부(170)는 생성된 메시지를 수신자 단말로 전송한다. 여기서 수신자 단말(미도시)은 메시지를 전송하는 얼굴 영역의 위치를 고려한 화상 통신 장치(100)와 같은 구성을 하고 있다. 즉, 얼굴 영역의 위치를 고려한 화상 통신 장치(100)는 송신측에서는 송신 장치로 수신측에서는 수신 장치로 이용될 수 있다.

[0033] 한편, 이상의 얼굴 영역의 위치를 고려한 화상 통신 장치(100)는 송신자로부터 음성 또는 텍스트 중 적어도 하나를 입력받는 텍스트 입력부(180) 또는 음성 입력부(190)를 더 포함할 수 있으며, 생성된 메시지는 입력된 음성 또는 텍스트 중 적어도 하나를 더 포함하여 수신자 단말에 전송될 수 있다.

[0034] 이하 본 발명의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 통신 방법에 대하여 자세하게 설명한다.

[0035] 도 2는 본 발명의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 통신 방법의 순서도이다.

[0036] 영상 입력부(110)는 촬영장치로부터 송신자의 행동을 촬영한 영상을 입력 받는다(S210). 이러한 영상은 동영상 촬영장치에 해당하는 방송용 무비 카메라 또는 가정용 캠코더를 통해 촬영되어 다수의 정지 화상이 시간의 흐름에 따라 연속적으로 배치된 동영상이거나 디지털 카메라에 의해 촬영된 정지 영상일 수 있다. 동영상이 소스로 사용되는 경우에는 동영상으로부터 캡쳐된 정지영상이 입력된다.

[0037] 얼굴 검출부(120)는 영상 입력부(110)에 입력된 영상으로부터 송신자의 얼굴 영역을 검출한다(S220).

[0038] 입력된 영상으로부터 얼굴을 검출하는데 있어서 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 이용할 수 있는 얼굴 검출 기술이 사용될 수 있다. 이러한 얼굴 검출 기술은 신체와 비교되는 얼굴의 모양과 얼굴의 피부톤에 기초한 것으로서 최근의 스마트 카메라에 널리 사용되고 있다. 몇 가지 알려진 얼굴 영역 검출 알고리즘의 예로는 Face Recognition Vendor Test, NIST Face Recognition Grand Challenge 및 Multiple Biometric

Grand Challenge 가 있다.

[0039] 다음으로, 실루엣 추출부(silhouette extractor)(130)는 입력된 영상(user image)으로부터 배경(background)을 분리하여 송신자의 실루엣(user silhouette)을 추출한다(S230).

[0040] 도 3은 본 발명의 실시예에 따른 도 2의S230 단계에서 실루엣을 추출하는 과정을 도시한 도면이다.

[0041] 입력된 영상으로부터 배경과 객체(사람)를 분리하기 위해, 사전에 배경 이미지를 저장해 놓을 경우, 객체와 배경의 분리가 더욱 용이하다. 도 3에 도시된 바와 같이, 배경 학습부(background learner)는 사전에 배경 이미지를 저장하여 배경 분리를 용이하게 할 수 있다.

[0042] 즉, 미리 저장된 배경 이미지와 입력된 영상 중 일 순간의 영상의 이미지 차이를 산출함으로써 송신자의 실루엣을 추출할 수 있다. 물론 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 이러한 방법 이외에서 다양한 기술적 수단을 통해 입력된 영상으로부터 배경과 송신자의 실루엣을 분리하는 것이 가능하다.

[0043] 성능 향상을 위해 이렇게 추출된 송신자의 실루엣으로부터 추가적으로 노이즈(noise)를 제거하는 것이 바람직하다.

[0044] 다음으로, 포즈 추정부(140)는 미리 촬영된 시범 영상에서 추출된, 사람의 얼굴 영역의 위치 별로 그룹핑된 복수의 대표 포즈가 저장된 데이터 베이스(150)를 이용하여, 상기 얼굴 검출부에 따른 송신자의 얼굴 영역의 위치에 대응되는 그룹을 선택하고, 상기 추출된 송신자의 실루엣과 가장 유사도가 높은 대표 포즈를 탐색 및 선택하고, 상기 선택된 대표 포즈(이하 선택 포즈)의 인덱스를 반환한다(S240).

[0045] 포즈 추정부(140)는 얼굴 영역 검출부(120)에서 검출된 얼굴 영역의 위치를 고려하여, 실루엣 추출부(130)에서 추출된 송신자의 실루엣과 가장 유사도가 높은 대표 포즈를 데이터 베이스(150)에서 탐색한다. 여기서 얼굴 영역의 위치는 추출된 송신자의 실루엣에서 송신자의 몸체를 기준으로 얼굴이 중앙, 좌측 또는 우측에 위치하는지 고려되는 것이다. 추가적으로 얼굴을 숙인 경우 또는 얼굴을 뒤로 젖힌 경우도 포함될 수 있다.

[0046] 데이터 베이스(150)는 시작 노드인 루트 노드 겸 부모 노드에서 순차적으로 2배수로 분기되는 자식 노드 형태의 트리 구조로 구성되고, 상기 부모 및 자식 노드 각각에는 유사도 판단에 필요한 제1 임계치 및 제2 임계치를 갖는 대표 포즈가 배치되는 복수의 의사 결정 트리(decision tree)를 포함하되, 상기 각각의 의사 결정 트리(decision tree)는 동일한 패턴의 얼굴 영역의 위치를 갖는 대표 포즈를 포함한다.

