

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5483863号  
(P5483863)

(45) 発行日 平成26年5月7日(2014.5.7)

(24) 登録日 平成26年2月28日(2014.2.28)

(51) Int. Cl. F I  
**HO4N 5/76 (2006.01)** HO4N 5/76 B  
**HO4N 5/91 (2006.01)** HO4N 5/91 Z

請求項の数 6 (全 17 頁)

(21) 出願番号	特願2008-290322 (P2008-290322)	(73) 特許権者	000001007
(22) 出願日	平成20年11月12日 (2008.11.12)		キヤノン株式会社
(65) 公開番号	特開2010-118868 (P2010-118868A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成22年5月27日 (2010.5.27)	(74) 代理人	100076428
審査請求日	平成23年11月11日 (2011.11.11)		弁理士 大塚 康德
前置審査		(74) 代理人	100112508
			弁理士 高柳 司郎
		(74) 代理人	100115071
			弁理士 大塚 康弘
		(74) 代理人	100116894
			弁理士 木村 秀二
		(74) 代理人	100130409
			弁理士 下山 治
		(74) 代理人	100134175
			弁理士 永川 行光

最終頁に続く

(54) 【発明の名称】 情報処理装置およびその制御方法

(57) 【特許請求の範囲】

【請求項1】

時系列に並んだ複数のフレーム画像を含む動画像データを入力する入力手段と、  
 入力された動画像データから所定の画像パターンに類似する被写体画像を含むフレーム  
 画像を検出する検出手段と、

前記検出手段により検出されたフレーム画像に含まれる前記被写体画像と類似する被写  
 体画像を含むフレーム画像を、前記検出されたフレーム画像の前後にあるフレーム画像か  
 ら検出する追跡手段と、

前記追跡手段により検出された連続したフレーム画像をそれぞれの被写体画像を含む画  
 像シーケンスとして決定し、当該画像シーケンスに対応する前記動画像データ内における  
 時間情報と関連付けて記憶する記憶手段と、

前記記憶手段に記憶された1以上の画像シーケンスの各々の開始時刻および終了時刻を  
 分割タイミングとして、前記動画像データを複数の時間区間に分割する分割手段と、

前記複数の時間区間の各々から代表フレーム画像候補を抽出する抽出手段と、

前記代表フレーム画像候補の中から、前記画像シーケンスのいずれかに含まれる被写体  
 画像をすべて含む代表フレーム画像候補の組み合わせのうち、より代表フレーム画像候補  
 の個数の少ない組み合わせであって、複数の画像シーケンスが重なる時間区間から抽出さ  
 れた代表フレーム画像候補を含む組み合わせを優先して、当該組み合わせを構成する代表  
 フレーム画像候補を、前記入力された動画像データの代表フレーム画像として選択する選  
 択手段と、

10

20

を備えることを特徴とする情報処理装置。

【請求項 2】

前記抽出手段は、前記複数の時間区間の各々に含まれるフレーム画像の各々の評価値を所定の評価ルールに基づいて算出し、当該評価値が最大または最小となるフレーム画像を代表フレーム画像候補として抽出することを特徴とする請求項 1 に記載の情報処理装置。

【請求項 3】

前記複数の時間区間の各々に含まれるフレーム画像が予め設定された 1 以上の構図テンプレートの何れかに対応する画像であるか否かを判定する構図判定手段をさらに備え、

前記抽出手段は、前記構図判定手段により前記 1 以上の構図テンプレートの何れかに対応する画像であると判定されたフレーム画像を含む時間区間からのみ前記代表フレーム画像を抽出することを特徴とする請求項 1 または 2 に記載の情報処理装置。

10

【請求項 4】

前記所定の画像パターンは人物の顔画像であり、

前記検出手段はフレーム画像に含まれる顔画像を検出することを特徴とする請求項 1 または 2 に記載の情報処理装置。

【請求項 5】

時系列に並んだ複数のフレーム画像を含む動画像データから 1 以上の代表フレーム画像を抽出する情報処理装置の制御方法であって、

入力手段が、動画像データを入力する入力工程と、

検出手段が、入力された動画像データから所定の画像パターンに類似する被写体画像を含むフレーム画像を検出する検出工程と、

20

追跡手段が、前記検出工程により検出されたフレーム画像に含まれる前記被写体画像と類似する被写体画像を含むフレーム画像を、前記検出されたフレーム画像の前後にあるフレーム画像から検出する追跡工程と、

