



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0040981  
(43) 공개일자 2015년04월15일

(51) 국제특허분류(Int. Cl.)  
H04N 7/15 (2006.01) H04L 29/06 (2006.01)  
H04N 7/14 (2006.01)  
(52) CPC특허분류  
H04N 7/152 (2013.01)  
H04L 65/1093 (2013.01)  
(21) 출원번호 10-2015-7005234  
(22) 출원일자(국제) 2013년06월05일  
심사청구일자 2015년02월27일  
(85) 번역문제출일자 2015년02월27일  
(86) 국제출원번호 PCT/EP2013/061544  
(87) 국제공개번호 WO 2014/032823  
국제공개일자 2014년03월06일  
(30) 우선권주장  
12182267.0 2012년08월29일  
유럽특허청(EPO)(EP)

(71) 출원인  
알까멜 루슨트  
프랑스 92100 불론뉴-비영꾸르 루뜨 들 라 렌느  
148/152  
(72) 발명자  
매릴리 엠마뉴엘  
프랑스 에프-91620 노자이 루트 드 빌쥐스메 상트  
르 드 빌라르소 알 벨 랩스 프랑스 내  
미호우브 알라에딘  
프랑스 에프-91620 노자이 루트 드 빌쥐스메 상트  
르 드 빌라르소 알까멜-루슨트 벨 랩스 프랑스 내  
아우타가르츠 압텔카멜  
프랑스 에프-91620 노자이 루트 드 빌쥐스메 상트  
르 드 빌라르소 알까멜-루슨트 벨 랩스 프랑스 내  
(74) 대리인  
장훈

전체 청구항 수 : 총 15 항

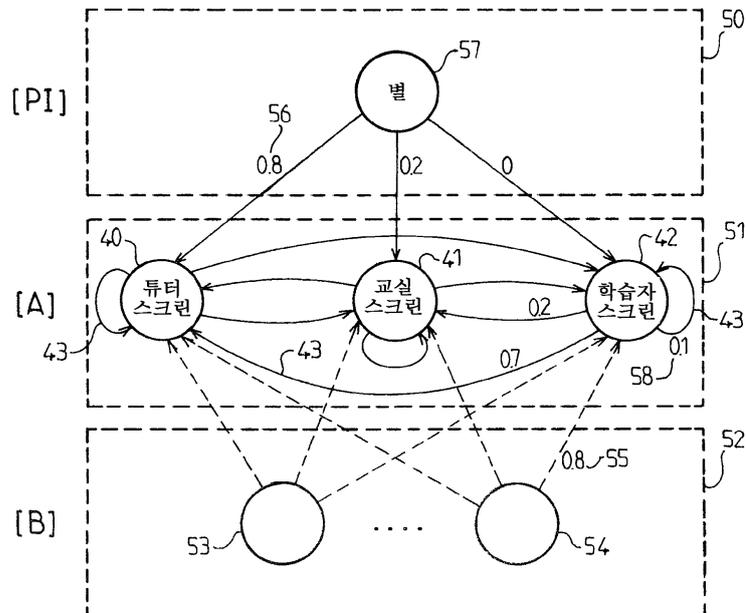
(54) 발명의 명칭 조직화 모델들을 구현하는 화상 회의 시스템들

(57) 요약

화상 회의시 출력 비디오 스트림을 생성하기 위한 방법은 화상 회의의 복수의 입력 비디오 스트림들을 수신하는 단계, 일련의 관찰 이벤트들(52, 53, 54)을 수신하는 단계로서, 관찰은 화상 회의의 참여자에 의해 행해진 동작들에 대응하는, 상기 관찰 이벤트들을 수신하는 단계, 복수의 조직화 모델들을 제공하는 단계, 조직화 모델들의

(뒷면에 계속)

대표도 - 도5



각각에 대하여, 수신된 일련의 관찰 이벤트들의 확률을 결정하는 단계, 가장 높은 확률에 대응하는 조직화 모델을 선택하는 단계, 선택된 조직화 모델을 사용하여: · 디스플레이 상태(51, 40, 41, 42)를 후보 디스플레이 상태로 선택하는 단계, · 수신된 일련의 관찰 이벤트들에 대하여 후보 디스플레이 상태의 조건부 확률을 결정하는 단계, · 가장 높은 조건부 확률을 갱신된 디스플레이 상태로서 제공하는 후보 디스플레이 상태를 결정하는 단계, 및 · 현재 디스플레이 상태 및 갱신된 디스플레이 상태를 포함하는 비디오 스트림을 생성하는 단계를 수행하기 위한 단계를 포함한다.

(52) CPC특허분류

*H04L 65/4038* (2013.01)

*H04N 7/147* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

화상 회의에서 출력 비디오(12) 스트림을 생성하는 방법에 있어서,

상기 화상 회의의 복수의 비디오 스트림들(11)을 수신하는 단계;

일련의 관찰 이벤트들(33)을 수신하는 단계로서, 상기 관찰 이벤트들은 상기 화상 회의의 참가자들에 의한 동작 모드에 대응하는 복수의 관찰가능한 동작들에 속하는, 상기 일련의 관찰 이벤트들(33)을 수신하는 단계,

복수의 조직화 모델들을 제공하는 단계로서, 각각의 모델은:

일 세트의 디스플레이 상태들(51)로서, 각각은 미리 결정된 스크린 템플릿을 포함하고, 각각의 스크린 템플릿은 상기 입력 비디오 스트림들의 선택된 서브세트를 포함하는, 상기 일 세트의 디스플레이 상태들(51),

상기 디스플레이 상태들 사이의 이동 확률들(43), 및

관찰가능한 동작들의 조건부 확률들을 상기 디스플레이 상태들의 함수로서 나타내는 관찰 확률들(55)을 포함하는, 상기 복수의 조직화 모델들을 제공하는 단계,

상기 조직화 모델들의 각각에 대하여, 수신된 상기 일련의 관찰 이벤트들의 확률을 결정하는 단계,

가장 높은 확률에 대응하는 조직화 모델을 선택하는 단계,

상기 선택된 조직화 모델을 사용하는 단계로서,

상기 조직화 모델의 각각의 디스플레이 상태(51)에 대하여, 상기 디스플레이 상태를 후보 디스플레이 상태로서 선택하는 단계,

과거 디스플레이 상태들 및 현재 디스플레이 상태를 포함하는 일 시퀀스의 디스플레이 상태들을 고려하여 상기 수신된 일련의 관찰 이벤트들에 대한 상기 후보 디스플레이 상태의 조건부 확률을 결정하는 단계;

가장 높은 조건부 확률을 제공하는 상기 후보 디스플레이 상태를 갱신된 디스플레이 상태로서 결정하는 단계, 및

상기 현재 디스플레이 상태에 연관된 상기 스크린 템플릿을 나타내는 제 1 시퀀스의 이미지들 및 상기 갱신된 디스플레이 상태에 연관된 상기 스크린 템플릿을 나타내는 제 2 시퀀스의 이미지들을 차례로 포함하는 비디오 스트림(12)을 생성하는 단계를 수행하기 위해, 상기 선택된 조직화 모델을 사용하는 단계를 포함하는, 화상 회의에서 출력 비디오 스트림을 생성하는 방법.

#### 청구항 2

제 1 항에 있어서,

상기 관찰 가능한 동작들은 제스처들, 머리 움직임들, 얼굴 표정들, 오디오 동작들, 키워드들의 발음, 프레젠테이션 슬라이드들에 관련된 동작들로 구성된 동작 카테고리들의 그룹에서 선택되는, 화상 회의에서 출력 비디오 스트림을 생성하는 방법.

#### 청구항 3

제 1 항에 있어서,

상기 관찰 가능한 동작들은:

손가락 들기, 손 들기,

머리 상하 움직임, 머리 좌우 움직임,

말하는 것 또는 잠자는 것에 대응하는 얼굴 표정 짓기,

소리내기, 침묵하기, 튜터에 의한 말하기, 참여자에 의한 말하기,  
청취자 또는 자막의 명칭을 발음하기,  
슬라이드 스위칭, 포인터 이동,  
질문 시작, 질문 종료,로 구성된 그룹에서 선택되는, 화상 회의에서 출력 비디오 스트림을 생성하는 방법.

#### 청구항 4

제 1 항에 있어서,

상기 입력 비디오 스트림들은: 개별적인 참여자들의 뷰들(703), 화자의 뷰들(701), 회의실의 뷰들(702), 및 프레젠테이션 슬라이드들의 뷰들로 구성된 그룹에서 선택되는, 화상 회의에서 출력 비디오 스트림을 생성하는 방법.

#### 청구항 5

제 1 항에 있어서,

스크린 템플릿(44)은 상기 대응하는 서브세트에 속하는 상기 입력 비디오 스트림들의 미리 규정된 정렬을 포함하는, 화상 회의에서 출력 비디오 스트림을 생성하는 방법.

#### 청구항 6

제 1 항에 있어서,

상기 이동 확률들은 이동 행렬로서 정렬되는, 화상 회의에서 출력 비디오 스트림을 생성하는 방법.

#### 청구항 7

제 1 항에 있어서,

관찰 확률들은 방출 행렬로서 정렬되는, 화상 회의에서 출력 비디오 스트림을 생성하는 방법.

