

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2015-19162
(P2015-19162A)

(43) 公開日 平成27年1月29日(2015.1.29)

(51) Int.Cl.		F I				テーマコード(参考)
HO4N	7/15	(2006.01)	HO4N	7/15	630Z	5B057
G06T	1/00	(2006.01)	G06T	1/00	340A	5C164

審査請求 未請求 請求項の数 27 O L (全 39 頁)

(21) 出願番号 特願2013-143898 (P2013-143898)
(22) 出願日 平成25年7月9日(2013.7.9)

(71) 出願人 00002897
大日本印刷株式会社
東京都新宿区市谷加賀町一丁目1番1号
(74) 代理人 100091476
弁理士 志村 浩
(72) 発明者 山王 卓
東京都新宿区市谷加賀町一丁目1番1号
大日本印刷株式会社内
Fターム(参考) 5B057 CA01 CA08 CA12 CA16 CC03
CD12 DA08 DA11 DA16 DC25
5C164 FA10 PA31 UB88P VA04S VA07P
VA35S VA54P

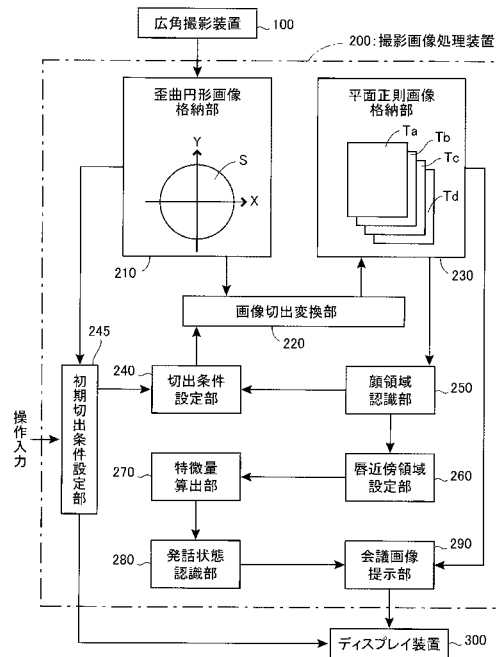
(54) 【発明の名称】 会議支援システム

(57) 【要約】

【課題】単純な構成で各参加者の発話状態を正確に認識する。

【解決手段】魚眼レンズを用いた広角撮影装置100により、会議の各参加者の顔を含む歪曲円形画像Sを撮影し、格納部210に格納する。ユーザが各人物の初期切出条件を設定すると、画像切出変換部220が各人物画像を切り出し、平面正則画像Tに変換して格納部230に格納する。顔領域認識部250は各平面正則画像Tから顔領域を認識し、唇近傍領域設定部260は顔領域上に唇近傍領域を設定する。特徴量算出部270は唇近傍領域内の輝度もしくは色を示す特徴量を算出し、発話状態認識部280は当該特徴量の時間変化に基づいて、個々の参加者の発話時間を決定する。会議画像提示部290は各人物画像とともに発話時間をディスプレイ装置300上に提示する。切出条件設定部240は、顔領域認識部250の認識結果に基づき、唇を中心とした切出条件の再設定を逐次行う。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

広範囲な画角をもった撮影により、会議の各参加者の顔を含む歪曲円形画像を取り込む
広角撮影装置と、

前記歪曲円形画像に基づいて、各参加者の人物画像および発話情報を作成する撮影画像
処理装置と、

各参加者の人物画像および発話情報を表示するディスプレイ装置と、

を備える会議支援システムであって、

前記撮影画像処理装置が、

前記広角撮影装置によって時間軸に沿って撮影された複数の歪曲円形画像を順次格納す
る歪曲円形画像格納部と、

各参加者について、切出中心点、切出方向、切出倍率を定める所定の切出条件をそれぞ
れ設定する切出条件設定部と、

前記ディスプレイ装置の画面上に前記歪曲円形画像を表示させ、表示中の歪曲円形画像
上で切出中心点の位置を指定するユーザの操作入力に基づいて、前記切出条件設定部に対
して、各参加者についての初期切出条件を設定する初期切出条件設定部と、

前記切出条件設定部に設定されている切出条件に基づいて、前記歪曲円形画像の一部
を切り出して変換し、各参加者についての平面正則画像を生成する画像切出変換部と、

前記平面正則画像を格納する平面正則画像格納部と、

前記平面正則画像について顔領域を認識する顔領域認識部と、

前記顔領域上に唇近傍領域を設定する唇近傍領域設定部と、

各参加者の平面正則画像について、前記唇近傍領域内の輝度もしくは色を示す特徴量
を算出する特徴量算出部と、

前記特徴量の時間変化に基づいて、個々の参加者ごとの発話状態を認識する発話状態認
識部と、

前記平面正則画像と前記発話状態に関連した発話情報とを含んだ会議画像を、時間軸に
沿って順次生成し、これを前記ディスプレイ装置に与えて動画として表示させる会議画像
提示部と、

を有しており、

前記切出条件設定部が、前記顔領域認識部の認識結果に基づいて、切出条件の再設定を
行う機能を有していることを特徴とする会議支援システム。

【請求項 2】

請求項 1 に記載の会議支援システムにおいて、

切出条件設定部が、顔領域の輪郭枠を基準として予め定められた所定位置にある点を再
設定基準点と定め、平面正則画像上の前記再設定基準点を歪曲円形画像上に逆変換して得
られる点を切出中心点とする再設定を行うことを特徴とする会議支援システム。

【請求項 3】

請求項 2 に記載の会議支援システムにおいて、

切出条件設定部が、唇の中心が位置すると予想される点を再設定基準点に定めること
を特徴とする会議支援システム。

【請求項 4】

請求項 1 に記載の会議支援システムにおいて、

切出条件設定部が、唇近傍領域設定部が設定した唇近傍領域の中心点を再設定基準点と
定め、平面正則画像上の前記再設定基準点を歪曲円形画像上に逆変換して得られる点を切
出中心点とする再設定を行うことを特徴とする会議支援システム。

【請求項 5】

請求項 1 ~ 4 のいずれかに記載の会議支援システムにおいて、

切出条件設定部が、顔領域の面積に応じた切出倍率をもった切出条件を再設定するこ
とを特徴とする会議支援システム。

【請求項 6】

10

20

30

40

50

請求項 1 ~ 5 のいずれかに記載の会議支援システムにおいて、
切出条件設定部が、切出中心点を通る歪曲円形画像の同心円を求め、当該同心円の接線方向が平面正則画像の水平方向となるような切出方向をもつ切出条件を再設定することを特徴とする会議支援システム。

【請求項 7】

請求項 1 ~ 6 のいずれかに記載の会議支援システムにおいて、
広角撮影装置が、一定の時間周期で撮影されたフレーム単位の静止画からなる歪曲円形画像を連続して取り込み、
切出条件設定部が、1 フレーム周期、もしくは、複数フレーム周期で切出条件の再設定を行うことを特徴とする会議支援システム。

10

【請求項 8】

請求項 1 ~ 7 のいずれかに記載の会議支援システムにおいて、
唇近傍領域設定部が、顔領域の輪郭枠を基準として予め定められた所定枠内の領域を唇近傍領域に設定することを特徴とする会議支援システム。

【請求項 9】

請求項 8 に記載の会議支援システムにおいて、
顔領域認識部が、矩形状の輪郭枠をもった顔領域を認識し、
唇近傍領域設定部が、前記輪郭枠の左辺の横方向に関する位置を 0 %、右辺の横方向に関する位置を 100 % とし、上辺の縦方向に関する位置を 0 %、下辺の縦方向に関する位置を 100 % としたときに、予め設定された所定の定数 a_1 、 a_2 、 b_1 、 b_2 (但し、 $0 < a_1 < a_2 < 100$ 、 $0 < b_1 < b_2 < 100$) に基づいて、左辺の横方向に関する位置が a_1 %、右辺の横方向に関する位置が a_2 % の位置にあり、上辺の縦方向に関する位置が b_1 %、下辺の縦方向に関する位置が b_2 % の位置にある矩形状の輪郭枠をもつ領域を唇近傍領域に設定することを特徴とする会議支援システム。

20

【請求項 10】

請求項 1 ~ 7 のいずれかに記載の会議支援システムにおいて、
唇近傍領域設定部が、個々の平面正則画像について、それぞれ顔領域内を所定の唇探索アルゴリズムを用いて探索し、探索された唇の周辺を含む領域を、当該平面正則画像についての唇近傍領域に設定することを特徴とする会議支援システム。

【請求項 11】

請求項 1 ~ 10 のいずれかに記載の会議支援システムにおいて、
平面正則画像が、3 原色 RGB の各画素値をもった画素の集合体によって構成され、
特徴量算出部が、唇近傍領域内の第 i 番目 ($1 \leq i \leq n$ 、但し、 n は唇近傍領域内の画素の総数) の画素の 3 原色の画素値をそれぞれ $R(i)$ 、 $G(i)$ 、 $B(i)$ としたときに、

$$= \sum_{i=1}^n (R(i) + G(i) + B(i)) / n$$

なる演算式に基づく演算により当該唇近傍領域内の輝度を示す特徴量を算出することを特徴とする会議支援システム。

30

【請求項 12】

請求項 1 ~ 10 のいずれかに記載の会議支援システムにおいて、
平面正則画像が、3 原色 RGB の各画素値をもった画素の集合体によって構成され、
特徴量算出部が、唇近傍領域内の第 i 番目 ($1 \leq i \leq n$ 、但し、 n は唇近傍領域内の画素の総数) の画素の原色 R の画素値を $R(i)$ としたときに、

$$= \sum_{i=1}^n (R(i)) / n$$

なる演算式に基づく演算により当該唇近傍領域内の色を示す特徴量を算出することを特徴とする会議支援システム。

40

【請求項 13】

請求項 1 ~ 10 のいずれかに記載の会議支援システムにおいて、
平面正則画像が、3 原色 RGB の各画素値をもった画素の集合体によって構成され、
特徴量算出部が、唇近傍領域内の第 i 番目 ($1 \leq i \leq n$ 、但し、 n は唇近傍領域内の画

50

素の総数)の画素の、前記3原色RGBの各画素値から求めた色相値を $H(i)$ としたときに、

$$= \frac{1}{n} \sum_{i=1}^n (H(i))$$

なる演算式に基づく演算により当該唇近傍領域内の色を示す特徴量を算出することを特徴とする会議支援システム。

【請求項14】

請求項1～13のいずれかに記載の会議支援システムにおいて、

発話状態認識部が、各参加者について、特徴量の変動幅が所定のしきい値以下となる期間を非発話期間、前記しきい値を超える期間を発話期間と認識することを特徴とする会議支援システム。

10

【請求項15】

請求項14に記載の会議支援システムにおいて、

発話状態認識部が、特徴量の時間的変動グラフの極大点と極小点とを変極点として逐次検出し、個々の変極点について、当該変極点の特徴量と直前の変極点の特徴量との偏差を求め、当該偏差が所定のしきい値以下の状態から前記しきい値を超える状態に転じた時点を非発話期間から発話期間への転換時点と認識し、当該偏差が前記しきい値を超えた状態から前記しきい値以下の状態に転じた時点を発話期間から非発話期間への転換時点と認識することを特徴とする会議支援システム。

【請求項16】

請求項1～15のいずれかに記載の会議支援システムにおいて、

会議画像提示部が、個々の参加者について、当該参加者の人物画像を含む平面正則画像を表示する参加者表示領域と、この参加者表示領域に紐づけて当該参加者の発話情報を表示する発話情報表示領域と、を有する会議画像を生成することを特徴とする会議支援システム。

20

【請求項17】

請求項16に記載の会議支援システムにおいて、

個々の参加者についての発話情報表示領域を、当該参加者についての参加者表示領域の近傍に配置し、互いに近傍に配置することにより、特定の参加者表示領域と特定の発話情報表示領域との紐付けを行うことを特徴とする会議支援システム。

【請求項18】

請求項16または17に記載の会議支援システムにおいて、

会議画像提示部が、発話情報表示領域に、各参加者が発話期間中か非発話期間中かを識別する識別マークを表示させることを特徴とする会議支援システム。

30

【請求項19】

請求項16または17に記載の会議支援システムにおいて、

会議画像提示部が、現在発話中の参加者についての発話情報表示領域についてのみ、発話中であることを示す識別マークを表示させることを特徴とする会議支援システム。

【請求項20】

請求項16または17に記載の会議支援システムにおいて、

会議画像提示部が、発話情報表示領域に、各参加者の累積発話時間を表示させることを特徴とする会議支援システム。

40

【請求項21】

請求項16または17に記載の会議支援システムにおいて、

会議画像提示部が、発話情報表示領域に、現在発話中の参加者についての発話継続時間を表示させることを特徴とする会議支援システム。

【請求項22】

請求項20または21に記載の会議支援システムにおいて、

会議画像提示部が、発話情報表示領域への時間表示を、数字もしくは円グラフまたはその双方で行うことを特徴とする会議支援システム。

【請求項23】

50

請求項 20 または 21 に記載の会議支援システムにおいて、
会議画像提示部が、発話情報表示領域に対して、円グラフを含む時間表示を行い、所定の時間ユニット U に対して前記円グラフの 360° を対応させ、表示対象となる時間 T を U で割り算したときの商を Q、剰余を R としたときに、Q 個の円と、1 個の扇形（中心角は、 $= 360^\circ \times R / U$ ）とを並べた表示を行うことを特徴とする会議支援システム。

【請求項 24】

請求項 1 ~ 23 のいずれかに記載の会議支援システムの一構成要素として記載されている撮影画像処理装置。

【請求項 25】

請求項 24 に記載の撮影画像処理装置としてコンピュータを機能させるプログラム。

【請求項 26】

請求項 24 に記載の撮影画像処理装置として機能する半導体集積回路。

【請求項 27】

請求項 1 ~ 23 のいずれかに記載の会議支援システムを複数台備え、個々の会議支援システムの構成要素となる撮影画像処理装置をネットワークで接続することにより相互に会議画像を送受信できるようにし、ディスプレイ装置に複数の会議画像を提示させるようにしたことを特徴とするテレビ会議支援システム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、会議支援システムに関し、特に、広範囲な画角をもった撮影を行うことにより、会議の各参加者の画像とともに各参加者の発話情報を提示するシステムに関する。

【背景技術】

【0002】

会議の各参加者の画像を撮影して、これをディスプレイ装置上に表示する会議支援システムは、古くから様々なタイプのものが提案されている。たとえば、下記の特許文献 1 には、会議の各参加者の画像を撮影するとともに、可変指向性マイクを用いて話者を特定し、発話中の人物画像を切り出して表示するテレビカメラ装置が開示されている。

【0003】

一方、音声ではなく、撮影画像から参加者の唇の動きを読み取ることにより話者を特定する技術も提案されている。たとえば、特許文献 2 には、参加者の撮影画像から顔の輪郭を抽出し、更に唇の動きを検出することにより発話者を特定するテレビ会議システムが開示されており、特許文献 3 には、参加者の顔画像から口唇部分の動きを捕捉して、発話者の前駆動作を検出することにより発話者を特定する技術が開示されている。

【0004】

また、特許文献 4 には、魚眼レンズによって会議の各参加者を撮影し、得られた歪曲円形画像上でユーザに所望の位置を指定させ、当該指定位置を中心とした切出処理により各参加者の平面正則画像を歪みなく表示する技術が開示されている。

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】特開平 11 - 331827 号公報

【特許文献 2】特開 2003 - 189273 号公報

【特許文献 3】特開 2004 - 118314 号公報

【特許文献 4】特開 2012 - 089954 号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

ディスプレイ上に会議の様子を提示する会議支援システムでは、各参加者の画像を表示

10

20

30

40

50

するだけでなく、各参加者の発話状態に関する情報を提示できると便利である。たとえば、現在、どの参加者が発言中であるのか、個々の参加者の発言時間はどの程度なのか、といった情報がディスプレイ上に提示されれば、視聴者や司会者は、会議を傍聴したり進行したりする上で、有用な情報を得ることができる。

【0007】

会議中に発話者を特定する方法としては、前掲の特許文献1に開示されているように、音声を利用した方法が古くから利用されている。ただ、音声を利用した方法を採用するには、個々の参加者の近傍に配置された指向性マイクなど、音声情報の収集のための専用機器が必要になり、システム全体が複雑化することは否めない。そこで近年は、特許文献2, 3に示す例のように、撮影画像上で各参加者の唇の動きを解析して発話者を特定する方法が提案されている。一方、各参加者の画像を撮影するための撮影系を単純化するためには、特許文献4に示す例のように、魚眼レンズなどの広角撮影装置を用いて各参加者の画像を取得する会議支援システムも提案されている。

【0008】

このように、広角撮影装置で撮影した画像を解析することにより発話者を特定する方法を採れば、発話者の特定にマイクは不要であり、また、撮影装置も1台で済むため、装置構成を単純化し小型化を図ることができる。しかしながら、広角撮影装置によって得られる画像には、光学的要因に基づく歪みが存在するため、各参加者の発話状態を正確に認識することが困難であるという問題がある。

【0009】

そこで本発明は、装置構成の単純化を図りつつ、各参加者の発話状態を正確に認識することが可能な会議支援システムを提供することを目的とする。

【課題を解決するための手段】

【0010】

(1) 本発明の第1の態様は、広範囲な画角をもった撮影により、会議の各参加者の顔を含む歪曲円形画像を取り込む広角撮影装置と、

歪曲円形画像に基づいて、各参加者の人物画像および発話情報を作成する撮影画像処理装置と、

各参加者の人物画像および発話情報を表示するディスプレイ装置と、

を備える会議支援システムにおいて、

撮影画像処理装置に、

広角撮影装置によって時間軸に沿って撮影された複数の歪曲円形画像を順次格納する歪曲円形画像格納部と、

各参加者について、切出中心点、切出方向、切出倍率を定める所定の切出条件をそれぞれ設定する切出条件設定部と、

ディスプレイ装置の画面上に歪曲円形画像を表示させ、表示中の歪曲円形画像上で切出中心点の位置を指定するユーザの操作入力に基づいて、切出条件設定部に対して、各参加者についての初期切出条件を設定する初期切出条件設定部と、