[0047] 즉, 데이터 베이스(150)는 복수의 의사 결정 트리를 포함하며, 여기서 복수의 의사 결정 트리는 송신자의 얼굴 영역 별로 구축될 수 있다. 전체 영상에서 송신자의 얼굴이 차지하는 위치 또는 송신자의 몸과 비교하여 송신자의 얼굴 영역이 차지하는 위치를 고려하여 시작 노드인 하나의 루트 노드를 정하고, 이하 분기되는 자식 노드에 배치되는 대표 포즈는 동일한 패턴의 얼굴 영역의 위치를 갖는다.

[0048] 데이터 베이스(150) 구축 및 대표 포즈 탐색에 있어서, 얼굴 영역의 위치를 고려하여 저장된 의사 결정 트리를 이용하여 대표 포즈를 추정하는 단계가 얼굴 영역의 위치를 분류하지 않고 대표 포즈를 추정하는 단계와 비교하여 소요 시간, 복잡도 및 정확성에 있어서 효과가 있다. 또한, 송신자의 얼굴 영역의 위치를 고려하여 시작 노드인 하나의 루트 노드의 결정에 있어서도 루트 노드의 대표 포즈 만으로 구성되는 새로운 의사 결정 트리 기법이 이용될 수 있다.얼굴 영역의 위치를 고려한 화상 통신 장치(100)는 인간의 행동을 포함하는 시범 영상으로부터 자동으로 생성된 복수 개의 대표 포즈가 미리 저장된 데이터 베이스를 이용하여 추출된 송신자 실루엣과 가장 유사한 대표 포즈를 선택함으로써 송신자의 행동에 대응하는 대표 포즈의 인덱스를 추정한다. 이 때, 미리 저장된 데이터 베이스는 다양한 데이터 포맷이 활용될 수 있으나, 본 실시예에서는 의사 결정 트리(decision tree)를 활용하는 방법을 이용한다.

[0049] 이하에서는 도 4 내지 도 7을 통하여 S240 단계에서 이용되는 데이터 베이스의 의사 결정 트리를 구성하는 과정을 설명한다.

[0050] 본 발명의 실시예에 따른 복수 개의 의사 결정 트리는 인간의 행동을 포함하는 시범 영상으로부터 자동으로 생성된 복수 개의 대표 포즈로 구성된다.

[0051] 보다 구체적으로, 인간의 행동을 포함하는 시범 영상으로부터 동일한 패턴의 송신자의 얼굴 영역의 위치를 갖는 복수 개의 후보 포즈가 추출되고, 추출된 후보 포즈로부터 복수 개의 대표 포즈가 자동으로 선택되며, 선택된 대표 포즈로부터 각 레벨의 부모 노드의 임계치를 기준으로 부모 노드의 자식 노드들이 양분되도록 구성된다. 이를 위해, 의사 결정 트리의 각 노드(node)에서 대표 포즈와 해당 임계치가 정의된다.

[0052] 추출된 후보 포즈로부터 복수 개의 대표 포즈의 선택은 Maximin 선택 알고리즘에 의한다. Maximin 선택 알고리즘은 복수 개의 원소를 갖는 집합에서 각각의 원소들의 특성에 따라 군집이 형성되는 경우 이러한 군집을 대표하는 원소를 선택할 수 있는 기법이다. 따라서, 선택되는 대표 원소는 각 군집의 특성을 대표하는 대표성을 가지며, 또한 대표 원소들 각각은 서로 차별적인 특성을 가지게 된다. 이러한 목적 하에 Maximin 선택 알고리즘을 본 발명의 실시예에 적용하면, 다양한 후보 포즈들로부터 상대적으로 적은 수의 대표 포즈를 선택할 수 있다. 이하에서는 의사 코드(pseudo-code)로 구현된 Maximin 선택 알고리즘을 참조하여 대표 포즈의 선택 방법을 보다 구체적으로 설명한다.

[0053] 도 4는 본 발명의 실시예에 따른 복수의 후보 포즈에서 복수의 대표 포즈를 선택하는데 사용되는 Maximin 알고리즘의 예시 도면이다.

[0054] 도 4에서, C는 후보 포즈의 집합을 의미하고, B는 후보 포즈들 중 대표가 아닌 포즈의 집합을 의미하며, T는 대표로 선택된 대표 행동의 집합을 의미한다. 즉, 집합 B는 집합 C에서 집합 T를 뺀 차집합에 해당한다. 알고리즘은 함수 호출시 후보 포즈 데이터베이스에 저장된 후보 포즈의 집합을 입력 받는다. 또한, 변수 actionCount는 최종적으로 선택하고자 하는 대표 행동의 개수를 의미하는 것으로, 사전에 미리 설정되는 것이 바람직하다.

[0055] 알고리즘의 최초 실행시 집합 B는 후보 포즈의 집합 C 전체를 할당 받으며, 알고리즘의 수행 과정을 통해 집합 B에서 선택된 대표 포즈는 집합 B에서 삭제되어 집합 T로 입력된다. 즉, 대표 포즈로 선택된 이후에는 더 이상 후보 포즈가 아니므로, 선택된 원소에 대한 집합 B로부터 집합 T로의 데이터 이동이 이루어진다.

[0056] 대표 포즈의 선택 과정은 다음과 같다. 우선, 집합 B에서 최초의 원소를 임의로 선택하여 대표 포즈로 설정한다. 즉, 선택된 원소를 집합 B에서 삭제한 후, 집합 T의 최초의 원소로서 저장한다. 이후 두 번째로 선택되는 대표 포즈에 대해서는 다음과 같은 일련의 과정을 수행하게 된다.