#### 청구항 8

화상 회의에서 출력 비디오 스트림을 생성하기 위한 화상 회의 제어 디바이스에 있어서,

상기 화상 회의의 복수의 입력 비디오 스트림들(11)을 수신하는 수단,

일련의 관찰 이벤트들(33)을 수신하는 수단으로서, 상기 관찰 이벤트들은 상기 화상 회의의 참여자들에 의해 행해진 동작들에 대응하는 복수의 관찰 가능한 동작들(52)에 속하는, 상기 수신 수단,

복수의 조직화 모델들을 저장하는 데이터 저장소(37)로서, 각각의 모델은:

일 세트의 디스플레이 상태들(51)로서, 각각은 미리 규정된 스크린 템플릿과 연관되고, 각각의 스크린 템플릿은 상기 입력 비디오 스트림들 중 선택된 서브세트를 포함하는, 상기 일 세트의 디스플레이 상태들(51),

상기 디스플레이 상태들 사이의 이동 확률들(43),

상기 관찰 가능한 동작들의 조건부 확률을 상기 디스플레이 상태들의 함수로서 나타내는 관찰 확률들(55)을 포함하는, 상기 데이터 저장소(37),

상기 조직화 모델들의 각각에 대하여, 수신된 상기 일련의 관찰 이벤트들의 확률을 결정하기 위한 수단,

가장 높은 확률에 대응하는 조직화 모델을 선택하기 위한 수단, 및

상기 선택된 조직화 모델을 사용하는 수단으로서:

상기 조직화 모델의 각각의 디스플레이 상태(51)에 대하여, 상기 디스플레이 상태를 후보 디스플레이 상태로서 선택하는 단계,

과거 디스플레이 상태들 및 현재 디스플레이 상태를 포함하는 일 시퀀스의 디스플레이 상태들을 고려하여 상기 수신된 일련의 관찰 이벤트들에 대하여 상기 후보 디스플레이 상태의 조건부 확률을 결정하는 단계,

가장 높은 조건부 확률을 제공하는 상기 후보 디스플레이 상태를 갱신된 디스플레이 상태로서 결정하는 단계, 및

상기 현재 디스플레이 상태와 연관된 상기 스크린 템플릿을 나타내는 제 1 시퀀스의 이미지들 및 상기 갱신된 디스플레이 상태와 연관된 상기 스크린 템플릿을 나타내는 제 2 시퀀스의 이미지들을 차례로 포함하는 비디오 스트림(12)을 생성하는 단계를 수행하기 위해, 상기 선택된 조직화 모델을 사용하는 수단을 포함하는, 화상 회의 제어 디바이스.

#### 청구항 9

제 8 항에 있어서,

상기 관찰 가능한 동작들은 제스처들, 머리 움직임, 얼굴 표정들, 오디오 동작들, 키워드들의 발음, 프레젠테이션 슬라이드들에 관련된 동작들로 구성된 동작 카테고리들의 그룹에서 선택되는, 화상 회의 제어 디바이스.

#### 청구항 10

제 8 항에 있어서,

상기 관찰 가능한 동작들은,

손가락 들기, 손 들기,

머리 상하 움직임, 머리 좌우 움직임,

말하는 것 또는 잠자는 것에 대응하는 얼굴 표정 짓기,

소리내기, 침묵하기, 튜터에 의한 말하기, 참여자에 의한 말하기,

청취자 또는 자막의 명칭을 발음하기,

슬라이드 스위칭, 포인터 이동,

질문 시작, 질문 종료,로 구성된 그룹에서 선택되는, 화상 회의 제어 디바이스.

#### 청구항 11

제 8 항에 있어서,

상기 입력 비디오 스트림들은: 개별적인 참여자들의 뷰들(703), 화자의 뷰들(701), 회의실의 뷰들(702), 및 프레젠테이션 슬라이드들의 뷰들로 구성된 그룹에서 선택되는, 화상 회의 제어 디바이스.

#### 청구항 12

제 8 항에 있어서,

스크린 템플릿(44)은 상기 대응하는 서브세트에 속하는 상기 입력 비디오 스트림들의 미리 규정된 정렬을 포함하는, 화상 회의 제어 디바이스.

#### 청구항 13

제 8 항에 있어서,

상기 이동 확률들은 이동 행렬로서 정렬되는, 화상 회의 제어 디바이스.

#### 청구항 14

제 8 항에 있어서,

관찰 확률들은 방출 행렬로서 정렬되는, 화상 회의 제어 디바이스.

#### 청구항 15

통신 네트워크(10)에 의해 복수의 단말들(2)에 접속되는, 제 8 항 내지 제 14 항 중 어느 한 항에 따른 화상 회의 제어 디바이스(1)를 포함하는 화상 회의 시스템에 있어서,

각각의 단말(2)은 입력 비디오 스트림(11)을 생성하기 위한 수단을 포함하고, 상기 통신 네트워크는 상기 단말들로부터 상기 제어 디바이스로 상기 비디오 스트림을 송신하고 상기 제어 디바이스에 의해 생성된 상기 출력 비디오 스트림(12)을 단말로 송신하도록 적응되는, 화상 회의 시스템.

**발명의 설명**

**기술 분야**

[0001] 본 발명은, 특히 다수 참여자 화상 회의 시스템에서, 다수의 사람들의, 다수의 장소들로부터 실감 화상을 생성하기 위한 방법들에 관한 것이다.

**배경 기술**

[0002] 통신 시스템들에서 대역폭 성능들의 증가에 따라, 화상 통신 시스템들은 사업 및 주거용 애플리케이션들 모두에서 점점 대중적이 되었다. 사실, 지리적으로 분산된 팀 협력의 경우에, 이들 시스템들은 팀 협력자들의 이동을 피하고 융통성을 증가시킨다.

[0003] 화상 통신 시스템들은 상이한 장소들의 사람들을 소집하기 위해 오디오 및 비디오 원격 통신들을 사용한다. 이는 개인 사무실들의 사람들 사이의 대화만큼 간단할 수 있거나, 또는 다수의 장소들의 큰 공간들에서 수 개의 다 지점 장소들을 포함할 수 있다. 시스템들은 점 대 점 및 다지점 통신들을 관리할 수 있다.

[0004] 알려진 시스템에서, 사용자들은 시청할 소스(비디오 스트림 또는 카메라)를 원격 제어에 의해 선택한다. 몇몇 시스템들은 이러한 정적 행동을 개선하고 활성 스피커를 자동으로 스위치 온한다. 이러한 동적 행동은 각각의 참여자의 오디오 정보에 기초한다. 솔루션을 고려하면, 개별 호출은 그의 요구들에 가장 알맞는 많은 포맷들 중 하나로부터 템플릿을 선택하는 것을 제안하는 새로운 능력을 사용자에게 도입했다. 이들의 템플릿들은 정적이고 화상 회의 동안 사용자의 몰입/주의를 증가시키는 것을 가능하게 하는 조직화에서 어떠한 역동성도 제공하지 않는다. 사용자에게 대한 화상 조직화의 프로그램 가능성 또는 어느 템플릿이 사용자 요구들에 매우 적합한지를 자동으로 선택하게 하는 지적 메커니즘은 존재하지 않는다. 시스코 솔루션(WebEx 및 Telepresence TX9000)에서, 활성 사용자는 메인 윈도우에 디스플레이된다. 고정된 수의 템플릿들은 화상 조직화에 이용가능하다. 그들 중 하나는 사용자에게 의해 선택된다. 비디오 스위칭 행동은 오디오 이벤트 검출에 의해 관리된다. 연구가 제안되었을 때, 유용한 정보의 약 70 퍼센트가 오디오 이벤트들로부터 누락된다.

**발명의 내용**

**해결하려는 과제**

[0005] 실감 통신을 개선하기 위해, 새로운 기술들은 규칙 엔진 및 규칙 템플릿들에 기초한 조직화기를 포함한다. 제 1 단계에서, 규칙 템플릿들 세트는 전문가에 의해 생성되고 단일 사용자에게 의해 변경되거나 강화될 수 없다.

**과제의 해결 수단**

[0006] 일 실시예에서, 본 발명은 화상 회의에서 출력 비디오 스트림을 생성하기 위한 방법을 제공하고, 상기 방법은:

- [0007] - 화상 회의의 복수의 입력 비디오 스트림들을 수신하는 단계,
- [0008] - 일련의 관찰 이벤트들을 수신하는 단계로서, 상기 관찰 이벤트들은 화상 회의의 참여자들에 의해 행해진 동작들에 대응하는 복수의 관찰 가능한 동작들에 속하는, 상기 일련의 관찰 이벤트들을 수신하는 단계,
- [0009] - 복수의 조직화 모델들을 제공하는 단계로서, 각각의 모델은:
  - [0010] o 일 세트의 디스플레이 상태들로서, 각각이 미리 결정된 스크린 템플릿과 연관되고, 각각의 스크린 템플릿은 입력 비디오 스트림들의 선택된 서브세트를 포함하는, 상기 일 세트의 디스플레이 상태들,
  - [0011] o 디스플레이 상태들 사이의 이동 확률들, 및
  - [0012] o 관찰 가능한 동작들의 조건부 확률들을 디스플레이 상태들의 함수로서 나타내는 관찰 확률들을 포함하는, 상기 복수의 조직화 모델들을 제공하는 단계,
- [0013] - 조직화 모델들의 각각에 대하여 수신된 일련의 관찰 이벤트들의 확률을 결정하는 단계,