切出条件設定部に設定されている切出条件に基づいて、歪曲円形画像の一部分を切り出して変換し、各参加者についての平面正則画像を生成する画像切出変換部と、

平面正則画像を格納する平面正則画像格納部と、

平面正則画像について顔領域を認識する顔領域認識部と、

顔領域上に唇近傍領域を設定する唇近傍領域設定部と、

各参加者の平面正則画像について、唇近傍領域内の輝度もしくは色を示す特徴量を算出する特徴量算出部と、

特徴量の時間変化に基づいて、個々の参加者ごとの発話状態を認識する発話状態認識部と、

平面正則画像と発話状態に関連した発話情報とを含んだ会議画像を、時間軸に沿って順次生成し、これをディスプレイ装置に与えて動画として表示させる会議画像提示部と、

を設け、

10

20

30

40

50

切出条件設定部に、顔領域認識部の認識結果に基づいて、切出条件の再設定を行う機能をもたせるようにしたものである。

【0011】

(2) 本発明の第2の態様は、上述した第1の態様に係る会議支援システムにおいて、切出条件設定部が、顔領域の輪郭枠を基準として予め定められた所定位置にある点を再設定基準点と定め、平面正則画像上の再設定基準点を歪曲円形画像上に逆変換して得られる点を切出中心点とする再設定を行うようにしたものである。

【0012】

(3) 本発明の第3の態様は、上述した第2の態様に係る会議支援システムにおいて、切出条件設定部が、唇の中心が位置すると予想される点を再設定基準点に定めるようにしたものである。

10

【0013】

(4) 本発明の第4の態様は、上述した第1の態様に係る会議支援システムにおいて、切出条件設定部が、唇近傍領域設定部が設定した唇近傍領域の中心点を再設定基準点と定め、平面正則画像上の再設定基準点を歪曲円形画像上に逆変換して得られる点を切出中心点とする再設定を行うようにしたものである。

【0014】

(5) 本発明の第5の態様は、上述した第1～第4の態様に係る会議支援システムにおいて、切出条件設定部が、顔領域の面積に応じた切出倍率をもった切出条件を再設定するようにしたものである。

20

【0015】

(6) 本発明の第6の態様は、上述した第1～第5の態様に係る会議支援システムにおいて、切出条件設定部が、切出中心点を通る歪曲円形画像の同心円を求め、当該同心円の接線方向が平面正則画像の水平方向となるような切出方向をもつ切出条件を再設定するようにしたものである。

【0016】

(7) 本発明の第7の態様は、上述した第1～第6の態様に係る会議支援システムにおいて、広角撮影装置が、一定の時間周期で撮影されたフレーム単位の静止画からなる歪曲円形画像を連続して取り込み、切出条件設定部が、1フレーム周期、もしくは、複数フレーム周期で切出条件の再設定を行うようにしたものである。

30

【0017】

(8) 本発明の第8の態様は、上述した第1～第7の態様に係る会議支援システムにおいて、唇近傍領域設定部が、顔領域の輪郭枠を基準として予め定められた所定枠内の領域を唇近傍領域に設定するようにしたものである。

【0018】

(9) 本発明の第9の態様は、上述した第8の態様に係る会議支援システムにおいて、顔領域認識部が、矩形形状の輪郭枠をもった顔領域を認識し、唇近傍領域設定部が、この輪郭枠の左辺の横方向に関する位置を0%、右辺の横方向に関する位置を100%とし、上辺の縦方向に関する位置を0%、下辺の縦方向に関する位置を100%としたときに、予め設定された所定の定数 a_1 、 a_2 、 b_1 、 b_2 （但し、 $0 < a_1 < a_2 < 100$ 、 $0 < b_1 < b_2 < 100$ ）に基づいて、左辺の横方向に関する位置が a_1 %、右辺の横方向に関する位置が a_2 %の位置にあり、上辺の縦方向に関する位置が b_1 %、下辺の縦方向に関する位置が b_2 %の位置にある矩形形状の輪郭枠をもつ領域を唇近傍領域に設定するようにしたものである。

40

【0019】

50

(10) 本発明の第10の態様は、上述した第1～第7の態様に係る会議支援システムにおいて、

唇近傍領域設定部が、個々の平面正則画像について、それぞれ顔領域内を所定の唇探索アルゴリズムを用いて探索し、探索された唇の周辺を含む領域を、当該平面正則画像についての唇近傍領域に設定するようにしたものである。

【0020】

(11) 本発明の第11の態様は、上述した第1～第10の態様に係る会議支援システムにおいて、

平面正則画像が、3原色RGBの各画素値をもった画素の集合体によって構成され、

特徴量算出部が、唇近傍領域内の第*i*番目(1 ≤ *i* ≤ *n*、但し、*n*は唇近傍領域内の画素の総数)の画素の3原色の画素値をそれぞれR(*i*)、G(*i*)、B(*i*)としたときに、

$$= \sum_{i=1}^n (R(i) + G(i) + B(i)) / n$$

なる演算式に基づく演算により当該唇近傍領域内の輝度を示す特徴量を算出するようにしたものである。

【0021】

(12) 本発明の第12の態様は、上述した第1～第10の態様に係る会議支援システムにおいて、

平面正則画像が、3原色RGBの各画素値をもった画素の集合体によって構成され、

特徴量算出部が、唇近傍領域内の第*i*番目(1 ≤ *i* ≤ *n*、但し、*n*は唇近傍領域内の画素の総数)の画素の原色Rの画素値をR(*i*)としたときに、

$$= \sum_{i=1}^n (R(i)) / n$$

なる演算式に基づく演算により当該唇近傍領域内の色を示す特徴量を算出するようにしたものである。

【0022】

(13) 本発明の第13の態様は、上述した第1～第10の態様に係る会議支援システムにおいて、

平面正則画像が、3原色RGBの各画素値をもった画素の集合体によって構成され、

特徴量算出部が、唇近傍領域内の第*i*番目(1 ≤ *i* ≤ *n*、但し、*n*は唇近傍領域内の画素の総数)の画素の、3原色RGBの各画素値から求めた色相値をH(*i*)としたときに、

$$= \sum_{i=1}^n (H(i)) / n$$

なる演算式に基づく演算により当該唇近傍領域内の色を示す特徴量を算出するようにしたものである。

【0023】

(14) 本発明の第14の態様は、上述した第1～第13の態様に係る会議支援システムにおいて、

発話状態認識部が、各参加者について、特徴量の変動幅が所定のしきい値以下となる期間を非発話期間、当該しきい値を超える期間を発話期間と認識するようにしたものである。

【0024】

(15) 本発明の第15の態様は、上述した第14の態様に係る会議支援システムにおいて、

発話状態認識部が、特徴量の時間的変動グラフの極大点と極小点とを変極点として逐次検出し、個々の変極点について、当該変極点の特徴量と直前の変極点の特徴量との偏差を求め、当該偏差が所定のしきい値以下の状態から当該しきい値を超える状態に転じた時点を非発話期間から発話期間への転換時点と認識し、当該偏差が当該しきい値を超えた状態から当該しきい値以下の状態に転じた時点を発話期間から非発話期間への転換時点と認識するようにしたものである。

【0025】

10

20

30

40

50

(16) 本発明の第16の態様は、上述した第1～第15の態様に係る会議支援システムにおいて、

会議画像提示部が、個々の参加者について、当該参加者の人物画像を含む平面正則画像を表示する参加者表示領域と、この参加者表示領域に紐づけて当該参加者の発話情報を表示する発話情報表示領域と、を有する会議画像を生成するようにしたものである。

【0026】

(17) 本発明の第17の態様は、上述した第16の態様に係る会議支援システムにおいて、

個々の参加者についての発話情報表示領域を、当該参加者についての参加者表示領域の近傍に配置し、互いに近傍に配置することにより、特定の参加者表示領域と特定の発話情報表示領域との紐付けを行うようにしたものである。

【0027】

(18) 本発明の第18の態様は、上述した第16または第17の態様に係る会議支援システムにおいて、

会議画像提示部が、発話情報表示領域に、各参加者が発話期間中か非発話期間中かを識別する識別マークを表示させるようにしたものである。

【0028】

(19) 本発明の第19の態様は、上述した第16または第17の態様に係る会議支援システムにおいて、

会議画像提示部が、現在発話中の参加者についての発話情報表示領域についてのみ、発話中であることを示す識別マークを表示させるようにしたものである。

【0029】

(20) 本発明の第20の態様は、上述した第16または第17の態様に係る会議支援システムにおいて、

会議画像提示部が、発話情報表示領域に、各参加者の累積発話時間を表示させるようにしたものである。

【0030】

(21) 本発明の第21の態様は、上述した第16または第17の態様に係る会議支援システムにおいて、

会議画像提示部が、発話情報表示領域に、現在発話中の参加者についての発話継続時間を表示させるようにしたものである。

【0031】

(22) 本発明の第22の態様は、上述した第20または第21の態様に係る会議支援システムにおいて、

会議画像提示部が、発話情報表示領域への時間表示を、数字もしくは円グラフまたはその双方で行うようにしたものである。

【0032】

(23) 本発明の第23の態様は、上述した第20または第21の態様に係る会議支援システムにおいて、

会議画像提示部が、発話情報表示領域に対して、円グラフを含む時間表示を行い、所定の時間ユニットUに対して円グラフの 360° を対応させ、表示対象となる時間TをUで割り算したときの商をQ、剰余をRとしたときに、Q個の円と、1個の扇形(中心角は、 $= 360^\circ \times R / U$)とを並べた表示を行うようにしたものである。

【0033】

(24) 本発明の第24の態様は、上述した第1～第23の態様に係る会議支援システムにおける撮影画像処理装置を構成したものである。

【0034】

(25) 本発明の第25の態様は、上述した第24の態様に係る撮影画像処理装置を、コンピュータにプログラムを組み込むことにより構成したものである。

【0035】

10

20

30

40

50

(26) 本発明の第26の態様は、上述した第24の態様に係る撮影画像処理装置を、半導体集積回路により構成したものである。

【0036】

(27) 本発明の第27の態様は、上述した第1～第23の態様に係る会議支援システムを複数台用意し、個々の会議支援システムの構成要素となる撮影画像処理装置をネットワークで接続することにより相互に会議画像を送受信できるようにし、ディスプレイ装置に複数の会議画像を提示させるようにしてテレビ会議支援システムを構成したものである。

【発明の効果】

【0037】

本発明の会議支援システムでは、広角撮影装置を用いた撮影により、会議の各参加者の顔を含む歪曲円形画像が取り込まれ、当該画像を解析することにより各参加者の発話状態の認識が行われる。そのため、撮影装置は1台で十分であり、特別な指向性マイクなども不要になり、装置構成の単純化を図ることができ、装置の小型化にも貢献することになる。

【0038】

そして、取り込まれた歪曲円形画像から、それぞれ各参加者の画像が切り出され、平面正則画像に変換した上で顔領域が認識され、更に唇近傍領域の設定が行われる。しかも、各参加者の画像の切出条件は、ユーザの操作入力に基づいて初期設定がなされた後、認識された顔領域に基づいて自動的に再設定される。このため、平面正則画像は、常に、唇の動きを認識するために好適な位置から切り出された歪みの少ない画像になり、各参加者の発話状態を正確に認識することが可能になる。

【0039】

特に、発話情報として各参加者の発話時間を表示させる実施形態を採れば、各参加者の発言度が明確になり、司会者等が会議を進行する上での目安として役立つ貴重な情報提示が可能になる。

【図面の簡単な説明】

【0040】

【図1】本発明の基本的な実施形態に係る会議支援システムの構成を示すブロック図である。

【図2】図1に示す会議支援システムを実際の会議室に設置した利用例を示す図である。

【図3】図1に示す会議支援システムにおいて、広角撮影装置100によって取り込まれた歪曲円形画像Sの具体例を示す平面図である。

【図4】図3に示す歪曲円形画像Sについて設定された切出条件（切出位置P（ x_0 ， y_0 ）および切出方向J）の一例を示す平面図である。

【図5】図4に示す切出条件に基づいて歪曲円形画像Sの一部を切り出し、これを変換することによって得られた平面正則画像Tの一例を示す平面図である。

【図6】歪曲円形画像Sの一部を切り出して平面正則画像Tに変換する基本原理を説明する斜視図である。

【図7】図1に示す会議支援システムにおいて、初期切出条件設定部245によってディスプレイ装置300に表示された初期設定画面の一例を示す平面図である。

【図8】図1に示す会議支援システムにおいて、会議画像提示部290によってディスプレイ装置300の画面に各参加者の人物画像を表示した状態を示す平面図である。

【図9】図1に示す会議支援システムにおいて、顔領域認識部250によって平面正則画像T上で顔領域Fの認識を行った例を示す平面図である。

【図10】図9に示す顔領域Fに基づいて、再設定基準点 $T_r(u_r, v_r)$ を唇の位置に設定した一例を示す平面図である。

【図11】図3に示す歪曲円形画像Sから、4人の参加者の画像を切り出すための切出条件の一例を示す平面図である。

【図12】図11に示す切出条件によって切り出された部分画像に基づいて作成された各参加者の人物画像（平面正則画像）を示す平面図である。

10

20

30

40

50

【図 1 3】図 1 に示す会議支援システムの唇近傍領域設定部 2 6 0 によって、顔領域 F 内に唇近傍領域 L を設定した一例を示す平面図である。

【図 1 4】図 1 3 に示す唇近傍領域 L の設定方法の一例を示す平面図である。

【図 1 5】図 1 3 に示す唇近傍領域 L の設定方法の別な一例を示す平面図である。

【図 1 6】図 1 に示す会議支援システムの特徴量算出部 2 7 0 によって算出された特徴量の技術的な意義を説明する平面図である。

【図 1 7】図 1 に示す会議支援システムの特徴量算出部 2 7 0 によって、唇近傍領域 L についての特徴量を算出するための数式の例をいくつか示す図である。

【図 1 8】図 1 に示す会議支援システムの唇近傍領域設定部 2 6 0 によって認識された唇領域 K のバリエーションを示す平面図である。

10

【図 1 9】図 1 に示す会議支援システムの発話状態認識部 2 8 0 によって、発話状態を認識する基本原理を示すグラフである。

【図 2 0】図 1 9 に示す基本原理に基づいて発話状態を認識するために、特徴量の変極点を決定する作業を示すグラフである。

【図 2 1】図 2 0 に示す方法で決定した変極点に基づいて、発話期間の認識を行う例を示すグラフである。

【図 2 2】図 1 に示す会議支援システムにおいて、会議画像提示部 2 9 0 によってディスプレイ装置 3 0 0 に表示された第 1 の会議画像表示画面 3 3 0 を示す平面図である。

【図 2 3】図 1 に示す会議支援システムにおいて、会議画像提示部 2 9 0 によってディスプレイ装置 3 0 0 に表示された第 2 の会議画像表示画面 3 4 0 を示す平面図である。

20

【図 2 4】図 1 に示す会議支援システムにおいて、会議画像提示部 2 9 0 によってディスプレイ装置 3 0 0 に表示された第 3 の会議画像表示画面 3 5 0 を示す平面図である。

【図 2 5】図 1 に示す会議支援システムにおいて、会議画像提示部 2 9 0 によってディスプレイ装置 3 0 0 に表示された第 4 の会議画像表示画面 3 6 0 を示す平面図である。

【図 2 6】図 1 に示す会議支援システムにおいて、会議画像提示部 2 9 0 によってディスプレイ装置 3 0 0 に表示された第 5 の会議画像表示画面 3 7 0 を示す平面図である。

【図 2 7】図 1 に示す会議支援システムを複数台備えたテレビ会議支援システムにおいて、ディスプレイ装置 3 0 0 に表示された複数の会議画像表示画面 3 8 0 , 3 9 0 を示す平面図である。

【図 2 8】図 1 に示す会議支援システムにおける撮影画像処理装置 2 0 0 の処理手順を示す流れ図である。

30

【発明を実施するための形態】

【0 0 4 1】

以下、本発明を図示する実施形態に基づいて説明する。

【0 0 4 2】

<<< § 1 . 本発明に係る会議支援システムの基本構成 >>>

図 1 は、本発明の基本的な実施形態に係る会議支援システムの構成を示すブロック図である。図示のとおり、この装置は、広角撮影装置 1 0 0、撮影画像処理装置 2 0 0、ディスプレイ装置 3 0 0 を備えた会議支援システムである。

【0 0 4 3】

40

広角撮影装置 1 0 0 は、広範囲な画角をもった撮影により、会議の各参加者の顔を含む歪曲円形画像を取り込む撮影装置であり、たとえば、魚眼レンズや全方位ミラーなどの光学系を装着することにより、3 6 0 ° の全方位の画像を撮影できる装置が実用化されている。魚眼レンズは複数のレンズの組み合わせによって構成され、屈折を利用して周囲からの入射光を撮像面に導く働きをする。一方、全方位ミラーは、反射を利用して周囲からの入射光を撮像面に導く働きをする。いずれも、半球状の視野内の外景からの光を集光して結像面に歪曲円形画像 S を生成する光学系であり、撮影によって得られる画像は、通常の正則矩形画像ではなく、歪曲した円形の画像になる。