[0057] 집합 B의 각각의 후보 포즈들을 집합 T의 각각의 대표 포즈들과 비교하여 그 차이값을 산출한다. 이 때, 집합 B의 원소마다 집합 T의 대표 포즈들과 비교를 수행하게 되므로, 산출되는 차이값의 수는 집합 B의 원소의 개수와 집합 T의 원소의 개수의 곱이 될 것이다. 이제, 이렇게 산출된 차이값들 중, 집합 B의 원소별로 최소값을 선택한다. 집합 B의 원소마다 하나의 최소값을 선택하게 되므로, 선택된 최소값의 개수는 집합 B의 원소의 개수와 같다. 이어서, 선택된 최소값들을 재차 비교하여, 이들 최소값 중에서 가장 큰 값을 갖는 집합 B의 원소를 대표 포즈로 선택한다. 당연히 선택된 대표 포즈는 집합 B에서 삭제되어 집합 T에 저장된다. 즉, 이미 선택되어 대표 포즈를 구성하는 집합 T와 가장 상이한 원소가 새로운 대표 포즈로 선택되게 된다. 이상과 같은 일련의 과정을 반복 수행함으로써 선택된 대표 포즈의 수가 최초에 설정된 변수 actionCount와 같아지면 알고리즘을 종료한다.

[0058] 이상의 과정을 통해 본 발명의 실시예에 따르면, 다양한 후보 포즈들 중에서 가장 특징적인 포즈들이 대표 포즈로서 자동으로 선택될 수 있다. 앞서 설명한 바와 같이 이상의 대표 포즈 선택 알고리즘에 따르면, 새로운 대표 포즈를 선택함에 있어서, 이미 선택되어 대표 포즈를 구성하는 집합 T와 가장 상이한 원소가 새롭게 선택되므로, 이러한 연산을 반복할수록 자동으로 최초의 후보 포즈 집합 중에서 가장 특징적이고 차별적인 영상들만이 대표 포즈로 선택될 수 있는 효과가 나타난다. 즉, 복수 개의 후보 포즈들에 대해 사용자가 일일이 대표 포즈를 시각적으로 확인하고 선택할 필요 없이 자동으로 대표 포즈에 대한 데이터베이스를 구축할 수 있다. 본 발명에 속하는 기술 분야에서 통상의 지식을 가진 자는 이상과 같이 복수의 대표 포즈를 선택하는 과정 및 부분 행동들을 분류하는 과정에 Maximin 선택 알고리즘 외에도 다양한 분류 알고리즘이 활용될 수 있음을 알 수 있다.

[0059] 이하에서는 도 5 내지 도 7을 통하여 선택된 대표 포즈를 이용하여 의사 결정 트리를 구성하는 방법에 대하여 예를 들어 설명한다.

[0060] 도 5는 선택된 대표 포즈를 이용하여 의사 결정 트리를 구성하는 방법을 설명하기 위한 도면이다.

[0061] 먼저, 복수의 후보 포즈에서 선택된 100장의 대표 포즈를 이용하여 하나의 의사 결정 트리를 구성하는 것으로 가정한다. 여기서 100장의 대표 포즈는 다음과 같은 과정을 통해 선택된다. 미리 촬영된 시범 영상에서 송신자의 얼굴 영역의 위치가 검출되고, 일정 패턴의 얼굴 영역의 위치를 갖는 시범 영상에서 송신자의 실루엣이 추출된다. 이렇게 추출된 복수 개의 송신자 실루엣은 후보 포즈에 해당되며, 복수 개의 후보 포즈 중에서 100 개의 대표 포즈가 상기 Maximin 선택 알고리즘에 의해 자동으로 선택된다.

[0062] 대표 포즈를 구성하는 경우에도 얼굴 검출부(120)는 시범 영상으로부터 얼굴 영역의 위치를 검출한다. 따라서, 100 장의 대표 포즈는 송신자의 얼굴 영역의 위치가 비슷한 패턴을 갖는 대표 포즈에 해당한다. 예를 들면 100

개의 대표 포즈에서 송신자의 얼굴 영역의 위치는 가운데에 놓인다.

[0063] 도 5에 도시된 바와 같이, 100개의 대표 포즈 중에서 제1행의 n^1 위치에는 임의의 대표 포즈를 위치 시킬 수 있다. 즉, 비슷한 패턴의 얼굴 영역의 위치의 대표 포즈 중에서 일반적인 동작의 포즈, 예를 들면 가운데 얼굴 영역의 위치를 하고 있으며 팔을 가지런히 앞으로 하고 있는 동작의 대표 포즈가 배치 될 수 있다. 그리고 나머지 n^2 내지 n^{100}의 대표 포즈들이 의사 결정 트리의 제2행에 임시로 배열된다.

[0064] 도 5에서, n^2 내지 n^{100}의 대표 포즈의 배열 순서는 대표 포즈의 유사도에 의한다. 즉 제2행의 99개의 대표 포즈를 n^1 의 대표 포즈와 비교하여 포즈 유사도가 가장 작은 대표 포즈를 n^2에 배열하고, 포즈 유사도가 가장 큰 대표 포즈가 n^{100}에 배열된다. 그러면, 제2행에는 n^1을 제외한 99개의 대표 포즈가 배열된다. 그리고 99개의 대표 포즈를 2개의 군(좌측군, 우측군)으로 나누는 작업을 하고 각각의 군에서 중간에 위치한 2개의 대표 포즈들(제2행 좌측 포즈, 제2행 우측 포즈)만 남기고, 나머지 97개의 대표 포즈들이 하위 제3행에 임시로 배열된다. 여기서, 유사도의 판단은 상관계수(correlation coefficient)에 의한다. 수학적으로 상관계수는 -1에서 1사이의 값을 가지므로 본 발명에서는 상관계수의 절대값으로 판단한다.