記憶制御手段が、前記追跡工程により検出された連続したフレーム画像をそれぞれの被写体画像を含む画像シーケンスとして決定し、当該画像シーケンスに対応する前記動画像データ内における時間情報と関連付けて記憶部に記憶する記憶制御工程と、

分割手段が、前記記憶部に記憶された 1 以上の画像シーケンスの各々の開始時刻および終了時刻を分割タイミングとして、前記動画像データを複数の時間区間に分割する分割工程と、

30

抽出手段が、前記複数の時間区間の各々から代表フレーム画像候補を抽出する抽出工程と、

選択手段が、前記代表フレーム画像候補の中から、前記画像シーケンスのいずれかに含まれる被写体画像をすべて含む代表フレーム画像候補の組み合わせのうち、より代表フレーム画像候補の個数の少ない組み合わせであって、複数の画像シーケンスが重なる時間区間から抽出された代表フレーム画像候補を含む組み合わせを優先して、当該組み合わせを構成する代表フレーム画像候補を、前記入力された動画像データの代表フレーム画像として選択する選択工程と、

を含むことを特徴とする情報処理装置の制御方法。

【請求項 6】

コンピュータを、請求項 1 乃至 4 の何れか一項に記載の情報処理装置の各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、動画像データから当該動画像データに含まれる映像を端的に表す代表フレームを抽出する技術に関するものである。

【背景技術】

【0002】

近年、デジタルカメラや、デジタルビデオカムコーダ等の普及により、個人でも大量の

50

静止画や動画を撮影するようになってきている。一般に、動画データはデータ量が膨大であり、かつ、静止画と異なり時間軸が加わっているため、その中の内容を簡単に把握することは難しい。そのため、内容の概略を知りたい場合や所望のシーンを探す場合には、例えば、利用者が機器を操作して早回しや巻き戻しなどを行っていた。そこで、動画の内容を短時間で把握するために、動画内のフレームの中から、動画の内容を良く示すような適切なフレームを選択し、これを動画の内容を示す情報として扱う、代表フレームの抽出に関する技術が提案されている。

【0003】

ところで、一般の家庭で、購入・利用されているデジタルカメラは子供の成長の記録など、家庭内のイベントや出来事などを記録するために用いられることが多いという傾向がある。そのため、撮影される被写体は人物であることが多い。

10

【0004】

したがって、このようなホームビデオで動画の内容把握を行いたい場合には、誰が登場人物であるかという情報がわかるような動画内のフレームを、代表画像として選ぶことが、非常に重要となる。また、家庭内のイベントや出来事などが主な対象なので、複数の登場人物がいる場合には、その人物が同時に写っていたという情報もあるほうが好ましい。

【0005】

そこで、特許文献1では、動画に映っている人物に着目し、人物の構成が変化した場合に動画のシーンを分割する。その際に、人物の構成を示す情報をシーンのそれぞれに付与し、その人物の構成情報をインデックス化する技術が開示されている。そして、この人物の構成を示す情報を利用することによりシーンを検索することが出来る。

20

【0006】

また、特許文献2では、オブジェクトの統計的な特徴に基づいた代表フレームの抽出技術が開示されている。これは、検出したいオブジェクトの画像を予め学習させておき、所定の方法でオブジェクトごとに辞書を用いて評価値を求め、フレーム単位で求めた各オブジェクトの評価値に基づいて動画インデックスを生成している。さらに、特許文献3では、会議中の撮影対象である話し手の顔の画像と、話した音声情報の時間区間(タイムライン)を関係付けて利用する技術が開示されている。