【0 0 4 4】

この広角撮影装置 1 0 0 は、動画撮影の機能を有しており、たとえば、3 0 フレーム /

50

秒といった撮影周期で、時間軸に沿って歪曲円形画像を連続撮影することができる。こうして撮影された複数の歪曲円形画像 S は、撮影画像処理装置 200 に取り込まれ、後述する所定の画像処理が施されることになる。この画像処理の目的は、取り込んだ歪曲円形画像 S に基づいて、各参加者の人物画像および発話情報を作成することにある。撮影画像処理装置 200 で作成された各参加者の人物画像および発話情報は、ディスプレイ装置 300 に与えられ、ディスプレイ画面上でユーザに提示される。

【0045】

図 2 は、この会議支援システムを実際の会議室に設置した利用例を示す図である。この例では、A 氏、B 氏、C 氏、D 氏の 4 名が、会議の参加者としてテーブルを囲むように着座しており、テーブル上には、広角撮影装置 100 およびマイクロフォン 150 が置かれて

10

【0046】

撮影画像処理装置 200 は、図 1 において、一点鎖線で囲って示す範囲内に描かれているブロック 210 ~ 290 によって構成されている。実際には、この撮影画像処理装置 200 は、コンピュータに専用の画像処理用プログラムを組み込むことにより構成することができ、図示されている個々のブロックは、そのようなコンピュータを個々の機能要素として捉えたものである。

【0047】

なお、ノートパソコンを利用して撮影画像処理装置 200 を構成した場合、当該ノートパソコンのディスプレイをそのままディスプレイ装置 300 として利用することができるので、結局、図 2 に示す撮影画像処理装置 200 およびディスプレイ装置 300 は、1 台のノートパソコンによって構成することができる。一方、マイクロフォン 150 も、ノートパソコンに内蔵されたものをそのまま利用することができる。したがって、実用上は、図 2 に示す会議支援システムは、撮影画像処理装置 200 としての機能を実行させるための専用のプログラムをインストールしたノートパソコン 1 台に、広角撮影装置 100 を接続した単純な構成によって実現可能である。

20

【0048】

また、図 2 では、広角撮影装置 100 およびマイクロフォン 150 を撮影画像処理装置 200 に対して有線接続した例が示されているが、もちろん、これらを実線接続するようにしてもかまわない。たとえば、Bluetooth (登録商標) などの無線通信を利用して、これらの間の接続を行うようにすれば、テーブルの上が繁雑になるのを避けることができる。

30

【0049】

もちろん、この撮影画像処理装置 200 を、半導体集積回路によって構成することも可能である。撮影画像処理装置 200 を 1 チップの LSI によって構成すれば、当該チップを広角撮影装置 100 の筐体内に組み込むことも容易である。

【0050】

なお、実用上は、図 2 に示す例のように、マイクロフォン 150 を用いて、各参加者の画像とともに音声を収録するのが好ましいが、本発明に係る会議支援システムでは、マイクロフォン 150 は必須の構成要素ではない。これは、本発明では、各参加者の唇の動きを解析して発話状態の認識を行うので、発話状態の認識に音声の収録は必要ないためである。したがって、図 1 に示す実施形態では、マイクロフォン 150 および音声を取り扱う構成要素は省略されており、ディスプレイ装置 300 には、音声なしの動画画像のみが提示される実施例になっている。もちろん、実用上は、この図 1 に示す装置に、更にマイクロフォン 150 を付加し、撮影画像処理装置 200 内に、音声信号を取り込んでディスプレイ装置 300 に提供する音声処理部を設けるようにするのが好ましい。

40

【0051】

以下、撮影画像処理装置 200 を、専用のプログラムをインストールしたノートパソコン

50

ンによって構成した実施例について、図 1 に各ブロックとして示された個々の構成要素の機能を説明する。

【 0 0 5 2 】

まず、歪曲円形画像格納部 2 1 0 は、広角撮影装置 1 0 0 によって時間軸に沿って撮影された複数の歪曲円形画像 S を順次格納する構成要素であり、コンピュータ用のメモリやハードディスク装置といった記憶装置によって構成される。前述したように、広角撮影装置 1 0 0 が、3 0 フレーム / 秒といった撮影周期で連続撮影する機能をもった全方位カメラであった場合、1 秒ごとに 3 0 枚の歪曲円形画像 S が取り込まれることになるので、そのような画像を必要な時間分だけ格納可能な容量をもった記憶装置を用意しておくようにする。もちろん、ディスプレイ装置 3 0 0 にリアルタイムで画像表示した後、会議の映像を保存しておく必要がない場合は、後述する発話状態の認識処理に必要なフレーム分を記憶するための必要最低限の記憶容量を確保しておき、不要になった過去のフレームについては逐次消去してゆくようにしてもかまわない。

10

【 0 0 5 3 】

魚眼レンズや全方位ミラーなどの光学系を用いた撮影によって得られる画像は、3 6 0 ° の画角をもった全方位画像になるが、当該画像自身は幾何学的な歪みを生じた円形の画像になる。そこで、本願では、広角撮影装置 1 0 0 で撮影された画像を「歪曲円形画像」と呼んでいる。図 3 は、図 2 に例示する 4 人による会議の様子を、テーブル中央に配置された広角撮影装置 1 0 0 によって撮影することによって得られた歪曲円形画像 S の一例を示す図である。

20

【 0 0 5 4 】

ここでは、説明の便宜上、この歪曲円形画像 S の中心点に原点 O をとり、図の右方向に X 軸、上方向に Y 軸をそれぞれとった二次元 X Y 直交座標系を定義する。図 2 に例示する撮影環境下では、図 3 に示す歪曲円形画像 S の中心に位置する原点 O は、広角撮影装置 1 0 0 の真上の天井面の 1 点に対応し、円形テーブルを取り巻いて着座した 4 名の参加者 A ~ D は、原点 O を取り巻く位置に写ることになる。この歪曲円形画像 S は、原点 O を中心とした半球面上の画像を X Y 平面に幾何学的に投影した画像に相当するため、図示のとおり、個々の人物画像は歪みを生じたものになる。特に、頭頂部（円の中心に近い部分）は三角形に尖るような変形を受け、胸の部分（円の外周に近い部分）は半径方向に圧縮された状態になる。

30

【 0 0 5 5 】

したがって、ディスプレイ画面上に個々の人物画像を表示するためには、この歪曲円形画像 S から、個々の人物に対応する一部分を切り出し、歪みを取り除くための幾何学的な画像変換を行う必要がある。たとえば、B 氏の人物画像をディスプレイの画面上に表示する場合は、図 4 に示すように、歪曲円形画像 S から B 氏の人物に対応する一部分を切出領域 E として切り出し、この切り出した画像に対して幾何学的な変換処理を施すことにより、図 5 に示すような歪みのない正則画像を得る必要がある。

【 0 0 5 6 】

図 4 に示す歪曲円形画像 S が、もともとは半球面上に形成された画像であるのに対して、図 5 に示す画像は、通常の平面上に形成された画像ということになる。そこで、本願では、図 4 に示す「歪曲円形画像 S」に対比させて、図 5 に示す画像を「平面正則画像 T」と呼ぶことにする。平面正則画像 T の輪郭形状は、必ずしも矩形である必要はないが、ディスプレイ画面に表示することを考慮すると、実用上は、矩形とするのが好ましい。そこで、以下、平面正則画像 T を矩形枠内の画像とする実施例を説明する。

40

【 0 0 5 7 】

結局、図 4 に示す扇形の切出領域 E 内の画像に対して幾何学的変換を行うことにより、図 5 に示す矩形形状の平面正則画像 T が得られることになる。ここでは、説明の便宜上、この平面正則画像 T については、図示のとおり、中心点に原点 G（×印で示す）をとり、図の右方向に U 軸、下方向に V 軸をそれぞれとった二次元 U V 直交座標系を定義する。したがって、切出領域 E 内の歪曲画像を矩形形状の平面正則画像 T に変換するプロセスは、二次

50

元 XY 座標系上に定義された画像を、二次元 UV 座標系上に定義された画像に変換する座標変換のプロセスということになる。

【0058】

図1に示す画像切出変換部220は、このような座標変換のプロセスを利用して、歪曲円形画像格納部210に格納されている歪曲円形画像Sの一部分を切り出して変換し、各参加者についての平面正則画像Tを生成する処理を行う構成要素である。もっとも、歪曲円形画像Sからその一部を切り出すためには、所定の切出条件を設定しておく必要がある。具体的な切出条件としては、「歪曲円形画像Sのどの位置から、どの方向に、どの倍率で切り出すか」という3条件(位置、方向、倍率)が必要である。

【0059】

そこで、ここで述べる実施形態の場合、切出条件として、「切出中心点P」、「切出方向J」、「切出倍率m」という3つのパラメータを定めるようにしている。図4に示す扇形の切出領域Eは、この3つのパラメータに基づいて歪曲円形画像S上に定義された領域である。図に×印で示す切出中心点Pは、 XY 座標系上の座標点 $P(x_0, y_0)$ で示される任意の点であり、この切出中心点Pの周囲が切り出しの対象領域ということになる。この実施例の場合、歪曲円形画像S上の切出中心点 $P(x_0, y_0)$ が、平面正則画像T上の原点Gに一致するような変換が行われるので、図5において、原点Gは平面正則画像Tの中心点になっている。

【0060】

切出方向Jは、平面正則画像Tの向きを定めるパラメータであり、ここで述べる実施形態の場合、切出方向Jは、U軸方向(平面正則画像T上での右方向)を示すパラメータになっている。図5に示す平面正則画像Tでは、人物画像が正しい向きに配置された正立像になっているが、これは、図4に示す歪曲円形画像S上において、切出方向Jを図示の方向(原点Oを中心として切出中心点 $P(x_0, y_0)$ を通る円についての反時計回りの接線方向)に設定したためである。切出方向Jを逆転させると、得られる平面正則画像Tは人物の天地が逆転したものになり、切出方向Jの向きを 90° 回転させると、得られる平面正則画像Tは人物が横向きのものになる。

【0061】

切出倍率mは、人物画像の拡大率(縮小率)を定めるパラメータであり、倍率mが大きくなればなるほど、平面正則画像T上で人物は拡大された状態になる。別言すれば、倍率mが大きくなればなるほど、切出領域Eの面積は小さくなり、当該小さい面積をもった領域内の画像が拡大して表示されることになる。

【0062】

切出条件設定部240は、上述した「切出中心点P」、「切出方向J」、「切出倍率m」という3つのパラメータをもった切出条件を設定する構成要素であり、画像切出変換部220は、この切出条件設定部240に設定されている切出条件に基づいて、歪曲円形画像格納部210に格納されている歪曲円形画像Sの一部分を切り出し、これに所定の幾何学的変換を施して、各参加者A~Dについての平面正則画像 $T_a \sim T_d$ を生成する機能を果たす。幾何学的変換の基本原理は、§2で説明する。

【0063】

平面正則画像格納部230は、こうして生成された平面正則画像 $T_a \sim T_d$ を格納する構成要素であり、コンピュータ用のメモリやハードディスク装置といった記憶装置によって構成される。前述したように、広角撮影装置100が、30フレーム/秒といった撮影周期で連続撮影する機能をもった全方位カメラであった場合、1秒ごとに30枚の歪曲円形画像Sが取り込まれることになるので、各参加者A~Dについての平面正則画像 $T_a \sim T_d$ も、これら個々のフレームごとに切り出されて逐次生成されることになる。したがって、平面正則画像格納部230には、そのような画像を必要な時間分だけ格納可能な容量をもった記憶装置を用意しておくようにする(もちろん、不要になった過去のフレームに関する画像については逐次消去してかまわない)。

【0064】

10

20

30

40

50

顔領域認識部 250 は、こうして平面正則画像格納部 230 に格納された各参加者についての平面正則画像 T a ~ T d について、人物の顔を構成する顔領域を認識する構成要素である。ここに示す実施形態の場合、人物の顔の部分を包摂する輪郭矩形を定め、この輪郭矩形内を顔領域として認識している。上例の場合、平面正則画像格納部 230 内には、30 フレーム / 秒の時間間隔で各参加者についての平面正則画像 T a ~ T d が順次得られるので、顔領域認識部 250 は、こうして得られた個々の平面正則画像 T a ~ T d について、それぞれ顔領域を認識する処理を施す。

【 0 0 6 5 】

なお、後に詳述するように、切出条件設定部 240 に設定される最初の切出条件は、ユーザの操作入力に基づいて、初期切出条件設定部 245 によって決定されるが、以後は、顔領域認識部 250 の認識結果に基づいて、切出条件の再設定が行われることになる。したがって、画像切出変換部 220 は、常に、最新の切出条件に基づいて、各参加者の人物画像の切出処理を行うことができる。

【 0 0 6 6 】

唇近傍領域設定部 260 は、顔領域認識部 250 が認識した顔領域上に唇近傍領域を設定する機能を果たす。この唇近傍領域は、顔画像内の唇部分およびその周囲部分を含む領域として設定される領域であり、唇の動きの有無を監視するための監視対象領域としての意味をもつ。

【 0 0 6 7 】

特徴量算出部 270 は、各参加者の平面正則画像 T a ~ T d について、それぞれ唇近傍領域内の輝度もしくは色を示す特徴量を算出する機能を果たし、発話状態認識部 280 は、こうして算出された特徴量の時間変化に基づいて、個々の参加者ごとの発話状態を認識する機能を果たす。そして、会議画像提示部 290 は、平面正則画像格納部 230 内の平面正則画像 T a ~ T d と、発話状態認識部 280 によって認識された発話状態に関連した発話情報と、を含んだ会議画像を、時間軸に沿って順次生成し、これをディスプレイ装置 300 に与えて動画として表示させる機能を果たす。なお、これらの各構成要素の具体的な処理内容については、§ 4 で詳述する。

【 0 0 6 8 】

かくして、図 1 に示す会議支援システムを用いれば、ディスプレイ装置 300 の画面上に、図 5 に示すような参加者の平面正則画像 T の動画とともに、当該参加者の発話情報（たとえば、発話期間中か否かを示す情報や発話時間を示す情報）が表示されることになる。

【 0 0 6 9 】

前述したとおり、図 2 に示すシステムでは、マイクロフォン 150 を用いて、各参加者の音声を収録しているが、本発明に係る会議支援システムでは、マイクロフォン 150 は必須の構成要素ではなく、発話情報を得るために音声情報は不要である。すなわち、1 台の広角撮影装置 100 とパソコン等によって構成される撮影画像処理装置 200 を用意するだけで、各参加者の発話情報を得ることができる。このため、装置構成の単純化を図ることができ、装置の小型化にも貢献することになる。また、切出条件設定部 240 が、顔領域認識部 250 の認識結果に基づいて、切出条件の再設定を行う機能を有しているため、常に、唇の動きを認識するのに最適な切出条件に基づいて人物画像を切り出すことができるようになり、各参加者の発話状態を正確に認識することが可能になる。

【 0 0 7 0 】

< < < § 2 . 画像の切出処理および変換処理の基本原理 > > >

続いて、ここでは、画像切出変換部 220 によって行われる画像の切出処理および変換処理の基本原理を簡単に説明しておく。図 6 は、歪曲円形画像 S の一部を切り出して平面正則画像 T に変換する基本原理を説明する斜視図であり、三次元 X Y Z 直交座標系における X Y 平面上に歪曲円形画像 S が形成されている例が示されている。ここでは、図示のとおり、Z 軸を図の上方にとり、Z 軸の正の領域側にドーム状の仮想球面 H（半球）を定義した例を示すことにする。

10

20

30

40

50

【 0 0 7 1 】

X Y 平面上に形成された歪曲円形画像 S は、座標系の原点 O を中心とした半径 R の円を構成する画像であり、Z 軸の正の領域側における 180° の画角をもった領域に存在する像を歪ませて記録したものに相当する。図 2 に例示するような撮影環境で、魚眼レンズを備えた広角撮影装置 100 をテーブル上に配置して撮影した場合、Z 軸は鉛直軸となり、たとえば、図 3 に例示したような歪曲円形画像 S が得られる。この歪曲円形画像 S には、Z 軸の正の領域側に存在するすべての像が記録されることになるが、その中心部分と周囲部分とでは、像の縮尺倍率が異なっており、記録された像の形状は歪んだものになる。特に、円周部分ほど画像は圧縮された状態になる。

【 0 0 7 2 】

実際の魚眼レンズは、複数の凸レンズや凹レンズを組み合わせた光学系によって構成されるが、その光学的な特性は、図 6 に示すような仮想球面 H によってモデル化できることが知られている。すなわち、歪曲円形画像 S の上面に、半径 R をもったドーム状の仮想球面 H (半球) を配置したモデルを考えれば、正射影方式の魚眼レンズの光学的特性は、仮想球面 H 上の任意の点に対して法線方向から入射する入射光線は、Z 軸に平行な入射光線として、X Y 平面向かう振る舞いをする、と考えてよい。逆言すれば、図 6 における歪曲円形画像 S 上の任意の点 $S_i(x_i, y_i)$ は、仮想球面 H 上の点 $Q_i(x_i, y_i, z_i)$ に対して、原点 O を通る一点鎖線 n_i に沿って入射した光線を、Z 軸に平行な方向に屈曲させたときの X Y 平面との交点であり、この点 $S_i(x_i, y_i)$ に位置する画素は、一点鎖線 n_i の延長線上に存在する物体上の 1 点を示していることになる。

【 0 0 7 3 】

結局、歪曲円形画像 S は、本来は平面上の画像ではなく、仮想球面 H という湾曲面上に形成された画像を、X Y 平面上に投影して得られる画像ということになる。§ 1 で述べたとおり、画像切出変換部 220 は、この歪曲円形画像 S 上の一部分を切り出して、平面正則画像 T に変換する処理を行うことになるが、そのときの切出条件の 1 つとして、X Y 平面上の 1 点として指定された切出中心点 $P(x_0, y_0)$ が用いられる。