[0065] 도 6은 도 5에 임시로 배열된 제2행의 대표 포즈 중에서 첫 번째 및 두 번 째로 배치되는 대표 포즈를 설명하기 위한 도면이다.

[0066] 도 6에서, 제2행의 첫 번째 노드(the 1^{st} orderd node)에 배치되는 좌측 대표 포즈 및 제2행의 두 번째 노드(the 2^{nd} ordered node)에 배치되는 우측 대표 포즈를 구하는 방법을 수학식으로 나타내면 다음과 같다.

수학식 1

[0067]
$$\text{자식 노드의 좌측 대표 포즈의 순서} = \left\lfloor \frac{O+1}{4} \right\rfloor$$

$$\text{자식 노드의 우측 대표 포즈의 순서} = \left\lfloor \frac{(O+1)*3}{4} \right\rfloor$$

[0068] 여기서, O는 자식 노드에 배열된 대표 포즈의 개수이고, 상기 $\lfloor\ \rfloor$ 연산자는 소수점 이하 자리를 버리는 버림 기호이다.

[0069] 예를 들면, 제2행의 자식 노드에 배열된 대표 포즈의 순서를 정함에 있어서, 자식 노드에 배열된 대표 포즈의 개수는 99개이다.

[0070] 따라서, 제2행 자식 노드의 좌측 대표 포즈의 순서 = $\lfloor (99+1)4 \rfloor$ = 25, 제2행 자식 노드의 우측 대표 포즈의 순서 = $\lfloor (99+1)*3/4 \rfloor$ = 75이므로, n^1에는 대표 포즈의 포즈 유사도의 올림차순 배열에 따른 배열에서 25번째 대표 포즈가 배치하고, n^2에는 같은 방법으로 75번째 대표 포즈가 배치된다.

[0071] 상기 과정을 간략하게 정리하면, 하나의 행이 내려갈 때마다 두 개의 군으로 대표 포즈들을 양분하되 양분된 군에서 중간에 위치하는 대표 포즈를 해당 행에 배치시키고 나머지 대표 포즈들은 하위 행으로 내려가서 배열된다. 그리고 상기 방법에 의한 임시 배열과 배치가 반복된다.

[0072] 도 7은 도 5에서 제3행에 배치될 자식 노드의 대표 포즈를 설명하기 위한 도면이다.

[0073] 제3행에 대해서는 상기 2행의 방법과 동일한 방법에 의해 n^4 내지 n^7의 대표 포즈가 배치된다. 즉, 하나의 부모 노드 하위에는 2개의 자식 노드가 생성되고, 각 행에 배치되는 대표 포즈의 수는 상위 행의 대표 포즈의 수의 2배가 된다. 여기서, 하위 행에서 자식 노드에 배치될 대표 포즈를 구하기 전에 후보가되는 대표 포즈의 순서를

다시 정해서 배열해야 한다.

[0074] 다음으로 S240 단계와 관련하여 상기 구성된 의사 결정 트리를 이용하여 포즈 인덱스를 추정하는 방법에 대하여 도 8을 통하여 설명한다.

[0075] 도 8은 본 발명의 실시예에 따른 의사 결정 트리를 이용하여 포즈 인덱스를 추정하는 방법을 설명하기 위한 도면이다.

[0076] 포즈 추정부(140)는, 제1 임계치를 이용하여 제1 임계치에 해당하는 대표 포즈가 선택 실루엣에 해당되는지 여부를 판단하고, 제1임계치에 해당하는 대표 포즈가 선택 실루엣에 해당되지 않을 경우, 제2 임계치를 이용하여 이하에 분기된 자식 노드에 위치하는 복수의 대표 포즈 중에서 유사도를 판단할 하나의 대표 포즈를 정한다.

[0077] 도 8에 도시된 바와 같이, 의사 결정 트리에서 루트 노드(n^1), 제2행의 좌측 자식 노드(n^2) 및 제 2행의 우측 자식 노드(n^3)에 대표 포즈가 배치되어 있다.

[0078] 각각의 대표 포즈에는 할당된 파라미터 값이 두 개씩 있는데, 첫째 파라미터는 제1임계치인 매칭값(matching value)이며, 둘째 파라미터는 제2 임계치인 중앙값(center value)이다. 여기서, 매칭값은 비교되는 송신자 실루엣과 대표 포즈의 유사 여부를 판단하는 임계치를 나타낸다. 즉, 대표 포즈의 포즈 유사도가 1인 경우, 대표 포즈는 비교 대상인 송신자 실루엣과 완전 동일한 경우에만 선택되고, 반대로 대표 포즈의 포즈 유사도가 0인 경우, 대표 포즈는 비교 대상인 송신자 실루엣과의 차이와 무관하게 선택될 수 있다. 그리고, 중앙값은 좌측 자식 노드 및 우측 자식 노드를 찾는 기준을 나타낸다. 여기서, 중앙값은 의사 결정 트리가 완성되면 자동으로 결정된다.

[0079] 도 8에서, 루트 노드의 대표 포즈의 매칭값은 0.93, 중앙값은 0.64이고, 좌측 자식 노드의 매칭값은 0.80, 중앙값은 0.47이고, 우측 자식 노드의 매칭값은 0.82, 중앙값은 0.64이다.