【特許文献1】特開2005-101906号公報

【特許文献2】特許第3312105号

30

【特許文献3】特開2006-129480号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

しかしながら、例えば特許文献1に記載の技術では、フレームごとに顔検出し、その結果のみに基づいて代表画像を抽出するため、動いている被写体の場合、顔検出に失敗する可能性がある。つまり、必ずしも顔がカメラの方向を向いているわけではないので、被写体となる人物がいても検出できない場合がある。このような現象は、特に子供が被写体の場合などに顕著である。また、複数の被写体がいる場合などでも向き合って話したりすることが多く、比較的発生しやすい現象となる。その結果、同一の人物が連続して撮影されていても、ある時その人物の検出に失敗すると、そこで構成が変化すると判断され、同一の登場人物について、何度も代表フレームが作成されてしまうという問題が生じてしまう。

40

【0008】

また、フレームごとの評価値のみを用いるため、主要でない人物であっても、一瞬でもフレームの真ん中に大きく写るなどしていれば、評価値が高くなり、代表画像として抽出されてしまう。また、多数の類似した代表フレームが抽出され、これをインデックスとした場合冗長になるという問題がある。さらに、複数の人物に対する撮影区間が互いに重なっている場合について考慮していないため、2人が会話しているような場面からも2枚の代表画像が抽出されてしまい冗長になる。なおかつ、顔領域のみを切り出しているので、

50

2人が会話している状況も把握できない。その結果、利用者は動画を内容を適切に把握することが出来なくなってしまう場合があった。

【0009】

本発明は上述の問題に鑑みなされたものであり、動画像データから当該動画像データの内容をより適切に表現している代表フレーム画像を抽出可能とする技術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】

上述の1以上の問題点を解決するため本発明の情報処理装置は以下の構成を備える。すなわち、情報処理装置において、時系列に並んだ複数のフレーム画像を含む動画像データを入力する入力手段と、入力された動画像データから所定の画像パターンに類似する被写体画像を含むフレーム画像を検出する検出手段と、前記検出手段により検出されたフレーム画像に含まれる前記被写体画像と類似する被写体画像を含むフレーム画像を、前記検出されたフレーム画像の前後にあるフレーム画像から検出する追跡手段と、前記追跡手段により検出された連続したフレーム画像をそれぞれの被写体画像を含む画像シーケンスとして決定し、当該画像シーケンスに対応する前記動画像データ内における時間情報と関連付けて記憶する記憶手段と、前記記憶手段に記憶された1以上の画像シーケンスの各々の開始時刻および終了時刻を分割タイミングとして、前記動画像データを複数の時間区間に分割する分割手段と、前記複数の時間区間の各々について代表フレーム画像を抽出する抽出手段と、前記代表フレーム画像候補の中から、前記画像シーケンスのいずれかに含まれる被写体画像をすべて含む代表フレーム画像候補の組み合わせのうち、より代表フレーム画像候補の個数の少ない組み合わせであって、複数の画像シーケンスが重なる時間区間から抽出された代表フレーム画像候補を含む組み合わせを優先して、当該組み合わせを構成する代表フレーム画像候補を、前記入力された動画像データの代表フレーム画像として選択する選択手段と、を備える。

【発明の効果】

【0012】

本発明によれば、動画像データから当該動画像データの内容をより適切に表現している代表フレーム画像を抽出可能とする技術を提供することができる。

【発明を実施するための最良の形態】

【0013】

以下に、図面を参照して、この発明の好適な実施の形態を詳しく説明する。なお、以下の実施の形態はあくまで例示であり、本発明の範囲を限定する趣旨のものではない。

【0014】

(第1実施形態)

<概要>

第1実施形態では、動画像データから顔画像を検索し、個々の人物が写っている時間区間および人物の組み合わせなどに基づいて、代表フレーム画像を選択する方法について説明する。

【0015】

<装置構成>

図1は、第1実施形態に係る情報処理装置の内部構成図である。

【0016】

情報処理装置は、CPU1001、ROM1002、CD-ROMドライブ1003、RAM1006、ハードディスクドライブ(HDD)1007、IEEE1394インターフェース(I/F)1010を含んでいる。そして、これらの各部はシステムバス1011を介して互いに通信可能なように接続されている。また、情報処理装置には、ユーザインタフェースとして、キーボード1004、マウス1005、ディスプレイ1008、プリンタ1009が接続されている。