【 0 0 7 4 】

図 6 に示すモデルにおいて、歪曲円形画像 S は、三次元 X Y Z 直交座標系の X Y 平面上に定義されているので、歪曲円形画像 S 自身は、二次元 X Y 直交座標系上に定義された画像である。そこで、この歪曲円形画像 S 上に定義された切出中心点 $P(x_0, y_0)$ を通り Z 軸に平行な直線と仮想球面 H との交点 $Q(x_0, y_0, z_0)$ を考える。この交点 Q は、いわば切出中心点 $P(x_0, y_0)$ の真上にある球面上対応点であり、その位置座標は (x_0, y_0, z_0) である。

【 0 0 7 5 】

次に、原点 O から球面上対応点 $Q(x_0, y_0, z_0)$ を通って外方へと向かう視線ベクトル n (点 Q における仮想球面 H の法線方向を示すベクトル) を定義し、この視線ベクトル n 上において、原点 O からの距離が半径 R の m 倍となる位置に点 $G(x_g, y_g, z_g)$ を定義し、図示のとおり、この点 $G(x_g, y_g, z_g)$ を原点とする二次元 U V 直交座標系を定める。そして、平面正則画像 T を、この二次元 U V 直交座標系上の画像として求めることにする。図 5 に示す平面正則画像 T は、このようにして得られた U V 平面上の画像である。

【 0 0 7 6 】

U V 座標系の原点となる点 $G(x_g, y_g, z_g)$ の位置は、図示のとおり、方位角と天頂角とによって特定することができる。ここで、方位角 $(0 < 360^\circ)$ は、切出中心点 $P(x_0, y_0)$ と X Y 座標系の原点 O とを結ぶ直線と Y 軸とのなす角であり、天頂角 $(0 < 90^\circ)$ は、U V 座標系の原点となる点 $G(x_g, y_g, z_g)$ と X Y 座標系の原点 O とを結ぶ直線と Z 軸とのなす角(鋭角)である。

【 0 0 7 7 】

このように、U V 平面は、方位角と天頂角と倍率 m とを指定することによって特定することができるが、U V 座標系を決定するには、更にもう 1 つの角度を指定する必要

10

20

30

40

50

がある。この角度 θ は、直線 OG を回転軸とした UV 座標系の向きを示すパラメータであり、図 6 の例では、 U 軸と D 軸とのなす角度として定義されている。ここで、 D 軸は、点 $G(x_g, y_g, z_g)$ を通り、 XY 平面に平行かつ直線 OG に直交する軸である。角度 θ は、 UV 座標系において、 U 軸方向を向いたベクトル U と D 軸方向を向いたベクトル D と、を定義したときに、ベクトル U とベクトル D とのなす角度として定義される角であり、通常、「平面傾斜角」と呼ばれている。

【0078】

結局、図 6 に示す平面正則画像 T を形成するための UV 座標系の位置および向きは、方位角 ϕ 、天頂角 θ 、平面傾斜角 α という 3 つの角度からなるパラメータと倍率 m を設定することにより一義的に決定される。この 3 つの角度は、一般にオイラー角と呼ばれている。切出条件設定部 240 において設定される切出条件は、前述したとおり、「切出中心点 $P(x_0, y_0)$ 」、「切出方向 J 」、「切出倍率 m 」という 3 つのパラメータによって構成されているが、これらのパラメータにより、上記オイラー角は一義的に決定される。

10

【0079】

すなわち、図 6 に示すように、歪曲円形画像 S 上で切出中心点 $P(x_0, y_0)$ の位置を決めてやれば、球面上対応点 $Q(x_0, y_0, z_0)$ が定まり、視線ベクトル n が定まるので、方位角 ϕ および天頂角 θ が決定される。また、切出倍率 m を決めてやれば、視線ベクトル n 上において原点 O から距離 $m \cdot R$ だけ離れた点として、 UV 座標系の原点 $G(x_g, y_g, z_g)$ が決定される。更に、図 4 に示す例のように、歪曲円形画像 S 上で切出方向 J の向きを決めてやれば、図 6 において、原点 $G(x_g, y_g, z_g)$ を通り切出方向 J に平行なベクトルとしてベクトル D を定義することができるので、平面傾斜角 α が決定される。図 4 に示す例のように、原点 O を中心として切出中心点 $P(x_0, y_0)$ を通る円についての接線方向に切出方向 J を設定すると、平面傾斜角 $\alpha = 0$ に設定することができ、ベクトル D は U 軸方向を向くことになる。

20

【0080】

こうして、オイラー角 ϕ 、 θ 、 α が決まると、二次元 XY 直交座標系と二次元 UV 直交座標系との間の幾何学的な関係が定まるので、 XY 座標系上における任意の点 $S_i(x_i, y_i)$ と UV 座標系上における対応点 $T_i(u_i, v_i)$ との間の幾何学的な対応関係も定まることになる。画像切出変換部 220 が行う変換処理は、 XY 座標系上に定義された歪曲円形画像 S 上の切出中心点 $P(x_0, y_0)$ を中心とした切出領域 E 内の歪曲画像を切り出して変形し、 UV 座標系上に平面正則画像 T を得ることにある。具体的には、図 6 に示すように、 UV 座標系上に得られる平面正則画像 T 上の 1 点 $T_i(u_i, v_i)$ に位置する画素の画素値を、これに対応する XY 座標系上の 1 点 $S_i(x_i, y_i)$ の近傍に位置する画素の画素値に基づいて決定することである。そのためには、座標 (u_i, v_i) と座標 (x_i, y_i) との対応関係を示す対応関係式が必要になる。

30

【0081】

このような対応関係式は、オイラー角 ϕ 、 θ 、 α および倍率 m を用いた三角関数を用いて一義的に定義することができ、画像切出変換部 220 は、この対応関係式を用いて、座標 (u_i, v_i) と座標 (x_i, y_i) との間の変換を行い、平面正則画像 T を生成する。なお、具体的な対応関係式については、たとえば、特許第 3012142 号公報や特許第 3012142 号公報等、様々な文献に開示されている公知事項であるため、ここでは詳しい説明は省略する。

40

【0082】

また、図 6 に示すモデルは、 XY 座標系上の歪曲円形画像 S から UV 座標系上の平面正則画像 T を生成する一手法の原理を示すものであり、本発明における画像切出変換部 220 が行う変換処理は、この図 6 に示すモデルに基づく方法に限定されるものではない。たとえば、特開 2010-062790 号公報や前掲の特開 2012-089954 号公報（特許文献 4）には、 UV 座標系を平面上に定義する代わりに、円柱面状に定義することにより、より歪みの少ない平面正則画像 T を生成する技術が開示されている。本発明における画像切出変換部 220 は、このように円柱面上に定義した UV 座標系を用いた変換処

50

理を行ってもかまわない。

【 0 0 8 3 】

< < < § 3 . 画像の切出処理および変換処理の詳細 > > >

ここでは、図 1 に示す会議支援システムの基本動作のうち、画像の切出処理および変換処理の詳細を具体例に即して説明する。§ 1 で述べたとおり、この会議支援システムを図 2 に示すような形態で利用すると、広角撮影装置 1 0 0 によって、図 3 に例示するような歪曲円形画像 S が撮影される。ユーザは、まず、このような歪曲円形画像 S を利用して、会議の個々の参加者を登録する作業を行う。この登録作業は、初期切出条件設定部 2 4 5 を利用して、切出条件設定部 2 4 0 に対して最初の切出条件を設定する作業に他ならない。

10

【 0 0 8 4 】

図 7 は、図 1 に示す会議支援システムにおいて、初期切出条件設定部 2 4 5 によってディスプレイ装置 3 0 0 に表示された初期設定画面 3 1 0 の一例を示す平面図である。図示のとおり、この初期設定画面 3 1 0 には、初期切出位置指定画面 3 1 1 と参加者登録画面 3 1 2 とが設けられている。初期切出位置指定画面 3 1 1 には、歪曲円形画像格納部 2 1 0 内に格納されている最新の歪曲円形画像 S (最新のフレーム)が表示される。結局、初期切出位置指定画面 3 1 1 には、会議の各参加者の様子がリアルタイムの動画として提示されることになる。

【 0 0 8 5 】

ユーザは、この初期切出位置指定画面 3 1 1 に表示されている歪曲円形画像 S 内の 1 点を指定することにより、1 人の参加者を登録することができる。参加者登録画面 3 1 2 は、このような参加者の登録作業を補助するための画面であり、図示のとおり、参加者名登録窓 3 1 3、参加者戻しボタン 3 1 4、登録完了ボタン 3 1 5、参加者送りボタン 3 1 6 が配置されている。

20

【 0 0 8 6 】

参加者送りボタン 3 1 6 は、第 i 番目の参加者の登録画面を第 $(i + 1)$ 番目の参加者の登録画面に送る操作を行うためのボタンであり、図示のように、第 1 番目の参加者の登録画面が表示されているときにマウスなどの入力機器でクリックすると、第 2 番目の参加者の登録画面が表示されることになる。一方、参加者戻しボタン 3 1 4 は、第 i 番目の参加者の登録画面を第 $(i - 1)$ 番目の参加者の登録画面に戻す操作を行うためのボタンである。ユーザは、ボタン 3 1 4、3 1 6 をクリックすることにより、任意番目の参加者の登録画面を表示させることができる。

30

【 0 0 8 7 】

特定の参加者を登録するには、参加者名登録窓 3 1 3 に登録対象となる参加者の氏名を入力し、初期切出位置指定画面 3 1 1 に表示されている歪曲円形画像 S 内の当該参加者の人物画像をマウスなどの入力機器でクリックすればよい。たとえば、第 1 番目の参加者として「A 氏 (左側の女性)」を登録するのであれば、図示の例のように、参加者名登録窓 3 1 3 に「A 氏」なる氏名を入力し、歪曲円形画像 S 内の A 氏の人物画像 (顔のあたりでよい) をクリックすればよい。初期切出条件設定部 2 4 5 は、歪曲円形画像 S 上のクリック位置を初期の切出中心点 $P(x_0, y_0)$ の位置として取り込み、これを切出条件設定部 2 4 0 に設定する。こうして、参加者全員 (図示の例の場合、A 氏 ~ D 氏の 4 名) についての登録作業が完了したら、登録完了ボタン 3 1 5 をクリックすれば、登録作業は完了である。

40

【 0 0 8 8 】

上述したとおり、この登録作業は、各参加者の人物画像についての初期切出条件を設定する作業に他ならない。結局、初期切出条件設定部 2 4 5 は、ディスプレイ装置 3 0 0 の画面上 (図 7 の例の場合は、初期切出位置指定画面 3 1 1) に、歪曲円形画像格納部 2 1 0 に格納されている歪曲円形画像 S を表示させ、表示中の歪曲円形画像 S 上で切出中心点の位置を指定するユーザの操作入力に基づいて、切出条件設定部 2 4 0 に対して、各参加者についての初期切出条件を設定することになる。

50

【 0 0 8 9 】

なお、切出条件を示すパラメータとしては、切出中心点 $P(x_0, y_0)$ の位置座標とともに、切出方向 J および切出倍率 m が必要である。そこで、ここに示す実施形態の場合、切出方向 J および切出倍率 m については、初期切出条件設定部 245 内に予めデフォルト値を格納しておき、初期切出条件としては、図 7 に示す初期設定画面 310 上でのユーザの操作入力（人物画像上の 1 点をクリックする入力）により決定された初期切出中心点 $P(x_0, y_0)$ と、デフォルト値として設定されている初期切出方向 J および初期切出倍率 m とを、切出条件設定部 240 に対して初期切出条件として設定するようにしている。

【 0 0 9 0 】

デフォルトの初期切出方向 J としては、図 4 に示す例のように、原点 O を中心として切出中心点 $P(x_0, y_0)$ を通る円についての反時計回りの接線方向を採用すればよい。図 2 に例示するように、広角撮影装置 100 を水平なテーブル面に載置する利用環境を前提とすれば、図 4 に示すような切出方向 J を設定し、この切出方向 J が U 軸方向となるような平面正則画像 T を生成すれば、図 5 に示すような人物の正立像が得られることになる。

10

【 0 0 9 1 】

一方、初期切出倍率 m としては、図 2 に例示するような一般的な会議における撮影環境を想定して、広角撮影装置 100 と各参加者との標準的な距離を定めておき、この標準的な距離に着座している参加者の上半身が平面正則画像 T として適切な倍率で表示されるような値をデフォルト値として定めておけばよい。

20

【 0 0 9 2 】

こうして、初期切出条件設定部 245 によって、切出条件設定部 240 に対する初期切出条件の設定処理が完了すれば、画像切出変換部 220 は、歪曲円形画像格納部 210 内に格納されている歪曲円形画像 S から、当該初期切出条件を用いて各参加者 $A \sim D$ の人物画像を切り出し、これを変換して平面正則画像 $T_a \sim T_d$ を生成することができ、会議画像提示部 290 は、こうして生成された平面正則画像 $T_a \sim T_d$ を各参加者 $A \sim D$ の人物画像としてディスプレイ装置 300 の画面上に表示することができる。

【 0 0 9 3 】

図 8 は、このようにしてディスプレイ装置 300 に表示された人物画像表示画面 320 の一例を示すものである。各平面正則画像 $T_a \sim T_d$ が各参加者 $A \sim D$ の上半身の人物画像になっており、その下には、図 7 に示す初期設定画面 310 で登録された各参加者の氏名（ A 氏～ D 氏）も併せて表示されている（実際には、後述するように、会議画像提示部 290 は、この人物画像表示画面 320 に、更に、各参加者の発話情報を付加した会議画像（たとえば、図 22 参照）を表示することになる）。

30

【 0 0 9 4 】

このように、最初の段階において、ユーザによる初期切出条件の設定操作を行っておけば、以後は、§ 1 で説明したとおり、各平面正則画像 $T_a \sim T_d$ から顔領域の認識が行われ、その認識結果に基づいて、切出条件の再設定が行われることになる。したがって、図 7 に示す初期設定画面 310 での登録位置（ユーザによるクリック位置）は、必ずしも顔の中央点といった正確な位置である必要はない。これは、切出条件の再設定が、認識された顔領域に基づいて行われるため、再設定された切出条件に基づいて再度の切出処理を行えば、顔領域を基準とした適切な位置で切り出された平面正則画像 T が得られるようになるためである。

40

【 0 0 9 5 】

このような切出条件の再設定処理を具体例を挙げながら説明しよう。たとえば、図 7 に示す初期設定画面 310 において、第 2 番目の参加者 B 氏の位置を登録するためにユーザがクリックした際に、クリック位置が顔の中心から若干外れて、髪の毛の位置をクリックしてしまった場合を考えてみる。このように、初期設定時のクリック位置が顔から若干ずれてしまうと、最初に得られる平面正則画像 T では、枠内の人物の割付位置が若干不適切

50

になる。図9は、B氏の髪の毛の位置をクリックしてしまった場合に得られる平面正則画像Tを示している。図に白い×印で示す点Gは、初期設定画面310において、ユーザがB氏の位置登録を行う際にクリックした位置に相当する。前述したとおり、初期切出条件設定部245は、クリック位置を初期の切出中心点P(x_0, y_0)の位置として設定するため、図9に示すように、このクリック位置を原点G(中心点)とした平面正則画像Tが生成され、平面正則画像格納部230に格納される。

【0096】

一方、顔領域認識部250は、この平面正則画像Tとして与えられた人物画像について、所定のアルゴリズムに基づく解析処理を実行し、顔領域Fを認識する。図9では、太い矩形枠内の領域として、顔領域Fの認識結果が例示されている。

10

【0097】

人物画像を解析して顔領域Fを抽出する方法としては、いくつかのアルゴリズムが知られている。たとえば、Haar-like検出器として知られている顔領域抽出アルゴリズムでは、顔の大きな明暗を特徴量として、学習データとの合致度を示す評価値を算出し、この評価値に基づいて顔を判定する手順が実行される。

【0098】

一方、HOG特徴量に基づく顔領域抽出アルゴリズムとして知られている方法では、人物の肩口までの画像を、正面、側面、背面に関して学習させ、この学習画像の局所領域におけるエッジの方向をHOG特徴量として算出し、解析対象となる画像について、学習した特徴量に近い特徴量をもつものを人物と認識し、その上方部分を顔領域と認識する手順が実行される。

20

【0099】

このように、人物画像について顔領域を認識する方法は様々なアルゴリズムが公知であり、既にデジカメ画像用の処理プログラムなどで実用化されているため、ここでは平面正則画像Tに基づいて顔領域Fを認識する具体的な方法についての説明は省略する。また、ここでは、説明の便宜上、図9に示す例のように、人物画像の髪の毛の部分も含めた頭部全体の外接矩形を顔領域Fの輪郭枠として抽出する例を述べることにするが、たとえば、髪の毛の部分を含まない肌の色の領域部分の外接矩形を顔領域Fの輪郭枠として抽出するアルゴリズムを用いてもかまわない。もちろん、顔領域Fの輪郭枠は、必ずしも矩形である必要はない。ただ、ここでは、基本的な実施形態として、矩形枠内の領域が顔領域Fとして抽出された場合を例にとって以下の説明を行うことにする。

30

【0100】

顔領域認識部250によって顔領域Fの認識が行われると、切出条件設定部240は、この顔領域Fの輪郭枠を基準として予め定められた所定位置にある点を再設定基準点 T_r と定め、平面正則画像T上の当該再設定基準点 T_r を元の歪曲円形画像S上に逆変換して得られる点を新たな切出中心点Pとする処理を行う。たとえば、図9に示す例のように、顔領域Fが矩形状の領域として抽出された場合、この矩形枠の中心点を再設定基準点 T_r と定めて新たな切出中心点Pを設定すれば、次のフレームでは、顔の中心点を切出中心点Pとする切り出しが行われることになる。

【0101】

ただ、ここに示す実施形態では、顔領域Fの中心を再設定基準点 T_r とする代わりに、唇の中心が位置すると予想される点を再設定基準点 T_r とする処理を行っている。その結果、次のフレームでは、唇の位置を切出中心点Pとする切り出しが行われることになる。