[0080] 도 8을 참고하여 사용자 실루엣과 가장 유사도가 높은 대표 포즈를 추정하여 보면 다음과 같다. 먼저, 얼굴의 영역의 위치를 고려하여 분류된 사용자 실루엣에서 전송 대상이 될 포즈 인덱스의 사용자 실루엣(이하 전송 실루엣)을 하나 선택한다. 다음으로, 전송 실루엣과 루트 노드의 대표 포즈와의 유사도를 측정한다. 여기서, 전송 실루엣과 루트 노드의 대표 포즈 간의 유사도가 루트 노드의 매칭값인0.98보다 크면 루트 노드의 대표 포즈를 선택하고, 0.98보다 작거나 같으면 자식 노드로 내려온다.

[0081] 자식 노드에서 상기 측정된 유사도가 루트 노드의 중간값인 0.64보다 작거나 같으면 좌측 자식 노드의 대표 포즈에 대해 판단하고, 0.64보다 크면 우측 자식 노드의 대표 포즈에 대해 판단한다. 다음으로, 좌측 대표 포즈 또는 우측 대표 포즈에 대해서 상기의 방법을 동일하게 적용하여 새로 구한 유사도를 자식 노드의 대표 포즈의 매칭값과 비교한다. 경우에 따라서는, 임의의 자식 노드의 대표 포즈를 선택하거나 의사 결정 트리 최종의 자식 노드의 대표 포즈를 선택할 수도 있다.

[0082] 이상과 같은 탐색 결과 획득된 노드에 저장된 대표 포즈의 인덱스를 송신자의 행동에 대응하는 대표 포즈의 인덱스로 결정한다.

[0083] 다시 도 2로 돌아와서 메시지 생성부(160)는 반환된 선택 실루엣의 인덱스를 포함하는 멀티 미디어 메시지를 생성한다(S250).

[0084] 송신자가 사용하는 화상 통신 장치에 따라 메시지 생성부(160)에서 발생되는 데이터의 종류는 매우 다양해질 수 있다. 즉, 생성되는 메시지는 선택 실루엣의 인덱스뿐만 아니라 텍스트 입력부(180)로 입력되는 텍스트, 음성 입력부(190)로 입력되는 음성을 포함할 수 있다.

[0085] 그리고, 메시지 전송부(170)는 생성된 메시지를 수신자 단말로 전송한다(S260).

[0086] 한편, 메시지를 수신한 수신자 단말이 대표 포즈에 대응하는 송신자의 액션을 추정하는데, 이하에서는 수신자 단말의 송신자 액션을 추정하는 방법에 대하여 설명한다.

[0087] 즉, 수신자 단말은 선택 포즈의 인덱스를 포함하는 수신 메시지를 기초로 다음 수학식 2의 행위 네트워크(behavior network)를 이용하여 상기 대표 포즈에 대응하는 송신자의 액션을 추정할 수 있다.

수학식 2

$$a^j \in A,\ A = \{a_1, a_2, \ldots\}$$

$$a^j = < p^i, d^j, c_1^j, c_2^j, \ldots >$$

[0088]

[0089] 여기서, a^j 는 j번째 대표 포즈로 추정되는 j번째 액션, p^j 는 j번째 대표 포즈, d^j 는 j번째 대표 포즈의 지속 시간, c^j_n 는 상기 가상 객체가 j번째 액션인 a^j 수행시에 요구되는 제n 조인트 각도, 즉 팔꿈치의 관절 각도를 나타낸다.

[0090] 수신자 단말에서 수신된 메시지를 수신자 단말의 표시장치(미도시)를 이용하여 표시하기 위해서, 미리 가상 객체(virtual human)에 의해 수행될 액션들이 정의된다. 포즈 인덱스(pose index)가 수학식 2와 같이 수신된 경우, 액션은 가상 객체가 포즈를 표현하는 가상 객체의 움직임이다.

[0091] 액션이 정의되면, 상기 수학식에 기초하여 포즈 인덱스가 수신될 때마다 액션을 선택하고 연속적으로 액션을 수행하기 위해 행위 네트워크(behavior network)가 정의된다.

[0092] 도 9은 본 발명의 실시예에 따른 행위 네트워크(behavior network)를 이용하여 액션 노드(action node)를 선택하는 과정을 설명하기 위한 예시 도면이다.

[0093] 수신자 단말은 송신자의 액션을 포함하는 액션 노드를 하기의 수학식 3의 행위 네트워크(behavior network)를 이용하여 선택한다.

[0094] 수신자 단말내에 포함된 행위 계획부(behavior planner)는 정의 단계에서 연속적인 액션을 수행하는 행위 네트워크(behavior network)를 다음과 같이 정의한다.

[0095] 첫째, 시작 포즈(start-pose) 및 목표 포즈(goal-pose)를 선정한다. 시작 포즈는 연속적인 액션을 생성하는 시작점이 되는 포즈이다. 목표 포즈는 목적이 되는 포즈이다. 생성된 연속적인 액션 중에서 마지막 액션을 위한 포즈가 목표 포즈가 된다. 따라서, 도 9에서 집합 P의 모든 포즈가 네트워크의 양쪽에 위치된다.

[0096] 다음으로, 집합 A의 모든 액션이 배치된다. 그리고 DAG(directed acyclic graph)를 사용한 트리로서 연속적인 액션의 시퀀스가 표현된다. 집합 A에서 하나의 액션을 포함하는 액션 노드가 최초로 정의된다. 그리고 도 9과 같이 액션 노드(action node)는 시작 포즈(start pose)와 목표 포즈(goal pose) 사이에 위치된다. 액션 노드가 조정되고, DAG가 각각의 액션 노드를 연결한다. 트리의 루프 형성을 방지하기 위하여, 도 9과 같이 특징적인 액션의 액션 노드가 액션 노드 a1과 같이 계속적으로 정의된다.