- [0014] - 가장 높은 확률에 대응하는 조직화 모델을 선택하는 단계,
- [0015] - 선택된 조직화 모델을 사용하는 단계로서:
- [0016]     o 조직화 모델의 각각의 디스플레이 상태에 대하여, 디스플레이 상태를 후보 디스플레이 상태로서 선택하는 단계,
- [0017]     o 과거 디스플레이 상태들 및 현재 디스플레이 상태를 포함하는 일련의 디스플레이 상태들을 고려하여 수신된 일련의 관찰 이벤트들에 대한 후보 디스플레이 상태의 조건부 확률을 결정하는 단계,
- [0018]     o 가장 높은 조건부 확률을 갱신된 디스플레이 상태로서 제공하는 후보 디스플레이 상태를 결정하는 단계, 및
- [0019]     o 현재 디스플레이 상태와 연관된 스크린 템플릿을 나타내는 제 1 시퀀스의 이미지들 및 갱신된 디스플레이 상태와 연관된 스크린 템플릿을 나타내는 제 2 시퀀스의 이미지들을 차례로 포함하는 비디오 스트림을 생성하는 단계를 수행하기 위한, 상기 선택된 조직화 모델을 사용하는 단계를 포함한다.
- [0020] 실시예들에 따라, 이러한 방법은 이하 피쳐들 중 하나 이상을 포함할 수 있다
- [0021] 방법의 실시예들에서, 관측 가능한 동작들은 제스처들, 머리 움직임들, 얼굴 표정들, 오디오 동작들, 키워드들의 발음, 프레젠테이션 슬라이드들에 관련된 동작들로 구성된 동작 카테고리들의 그룹에서 선택된다.
- [0022] 방법의 실시예들에서, 관측 가능한 동작들은:
- [0023] - 손가락 들기, 손 들기,
- [0024] - 머리 상하 움직임, 머리 좌우 움직임,
- [0025] - 말하는 것 또는 잠자는 것에 대응하는 얼굴 표정 짓기,
- [0026] - 소리내기, 침묵하기, 튜터에 의한 말하기, 참여자에 의한 말하기,
- [0027] - 청취자 또는 자막의 명칭을 발음하기,
- [0028] - 슬라이드 스위칭, 포인터 이동,
- [0029] - 질문 시작, 질문 종료,로 구성된 그룹에서 선택된다.
- [0030] 방법의 실시예들에서, 입력 비디오 스트림들은: 개별적인 참여자들의 뷰들, 화자의 뷰들, 회의실의 뷰들, 및 프레젠테이션 슬라이드들의 뷰들로 구성된 그룹에서 선택된다.
- [0031] 방법의 실시예들에서, 스크린 템플릿은 대응하는 서버세트에 속하는 입력 비디오 스트림들의 미리 규정된 정렬을 포함한다.
- [0032] 방법의 실시예들에서, 이동 확률들은 이동 행렬로서 정렬된다.
- [0033] 방법의 실시예들에서, 관찰 확률들은 방출 행렬로서 정렬된다.
- [0034] 일 실시예에서, 본 발명은 화상 회의에서 출력 비디오 스트림을 생성하기 위한 화상 회의 제어 디바이스를 또한 제공하고, 상기 디바이스는:
- [0035] - 화상 회의의 복수의 입력 비디오 스트림들을 수신하는 수단,
- [0036] - 일련의 관찰 이벤트들을 수신하는 수단으로서, 상기 관찰 이벤트들은 화상 회의의 참여자들에 의해 행해진 동작들에 대응하는 복수의 관찰 가능한 동작들에 속하는, 상기 수신 수단,
- [0037] - 복수의 조직화 모델들을 저장하는 데이터 저장소로서, 각각의 모델은:
- [0038]     o 일 세트의 디스플레이 상태들로서, 각각은 미리 규정된 스크린 템플릿과 연관되고, 각각의 스크린 템플릿은 입력 비디오 스트림들 중 선택된 서버세트를 포함하는, 상기 일 세트의 디스플레이 상태들,
- [0039]     o 디스플레이 상태들 사이의 이동 확률들,
- [0040]     o 관찰 가능한 동작들의 조건부 확률들을 디스플레이 상태들의 함수로서 나타내는 관찰 확률들을 포함하는, 상기 데이터 저장소,

- [0041] - 조직화 모델들의 각각에 대하여, 수신된 일련의 관찰 이벤트들의 확률을 결정하기 위한 수단,
- [0042] - 가장 높은 확률에 대응하는 조직화 모델을 선택하기 위한 수단, 및
- [0043] - 선택된 조직화 모델을 사용하는 수단으로서:
- [0044]     o 조직화 모델의 각각의 디스플레이 상태에 대하여, 디스플레이 상태를 후보 디스플레이 상태로서 선택하는 단계,
- [0045]     o 과거 디스플레이 상태들 및 현재 디스플레이 상태를 포함하는 일 시퀀스의 디스플레이 상태들을 고려하여 수신된 일련의 관찰 이벤트들에 대한 후보 디스플레이 상태의 조건부 확률을 결정하는 단계,
- [0046]     o 가장 높은 조건부 확률을 제공하는 후보 디스플레이 상태를 갱신된 디스플레이 상태로서 결정하는 단계, 및
- [0047]     o 현재 디스플레이 상태와 연관된 스크린 템플릿을 나타내는 제 1 시퀀스의 이미지들 및 갱신된 디스플레이 상태와 연관된 스크린 템플릿을 나타내는 제 2 시퀀스의 이미지들을 차례로 포함하는 비디오 스트림을 생성하는 단계를 수행하기 위한, 상기 선택된 조직화 모델을 사용하는 수단을 포함한다.
- [0048] 실시예들에 따라, 이러한 화상 회의 제어 디바이스는 이하의 특징들 중 하나 이상을 포함할 수 있다.
- [0049] 화상 회의 제어 디바이스의 실시예들에서, 관찰 가능한 동작들은 제스처들, 머리 움직임들, 얼굴 표정들, 오디오 동작들, 키워드들의 발음, 프레젠테이션 슬라이드들에 관련된 동작들로 구성된 동작 카테고리들의 그룹에서 선택된다.
- [0050] 화상 회의 제어 디바이스의 실시예들에서, 관찰 가능한 동작들은:
- [0051] - 손가락 들기, 손 들기,
- [0052] - 머리 상하 움직임, 머리 좌우 움직임,
- [0053] - 말하는 것 또는 잠자는 것에 대응하는 얼굴 표정 짓기,
- [0054] - 소리내기, 침묵하기, 튜터에 의한 말하기, 참여자에 의한 말하기,
- [0055] - 청취자 또는 자막의 명칭을 발음하기,
- [0056] - 슬라이드 스위칭, 포인터 이동,
- [0057] - 질문 시작, 질문 종료,로 구성된 그룹에서 선택된다.
- [0058] 화상 회의 제어 디바이스의 실시예들에서, 입력 비디오 스트림들은 개별적인 참여자들의 뷰들, 화자의 뷰들, 회의실의 뷰들, 및 프레젠테이션 슬라이드들의 뷰들로 구성된 그룹에서 선택된다.
- [0059] 화상 회의 제어 디바이스의 실시예들에서, 스크린 템플릿은 대응하는 서브세트에 속하는 입력 비디오 스트림들의 미리 규정된 정렬을 포함한다.
- [0060] 화상 회의 제어 디바이스의 실시예들에서, 이동 확률들은 이동 행렬로서 정렬된다.
- [0061] 화상 회의 제어 디바이스의 실시예들에서, 관찰 확률들은 방출 행렬로서 정렬된다.
- [0062] 실시예들에서, 본 발명은 또한 통신 네트워크에 의해 복수의 단말들에 접속된 화상 회의 제어 디바이스를 포함하는 화상 회의 시스템을 제공하고, 각각의 단말은 입력 비디오 스트림을 생성하기 위한 수단을 포함하고, 통신 네트워크는 단말들로부터 제어 디바이스로 비디오 스트림을 송신하고 제어 디바이스에 의해 생성된 출력 비디오 스트림을 단말로 송신하도록 적응된다.
- [0063] 일 실시예에서, 본 발명은 복수의 입력 비디오 스트림 및 일련의 입력 관찰 이벤트들을 포함하는 화상 회의에서 비디오 스트림들의 조직화 모델을 생성하는 방법을 또한 제공하고, 상기 관찰 이벤트들은 복수의 관찰 가능한 동작들에 속하고, 조직화 모델은:
- [0064]     o 일 세트의 디스플레이 상태들로서, 각각은 미리 결정된 스크린 템플릿과 연관되고, 각각의 스크린 템플릿은 화상 회의의 입력 비디오 스트림들의 선택된 서브세트를 포함하는, 상기 일 세트의 디스플레이 상태들
- [0065]     o 디스플레이 상태들 사이의 이동 확률들, 및