40

【0102】

図10は、図9に示す顔領域Fに基づいて、再設定基準点 $T_r(u_r, v_r)$ を唇の位置に設定した一例を示す平面図である。一般に、どのような人物画像であっても、正面向きの顔の輪郭枠に対する唇の相対位置は、ほぼ類似している。具体的には、一般的な人物の唇は、横方向に関しては、顔の輪郭枠のほぼ中心線上に位置し、縦方向に関しては、顔の輪郭枠の幅を100%としたときに、上から70~90%の付近に位置する。

【0103】

50

そこで、図10に示す例では、図に破線で示す中心鉛直線L1と唇位置水平線L2との交点を、唇の中心が位置すると予想される点と考え、再設定基準点 $T_r(u_r, v_r)$ に定めている。ここで、中心鉛直線L1は、顔領域Fの輪郭枠（この例では矩形枠）の横方向に関する中心線であり、唇位置水平線L2は、顔領域Fの輪郭枠の縦方向寸法を100%としたときに、上端から80%の位置に引いた水平線である。もちろん、唇位置水平線L2の位置は、必ずしも上端から80%に設定する必要はなく、一般的には、上端から70~90%の範囲内の所定値に設定すればよい。

【0104】

要するに、この再設定基準点 $T_r(u_r, v_r)$ は、「唇の中心が位置すると予想される点」に設定すればよいので、実用上は、顔領域認識部250が顔領域Fを認識する際に採用した認識アルゴリズムを考慮して（たとえば、髪の毛の部分を含めた領域を顔領域とするか否か等の事情を考慮して）、統計的な予想点に設定するようにすればよい。

10

【0105】

もちろん、実際には、たまたま人物が横を向いた場合など、予想どおりにならない場合もあり、その場合、再設定基準点 $T_r(u_r, v_r)$ は、唇の中心位置から外れることになる。ただ、そのような場合でも、次のフレームの切出位置が、唇を中心とした位置から若干外れるだけであり、大きな問題は生じない。図10に示す例では、上記設定により、唇のほぼ中心位置に再設定基準点 $T_r(u_r, v_r)$ が定義されている。

【0106】

さて、こうして平面正則画像T上に再設定基準点 $T_r(u_r, v_r)$ が定義されたら、当該再設定基準点 $T_r(u_r, v_r)$ を歪曲円形画像S上に逆変換して対応点を求める。図10に示す例の場合、平面正則画像Tの中心点 $G(u_0, v_0)$ は、UV座標系の原点であり、 $u_0 = v_0 = 0$ であるが、再設定基準点 $T_r(u_r, v_r)$ は中心点Gから外れているため、その座標値 (u_r, v_r) は所定の値をもっている。そこで、§2で求めた変換処理により、UV座標系からXY座標系への変換を行えば、歪曲円形画像Sにおいて、再設定基準点 $T_r(u_r, v_r)$ の対応点の座標 (x_r, y_r) を求めることができる。切出条件設定部240は、当該座標 (x_r, y_r) をもった対応点を、次のフレームについての切出中心点 $P(x_0, y_0)$ とする切出条件の再設定を行う。

20

【0107】

こうして、次のフレームの歪曲円形画像Sからは、新たな切出中心点 $P(x_0, y_0)$ の位置（すなわち、前のフレームにおける唇の位置）を中心とした切り出しが行われるので、得られる平面正則画像Tは、図10に示す再設定基準点 $T_r(u_r, v_r)$ の位置を中心とした人物画像になる。したがって、図7に示す初期設定画面310でのユーザによるクリック操作が大まかな位置を示すものであっても、切出条件設定部240の再設定機能により、常に、唇の位置を中心とした切出処理が行われるような補正機能が働くことになる。

30

【0108】

もちろん、この補正機能は、会議中に参加者が動いた場合にも有効である。通常、会議中、各参加者は定位置を維持するのが一般的であるが、顔の位置はわずかながら変化するのであろうし、場合によっては、着座位置を移動させることもある。その場合でも、顔領域認識部250が、個々のフレームごとに顔領域の認識を行い、切出条件設定部240が、その認識結果に基づいて次のフレームについての切出条件を再設定するので、常に最適な切出条件を維持することができる。

40

【0109】

特に、図10に示す実施形態では、唇の中心が位置すると予想される点を再設定基準点 $T_r(u_r, v_r)$ に定め、これに対応する位置に新たな切出中心点 $P(x_0, y_0)$ を再設定しているため、常に、発話状態の認識に用いられる唇近傍領域を中心とした切り出しが可能になり、発話状態の認識処理の精度を向上させるメリットが得られる。

【0110】

図4に示す歪曲円形画像S上のB氏の人物画像と、図5に示す平面正則画像T上のB氏

50

の人物画像とを比較すると、前者に含まれていた歪みが、後者では是正されていることがわかる。しかしながら、この平面正則画像 T は、歪みを完全に是正した画像にはなっていない。これは原理的に、歪曲円形画像 S の歪みを完全に是正した平面正則画像 T を作成することができないためである。すなわち、歪曲円形画像 S は、元来、図6に示す仮想球面 H 上に形成された球面画像であるため、これを完全に平面的な画像に置き換えることはできないのである。その結果、図5に示す平面正則画像 T では、原点 G （画像の中心点：切出中心点 $P(x_0, y_0)$ に対応した点）の近傍については歪みは少ないが、原点 G から離れた周囲部分ほど歪みが大きくなる。

【0111】

図10に示す実施形態を採用すれば、常に唇の近傍領域が平面正則画像 T の中心に位置することになるので、唇の近傍は最も歪みが少ない領域ということになる。このため、発話状態の認識処理の対象となる唇近傍領域内の画像として、歪みの少ない画像を利用することができるようになり、発話状態の認識処理の精度を向上させることができる。

10

【0112】

図11は、図3に示す歪曲円形画像 S から、4人の参加者の画像を切り出すための切出条件の一例を示す平面図であり、図12は、この図11に示す切出条件によって切り出された部分画像に基づいて作成された各参加者 $A \sim D$ の人物画像（平面正則画像 $T_a \sim T_d$ ）を示す平面図である。いずれも、図10に例示するように、唇の中心が位置すると予想される点を再設定基準点 $T_r(u_r, v_r)$ に定めた実施形態を採用した例である。したがって、図11において白抜きの \times 印で示された参加者 $A \sim D$ についての切出中心点 $P_a \sim P_d$ の位置は、いずれもほぼ唇の中心位置に設定されており、それぞれ図示されている扇形の切出領域 $E_a \sim E_d$ から部分画像が切り出され、これらを変換することにより、図12に示すような平面正則画像 $T_a \sim T_d$ が作成される。

20

【0113】

その結果、平面正則画像 $T_a \sim T_d$ 上において、各参加者 $A \sim D$ の唇は、ほぼその中心（ UV 座標系の原点 G ）に位置し、発話状態の認識処理に利用される唇に関しては、歪みの少ない最適な画像変換が行われていることになる。こうして作成された平面正則画像 $T_a \sim T_d$ に基づいて、顔領域認識部250が顔領域 $F_a \sim F_d$ を抽出する処理を行い、切出条件設定部240が、これら顔領域 $F_a \sim F_d$ の輪郭枠（図12の例では太線で示す矩形枠）を基準として、所定位置（たとえば、図10に示されているような輪郭枠に対する

30

【0114】

なお、図11に示す歪曲円形画像 S から、各切出領域 $E_a \sim E_d$ 内の部分画像を切り出して、これを幾何学的に変換することにより各平面正則画像 $T_a \sim T_d$ を作成するためには、図12の下段に示すとおり、切出中心点、切出方向、切出倍率という3つのパラメータからなる切出条件が必要である。たとえば、参加者 B の平面正則画像 T_b を作成するためには、切出中心点 P_b 、切出方向 J_b 、切出倍率 m_b という3つのパラメータを定める必要がある。

40

【0115】

ここで、切出中心点 P_b は、図11に白抜きの \times 印で示された点であり、前のフレームについて得られた再設定基準点 $T_r(u_r, v_r)$ の対応点として定まる点であることは、既に述べたとおりである。一方、ここに示す実施形態の場合、切出方向 J_b および切出倍率 m_b としては、予め定められたデフォルト値をそのまま利用している。たとえば、切出方向 J_b としては、図11に示すとおり、原点 O を中心として切出中心点 P_b を通る円 C_b についての反時計回りの接線方向を設定している。他の参加者についての切出方向 J_a, J_c, J_d についても同様である。

【0116】

これは、図2に例示するように、広角撮影装置100を水平なテーブル面に載置する利

50

用環境を考慮したものであり、このような利用環境を前提とする限り、上述したデフォルトの切出方向を設定することにより、図12に例示するように、人物が正立した状態の平面正則画像T_a~T_dが得られることになる。要するに、切出条件設定部240は、切出中心点P(x₀, y₀)を通る歪曲円形画像Sの同心円を求め、当該同心円の接線方向が平面正則画像Tの水平方向となるような切出方向Jをもつ切出条件を再設定すればよい。

【0117】

もちろん、必要があれば、切出条件設定部240が、顔領域認識部250の認識結果に基づいて切出方向Jの再設定を行うようにすることもできる。一般的な顔領域抽出アルゴリズムでは、両眼の位置の認識が行われるので、たとえば、「右目の中心点から左目の中心点に向かう方向を新たな切出方向Jとする」というような規則を定めておけば、切出条件設定部240は、個々のフレームごとに、顔領域認識部250の認識結果に基づいて新たな切出方向Jを設定することができる。また、ユーザの指示入力に基づいて、デフォルトの切出方向Jを変更するような運用も可能である。

10

【0118】

一方、切出倍率mとしては、広角撮影装置100と各参加者との標準的な距離を定めおき、この標準的な距離に着座している参加者の上半身が適切な倍率で表示されるような値をデフォルト値として定めている。図12に示す例では、参加者A~Dの人物像は、ほぼ理想的な倍率で表示されているが、これは、図2に示す撮影環境において、各参加者A~Dが、広角撮影装置100から標準的な距離を保って着座しているためである。実際には、広角撮影装置100から遠くに着座した参加者は縮小表示され、近くに着座した参加者は拡大表示されることになる。

20

【0119】

参加者の着座位置にかかわらず、常に最適な倍率で人物画像を表示させたい場合には、切出倍率mとしてデフォルト値を用いる代わりに、切出条件設定部240が、顔領域Fの面積に応じた切出倍率mを切出条件として再設定するようにすればよい。たとえば、平面正則画像Tの全面積に対する顔領域Fの面積の割合の基準値を40%に設定しておき、顔領域Fの面積が40%に満たなかった場合には切出倍率mを増加補正し、顔領域Fの面積が40%を超えていた場合には切出倍率mを減少補正する、というような調整を行い、顔領域Fの面積割合が常に40%を維持するように切出倍率mを再設定すればよい。もちろん、ユーザの指示入力に基づいて、切出倍率mを直接指定させたり、上記面積割合の基準値を指定させたりすることもできる。

30

【0120】

以上、広角撮影装置100が、たとえば、30フレーム/秒といった一定の時間周期で撮影されたフレーム単位の静止画からなる歪曲円形画像Sを連続して取り込み、切出条件設定部240が、1フレーム周期で切出条件の再設定を行う実施形態を述べた。この実施形態では、広角撮影装置100が撮影した個々のフレームごとに、それぞれ新たな切出条件の設定(直前のフレームについて行われた顔領域の認識結果に基づく設定)が行われることになる。

【0121】

ただ、本発明を実施する上では、必ずしも1フレーム周期で切出条件の再設定を行う必要はなく、複数フレーム周期で切出条件の再設定を行うようにしてもかまわない。たとえば、10フレームごとに切出条件の再設定を行うようにすれば、切出条件再設定の処理負担を1/10に軽減することができる。広角撮影装置100が、30フレーム/秒という程度の動画撮影を行えば、10フレームごとに切出条件の再設定を行ったとしても、切出条件の精度が大幅に低下することはない。

40

【0122】

<<< §4. 発話状態認識処理の詳細 >>>

続いて、図1に示す会議支援システムにおける唇近傍領域設定部260、特徴量算出部270、発話状態認識部280の連携動作によって行われる発話状態の認識処理の詳細を具体例に即して説明する。

50

【 0 1 2 3 】

§ 1 で述べたとおり、唇近傍領域設定部 2 6 0 は、顔領域認識部 2 5 0 によって認識された顔領域 F 上に唇近傍領域 L を設定する処理機能を有している。図 1 3 に示す顔領域 F の画像は、図 1 2 に示す参加者 B について認識された顔領域 F b に対応するものであり、この顔領域 F 内には、実線で描かれた矩形の唇近傍領域 L が設定された例が示されている。この唇近傍領域 L は、顔画像 F 内の唇部分およびその周囲部分を含む領域として設定される領域であり、唇の動きの有無を監視するための監視対象領域として機能する。したがって、必ずしも唇を中央部分に含むような正確な領域として設定する必要はなく、唇の動きの監視に利用可能な領域であれば、たとえ唇の一部が食み出してしまうような領域設定が行われても問題はない。

10

【 0 1 2 4 】

唇近傍領域設定部 2 6 0 によって唇近傍領域 L を設定する最も簡単な方法は、顔領域 F の輪郭枠を基準として予め定められた所定枠内の領域を唇近傍領域 L とする方法である。図 1 4 は、このような設定方法の一例を示す平面図である。図の外側の矩形は、顔領域認識部 2 5 0 によって認識された矩形の顔領域 F の輪郭枠であり、内側の矩形は、この方法で設定された唇近傍領域 L の輪郭枠である。輪郭枠 L は、輪郭枠 F を基準として所定の相対サイズを有する矩形であり、輪郭枠 F を基準とした所定の相対位置に配置されている。

【 0 1 2 5 】

具体的には、この図 1 4 に示す例の場合、顔領域の矩形輪郭枠 F の左辺の横方向に関する位置を 0 %、右辺の横方向に関する位置を 1 0 0 % とし、上辺の縦方向に関する位置を 0 %、下辺の縦方向に関する位置を 1 0 0 % としたときに、予め設定された所定の定数 a_1 、 a_2 、 b_1 、 b_2 (但し、 $0 < a_1 < a_2 < 100$ 、 $0 < b_1 < b_2 < 100$) に基づいて、左辺の横方向に関する位置が a_1 %、右辺の横方向に関する位置が a_2 % の位置にあり、上辺の縦方向に関する位置が b_1 %、下辺の縦方向に関する位置が b_2 % の位置にある矩形の輪郭枠 L をもつ領域を唇近傍領域 L に設定している。特に、図示の例は、 $a_1 = 25$ 、 $a_2 = 75$ 、 $b_1 = 70$ 、 $b_2 = 90$ に設定した例になっている。

20

【 0 1 2 6 】

もちろん、顔の輪郭に対する唇の相対位置や相対的な大きさには個人差があり、また、発話中は唇部分が開口することになるので、図 1 4 に示すような条件で唇近傍領域 L を一義的に定義した場合、必ずしも唇が唇近傍領域 L 内に収まるとは限らない。特に、参加者が横向きになった場合、唇の位置は唇近傍領域 L の中央位置から外れる可能性が高く、場合によっては、唇近傍領域 L から唇の一部が食み出すこともある。

30

【 0 1 2 7 】

たとえば、図 1 3 に破線で示す矩形枠 L は、顔領域 F 内に、図 1 4 に示す条件で唇近傍領域を定義した場合の輪郭枠を示している。この場合、唇近傍領域 L は、鼻の下端部分まで含む領域として定義され、唇は、当該唇近傍領域 L 内の中央ではなく、若干左下に偏った位置に配置される。ただ、上述したとおり、唇近傍領域 L は、唇の動きを監視する上での監視対象領域として機能するものであるため、図 1 4 に示すような一義的な条件設定により、図 1 3 に破線で示すような唇近傍領域 L が定義されたとしても大きな問題は生じない。

40

【 0 1 2 8 】

もっとも、発話状態の認識処理の精度を高めるために、より正確な唇近傍領域 L の設定を行いたい場合には、唇近傍領域設定部 2 6 0 が、個々の平面正則画像について、それぞれ顔領域 F 内を所定の唇探索アルゴリズムを用いて探索し、探索された唇の周辺を含む領域を、当該平面正則画像についての唇近傍領域に設定するようにすればよい。

【 0 1 2 9 】

図 1 5 は、このような方法で顔領域 F 内に唇近傍領域 L を設定した例を示す平面図である。図にハッチングを施して示す領域 K は、所定の唇探索アルゴリズムを用いて探索された唇領域 (この例の場合は、唇の外接矩形) であり、ほぼその中心点に唇の中心点 P k が

50

位置する。唇近傍領域 L は、こうして探索された唇領域 K とその周辺を含む領域（たとえば、唇領域 K の四辺を外側に所定寸法だけ広げた矩形領域）として設定すればよい。唇領域 K の周辺を含めた領域を唇近傍領域 L とするのは、発話中には唇の輪郭が外側に広がると予想されるためである。

【0130】

なお、唇探索アルゴリズムとしては、たとえば、AAM (Active Appearance Model) を用いた手法などが知られている。この手法では、予め唇のテクスチャを学習させておき、顔領域 F 内の唇が存在しそうな領域を探索し、学習画像との間で特徴量の比較を行い、更に、エッジの情報を考慮して、唇領域 K の決定が行われる。また、特開 2005-048150 号公報には、鼻の鼻孔にはかからない部分の色を肌の基準色として認定し、この肌の基準色に対して所定の基準以上の色差をもつ領域を唇領域 K と認識するアルゴリズムが開示されている。したがって、本願では、具体的な唇探索アルゴリズムの詳細についての説明は省略する。