【0017】

10

20

30

40

50

CPU1001は、画像処理装置全体の動作制御を司り、例えばROM1002などにあらかじめ記憶された処理プログラムを読み出して実行することで図2で後述する各機能部を実現する。ROM1002は、CPU1001により実行されることにより後述の制御動作を行なうプログラムなどが格納される。RAM1006は、後述する顔シーケンス情報などの一時的なデータを格納する。また、CD-ROMドライブ1003は、CD-ROM1013に格納された制御プログラムを読み取り、当該制御プログラムをRAM1006に格納することが出来る。また、HDD1007には、IEEE1394I/F1010を經由してカムコーダ1012から読み取った動画データデータを記憶する。

【0018】

なお、以下の説明においては、情報処理装置とカムコーダ1012とはIEEE1394I/Fを介して接続され相互に通信可能であるものとする。

【0019】

図2は、第1実施形態に係る情報処理装置の機能ブロック図である。また、図3は、各機能部内部の詳細機能ブロックを示す図である。

【0020】

映像入力部2010は、IEEE1394インターフェース1010を介してカムコーダ1012から入力となる映像として動画データの情報を入力する。なお、映像入力部2010は、動画データを読み込み可能なものであれば、任意のインターフェース機器であってよい。なお、動画データには、時系列に並んだ複数のフレーム画像が格納されている。

【0021】

顔シーケンス生成部2020は、映像入力部2010で入力した映像を解析し、顔が写っている映像期間において各フレームから顔画像を抽出し、顔シーケンス(画像シーケンス)として出力する。なお、ここで顔シーケンスとは、連続した映像期間から抽出された顔画像および、その付帯情報の集まりを言う。付帯情報としては、顔画像を抽出したフレームの時間位置、そのフレームにおける顔画像を切り取った領域の情報などがある。

【0022】

顔シーケンス生成部2020は、画像メモリ2021、顔検出部2022、顔追跡部2023、顔シーケンス記憶部2024より構成される。画像メモリ2021は、映像入力部2010から出力された動画データをフレームごとに一時的にRAM1006へ記憶する。ここで読み込まれた画像データは、2次元配列のデータであり、例えば各々が8ビットの画素により構成されるRGBの3面により構成される。このとき、画像データがMP EG、JPE G等の方式により圧縮されている場合は、画像データを所定の解凍方式にしたがって解凍し、RGB各画素により構成される画像データとする。

【0023】

顔検出部2022は、動画データの所定のフレームから人物顔パターンの検出を行い、その検出結果を出力する。すなわち、動画データの所定フレーム間隔ごとに各フレームから顔検出を行う。ここでは、以下の参考文献1で提案されているニューラル・ネットワークにより画像中の顔パターンを検出する方法を適用した場合について説明する。

【0024】

まず、顔の検出を対象とする画像データをメモリに読み込み、顔と照合する所定の領域を読み込んだ画像中から切り出す。そして、切り出した領域の画素値の分布を入力としてニューラル・ネットワークによる演算で一つの出力を得る。このとき、ニューラル・ネットワークの重み、閾値が膨大な顔画像パターンと非顔画像パターンによりあらかじめ学習されており、例えば、ニューラル・ネットワークの出力が0以上なら顔、それ以外是非顔であると判別する。

【0025】

図4は、ニューラル・ネットワークの手法により画像中から顔を検出する様子を例示的に示す図である。特に、ニューラル・ネットワークの入力である顔と照合する画像パターンの切り出し位置を、画像全域に対して縦横順次に走査する様子を示している。なお、様

10

20

30

40

50

々な大きさの顔の検出に対応するため、図4に示すように読み込んだ画像を所定の割合で順次縮小し、それに対して前述した顔検出の走査を行うように構成すると好適である。

【0026】

なお、画像中から顔を検出する方法は上で説明したニューラル・ネットワークによる方法に限定されるものではなく、例えば参考文献2に挙げられている各種方式が適用可能である。