- [0066] ○ 관찰 가능한 동작들의 조건부 확률들을 디스플레이 상태들의 함수로서 나타내는 관찰 확률을 포함하고,
- [0067] 상기 방법은:
- [0068] - 사용자 입력 인터페이스를 제공하는 단계로서, 상기 사용자 입력 인터페이스는:
- [0069] ○ 모델의 디스플레이 상태들에 연관된 스크린 템플릿들에 따라 정렬된 상기 비디오 스트림들을 디스플레이하기 위한 스크린 템플릿 디스플레이 수단,
- [0070] ○ 현재 관찰 이벤트를 디스플레이하기 위한 관찰 이벤트 디스플레이 수단, 및
- [0071] ○ 사용자가 디스플레이된 미리 규정된 스크린 템플릿들 중에서 하나의 스크린 템플릿을 선택할 수 있게 하기 위한 사용자 선택 수단을 포함하는, 상기 사용자 입력 인터페이스 제공 단계,
- [0072] - 동기화된 방식으로, 스크린 템플릿 디스플레이 수단에 의해 미리 규정된 스크린 템플릿들에 따라 정렬된 입력 비디오 스트림들을 디스플레이하는 단계,
- [0073] - 입력 비디오 스트림들과 동기화된 방식으로, 관찰 이벤트 디스플레이 수단에 의해 현재 관찰 이벤트들은 디스플레이하는 단계,
- [0074] - 시간에서 연속하는 인스턴트들에서 사용자에게 의해 선택된 현재 스크린 템플릿들에 따라, 입력 비디오 스트림들의 디스플레이 동안, 시간에서 상기 연속하는 인스턴트들의 현재 디스플레이 상태들의 하나의 시퀀스를 입력 비디오 스트림들과 동기화된 방식으로 기록하는 단계,
- [0075] - 두 개의 연속하는 디스플레이 상태들 사이에서 각각 발생하는 이동 발생들의 수들을 결정하는 단계로서, 연속하는 디스플레이 상태들은 상이하거나 동일한, 상기 결정 단계,
- [0076] - 이동 발생들의 수들로부터 모든 디스플레이 상태들 사이의 이동 확률들을 결정하는 단계,
- [0077] - 입력 비디오 스트림들의 디스플레이 동안 관찰 가능한 동작들의 각각에 대해 발생하는 관찰 이벤트들의 수들, 각각의 관찰 가능한 동작 및 각각의 디스플레이 상태에 대해 사용되는 상이한 이벤트 카운터, 및 관찰 이벤트가 발생할 때 관찰 이벤트가 선택된 현재 디스플레이 상태의 함수로서 발생할 때마다 선택 및 증분되는 발생 카운터를 결정하는 단계,
- [0078] - 관찰 확률들을 관찰 이벤트들의 수들의 함수로서 결정하는 단계, 및
- [0079] - 데이터 저장소에 조직화 모델을 저장하는 단계를 포함한다.
- [0080] 실시예들에 따라, 이러한 방법은 이하의 특징들 중 하나 이상을 포함할 수 있다.

[0081] 방법의 실시예들에서, 상태(i) 및 상태(j) 사이의 이동 확률( $a_{ij}$ )은 수식 
$$a_{ij} = \frac{occ_{ij}}{\sum_{h=1}^N occ_{ih}}$$
 을 계산함으로써 결정되고,  $a_{ij}$ 는 디스플레이 상태(i)로부터 디스플레이 상태(j)로의 이동의 확률이고,  $occ_{ij}$ 는 디스플레이 상태(i)로부터 디스플레이 상태(j)로의 이동 발생들의 수이고,  $occ_{ih}$ 는 상태(i)로부터 상태(h)로의 이동 발생들의 수이고, N은 디스플레이 상태들의 총 수이다.

[0082] 방법의 실시예들에서, 관찰 확률( $b_{ik}$ )은 식 
$$b_{ik} = \frac{occObs_{ik}}{\sum_{h=1}^M occObs_{ih}}$$
 을 계산함으로써 결정되고,  $b_{ik}$ 는 디스플레이 상태(i)에서 주어진 관찰 가능한 동작(k)의 확률이고,  $occObs_{ik}$ 는 디스플레이 상태(i)에서 발생된 관찰 가능한 동작(k)에 속하는 관찰 이벤트들의 수이고,  $occObs_{ih}$ 는 디스플레이 상태(i)에서 발생된 관찰 가능한 동작(h)에 속하는 관찰 이벤트들의 수이고, M은 관찰 가능한 동작들의 총 수이다.

- [0083] 방법의 실시예들에서, 방법은:
- [0084] - 생성된 조직화 모델과 데이터 저장소에 저장된 미리 규정된 조직화 모델 사이의 거리를 측정하는 단계,
- [0085] - 거리를 임계치와 비교하는 단계를 추가로 포함하고,

- [0086] - 생성된 조직화 모델을 저장하는 단계는 거리가 상기 임계치보다 큰 경우에만 행해진다.
- [0087] 방법의 실시예들에서, 관찰 가능한 동작들은 제스처들, 머리 움직임들, 얼굴 표정들, 오디오 동작들, 키워드들의 발음, 프레젠테이션 슬라이드들에 관련된 동작들로 구성된 동작 카테고리들의 그룹에서 선택된다.
- [0088] 방법의 실시예들에서, 관찰 가능한 동작들은:
  - [0089] - 손가락 들기, 손 들기,
  - [0090] - 머리 상하 움직임, 머리 좌우 움직임,
  - [0091] - 말하는 것 또는 잠자는 것에 대응하는 얼굴 표정 짓기,
  - [0092] - 소리내기, 침묵하기, 튜터에 의한 말하기, 참여자에 의한 말하기,
  - [0093] - 청취자 또는 자막의 명칭을 발음하기,
  - [0094] - 슬라이드 스위칭, 포인터 이동,
  - [0095] - 질문 시작, 질문 종료,로 구성된 그룹에서 선택된다.
- [0096] 방법의 실시예들에서, 입력 비디오 스트림들은: 개별적인 참여자들의 뷰들, 화자의 뷰들, 회의실의 뷰들, 및 프레젠테이션 슬라이드들의 뷰들로 구성된 그룹에서 선택된다.
- [0097] 방법의 실시예들에서, 스크린 템플릿은 대응하는 서브세트에 속하는 입력 비디오 스트림들의 미리 결정된 정렬을 포함한다.
- [0098] 방법의 실시예들에서, 이동 확률들은 이동 행렬로서 정렬된다.
- [0099] 방법의 실시예들에서, 관찰 확률들은 방출 행렬로서 정렬된다.
- [0100] 일 실시예에서, 본 발명은 복수의 입력 비디오 스트림들을 포함하는 화상 회의의 비디오 스트림들 및 일련의 입력 관찰 이벤트들의 조직화 모델을 생성하기 위한 화상 회의 학습 모듈을 또한 제공하고, 상기 관찰 이벤트들은 복수의 관찰 가능한 동작들에 속하고, 상기 조직화 모델은:
  - [0101] o 일 세트의 디스플레이 상태들로서, 각각이 미리 규정된 스크린 템플릿과 연관되고, 각각의 스크린 템플릿은 화상 회의의 입력 비디오 스트림들의 선택된 서브세트를 포함하는, 상기 일 세트의 디스플레이 상태들,
  - [0102] o 디스플레이 상태들 사이의 이동 확률들, 및
  - [0103] o 관찰 가능한 동작들의 조건부 확률들을 디스플레이 상태들의 함수로서 나타내는 관찰 확률들을 포함하고,
- [0104] 화상 회의 학습 모듈은:
  - [0105] - 사용자 입력 인터페이스로서,
    - [0106] o 디스플레이 상태들과 연관된 스크린 템플릿들에 따라 정렬된 상기 비디오 스트림들을 동기화된 방식으로 디스플레이하기 위한 스크린 템플릿들 디스플레이 수단,
    - [0107] o 입력 비디오 스트림들과 동기화된 방식으로 현재 관찰 이벤트를 디스플레이하기 위한 관찰 이벤트들 디스플레이 수단, 및
    - [0108] o 사용자가 디스플레이된 미리 규정된 스크린 템플릿들 중에서 하나의 스크린 템플릿을 선택할 수 있게 하기 위한 사용자 선택 수단을 포함하는, 상기 사용자 입력 인터페이스,
  - [0109] - 시간상 연속하는 인스턴트들에서 사용자 선택 수단을 통해 사용자에게 의해 선택된 현재 스크린 템플릿들에 따라, 입력 비디오 스트림들의 디스플레이 동안, 시간상 상기 연속하는 인스턴트들에서 현재 디스플레이 상태들의 시퀀스를 입력 비디오 스트림들과 동기화된 방식으로, 기록하기 위한 수단,
  - [0110] - 두 개의 연속하는 디스플레이 상태들 사이에 각각 발생된 이동 발생들의 수들을 결정하기 위한 수단으로서, 연속하는 디스플레이 상태들은 상이하거나 동일한, 상기 결정 수단,
  - [0111] - 이동 발생들의 수들로부터 모든 디스플레이 상태들 사이의 이동 확률들을 결정하기 위한 수단,
  - [0112] - 입력 비디오 스트림들의 디스플레이 동안 관찰 가능한 동작들의 각각에 대해 발생된 관찰 이벤트들의 수들, 각각의 관찰 가능한 동작 및 각각의 디스플레이 상태에 대해 사용되는 상이한 이벤트 카운터, 관찰 이벤트가 발

생활 때 관찰 이벤트가 선택된 현재 디스플레이 상태들의 함수로서 발생할 때마다 선택 및 증분되는 발생 카운터를 결정하기 위한 수단,

[0113] - 관찰 확률들을 관찰 이벤트들의 수들의 함수로서 결정하기 위한 수단, 및

[0114] - 관찰 모델을 저장하기 위한 데이터 저장소를 포함한다.