10

【0131】

結局、図 14 に例示した実施形態のように、顔領域 F の輪郭枠を基準とした相対位置として一義的に唇近傍領域 L を定義した場合は、いずれのフレームについても、顔領域 F に対して常に同じ相対位置にある唇近傍領域 L が定義されることになる。これに対して、図 15 に例示した実施形態のように、顔領域 F 内を所定の唇探索アルゴリズムを用いて探索して唇領域 K を決定し、その周辺を含む領域を唇近傍領域 L とする場合は、個々のフレームごとに、顔領域 F に対して固有の相対位置にある唇近傍領域 L が定義されることになる。後者の実施形態を採用すると、唇探索アルゴリズムを実行する処理負担が増えることになるが、常に、唇をほぼ中央位置に含んだ唇近傍領域 L を定義することができ、発話状態の認識処理の精度を向上させることができる。

20

【0132】

なお、唇探索アルゴリズムを用いて探索した唇領域 K を利用して唇近傍領域 L を設定する実施形態を採る場合は、こうして設定された唇近傍領域 L の情報を利用して、切出条件の再設定を行うようにすることも可能である。たとえば、切出条件設定部 240 が新たな切出中心点 $P(x_0, y_0)$ の位置を決定する処理を行う場合、§3 では、図 10 に示すように、顔領域 F の輪郭枠を基準として特定の位置に再設定基準点 $T_r(u_r, v_r)$ を定義し、この再設定基準点 T_r を歪曲円形画像 S 上に逆変換して得られる点を新たな切出中心点 $P(x_0, y_0)$ とする手順を述べた。

30

【0133】

これに対して、唇近傍領域設定部 260 が唇探索アルゴリズムで探索した唇領域 K を利用して唇近傍領域 L を設定する場合は、設定した当該唇近傍領域 L の中心点を再設定基準点 $T_r(u_r, v_r)$ と定め、平面正則画像 T 上の当該再設定基準点 T_r を歪曲円形画像 S 上に逆変換して得られる点を新たな切出中心点 $P(x_0, y_0)$ とすることもできる。そうすれば、実際の人物画像に即したより正確な唇の位置を新たな切出中心点 $P(x_0, y_0)$ とする再設定が可能になる。

【0134】

こうして、唇近傍領域設定部 260 によって唇近傍領域 L が設定されたら、特徴量算出部 270 が、当該唇近傍領域 L 内の輝度もしくは色を示す特徴量を算出する処理を行う。図 16 は、この特徴量算出部 270 によって算出された特徴量の技術的な意義を説明する平面図である。図 16(a) も図 16(b) も、同じ面積をもった唇近傍領域 L 内の画像であるが、前者は唇を閉じた状態、後者は唇を開いた状態の画像である。両者を比較すると、後者では、口腔内の面積割合が増えるため、輝度が低下することになる。また、両者では、色合いにも変化が見られる。

40

【0135】

そこで、この唇近傍領域 L 内の輝度もしくは色を示すパラメータとして特徴量を定義すれば、唇の開閉によって、当該特徴量には顕著な変動が生じることになる。本発明では、このような特徴量の変動に基づいて、当該人物が発話状態にあるか否かを判定する。

50

【0136】

図17は、特徴量算出部270によって、唇近傍領域Lについての特徴量を算出するための数式の例をいくつか示す図である。ここでは、説明の便宜上、図の上段に示すように、唇近傍領域Lが複数n個（左上隅の画素W(1)～右下隅の画素W(n)）の画素の集合によって構成され、第i番目の画素W(i)が、R色成分R(i)、G色成分G(i)、B色成分B(i)を有する画素であるものとして、以下の説明を行うことにする。

【0137】

別言すれば、図1に示すシステムにおいて、広角撮影装置100は、3原色RGBの各色成分をもったカラー画像を撮影する機能を有しており、平面正則画像格納部230に得られる平面正則画像Tおよびその一部を構成する唇近傍領域L内の画像は、3原色RGBの各画素値をもった画素の集合体によって構成されているものとする。

10

【0138】

まず、図17(a)に示す第1の実施形態は、唇近傍領域L内の輝度を示すパラメータとして特徴量を定義した例である。この例では、特徴量算出部270は、唇近傍領域L内の第i番目(1 ≤ i ≤ n、但し、nは唇近傍領域内の画素の総数)の画素の3原色の画素値をそれぞれR(i)、G(i)、B(i)としたときに、

$$= \sum_{i=1}^n (R(i) + G(i) + B(i)) / n$$

なる演算式に基づく演算により当該唇近傍領域L内の輝度を示す特徴量を算出する。要するに、3原色の画素値の総和の全画素についての平均値を特徴量として利用することになる。

20

【0139】

一方、図17(b)に示す第2の実施形態は、唇近傍領域L内の色、特に、R色成分を示すパラメータとして特徴量を定義した例である。この例では、特徴量算出部270は、唇近傍領域L内の第i番目(1 ≤ i ≤ n、但し、nは唇近傍領域内の画素の総数)の画素の原色Rの画素値をR(i)としたときに、

$$= \sum_{i=1}^n (R(i)) / n$$

なる演算式に基づく演算により当該唇近傍領域内の色を示す特徴量を算出する。要するに、原色Rにのみ着目し、全画素についての原色Rの画素値の平均値を特徴量として利用することになる。

【0140】

そして、図17(c)に示す第3の実施形態は、唇近傍領域L内の色、特に、色相を示すパラメータとして特徴量を定義した例である。この例では、特徴量算出部270は、唇近傍領域内の第i番目(1 ≤ i ≤ n、但し、nは唇近傍領域内の画素の総数)の画素の、3原色RGBの各画素値から求めた色相値をH(i)としたときに、

$$= \sum_{i=1}^n (H(i)) / n$$

なる演算式に基づく演算により当該唇近傍領域内の色を示す特徴量を算出する。ここで、3原色の各画素値R(i)、G(i)、B(i)から色相値H(i)を求める式は、図17(c)下段に示すとおりである。要するに、全画素についての色相値の平均値を特徴量として利用することになる。

30

【0141】

本願発明者が行った実験(日本人を被験者としたもの)によれば、上記3つの実施形態のいずれを採用した場合も、唇の動きが特徴量の変動に十分に反映されることが確認できた。もちろん、会議の参加者の人種や肌の色、会議場の照明環境などによっても、最適な特徴量の定義方法は変わってくるので、実用上は、何通りかの定義方法を選択できるような構成にしておくのが好ましい。

40

【0142】

なお、図15に示す実施形態を採用すると、個々のフレームごとに、それぞれ唇領域Kの大きさや形状が異なることになる。たとえば、図18(a)は、ある人物が正面向きで唇を閉じた状態において抽出された唇領域Kを示しているが、同一人物が唇を開いた状態では、図18(b)に示すような唇領域Kが抽出される。また、同一人物が斜め横向きになる

50

と、唇の横幅は縮むことになるので、唇を閉じた状態では図 18 (c) のような唇領域 K が抽出され、唇を開いた状態では図 18 (d) に示すような唇領域 K が抽出される。

【 0 1 4 3 】

このように、抽出される唇領域 K の大きさや形状が変動すると、唇領域 K を基準に設定される唇近傍領域 L の大きさや形状も変動することになるが、図 17 に例示したとおり、特徴量の算出処理では、唇近傍領域 L を構成する全 n 個の画素についての平均値が用いられるため、唇近傍領域 L を構成する画素数が変動しても大きな問題は生じない。ただ、特徴量の算出精度をより高めるには、個々の人物ごとに理想的な唇近傍領域 L を設定し、しかもフレームごとに変動しないようにするのが好ましい。そのためには、たとえば、図 7 に示すような初期切出位置指定画面 310 を表示して初期切出条件設定作業を行う際に、全参加者に正面（広角撮影装置 100 の方向）を向き、唇を閉じた状態で所定時間（たとえば、10 秒間）だけ静止してもらい、その間に得られた個々のフレームから抽出した唇領域 K の平均に基づいて唇近傍領域 L を設定し、以後、こうして設定された同一の唇近傍領域 L を利用するようにしてもよい。

10

【 0 1 4 4 】

こうして、特徴量算出部 270 では、個々の参加者について、それぞれ各フレーム単位で特徴量が算出されるので、発話状態認識部 280 は、個々の参加者ごとに、その特徴量の時間変化に基づいて発話状態を認識する処理を行うことができる。図 19 は、この発話状態認識部 280 によって、発話状態を認識する基本原理を示すグラフであり、横軸は時間 t、縦軸は特徴量（図 17 (a) , (b) , (c) のいずれの定義法を採用してもよい）を示している。

20

【 0 1 4 5 】

基本的には、発話状態認識部 280 は、各参加者について、特徴量の変動幅が所定のしきい値以下となる期間を非発話期間、当該しきい値を超える期間を発話期間と認識すればよい。図 19 に示す例の場合、時点 $t_1 \sim t_2$, $t_3 \sim t_4$, $t_5 \sim t_6$ の各期間において、特徴量の変動幅がしきい値を超えているため、これらの期間は発話期間と認識され、時点 $t_0 \sim t_1$, $t_2 \sim t_3$, $t_4 \sim t_5$, $t_6 \sim$ の各期間において、特徴量の変動幅がしきい値以下となっているため、これらの期間は非発話期間と認識される。なお、認識された発話期間もしくは非発話期間の長さが所定の基準値（たとえば、1 秒間）に満たない場合は、何らかのノイズの影響で当該期間が誤認識された可能性が高いので、それらの誤認識期間は無視するような運用を行うのが好ましい。

30

【 0 1 4 6 】

以下に、特徴量の変動幅がしきい値を超えているか否かを判断するための具体的な方法の一例を例示しておく。ここでは、広角撮影装置 100 が、30 フレーム / 秒のレートで動画撮影を行い、歪曲円形画像格納部 210 内には、1 秒間に 30 フレーム分の歪曲円形画像 S が格納され、平面正則画像格納部 230 内には、個々の参加者について、それぞれ 1 秒間に 30 フレーム分の平面正則画像 T が格納されるものとしよう。

【 0 1 4 7 】

特徴量は、個々の参加者について、個々のフレームごとに算出されるので、ある 1 人の参加者についての特徴量の時間的変遷は、たとえば、図 20 に示すようなものになる。ここで、横軸は時間軸に相当し、 $f_1 \sim f_{15}$ は、平面正則画像 T のフレーム番号を示す（撮影レートが 30 フレーム / 秒の場合、図示のグラフは、約 0.5 秒の期間を示すことになる）。グラフ上の黒点は、時間軸上に離散的（1 / 30 秒おき）に定義された特徴量のサンプル点を示しており、その上の「+」もしくは「-」の符号は、前のサンプル点の値に対する増減を示している。

40

【 0 1 4 8 】

このように、特徴量の時間的変動グラフ上に新たなサンプル点が得られた時点で、直前のサンプル点に対する増減を調べるようにすれば、増減の関係が反転したサンプル点の直前のサンプル点として、極大点と極小点とを定義することができる。たとえば、図示の例の場合、フレーム f_5 において、符号が「+」から「-」に反転しているので、その直

50

前のフレーム f 4 におけるサンプル点が極大点となり、フレーム f 1 1 において、符号が「 - 」から「 + 」に反転しているため、その直前のフレーム f 1 0 におけるサンプル点が極小点になる。ここでは、このような極大点および極小点を併せて、変極点と呼ぶことにする。

【 0 1 4 9 】

特徴量 に変動が生じていれば、時間軸に沿って、このような変極点が順次出現することになる。そこで、これら変極点を利用して、図 2 1 に示すような方法で、発話期間の認識を行えばよい。図 2 1 に示すグラフは、図 2 0 に示すグラフの時間軸を圧縮したものに相当し、グラフ上の点 Q 0 ~ Q 1 0 は、それぞれ変極点を示している。より具体的には、×印で示す偶数番号の変極点 Q 0 , Q 2 , Q 4 , ... は極小点であり、黒丸で示す奇数番号の変極点 Q 1 , Q 3 , Q 5 , ... は極大点である。

10

【 0 1 5 0 】

そして、個々の変極点について直前の変極点との間で特徴量 の偏差 d を求めるようにする。たとえば、変極点 Q 1 については直前の変極点 Q 0 との間で、特徴量 の偏差 d 1 を求め、変極点 Q 2 については直前の変極点 Q 1 との間で、特徴量 の偏差 d 2 を求め、... といった具合である。そして、この各偏差 d 1 , d 2 , d 3 , ... が、所定のしきい値 t h を超えるか否かを判定し、偏差 d がしきい値 t h 以下の状態からしきい値 t h を超える状態に転じた時点 t 4 を非発話期間から発話期間への転換時点と認識し、偏差 d がしきい値 t h を超えた状態からしきい値 t h 以下の状態に転じた時点 t 9 を発話期間から非発話期間への転換時点と認識すればよい。

20

【 0 1 5 1 】

<<< § 5 . 会議画像提示処理の詳細 >>>

さて、上述したように、発話状態認識部 2 8 0 は、個々の参加者について、発話中か否かの状態もしくは発話時間を示す発話情報を取得することができる。そこで、会議画像提示部 2 9 0 は、個々の参加者の人物画像とともに当該参加者についての発話情報を、ディスプレイ装置 3 0 0 の画面に表示する処理を行う。具体的には、会議画像提示部 2 9 0 は、個々の参加者について、当該参加者の人物画像を含む平面正則画像を表示する参加者表示領域と、この参加者表示領域に紐づけて当該参加者の発話情報を表示する発話情報表示領域と、を有する会議画像を生成し、この会議画像をディスプレイ装置 3 0 0 の画面に表示する。以下に、このような会議画像表示画面のいくつかのバリエーションを列挙しておく。

30

【 0 1 5 2 】

(1) 第 1 のバリエーションに係る会議画像表示画面

図 2 2 に示す第 1 の会議画像表示画面 3 3 0 には、参加者表示領域 3 3 1 と発話情報表示領域 3 3 2 (図では、説明の便宜上、破線で囲って示す) とが設けられている。この例では、参加者表示領域 3 3 1 は、個々の参加者 A ~ D の人物画像 (図 8 に示す平面正則画像 T a ~ T d) を表示するための領域 3 3 1 A ~ 3 3 1 D によって構成されている。一方、発話情報表示領域 3 3 2 は、個々の参加者 A ~ D のそれぞれに対応した発話情報表示領域 3 3 2 A ~ 3 3 2 D によって構成されており (図では破線で示すが、実際には、この破線は必ずしも表示する必要はない。) 、現在発話中の参加者についての発話情報表示領域 3 3 2 B についてのみ、発話中であることを示す識別マークを表示させている。この例では、発話中識別マーク M として、マイクの図柄をあしらったアイコンを用いており、この発話中識別マーク M が表示されている参加者 (図示の例の場合は B 氏) が現在発話中の参加者であることが一目で認識できる。



40

【 0 1 5 3 】

発話状態認識部 2 8 0 により、個々の参加者が現在発話中か否かを認識することができるので、会議画像提示部 2 9 0 は、現在発話中と認識された参加者についての発話情報表示領域 (図示の例では、領域 3 3 2 B) にのみ、発話中識別マーク M を表示する処理を行えばよい。このような表示形態を採用すれば、発話者が交替すると、発話中識別マーク M が移動するので、現在誰が発話中であるのかを容易に把握することが可能になる。

50

【0154】

なお、図22に示す例では、非発話者の発話情報表示領域には発話中識別マークMを表示せず、発話者の発話情報表示領域にのみ発話中識別マークMを表示する運用を採っているが、基本的には、発話情報表示領域には、各参加者が発話期間中か非発話期間中かを識別する識別マークを表示させればよい。たとえば、発話期間中を示す識別マークとして「」印、非発話期間中を示す識別マークとして「x」印を表示するようにすれば、「」か「x」かで、各参加者の発話状態を認識することができる。図22に示す例は、発話期間中を示す識別マークとしてマイクの図柄をあしらったアイコンを用い、非発話期間中を示す識別マークとして無印のマークを用いた例ということができる。

【0155】

(2) 第2のバリエーションに係る会議画像表示画面

図23に示す第2の会議画像表示画面340には、参加者表示領域341と発話情報表示領域342とが設けられている。参加者表示領域341に、個々の参加者A～Dの人物画像が表示される点は上述の例と同じであるが、発話情報表示領域342には、各参加者A～Dの会議開始後（初期切出条件の設定後）の累積発話時間が数値により表示されている。

【0156】

図示の例では、累積発話時間として「時：分：秒」の書式を採用しているので、たとえば、参加者Aの累積発話時間は、1時間25分3秒ということになる。このように、各参加者の累積発話時間が表示されれば、誰が長時間発言し、誰の発言が少ないのかが一目瞭然であり、会議の円滑な進行を図る上で有益な情報提示が行われることになる。たとえば、図示の例では、B氏は15分40秒しか発言していないので、以後、司会者は、B氏に積極的に発言を求めるような進行を心がけることができる。あるいは、自分が長時間発言していることを認識したA氏は、以後は、自発的に発言を控えるよう自制することもできる。

【0157】

(3) 第3のバリエーションに係る会議画像表示画面

図24に示す第3の会議画像表示画面350には、参加者表示領域351と発話情報表示領域352とが設けられている。参加者表示領域351に、個々の参加者A～Dの人物画像が表示される点は上述の例と同じであるが、発話情報表示領域352には、現在発話中の参加者についての発話継続時間が表示されている。

【0158】

図示の例では、発話継続時間として「分：秒」の書式を採用しており、現在、参加者Bが発話中であり、発話開始後の継続時間が2分13秒であることが認識できる。ここでB氏が発言を終了し、代わりにD氏が発言を開始したとすると、B氏の発話継続時間は0にリセットされ、D氏の発話継続時間が0からカウントアップされる。しかも、発話継続時間の表示は、現在発話中の者についてのみ行われるので、D氏の発言に移行した後は、D氏の人物画像の上方にD氏の発話継続時間が表示され、他の参加者の人物画像の上方には、何ら時間表示はなされなくなる。