[0097] 다음으로, 모든 DAG의 전이 확률(transition probbility)이 정의된다. 각각의 액션에서 다른 액션으로의 전이 확률의 합은 100으로 일반화된다.

[0098] 마지막으로, 포즈 인덱스 수신 후, 시작 포즈는 처음으로 수행될 수 있는 액션에 연결된다. 목표 포즈는 또한 마지막으로 수행될 수 있는 액션에 연결된다. 도 9에서, 시작 포즈(p_1, p_2)는 각각의 대응 액션인 a_1 및 a_2에 연결되어 있다. 목표 포즈 중에서 p_1은 하나 이상인, 두 개의 액션 노드(m_1 및 m_2)에 연결된다. 두 개의 노드의 두 개의 액션은 최종적으로 수행된다. 따라서, 두 노드는 목표 포즈(p1)에 연결된다. p_2의 경우, a_2를 포함하는 두 개의 액션 노드 사이의 하나의 노드만이 연결된다. 네트워크를 구성하는 액션 노드는 다음 수학식 3과 같이 정의된다.

수학식 3

$$m^k = < a^j, s^k, g^k, o_1^k, o_2^k, \ldots, q_1^k, q_2^k >$$

[0099]

[0100] 여기서, m^k 는 a^j를 포함하는 액션 노드, s^k 및 g^k 는 액션 노드(m^k)와 연결된 시작 포즈 및 목표 포즈의 인덱스, o^k_x 는 액션노드(m^k)가 x 번째로 연결된 다른 액션 노드, q^k_x 는 o^k_x 로의 전이 확률을 각각 나타낸다.

[0101] 액션 노드의 네트워크 구성은 행위 네트워크(behavior network)로 정의된다.

[0102] 행위 네트워크(behavior network)는 구현 단계에서 하나의 액션이 선택되는 경우에 사용된다. (t-1) 시간에 수신된 포즈 인덱스를 시작 포즈로 사용하고, (t) 시간에 수신된 포즈 인덱스를 목표 포즈의 인덱스로 사용함으로써, 행위 계획부는 다음과 같이 연속적인 액션을 생성한다.

[0103] 첫째, 수개의 시작 포즈 중에서, (t-1) 시간에 수신된 포즈 인덱스의 포즈가 선택된다. 다음으로, 수개의 목표 포즈 중에서, (t) 시간에 수신된 포즈 인덱스의 포즈가 선택된다. 다음으로, 선택된 시작 포즈에서 선택된 목표 포즈로 이동이 가능한 모든 액션 및 연결이 선택된다. 다음으로, 각각의 액션 노드 이외의 선택된 노드에 대한 전이 확률이 100으로 일반화된다. 다음으로, 선택된 연결 중에서, 하나의 연결이 확률을 통해 선택된다. 하나의 액션 노드만이 연결되는 경우에는 액션 노드는 직접 선택될 수 있다. 연결을 선택하는 작업은 액션 노드가 목표 포즈에 도달할 때까지 반복해서 계속된다. 마지막으로, 행위는 모든 방문 액션 노드(visiting action node)를 연결함으로써 완성되고, 하나의 행위로서 정의된다. 다음으로 상기 정의된 행위는 수행된다.

[0104] 예를 들면, 도 9에서 포즈 인덱스(인덱스 2)가 (t-1) 시간 및 (t) 시간에 수신된 경우, 상기의 연결 행위가 생성 된다.

[0105] 각각의 m_4 및 m_2는 시작 포즈 및 목표 포즈로 연결되면서 활성화 된다. 다른 활성화된 두 개의 액션 노드인 m_7 및 m 은 m_4 에서 m_6의 연결에 존재한다. 상기 행위 네트워크(behavior network)로부터, 액션 노드는 m_4로부터의 연결에 따라 m_4-m_7-m_6 또는 m_4-m-m_6 연결에 따라 방문된다. 따라서, a_2-a_3-a_2 또는 a_2-a-a_2 순서를 갖는 행위가 생성되고 수행된다.

[0106] 이상의 단계들은 적어도 하나의 프로세서를 통해 수행될 수 있다. 이러한 프로세서는 일련의 데이터 베이스를 자동으로 구축하고, 수신된 대표 포즈의 인덱스를 이용하여 데이터 베이스들로부터 가상 객체에 대한 일련의 행위 계획을 생성하고 재생한다. 나아가 이상의 수신 장치는 상기된 프로세서가 일련의 연산을 수행하는 과정에서 연산 그 자체 또는 임시 저장을 위한 공간으로서 사용되는 메모리를 더 포함할 수 있다. 물론, 이들 연산들을 상기된 프로세서와 메모리를 통해 수행하기 위해 부가적인 소프트웨어 코드가 활용될 수 있음은 당연하다.

[0107] 한편, 본 발명의 실시예들은 컴퓨터로 읽을 수 있는 기록 매체에 컴퓨터가 읽을 수 있는 코드로 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다.

[0108] 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현하는 것을 포함한다. 또한, 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다. 그리고 본 발명을 구현하기 위한 기능적인 프로그램, 코드 및 코드 세그먼트들은 본 발명이 속하는 기술 분야의 프로그래머들에 의해 용이하게 추론될 수 있다.