[0115] 실시예들에 따라, 이러한 화상 회의 학습 모듈은 이하의 피쳐들 중 하나 이상을 포함할 수 있다.

[0116] 화상 회의 학습 모듈의 실시예들에서, 상태(i)와 상태(j) 사이의 이동 확률( $a_{ij}$ )은 식 
$$a_{ij} = \frac{occ_{ij}}{\sum_{h=1}^N occ_{ih}}$$
 을 계산함으로써 결정되고,  $a_{ij}$ 는 디스플레이 상태(i)로부터 디스플레이 상태(j)로의 이동의 확률이고,  $occ_{ij}$ 는 디스플레이 상태(i)로부터 디스플레이 상태(j)로의 이동 발생들의 수이고,  $occ_{ih}$ 는 상태(i)로부터 상태(h)로의 이동 발생들의 수이고, N은 디스플레이 상태들의 총 수이다.

[0117] 화상 회의 학습 모듈의 실시예들에서, 관찰 확률( $b_{ik}$ )은 식 
$$b_{ik} = \frac{occObs_{ik}}{\sum_{h=1}^M occObs_{ih}}$$
 을 계산함으로써 결정되고,  $b_{ik}$ 는 주어진 디스플레이 상태(i)에서 발생된 관찰 가능한 동작(k)의 확률이고,  $occObs_{ik}$ 는 디스플레이 상태(i)에서 발생된 관찰 가능한 동작(k)에 속하는 관찰 이벤트들의 수이고,  $occObs_{ih}$ 는 디스플레이 상태(i)에서 발생된 관찰 가능한 동작(h)에 속하는 관찰 이벤트들의 수이고, M은 관찰 가능한 동작들의 총 수이다.

[0118] 화상 회의 학습 모듈의 실시예들에서, 모듈은:

[0119] - 생성된 조직화 모델과 데이터 저장소에 저장된 미리 규정된 조직화 모델 사이의 거리를 측정하기 위한 수단, 및

[0120] - 거리를 임계치와 비교하기 위한 수단을 추가로 포함하고.

[0121] - 데이터 저장소(37)는 거리가 상기 임계치보다 큰 경우에만 생성된 조직화 모델을 저장한다.

[0122] 화상 회의 학습 모듈의 실시예들에서, 사용자 입력 인터페이스는 확인 버튼의 작동에 응답하여 이동 확률들 및 관찰 확률들의 결정을 트리거링하기 위한 확인 버튼을 추가로 포함한다.

[0123] 화상 회의 학습 모듈의 실시예들에서, 관찰 가능한 동작들은 제스처들, 머리 움직임들, 얼굴 표정들, 오디오 동작들, 키워드들의 발음, 프레젠테이션 슬라이드들에 관련된 동작들로 구성된 동작 카테고리들의 그룹에서 선택된다.

### 발명의 효과

[0124] 본 발명은 조직화 모델들을 구현하는 개선된 화상 회의 시스템들을 제공한다.

### 도면의 간단한 설명

[0125] 도 1은 화상 회의 시스템의 개략적인 기능도.

도 2는 도 1의 시스템에서 사용될 수 있는 사용자 단말의 개략도.

도 3은 도 1의 시스템에서 사용될 수 있는 HMM 조직화기의 개략적인 기능도.

도 4는 HMM 모델의 일 실시예에서 상태들 및 상태 이동들의 개략도.

도 5는 관찰 가능한 동작들을 또한 도시하는 도 4의 HMM 모델의 다른 뷰를 도시하는 도면.

도 6은 HMM 조직화기의 다른 실시예의 개략적인 뷰를 도시하는 도면.

도 7은 사용자 학습 인터페이스의 기능도.

도 8은 HMM 조직화기의 다른 실시예의 개략적인 뷰를 도시하는 도면.

도 9는 HMM 조직화기의 다른 실시예의 개략적인 뷰를 도시하는 도면.

**발명을 실시하기 위한 구체적인 내용**

- [0126] 본 발명의 이들 및 다른 양태들은 도면들을 참조하여 예로서, 이후 기술된 실시예들을 참조하여 명확해지고 명백해질 것이다.
- [0127] 이하에 기술된 화상 회의 시스템은 강화된 실감 통신 경험을 제공하기 위해 다수의 소스들을 다룰 수 있다.
- [0128] 화상 회의 시스템은 적어도 두 개의 위치들 사이의 오디오 및 화상 통신들을 공유할 수 있는 원격 통신 시스템이다. 개별적인 위치들에서 사람들 사이의 이러한 실시간 접촉은 사회적 상호 작용을 증가시킨다. 도 1을 참조하면, 화상 회의 시스템의 일 실시예는 본 명세서에서 조직화기로 칭하는 비디오 제어기(1) 및 복수의 단말들(2)을 포함한다. 이들 단말들은 통신 네트워크(10)에 의해 조직화기에 접속된다. 통신 네트워크는 오디오 및 비디오 스트림들을 송신하도록 적응된다. 본 문맥에서, 조직화기는 단말들(2)에 의해 전송된 상이한 실시간 입력 비디오 스트림들(11)을 관리할 수 있다. 출력 비디오를 생성하기 위해, 상이한 믹싱 방법들이 존재한다. 이러한 개시는 조직화기에 의해 실행된 동적 믹싱 방법을 제안한다. 해결책은 화상 회의에 참여하는 사람들의 상이한 카메라로부터 유래한 멀티미디어 스트림들 및 상이한 비디오 분석자(32s) 및 메타데이터 수집자로부터 유래한 비디오 이벤트들의 메타데이터를 입력들로서 수신한다. 조직화기는 조직화 모델들 및 스크린 템플릿들(44)에 따라 입력 비디오 스트림들(11)을 믹싱하고 이를 단말들(2)에 전송하는 하나 이상의 출력 비디오 스트림들(12)을 생성한다.
- [0129] 화상 회의 시스템에서, 단말들(2)은 동일한 빌딩에서 또는 전세계를 통해 상이한 장소들에 위치된다. 실감 화상 회의를 생성할 수 있기 위해서, 각각의 단말(2)은 몇몇 캡처 수단을 포함한다. 도 2를 참조하면, 단말(2)은 카메라(21) 및 마이크폰(22)과 같은 오디오 및 비디오 캡처 수단을 포함한다. 이들 수단들은 각각의 입력 비디오 스트림(11)을 생성하기 위해 사용된다. 단말(2)은 조직화기(1)에 의해 생성된 출력 비디오 스트림(12)을 시청하기 위한 디스플레이(23)를 또한 포함한다.
- [0130] 도 3을 참조하면, 조직화기(1)는 히든 마르코프 모델들(HMM)에 기초하여 특정 기능들(예를 들면, 학습 메커니즘들, 시나리오 인식...)을 수행한다.
- [0131] 조직화기(1)는 입력들로서:
- [0132] - 예를 들면, 카메라들/웹캠들로부터 유래하는 비디오 스트림들(11), 및
  - [0133] - 예를 들면, 비디오 및 오디오 분석자(32s) 출력들 또는 메타데이터 수집자로부터 유래한 비디오 및 오디오 이벤트들의 메타데이터를 취한다.
- [0134] 입력 비디오 스트림들(11)은 또한 분석자(32)에 송신된다. 비디오 분석자(32)는 제스처들, 자세들, 얼굴들과 같은 비디오 이벤트들을 검출한다. 오디오 분석자(32)는 누가 말하고 있는지, 키워드들, 칩목, 및 소음 레벨과 같은 오디오 이벤트들을 검출한다.
- [0135] 조직화기에 의해 생성된 출력 비디오 스트림(12)은 비디오 믹서(34)에 의해 믹싱된다. 비디오 믹서(34)는 HMM 엔진(35)의 결과들을 사용하여 이하에 더 설명되는 바와 같이 미리 규정된 스크린 템플릿들에 따라 수신된 입력 비디오 스트림들(11)에서 믹싱한다. 스크린 템플릿들(44)은 스크린 템플릿 저장소(38)에 저장된다. HMM 엔진(35)에 의해 수행된 프로세스들은 이제 도 4 및 도 5를 참조하여 기술될 것이다.
- [0136] 도 4를 참조하면, 스크린 템플릿(44)은 스크린상의 적어도 하나의 입력 비디오 스트림의 미리 규정된 배치이다. 템플릿(44)은 정보의 상이한 소스들 사이에 스크린을 조직하고 때때로 분할하도록 구성된다. 도 4의 예에서, 화상 회의의 환경은 가상 교실이다. 세 가지 스크린 템플릿들(44): 튜터의 단일 뷰를 나타내는 튜터 스크린 템플릿들(701), 및 참여자들의 뷰들의 모자이크를 갖는 가상 교실 스크린 템플릿(702) 및 예를 들면 질문하기를 원하는 참여자를 보여주는 학습자 스크린 템플릿(703)이 존재한다. HMM에서, 각각의 스크린 템플릿(44)은 디스플레이 상태와 연결된다. 도 4의 이러한 HMM 예에서, 세 개의 디스플레이 상태들(튜터 스크린 상태(40), 교실 스크린 상태(41), 및 학습자 스크린 상태(42))이 존재한다. HMM 모델의 이동 행렬(A)은 이들 상태들 사이의 이동들(43)을 규정한다.
- [0137] 모델의 다른 상세들을 제공하기 위해, 도 5는 초기 스크린 상태(57) 및 상기 언급된 상태들(40, 41, 42)을 또한 나타낸다. 이러한 도면은 또한 복수의 관찰 가능한 동작들:
- [0138] - 튜터가 말하고 있는 동작(53), 및