【0159】

このように、現在発話中の参加者についての発話継続時間を表示する形態は、1回の発言時間を所定の設定時間内に制限するルールを採用した会議において有効である。たとえば、1回の発言時間を3分以内に制限するルールが定められた会議では、発話情報表示領域352に表示される発話継続時間により、当該ルールが遵守されているか否かを監視することができる。もちろん、発話継続時間の代わりに、制限時間までの残存時間を表示してもよい。

【0160】

(4) 第4のバリエーションに係る会議画像表示画面

図25に示す第4の会議画像表示画面360の表示内容は、基本的には、図23に示す第2の会議画像表示画面340の表示内容と同じである。すなわち、参加者表示領域36

10

20

30

40

50

1には、個々の参加者A～Dの人物画像が表示され、発話情報表示領域362には、各参加者A～Dの累積発話時間が数値により表示されている。

【0161】

ただ、図23に示す会議画像表示画面340では、個々の参加者についての発話情報表示領域（時間が表示されている領域）を、当該参加者についての参加者表示領域（人物画像が表示されている領域）の近傍に配置し、互いに近傍に配置することにより、特定の参加者表示領域と特定の発話情報表示領域との紐付けを行っている。たとえば、会議画像表示画面340において、参加者Aの累積発話時間が、1時間25分3秒であることが認識できるのは、参加者Aの人物画像の近傍に「1：25：03」なる時間表示がなされているためであり、両者は近傍に配置されているという理由で相互に紐付けがなされている。

10

【0162】

これに対して、図25に示す会議画像表示画面360では、左側に参加者表示領域361が配置され、右側に発話情報表示領域362が配置されており、個々の参加者についての参加者表示領域（人物画像が表示されている領域）と、当該参加者についての発話情報表示領域（時間が表示されている領域）との間の紐付けは、参加者名の表記によって行われている。たとえば、参加者Aの累積発話時間が、1時間25分3秒であることが認識できるのは、参加者Aの人物画像と「1：25：03」なる時間表示とが、「A氏」という参加者名を介して紐付けされていることになる。

【0163】

このように、会議画像提示部290によって提示される会議画像上では、個々の参加者についての人物画像（平面正則画像）と、当該参加者の発話情報とが、何らかの方法で紐付けされていれていれば足り、必ずしも紐付けされた両者を近傍に配置する必要はない。

20

【0164】

(5) 第5のバリエーションに係る会議画像表示画面

図26に示す第5の会議画像表示画面370の表示内容も、基本的には、図23に示す第2の会議画像表示画面340の表示内容と同じである。すなわち、参加者表示領域371には、個々の参加者A～Dの人物画像が表示され、発話情報表示領域372には、各参加者A～Dの累積発話時間が表示されている。

【0165】

ただ、累積発話時間の表示は、数字ではなく円グラフMa～Mdによって行われている。すなわち、所定の時間ユニットUに対して円グラフの360°を対応させ、表示対象となる時間TをUで割り算したときの商をQ、剰余をRとしたときに、Q個の円と、1個の扇形（中心角は、 $= 360^\circ \times R / U$ ）とを並べることにより時間表示を行っている。図示の例の場合、時間ユニットU = 1時間に設定されており、円グラフの360°が1時間に対応づけられている。したがって、1個の完全な円グラフは1時間を表しており、不完全な円グラフは、その中心角により、 $/ 360^\circ$ 時間を表している。

30

【0166】

結局、図26に示す発話情報表示領域372に円グラフMa～Mdとして表示されている累積発話時間は、図23に示す発話情報表示領域342に数字で表示されている累積発話時間と全く同じ内容であるが、時間の表示形式が異なっていることになる。このように円グラフを用いて時間表示を行えば、より直感的、視覚的、アナログ的な時間把握が可能になる。もちろん、発話情報表示領域372への時間表示を、数字と円グラフの双方で行うようにしてもかまわない。

40

【0167】

<<< §6. テレビ会議支援システム >>>

これまで述べてきた実施形態は、図2に例示するように、1台の会議支援システムを1つの会議室に設置し、当該会議室内の各参加者についての人物画像や発話情報をディスプレイ上に提示するものであったが、このような会議支援システムを複数台用意して、個々の会議支援システムの構成要素となる撮影画像処理装置200をネットワークで接続することにより相互に会議画像を送受信できるようにすれば、ディスプレイ装置300に複数

50

の会議画像を提示させることが可能になり、テレビ会議支援システムとして利用できるようになる。

【0168】

たとえば、図1に示す構成をもった会議支援システムを2台用意し、1台を大阪の会議室に設置し、もう1台を東京の会議室に設置し、相互の撮影画像処理装置200をインターネットを介して接続し、相互に会議画像を送受信できるようにすれば、それぞれのディスプレイ装置300には、図27に示すような会議画像表示画面を表示させることができる。

【0169】

図27において、上下2段にわたって配置された画面380, 390は、それぞれが地区別の会議画像表示画面であり、上段が大阪地区、下段が東京地区を示している。上段の大阪地区の会議画像表示画面380には、地区表示381、参加者表示領域382、発話情報表示領域383が設けられており、下段の東京地区の会議画像表示画面390には、地区表示391、参加者表示領域392、発話情報表示領域393が設けられている。いずれも、その表示内容は、基本的には、図23に示す会議画像表示画面340の表示内容とほぼ同じである。

【0170】

もちろん、3台以上の会議支援システムを相互接続することにより、3カ所以上に分散した会議室を利用したテレビ会議を支援するためのテレビ会議支援システムを構成することも可能である。

【0171】

<<< §7. 撮影画像処理装置の処理手順 >>>

最後に、図28の流れ図を参照しながら、図1に示す会議支援システムにおける撮影画像処理装置200の処理手順をまとめておく。

【0172】

まず、ステップS1において、初期切出条件の設定処理が行われる。この処理は、初期切出条件設定部245の機能によって行われる処理であり、たとえば、図7に示すような初期設定画面310がディスプレイ装置300に表示され、§3で述べたように、ユーザの操作入力に基づいて、各参加者の人物画像についての初期切出条件が、切出条件設定部240に対して設定される。

【0173】

続くステップS2では、新しいフレーム画像(広角撮影装置100が撮影した歪曲円形画像S)が取り込まれ、歪曲円形画像格納部210に格納される。そして、ステップS3では、画像切出変換部220が、切出条件設定部240にその時点で格納されている切出条件に基づいて、取り込んだ歪曲円形画像Sから個々の人物画像を切り出し、これを平面正則画像Tに変換した上で、平面正則画像格納部230に格納する画像切出変換処理を実行する。

【0174】

次のステップS4では、顔領域認識部250により、各参加者についての平面正則画像Tから顔領域Fを認識する処理が実行され、ステップS5では、更に、唇近傍領域設定部260により、顔領域F内に唇近傍領域Lを設定する処理が行われる。そして、ステップS6では、特徴量算出部270により、唇近傍領域Lから特徴量を算出する処理が行われ、ステップS7では、発話状態認識部280により、個々の参加者ごとの発話状態を認識する処理が行われる。この発話状態の認識は、特徴量の時間変化に基づいて行われるため、ステップS6で算出した特徴量は、発話状態の認識に必要な期間だけ保存しておくようにする。

【0175】

そして、ステップS8では、ステップS3で得られた平面正則画像(人物画像)とステップS7で認識された発話状態に関連した発話情報とを含んだ会議画像を作成し、ディスプレイ装置に提示する処理が行われる。

10

20

30

40

50

【 0 1 7 6 】

以上述べたステップ S 2 ~ S 8 の処理は、1 フレーム分の撮影画像に対して行われる一巡処理であり、当該一巡処理が、ステップ S 9 を経て、会議が終了するまで繰り返し実行されることになる。その際、ステップ S 1 0 において切出条件の再設定を行うか否かが判断され、再設定を行う場合は、ステップ S 1 1 において、切出条件設定部 2 4 0 による切出条件再設定処理、すなわち、ステップ S 4 における顔領域認識処理の認識結果に基づいて、新たな切出条件を設定する処理が行われる。

【 0 1 7 7 】

切出条件再設定処理を毎フレームごとに行う場合は、ステップ S 1 0 では常に肯定的な判断を行うようにすればよい。また、たとえば、切出条件再設定処理を 1 0 フレームごとに行う場合は、ステップ S 1 0 では 1 0 回ごとに 1 回の周期で肯定的な判断を行うようにすればよい

10

【 符号の説明 】

【 0 1 7 8 】

1 0 0	: 広角撮影装置	
1 5 0	: マイクフォン	
2 0 0	: 撮影画像処理装置	
2 1 0	: 歪曲円形画像格納部	
2 2 0	: 画像切出変換部	
2 3 0	: 平面正則画像格納部	20
2 4 0	: 切出条件設定部	
2 4 5	: 初期切出条件設定部	
2 5 0	: 顔領域認識部	
2 6 0	: 唇近傍領域設定部	
2 7 0	: 特徴量算出部	
2 8 0	: 発話状態認識部	
2 9 0	: 会議画像提示部	
3 0 0	: ディスプレイ装置	
3 1 0	: 初期設定画面	
3 1 1	: 初期切出位置指定画面	30
3 1 2	: 参加者登録画面	
3 1 3	: 参加者名登録窓	
3 1 4	: 参加者戻しボタン	
3 1 5	: 登録完了ボタン	
3 1 6	: 参加者送りボタン	
3 2 0	: 人物画像表示画面	
3 3 0	: 会議画像表示画面	
3 3 1	: 参加者表示領域	
3 3 1 A ~ 3 3 1 D	: A 氏 ~ D 氏の人物画像 (平面正則画像)	
3 3 2	: 発話情報表示領域	40
3 3 2 A ~ 3 3 2 D	: A 氏 ~ D 氏用の発話情報表示領域	
3 4 0	: 会議画像表示画面	
3 4 1	: 参加者表示領域	
3 4 2	: 発話情報表示領域	
3 5 0	: 会議画像表示画面	
3 5 1	: 参加者表示領域	
3 5 2	: 発話情報表示領域	
3 6 0	: 会議画像表示画面	
3 6 1	: 参加者表示領域	
3 6 2	: 発話情報表示領域	50

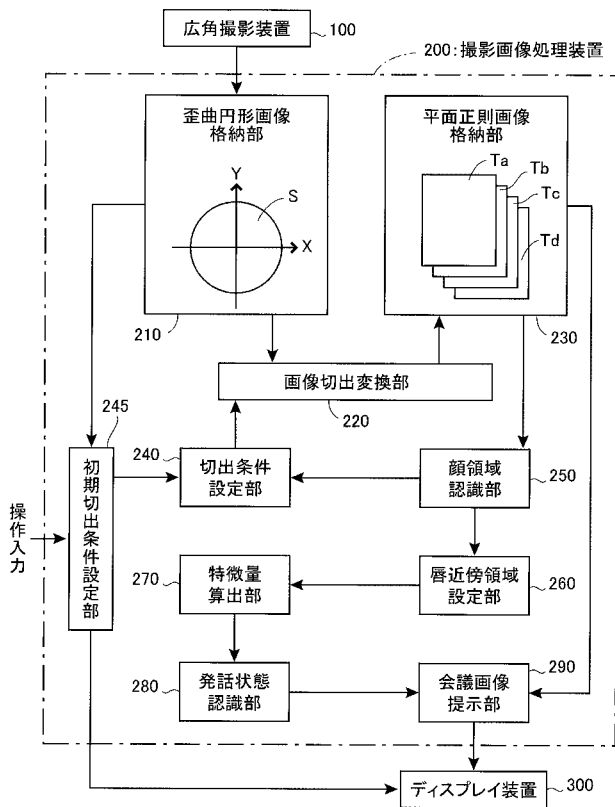
370	: 会議画像表示画面	
371	: 参加者表示領域	
372	: 発話情報表示領域	
380	: 地区別の会議画像表示画面	
381	: 地区表示	
382	: 参加者表示領域	
383	: 発話情報表示領域	
390	: 地区別の会議画像表示画面	
391	: 地区表示	
392	: 参加者表示領域	10
393	: 発話情報表示領域	
A ~ D	: 会議の各参加者	
a1, a2	: 横方向の位置指標となる定数	
B(i)	: 第 i 番目の画素の原色 B の画素値	
b1, b2	: 縦方向の位置指標となる定数	
Cb	: 同心円	
D	: 点 G を通り、XY 平面に平行かつ直線 OG に直交する軸	
d1 ~ d10	: 偏差	
E	: 切出領域	
Ea ~ Ed	: A 氏 ~ D 氏用の切出領域	20
F	: 顔領域 (顔領域の輪郭枠)	
Fa ~ Fd	: A 氏 ~ D 氏用の顔領域	
f1 ~ f15	: フレーム番号	
G	: 平面正則画像 T の中心点 (二次元 UV 直交座標系の原点)	
G(u0, v0)	: 平面正則画像 T の中心点 (二次元 UV 直交座標系の原点)	
G(xg, yg, zg)	: 二次元 UV 直交座標系の原点	
G(i)	: 第 i 番目の画素の原色 G の画素値	
H	: 仮想球面	
H(i)	: 第 i 番目の画素の色相値	
J	: 切出方向	30
Ja ~ Jd	: A 氏 ~ D 氏用の切出方向	
K	: 唇領域	
L, L'	: 唇近傍領域 (唇近傍領域の輪郭枠)	
L1	: 顔領域 F の輪郭枠の中心鉛直線	
L2	: 顔領域 F の唇位置水平線	
M	: 発話中識別マーク	
Ma ~ Md	: 時間表示用円グラフ	
m	: 切出倍率	
ma ~ md	: A 氏 ~ D 氏用の切出倍率	
n	: 視線ベクトル / 唇近傍領域 L 内の画素数	40
ni	: O と Qi とを結ぶ直線	
O	: 歪曲円形画像 S の中心点 (二次元 XY 直交座標系の原点)	
P(x0, y0)	: 切出中心点	
Pa ~ Pd	: A 氏 ~ D 氏用の切出中心点	
Pk	: 唇の中心点	
Q(x0, y0, z0)	: 球面上対応点	
Qi(xi, yi, zi)	: 球面上対応点	
Q0 ~ Q10	: 変極点	
R	: 歪曲円形画像 S の半径 (仮想球面 H の半径)	
R(i)	: 第 i 番目の画素の原色 R の画素値	50

- S : 歪曲円形画像
- $S_i(x_i, y_i)$: 二次元XY直交座標系上の歪曲円形画像S内の点
- S1 ~ S11 : 流れ図の各ステップ
- T : 平面正則画像 (人物画像)
- Ta ~ Td : 平面正則画像 (A氏 ~ D氏の人物画像)
- $Tr(ur, vr)$: 再設定基準点
- t : 時間軸
- t1 ~ t9 : 時間軸t上の時点
- th : 偏差のしきい値
- U : 平面正則画像T上の座標軸 (二次元UV直交座標系の座標軸)
- ur : 再設定基準点TrのU座標値
- V : 平面正則画像T上の座標軸 (二次元UV直交座標系の座標軸)
- vr : 再設定基準点TrのV座標値
- W(1) ~ W(n) : 唇近傍領域L内の画素
- X : 歪曲円形画像S上の座標軸 (二次元XY直交座標系の各座標軸)
- x0 : 切出中心点PのX座標値
- Y : 歪曲円形画像S上の座標軸 (二次元XY直交座標系の各座標軸)
- y0 : 切出中心点PのY座標値
- Z : 三次元XYZ直交座標系の座標軸
- : 方位角
- : 天頂角
- : 特徴量
- : 平面傾斜角
- : 円グラフの中心角