[0109] 이와 같이 본 발명의 실시예에 따른 얼굴 영역의 위치를 고려한 화상 통신 장치 및 그 방법에 따르면, 사용자의 얼굴 영역의 위치를 고려하여 실제 영상에 대응하는 대표 포즈의 인덱스 값을 수신자에게 전송함으로써 가상 객체에 의해 표현할 송신자의 행동을 탐색하는 알고리즘이 간단해져서 탐색 시간을 줄일 수 있다. 또한, 행위 네트워크(behavior network)를 통해 가상 객체에 의한 송신자의 행동을 실제에 가깝게 구현하여 부드럽고 자연스럽게 수신자 단말에서 구현할 수 있다.

[0110] 이제까지 본 발명에 대하여 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 따라서 본 발명의 범위는 전술한 실시예에 한정되지 않고 특허청구범위에 기재된 내용 및 그와 동등한 범위 내에 있는 다양한 실시 형태가 포함되도록 해석되어야 할 것이다.

부호의 설명

[0111] 100: 얼굴 영역의 위치를 고려한 화상 통신 장치,

110: 영상 입력부, 120: 얼굴 영역 검출부,

130: 실루엣 추출부, 140: 포즈 추정부,

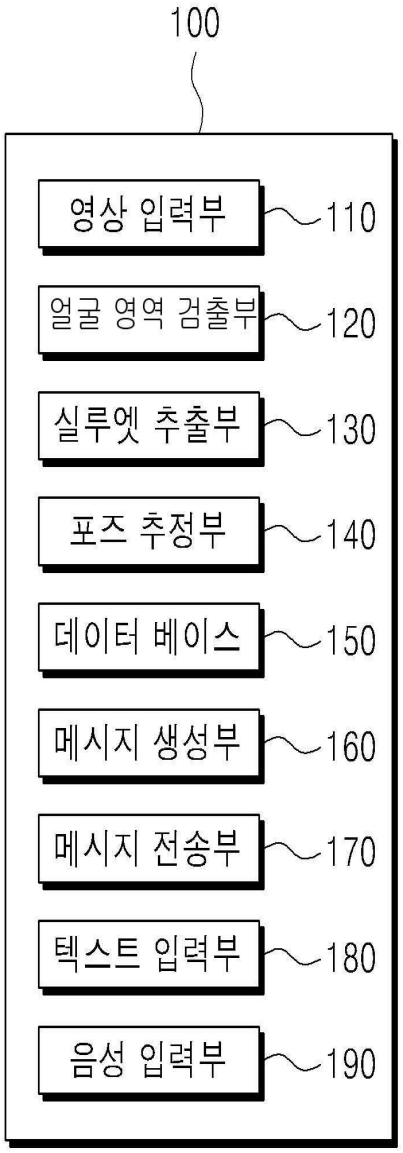
150: 데이터 베이스, 160: 메시지 생성부,

170: 메시지 전송부, 180: 텍스트 입력부,

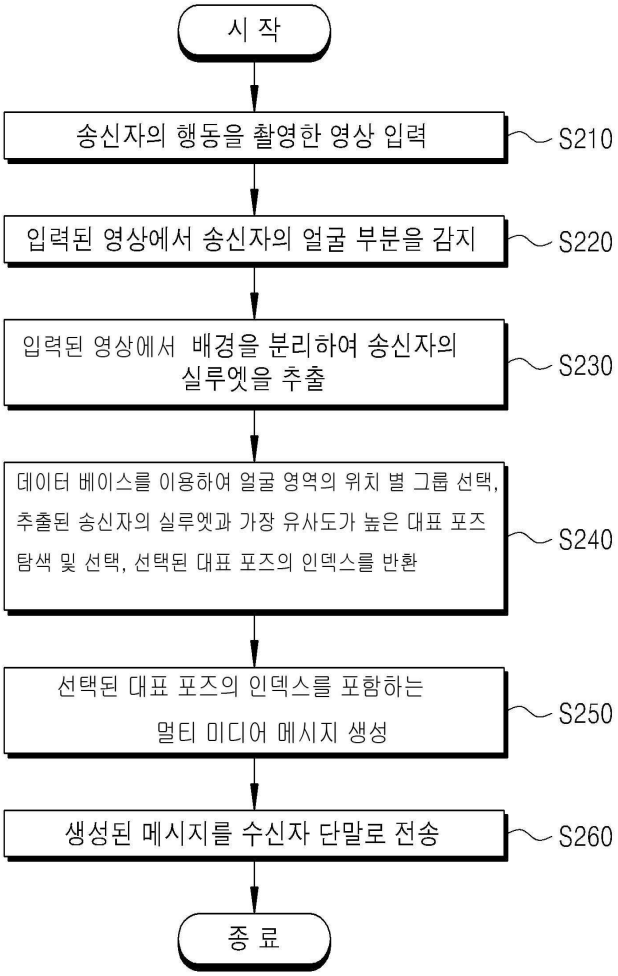
190: 음성 입력부

도면

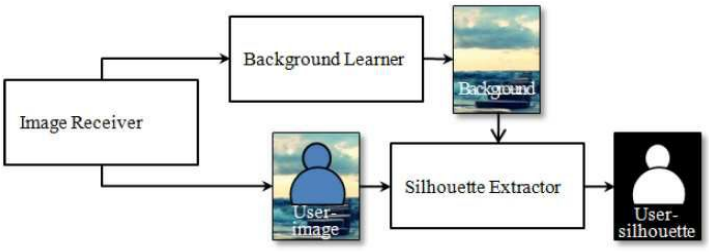
도면1



도면2


시 작
송신자의 행동을 촬영한 영상 입력
S210
입력된 영상에서 송신자의 얼굴 부분을 감지
S220
입력된 영상에서 배경을 분리하여 송신자의 실루엣을 추출
S230
데이터 베이스를 이용하여 얼굴 영역의 위치 별 그룹 선택, 추출된 송신자의 실루엣과 가장 유사도가 높은 대표 포즈 탐색 및 선택, 선택된 대표 포즈의 인덱스를 반환
S240
선택된 대표 포즈의 인덱스를 포함하는 멀티 미디어 메시지 생성
S250
생성된 메시지를 수신자 단말로 전송
S260
종 료