- [0139] - 손을 들고 있는 동작(54),을 도시한다.
- [0140] 이들은 분석자(32)에 의해 검출될 수 있는 관찰 가능한 동작들의 예들이다.
- [0141] 일 실시예에서, HMM 엔진(35)은 관찰가능한 동작들을 처리한다(16). 이들 관찰 가능한 동작은 두 개의 제스처들(손가락 들기, 손 들기), 두 개의 움직임들(머리 상하 이동, 머리 좌우 이동), 두 개의 얼굴 표정들(말하기(얼굴+ 말하기/입술들이 움직이고 있음), 또는 잠자기(눈들이 없음/눈을 감음/얼굴이 스크린 뒤에 없음), 두 개의 키워드 동작들(청취자의 이름 또는 서브타이틀을 받음하기), 네 개의 오디오 동작들(튜터에 의한 말하기, 청취자에 의한 말하기, 소음 생성, 침묵하기), 두 개의 슬라이드 동작들(슬라이드를 스위칭, 포인터를 움직임), 및 두 개의 서브 이벤트들(질문 시작, 질문 종료)을 실행한다.
- [0142] 도 5는 또한 결정된 디스플레이 상태에서 발생할 관찰 이벤트의 확률들(55)을 도시한다. 각각의 결합[관찰 이벤트, 디스플레이 상태]에 대한 하나의 확률이 존재한다. 도 5는 두 개의 상태들 사이의 각각의 이동(43)에 연관된 확률들(58) 및 초기화 확률들(56)을 또한 도시한다.
- [0143] 허든 마르코프 모델(HMM)은 초기화 행렬(50), 이동 행렬(51) 및 방출 행렬(52)로 표현된다. 이러한 별개의 HMM 방법은 동적 믹싱 행동의 기초를 제공한다. HMM 방법을 기술하기 위해, 다음의 개념들이 규정된다:
- [0144]  $Q = \{q_1, q_2, \dots, q_N\}$ : 일 세트의 디스플레이 상태들; 각각의 상태는 스크린 템플릿들을 나타낸다.
- [0145]  $N$  = 디스플레이 상태들의 수
- [0146]  $V = \{v_1, v_2, \dots, v_N\}$ : 관찰 가능한 동작들의 세트
- [0147]  $M$  = 관찰 가능한 동작들의 수
- [0148]  $T$  = 관찰 시퀀스의 길이
- [0149]  $O$  = 관찰 이벤트들의 관찰된 시퀀스
- [0150]  $S = \{s_t\}$ ;  $s_t$ 는  $t$  시간의 디스플레이 상태
- [0151] 모델은 식  $\lambda = (A, B, \pi)$ 에 의해 완전히 규정되고 또한 조직화 모델이라고 불린다.
- [0152]  $A$ 는 이동 행렬이고,  $B$ 는 방출 행렬이고,  $\pi$ 는 초기화 행렬이다. 모델에서,  $A$ 는 디스플레이 상태들, 즉, 다양한 카메라 뷰들 사이의 이동 확률들을 포함하고;  $B$ 는 현재 디스플레이 상태를 인식하는 각각의 관찰 가능한 동작의 방출 확률들을 포함하고;  $\pi$ 는 디스플레이 상태가 제 1 장소에서 보여질 확률을 포함한다. 세 개의 행렬들은 다음과 같이 수학적으로 기술된다:
- [0153] 
$$A = \{a_{ij} | a_{ij} = \Pr(s_{t+1} = q_i | s_t = q_j)\} \quad (1)$$
- [0154] 
$$B = \{b_{jk} | b_{jk} = \Pr(o_t = v_k | s_t = q_j)\} \quad (2)$$
- [0155] 
$$\pi = \{\pi_i | \pi_i = \Pr(s_1 = q_i)\} \quad (3)$$
- [0156] 상기에 기술된 조직화 모델은 도 3에 기술된 조직화기(1)의 HMM 엔진(35)에 의해 사용된다. HMM 엔진(35)의 목표는, 조직화 모델( $\lambda$ ) 및 관찰 시퀀스( $O$ )를 사용하여, 가장 적절한 스크린 템플릿들을 예측하는 것이다. 관찰 시퀀스( $O$ )는 분석자(32)에 의해 제공된다. HMM 엔진(35)의 기능은 디코딩 기능이다. 이러한 기능은 관찰 시퀀스 및 HMM 모델이 이루어졌을 때 가장 가능성 있는 디스플레이 상태들의 시퀀스를 얻도록 구성된다. 최상의 디스플레이 상태 시퀀스( $Q_{optimal}$ )를 찾기 위해, 다음 수식의 해를 구한다:
- [0157] 
$$Q_{optimal} = \arg \max_Q \Pr(Q | \lambda, O) \quad (4)$$
- [0158] 식(4)을 풀기 위해, HMM 엔진(35)은 비터비 알고리즘을 사용한다. 때가 경과함에 따라, 디코딩은 HMM 엔진(35)에 의해 주어진 클럭 레이트로 행해진다. 디코딩은 때가 경과함에 따라 상태들의 시퀀스를 야기한다. HMM 엔진(35)은 비디오 믹서(34)를 통해 비디오를 조직화한다.
- [0159] 상기 디코딩 프로세스에서, 도 4 및 도 5에 도시되는 단일 HMM 모델이 이용되었다. 다른 실시예에서, 조직화기

(1)는 복수의 조직화 모델들을 갖는다.

[0160] 더 양호한 융통성을 추가하기 위해, 상기 목적을 위해, 조직화기(1)는 HMM 모델 저장소(37)를 포함한다. 이러한 저장소(37)는 복수의 미리 규정된 조직화 모델들을 저장한다. 일 실시예에서, 사용자가 현재 화상 회의 세션에서 HMM 엔진(35)에 의해 사용된 조직화 모델( $\lambda$ )을 선택하는 것이 가능하다.

[0161] 실감 인식을 증가시키기 위해, 조직화기(1)의 다른 실시예는 HMM 엔진(35)에 의해 사용된 조직화기 모델의 동적 선택을 또한 제안한다. 조직화기(1)는 화상 회의 콘텍스트 또는 시나리오 및 사용자 프로파일에 가장 잘 맞는 화상 조직화 모델을 인식할 수 있다. 분류자(36)의 목표는 HMM 저장소(37)에서 이용가능한 어느 조직화 모델( $\lambda$ )이 현재 사용 경우에 가장 적절한지를 동적으로 식별하는 것이다.

[0162] 처음에, 제 1 수신된 비디오 및 오디오 관찰 이벤트들에 기초하여, 분류자(36)는 관찰 이벤트들의 시간적 시퀀스에 가장 잘 맞는 HMM 조직화 모델을 선택한다. 화상 회의 세션 동안, 분류자(36)는 다른 것이 관찰 이벤트들의 시간적 시퀀스에 더 적합한 경우 HMM 모델을 변경할 수 있다.

[0163] 올바른 모델을 선택하는 이러한 기능은 인식 기능이다: 관찰 시퀀스 및 상이한 HMM 모델이 주어지면, 분류자(36)는 이들 관찰들에 가장 잘 부합하는 HMM 조직화 모델을 선택한다.  $n$  개의 모델들( $\lambda_{i,i=1..n}$ )에 대하여, 분류자(36)는:

$$optimal = arg \max_i Pr(O|\lambda_i) \quad (5)$$

[0164]

인 최적 모델( $\lambda_{optimal}$ )을 선택한다.

[0165]

분류자(36)는 포워드 알고리즘 또는 백워드 알고리즘에 의해 이러한 기능을 실행한다.

[0166]

이러한 실시예에서, 조직화기(1)는 스마트 화상 조직화 기능들을 제공할 수 있다. 시스템은 더 융통성 있고 더 동적이다.

[0167]

[0168] 다른 실시예에서, 새로운 조직화 모델들을 생성함으로써 조직화 능력들을 강화하는 것이 또한 가능하다. 사용자가 새로운 조직화 모델들을 생성할 수 있게 하기 위해, 도 6에 도시된 조직화(1)의 다른 실시예는 학습 기능을 포함한다.

[0169] 학습 프로세스는 비전문 사용자가 임의의 기술 숙련들 없이 그들의 사용들에 기초하여 그들 자신의 조직화 모델들을 생성할 수 있게 한다. 이는 3 개의 모듈들: 사용자 학습자 인터페이스(700), 사용자 활성 기록자(602), 및 HMM 발생자(603)에 의해 실행된다.

[0170] 실시간으로 및 관찰 이벤트들에 의존하여, 사용자는 어느 메인 비디오 스트림이 조직화기(1)에 의해 디스플레이 되어야 하는지를 선택한다. 학습 모듈(601)은 때가 경과함에 따라 사용자에게 의해 선택된 디스플레이 상태들 및 관찰 이벤트들을 기록하고 새로운 HMM 모델을 생성하거나 또는 사용자의 선택들에 기초하여 연관된 확률들에 의해 기존 모델을 갱신한다.