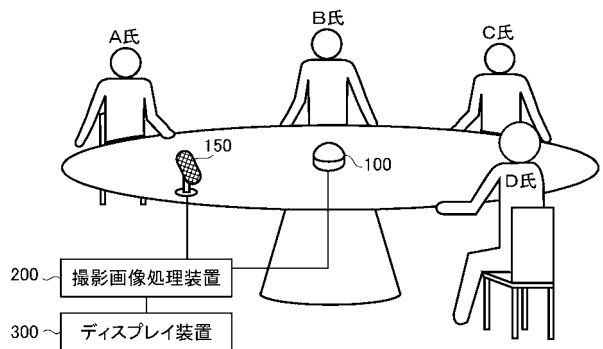
10

20

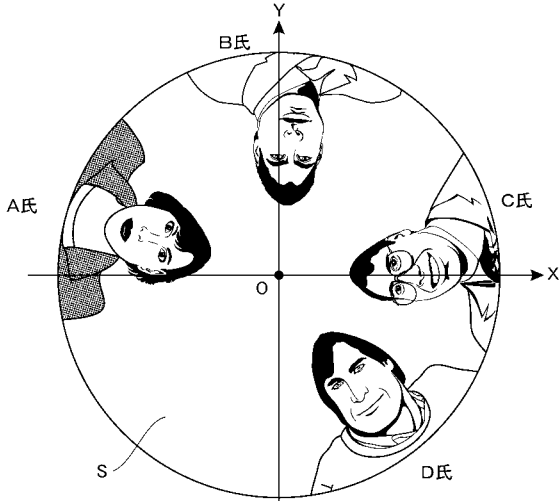
【図1】



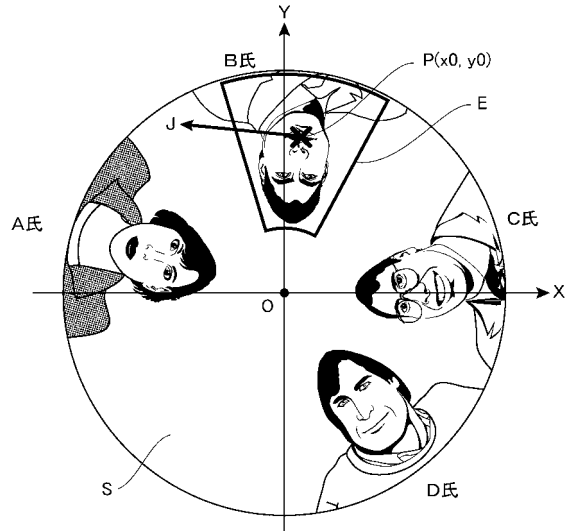
【図2】



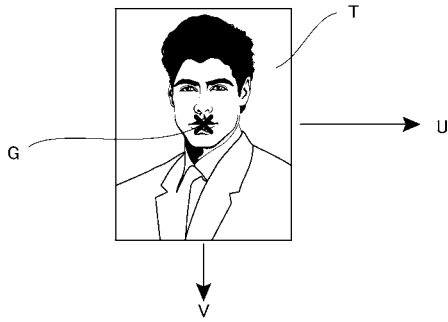
【 図 3 】



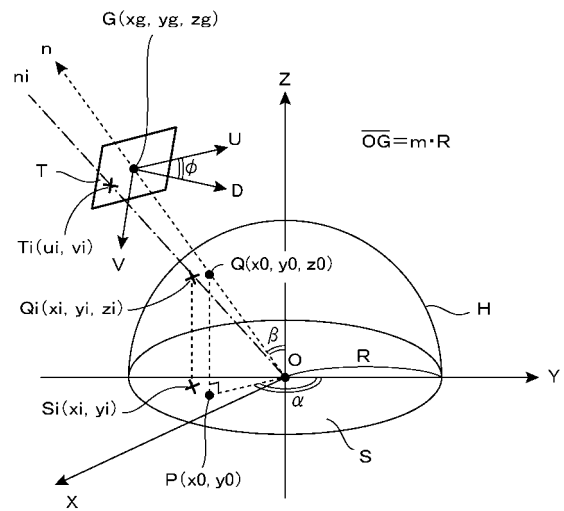
【 図 4 】



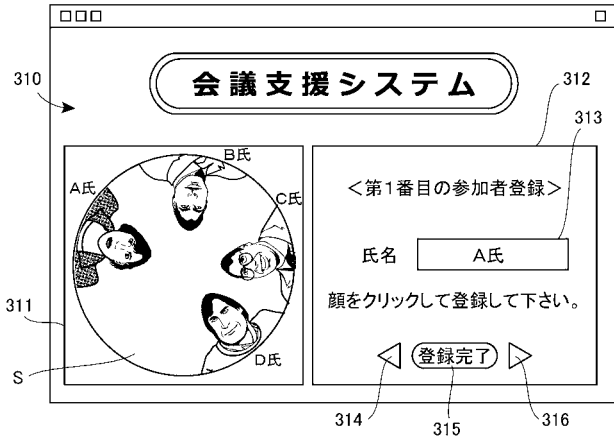
【 図 5 】



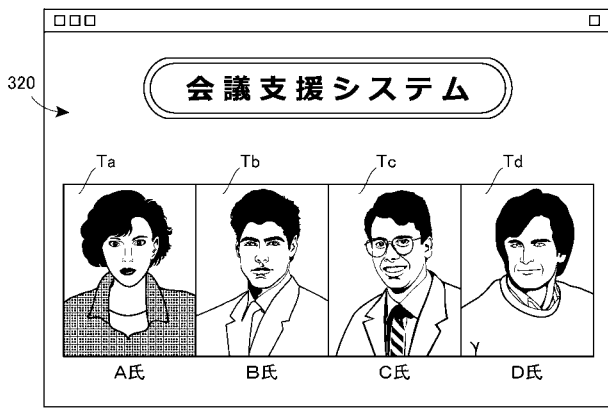
【 図 6 】



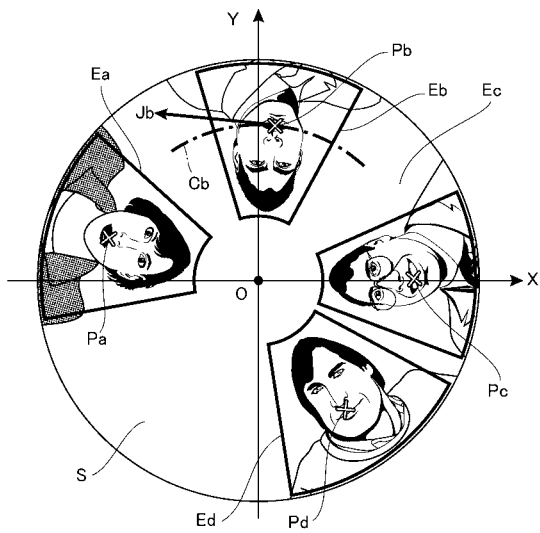
【 図 7 】



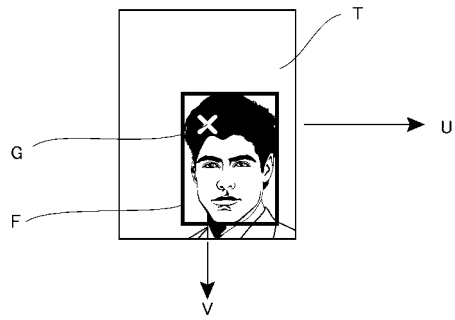
【 図 8 】



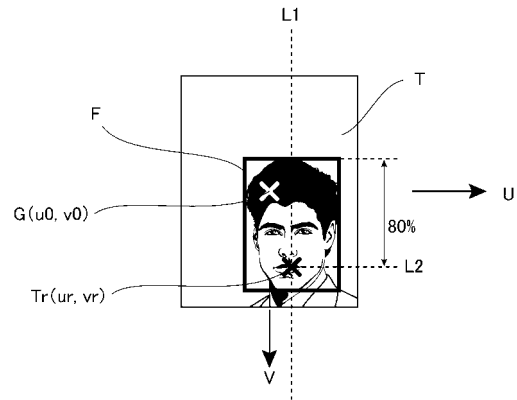
【 図 1 1 】



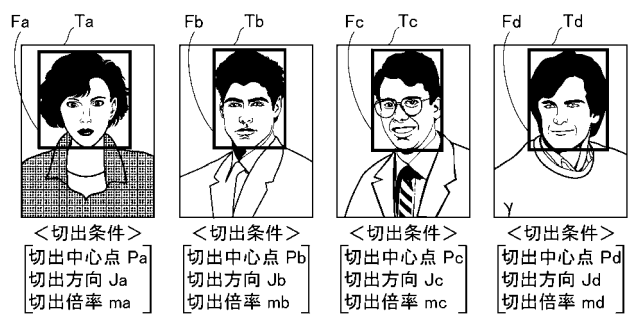
【 図 9 】



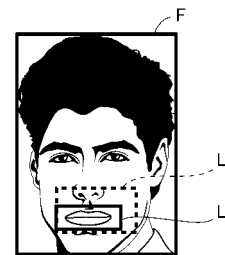
【 図 1 0 】



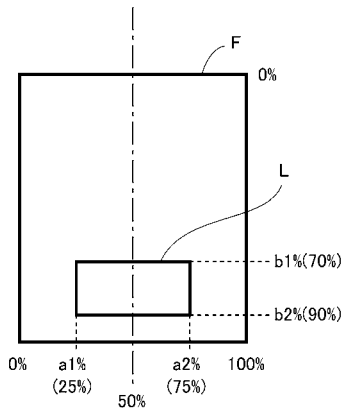
【 図 1 2 】



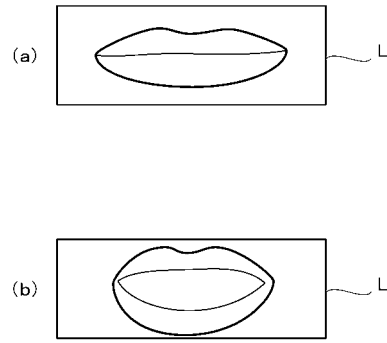
【 図 1 3 】



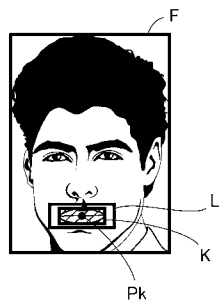
【 図 1 4 】



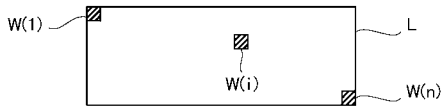
【 図 1 6 】



【 図 1 5 】



【 図 1 7 】



R(i) : 画素 W(i) の R 色成分の画素値
 G(i) : 画素 W(i) の G 色成分の画素値
 B(i) : 画素 W(i) の B 色成分の画素値

(a) 特徴量 $\xi = \sum_{i=1}^n (R(i) + G(i) + B(i)) / n$
 (唇近傍領域 L 内の輝度を示すパラメータ)

(b) 特徴量 $\xi = \sum_{i=1}^n R(i) / n$
 (唇近傍領域 L 内の R 色成分を示すパラメータ)

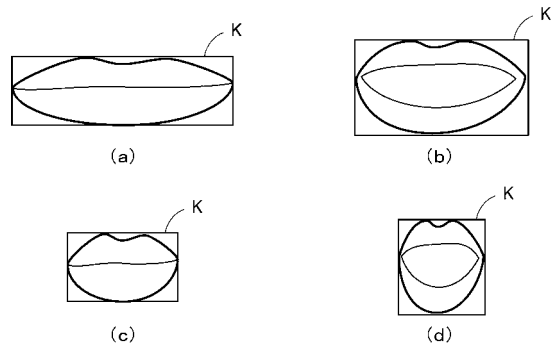
(c) 特徴量 $\xi = \sum_{i=1}^n H(i) / n$
 (唇近傍領域 L 内の色相を示すパラメータ)

但し、

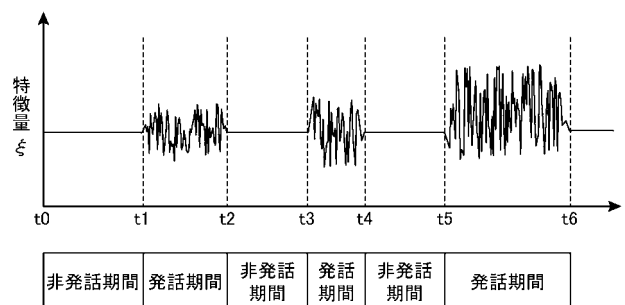
$$H(i) = \begin{cases} 60 \times \frac{G(i)-B(i)}{MAX-MIN} + 0, & \text{if MAX} = R \\ 60 \times \frac{B(i)-R(i)}{MAX-MIN} + 120, & \text{if MAX} = G \\ 60 \times \frac{R(i)-G(i)}{MAX-MIN} + 240, & \text{if MAX} = B \end{cases}$$

(R(i), G(i), B(i) のうち、最大のものを MAX, 最小のものを MIN とする)

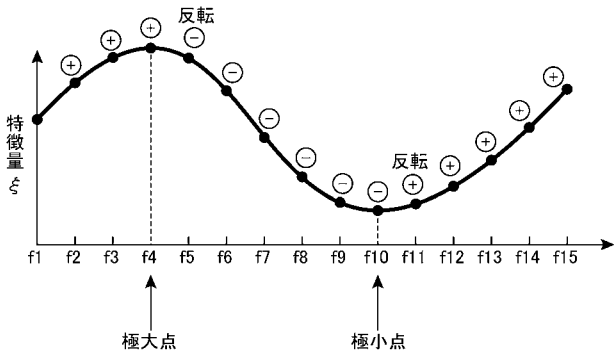
【 図 1 8 】



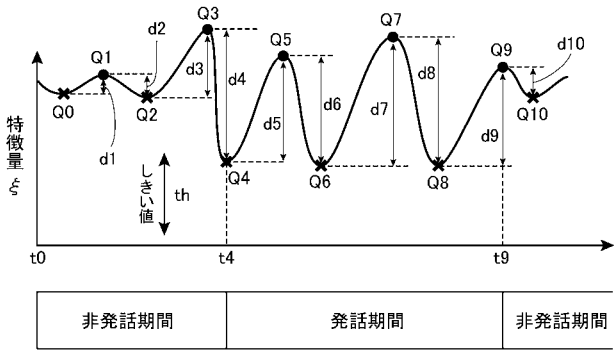
【 図 1 9 】



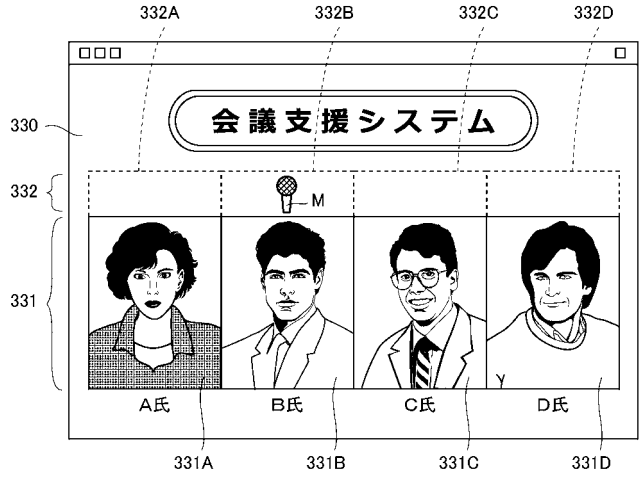
【図 2 0】



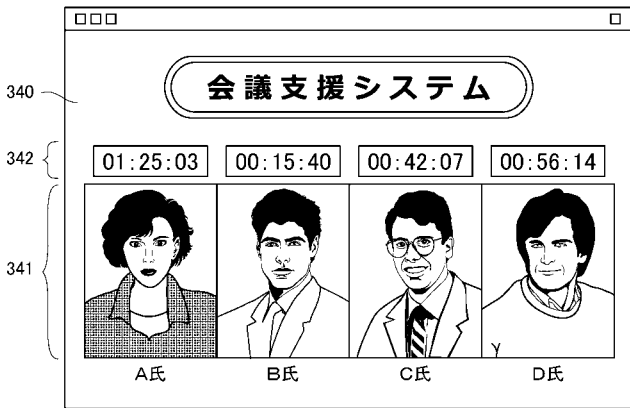
【図 2 1】



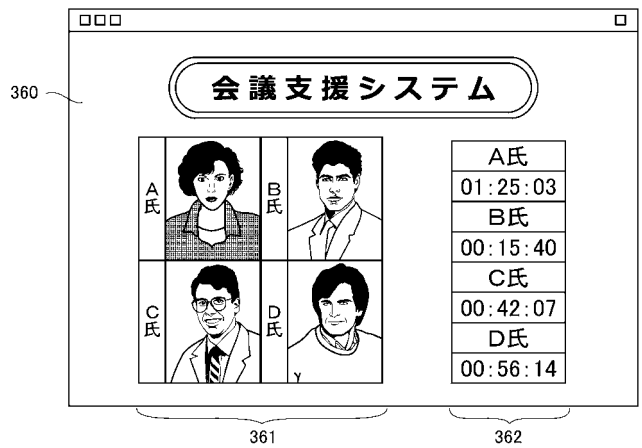
【図 2 2】



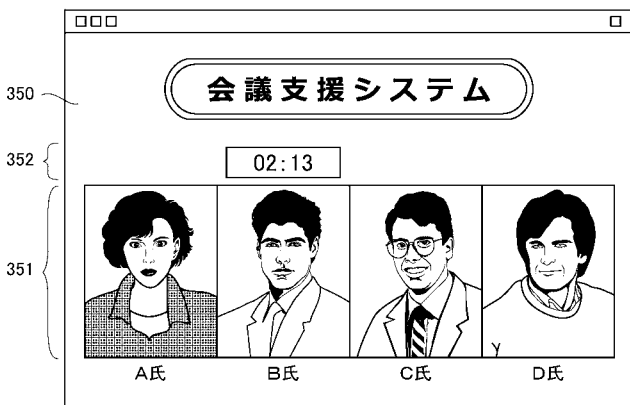
【図 2 3】



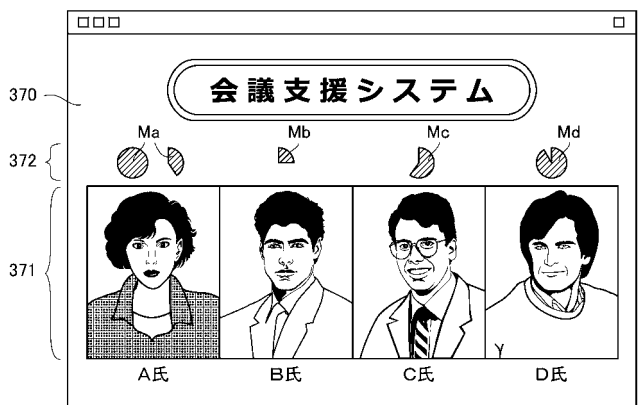
【図 2 5】



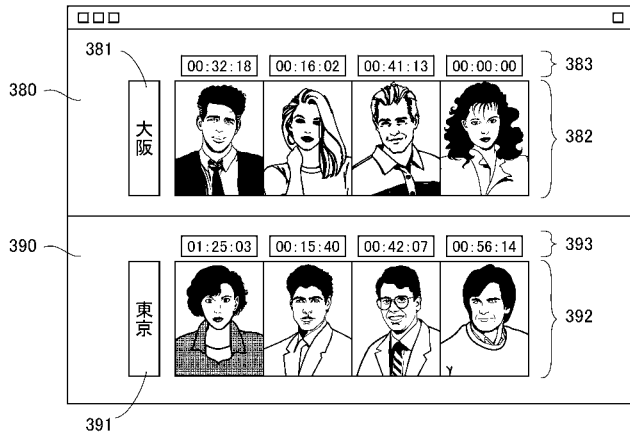
【図 2 4】



【図 2 6】



【 図 2 7 】



【 図 2 8 】