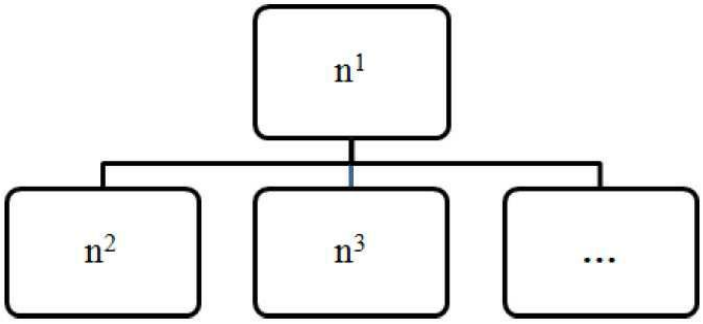
도면3


Background Learner
Background
Image Receiver
User-image
Silhouette Extractor
User-silhouette

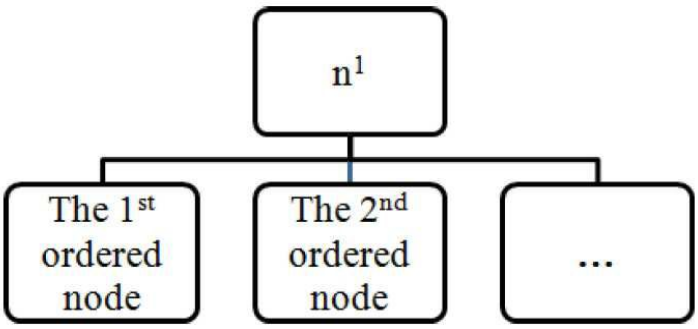
도면4

```
FUNCTION MaximinSelection with C, actionCount
BEGIN
SET B = C
SET T = {b¹}
SET B = B - {b¹}
FOR actionIndex = 2 TO actionCount
    SET max = 0
    FOR b in B
        SET min = ∞
        FOR t in T
            IF min≥EuclideanDistance(t, b) THEN
                SET minElement = b
                SET min = EuclideanDistance(t, b)
            END IF
        END FOR
        IF max ≤ min THEN
            SET max = min
            SET maxElement = minElement
        END IF
    END FOR
    SET B = B - {maxElement} SET T = T + {maxElement}
END FOR
RETURN TEND
```

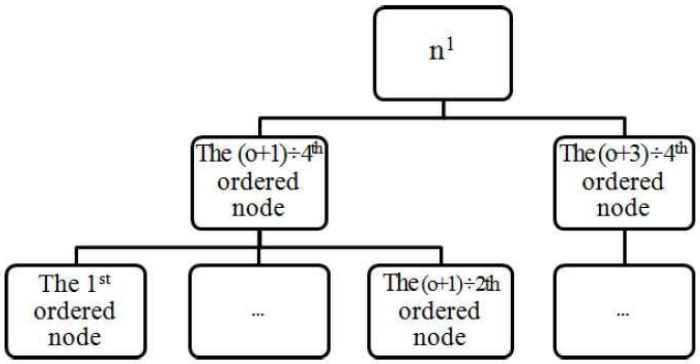
도면5



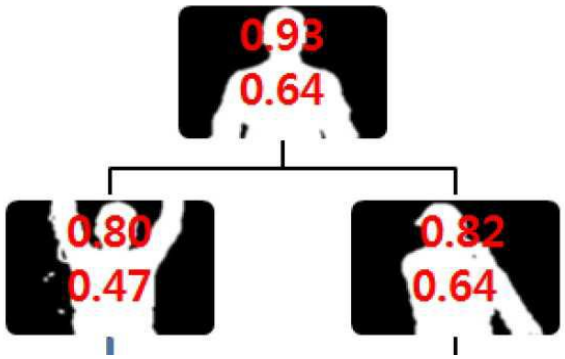
도면6



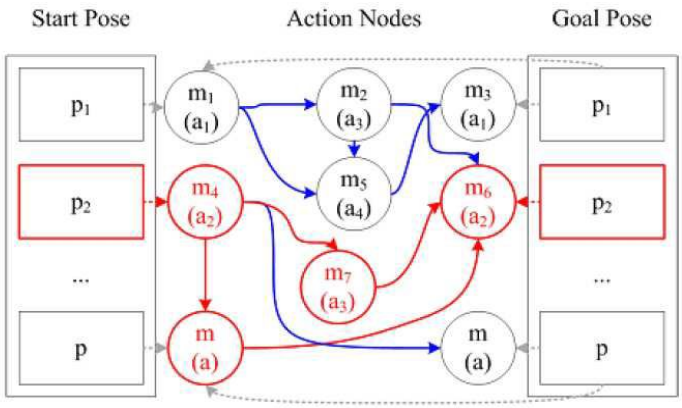
도면7


n1
The (o+1)÷4th ordered node
The (o+3)÷4th ordered node
The 1st ordered node
...
The (o+1)÷2th ordered node
...

도면8


0.93
0.64
0.80
0.47
0.82
0.64

도면9


Start Pose
Action Nodes
Goal Pose
p1
p2
...
p
m1 (a1)
m2 (a3)
m3 (a1)
m4 (a2)
m5 (a4)
m6 (a2)
m7 (a3)
m (a)
m (a)
p1
p2
...
p