[0171] 도 7을 참조하면, 그래픽 사용자 학습자 인터페이스(700)의 예는 상이한 입력 비디오 스트림들(11)을 보여주는 상이한 스크린 템플릿들을 디스플레이한다. 이러한 예는 세 개의 디스플레이 상태들, 튜터 스크린(701), 교실의 일반 뷰의 스크린(702), 및 특정 학습자의 스크린(703)을 제안한다. 관찰 이벤트 윈도우(704)는 때가 경과함에 따라 현재 관찰 이벤트들을 디스플레이한다.

[0172] 사용자 학습 인터페이스(700)는 사용자가 상이한 스크린들 사이에서 선택을 할 수 있게 하기 위한 버튼들(705)과 같은 몇몇 입력 수단을 또한 포함한다. 버튼(706)은 새로운 기록 시퀀스를 시작하는 역할을 한다. 버튼(707)은 기록 시퀀스를 중단 및 확인하기 위해 제공한다. 버튼(707)의 작동은 학습 모듈(601)이 사용자에게 의해 행해진 선택들을 기록하고 이후 대응하는 조직화 모델을 생성하게 한다.

[0173] 훈련 프로세스에서, 일어나는 각각의 관찰 이벤트들에 대하여, 사용자는 스크린 템플릿을 선택하도록, 즉, 사실상 생성될 HMM 모델의 대응하는 디스플레이 상태를 선택하도록 요청된다.

[0174] 사용자가 기록 시퀀스를 시작할 때, 비디오 스트림들이 디스플레이된다. 관찰 이벤트가 발생할 때, 사용자는 스크린 버튼들(705)을 통해 스크린을 선택하도록 요청되고 최종적으로 사용자는 버튼(707)을 통해 그의 선택들을 확인한다. 사용자 입력들은 기록되어 HMM 저장소(37)에 저장될 수 있는 HMM 조직화 모델( $\lambda$ )로 변환된다. 학습 모듈(601)은 또한 기존 모델을 갱신할 수 있다.

[0175] 모델 생성 피치는 실감 통신 품질 결과를 개선하는 것에 매우 관심이 있다. 그러나, 이미 기존 모델과 매우 유사한 모델을 저장하는 것은 유용하지 않을 수 있다. 일 실시예에서, 학습 모듈(601)은 새로운 모델과 HMM 저장소(37)에 이미 저장된 모델들 사이의 거리를 측정할 수 있다. 학습 모듈(601)은 쿨백 라이브너 거리를 통해 상이한 HMM들 모델 사이의 차이를 측정한다. 요약하면, 사용자는 기존 조직화 모델을 개인화할 수 있다. 그러나, 사용자는 새로운 조직화기 모델을 또한 생성할 수 있다: 모듈은 사용자에게 의해 행해진 선택을 기록하고 이들 관찰들로부터 새로운 HMM 모델을 생성한다. 이후 쿨백 라이브너 거리는 이러한 템플릿이 저장되고 확인되기 위해 기존 템플릿들과 충분히 상이한지를 결정하기 위해 사용된다.

[0176] 상기에 기술된 바와 같이, 이를 생성하기 위해 모델 파라미터들  $\lambda = (A, B, \pi)$ 을 초기화할 필요가 있다. 학습 모듈(601)에 의해 실행된 프로세스는 다음의 단계들을 포함한다:

[0177] 1. 초기화 행렬 훈련

[0178] 초기화 행렬( $\pi$ )의 훈련은 초기화 확률을 통해 행해진다: 사용자에게 의해 선택된 제 1 상태는 1로 설정되고 다른 것들은 0으로 설정된다.

[0179] 2. 이동 행렬 훈련

[0180] 훈련 프로세서에서, 각각의 관찰에 대하여, 사용자는 스크린 템플릿들 사이에서 선택하도록 요청될 것이다. 결과로서 디스플레이 상태들의 일 시퀀스가 기록될 것이다.

[0181] 이동 행렬(A)의 훈련의 알고리즘은 4 개의 단계들로 구성된다:

[0182] 단계1: 입력된 HMM에 대하여 디스플레이 상태들의 수를 획득한다.

[0183] 단계2: 디스플레이 상태들 사이의 모든 가능한 이동들을 포함하는 비교 행렬을 생성한다.

[0184] 단계3: 상태들 시퀀스를 브라우징하고 발생 행렬에서 카운터들을 증분시킨다. 발생 행렬은 두 개의 상태들(i, j) 사이의 각각의 이동에 대한 발생을 포함하는 행렬이다. 비교 행렬, 발생 행렬, 및 이동 행렬(A)은 동일한 크기들  $N \times N$ 을 갖는다.

[0185] 단계4: 발생 행렬, 이동 행렬은 다음과 같이 계산된다; 각각의 라인에 대하여, 각각의 값을 이러한 라인의 합계로 나눈다.

[0186] 이는 다음의 식으로 요약된다:

$$a_{ij} = \frac{occ_{ij}}{\sum_{h=1}^N occ_{ih}} \quad (6)$$

[0187] [0188]  $occ$ 는 발생 행렬 계수이다.

[0189] 3. 방출 행렬 훈련

[0190] 각각의 상태에 대하여, 모듈은 각각의 관찰 가능한 동작의 관찰 이벤트들을 개별적으로 카운트한다. 이후, 이러한 수는 동일한 디스플레이 상태에서 발생된 관찰 이벤트들의 총 수로 나뉘진다. 이는 다음 식으로 요약된다:

$$b_{ik} = \frac{occObs_{ik}}{\sum_{h=1}^M occObs_{ih}} \quad (7)$$

[0191] [0192]  $occObs$ 는, 크기들  $N \times M$ 을 갖는, 각각의 관찰 가능한 동작 및 각각의 디스플레이 상태에 대한 발생 행렬을 나타낸다.

[0193] 도 6을 참조하면, 이제 학습 모듈(601), 사용자 훈련 인터페이스(700), 사용자 활동들 기록기(602) 및 HMM 발생기(603)를 포함하는 일 실시예를 기술한다. 학습 모듈(601)은 사용자 훈련 인터페이스(700)를 통해 사용자 입력들을 수신하고 이러한 사용자의 결정들을 사용자 활동들의 기록기(602)에 의해 기록하고, HMM 모델을 HMM 발생기(603)에 의해 계산한다. 결과는 HMM 모델 저장소(37)에 저장된다. 도 6에 도시된 조직화기(1)의 다른 모듈들은 도 3의 모듈들과 유사하다.

[0194] 도 8을 참조하면, 조직화기(1)의 다른 실시예는 학습 모듈(601)을 수 개의 인스턴스들(80)을 지원하는 중앙화된 비디오 믹서(34)와 통합시킨다. 도 6의 실시예와 대조적으로, 비디오 믹서(34) 모듈은 중앙화된 방식으로 디스

플레이하는 비디오의 상이한 인스턴스들(80)을 지원한다. 각각의 사용자는 그 자신의 비디오 조직화를 생성 및 개인화하고 개인화된 조직화 비디오 스트림을 수신할 수 있다. 비디오 조직화는 수 개의 비디오 믹서 인스턴스들(80)에서 행해진다. 사용자들은 이들을 바로 볼 수 있다(즉, 사용자 디바이스들 상에 화상 조직화 없이). "사용자 저장소"(81) 모듈은 상이한 사용자들을 관리하기 위해 사용된다(id, 프로파일, 조직화 모델, 등...).

[0195] 도 9를 참조하면, 조직화기(1)의 일 실시예는 학습 모듈(601)을 포함하고, 반면에 비디오 믹서들(34) 및 HMM 엔진들(35)은 원격 단말들(2)에 분산된다. 이러한 구현은 서버상의 너무 많은 처리를 피하기 위해 사용자에게 더 가까운 조직화를 실행할 수 있게 한다. 조직화기(1)에 의해 선택된 HMM 조직화 모델은 사용자 단말(2)상에 업로딩된다. 로컬 비디오 조직화기(902)는 이러한 조직화 모델을 사용하여 서버로부터 유래한 비디오 스트림들을 구성한다. 로컬 비디오 조직화기(902)는 로컬 비디오 믹서(934) 및 HMM 엔진(935)을 포함한다. 로컬 비디오 조직화기(902)는 도 2에 또한 도시된다. 단지 로컬 비디오 조직화기들에 의해 요청된 비디오 스트림들만이 중앙 비디오 믹서(34)에 의해 전송된다. 사용자는 그의 자신의 모델을 로컬로 개인화하거나 규정하고 이들을 중앙 서버상에 저장 및 공유할 수 있다. 이러한 경우에, 로컬 조직화기는 중앙 HMM 관리자, 엔진, 믹서, 템플릿, 및 학습자와 상호작용한다.

[0196] 제어 유닛들과 같은 요소들은 예를 들면, 예를 들면, ASIC과 같은 하드웨어 수단들, 또는 하드웨어 및 소프트웨어 수단의 조합, 예를 들면, ASIC 및 FPGA, 또는 적어도 하나의 마이크로프로세서 및 그에 위치한 소프트웨어 모듈을 갖는 적어도 하나의 메모리일 수 있다.

[0197] 본 발명은 기술된 실시예들로 제한되지 않는다. 첨부된 청구항들은 설명된 여기의 기본 교시 내에 명백하게 포함되는 본 기술 분야의 숙련자에게 생각될 수 있는 모든 변경들 및 대안적인 구성들을 구현하는 것으로 해석될 것이다.

[0198] 동사 "포함하기 위해" 및 그의 활용의 사용은 청구항들에 기재되는 것과 다른 요소들 또는 단계들의 존재를 배제하지 않는다. 또한, 요소 또는 단계 앞에 있는 단수 표현의 사용은 이러한 요소들 또는 단계들의 복수의 존재를 배제하지 않는다.

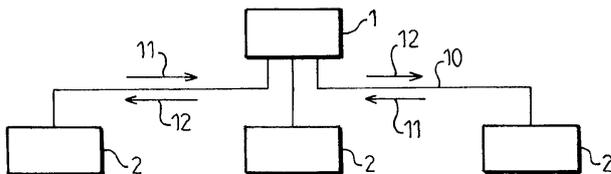
[0199] 청구항들에서, 괄호들 사이에 배치된 임의의 참조 기호들은 청구항들의 범위를 제한하는 것으로 해석되지 않아야 한다.

**부호의 설명**

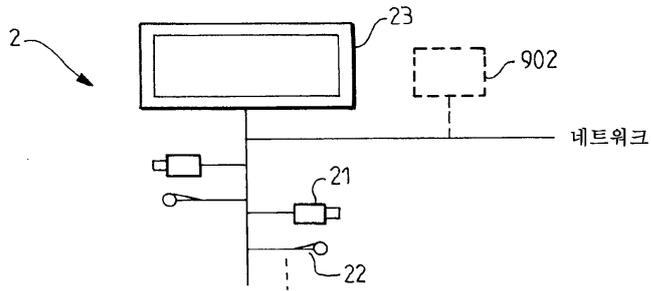
- [0200] 11 : 비디오 스트림들
- 32 : 분석자
- 34 : 중앙 비디오 믹서
- 35 : HMM 엔진
- 36 : 분류자
- 37 : HMM 모델
- 81 : 사용자 저장소
- 602 : 사용자 활동들/믹싱 기록기
- 603 : HMM 발생기
- 700 : 사용자 학습 인터페이스
- 934 : 로컬 비디오 믹서
- 935 : HMM 엔진

**도면**

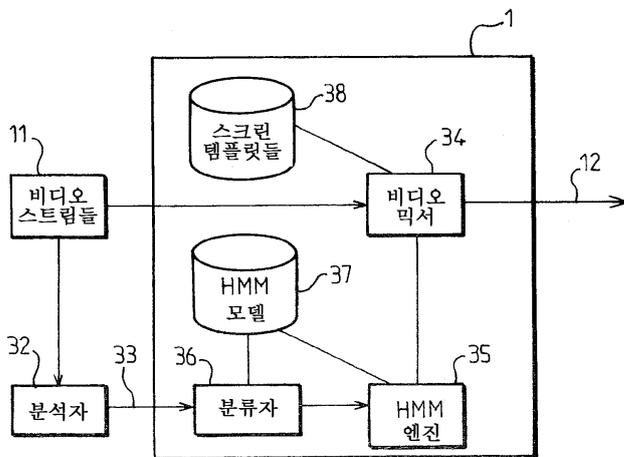
**도면1**



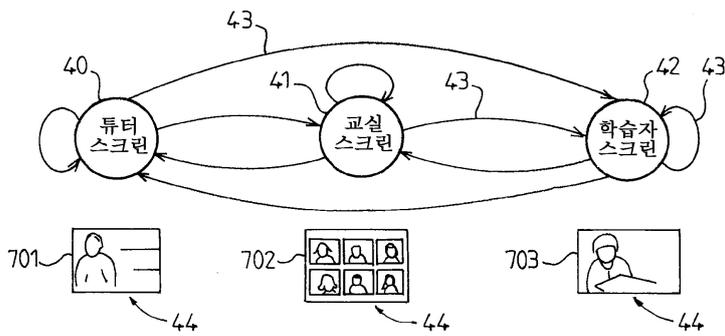
도면2



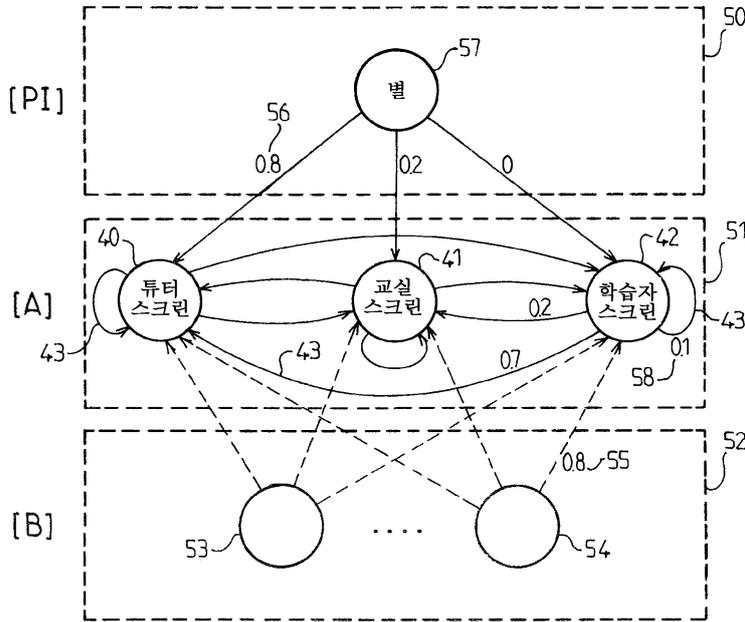
도면3



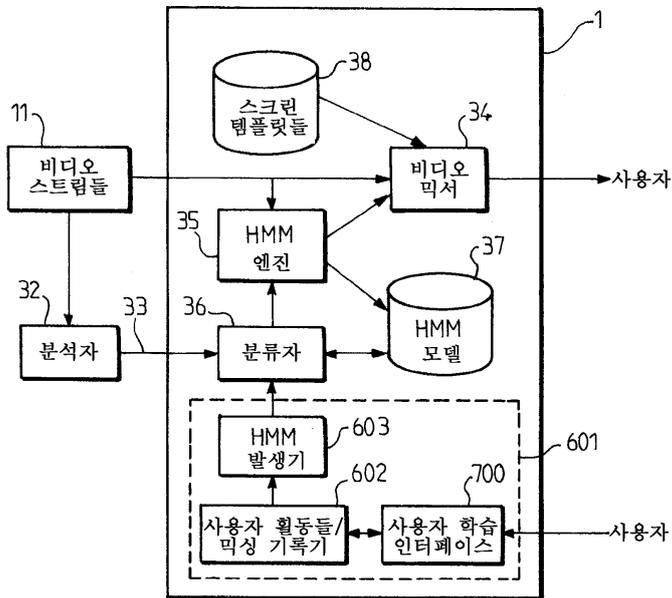
도면4



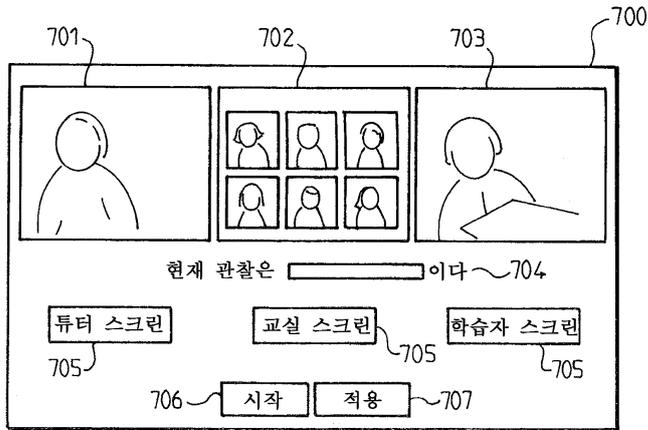
도면5



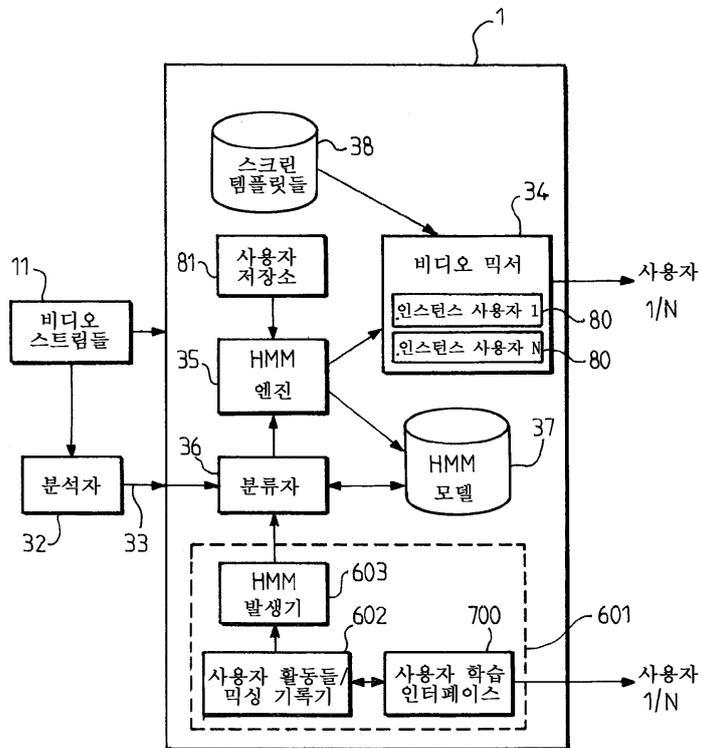
도면6



도면7



도면8



도면9