参考文献1: Rowley et al, "Neural network-based face detection", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL.20, NO.1, JANUARY 1998

参考文献2: Yang et al, "Detecting Faces in Images: A Survey", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL.24, NO.1, JANUARY 2002

10

【0027】

顔追跡部2023は、顔検出部2022で検出された人物顔パターンを後続するフレーム中から探索し、その追跡結果から顔領域の情報と顔シーケンスの区間を出力する。

【0028】

図5は、動画像データから顔シーケンスを生成する様子を例示的に示す図である。

【0029】

顔検出の結果、図5に示すように動画像データの所定間隔のフレームにおいて顔を検出した結果、フレーム中から切り取られた矩形の画像パターンが顔画像として抽出される。そして、検出された顔画像の各々について後続するフレームで追跡を行い、その結果を顔シーケンスとして出力する。その際、顔画像を抽出したフレームの動画像データ内における時間情報(時間位置)、そのフレームにおける顔画像を切り取った領域の情報、などの付帯情報も関連付けて出力する。

20

【0030】

<装置の動作>

・顔シーケンスの検出処理

図6は、顔シーケンスを検出するためのフローチャートである。これは、顔追跡部2023の動作に相当する。

【0031】

ステップS5001により処理を開始し、まず、ステップS5002では、顔パターンの領域情報をもとに後続のフレームにおいて顔パターンの探索を行う探索領域を設定する。顔検出部2022において顔を検出したフレームの次フレームから探索を行う場合には、顔検出結果である顔の領域に対して水平、垂直位置について所定量だけその中心位置をずらした近傍の矩形領域を顔の探索範囲とする。さらに後続するフレームについて探索を行う場合には、同様に顔の追跡結果である顔の領域を利用する。

30

【0032】

ステップS5003では、探索領域内で切り取られた領域と探索する顔パターンとの相関をもとに顔の追跡を行う。例えば、探索領域として設定された中心位置を中心として探索する顔パターンと同じ大きさの矩形領域を順次切出し、切り出した領域と探索する顔パターンとの輝度分布をテンプレートとした相関値を算出する。そして、相関値が最も高い領域を顔パターンの追跡結果として、その相関値とともに出力する。なお、ここで顔パターンの追跡のために輝度分布の相関値を用いたが、RGBごとの画素値分布の相関を用いてもよい。また、領域内での輝度分布やRGB値のヒストグラムなど画像特徴量の相関を用いてもよい。

40

【0033】

ステップS5004では、顔の追跡処理で出力された相関値が所定の値以上であるかどうかで、顔が追跡できたかどうかを判定する。所定の値以上の場合には、類似度が高いので顔が正確に追跡できたと判定しステップS5005に進む。一方、所定値以下の場合には、類似度が低いので顔が追跡できなかったと判断し、顔の追跡を終了する(S5008)。

【0034】

50

ステップS5005では、追跡した先に顔画像があるかどうかを判定する。顔画像がある場合にはステップS5006に進み、無い場合にはステップS5007に進む。

【0035】

ステップS5006では、ステップS5005で検出した顔画像を、現在追跡中の顔であると見なし顔シーケンスとして統合を行う。この際に、顔情報の類似度計算を行い、統合するかどうかの判定を併せて行っても良い。

【0036】

ステップS5007では、顔の追跡を行なう対象とするフレームを後続するフレームに変更しステップS5002に戻る。以上の処理を繰り返し行うことで検出した顔ごとの顔シーケンスを得る。

【0037】

なお、ここでは、顔検出部2022で検出された人物顔パターンごとに後続するフレーム中から顔パターンを探索し追跡するよう説明した。しかし、顔パターンが検出されたフレーム画像に先行するフレーム画像を対象として顔パターンを探索し追跡するよう構成してもよい。その他、例えば動画像から動きベクトル分布を求め、当該動きベクトル分布を手がかりにして顔の追跡を行うようにしてもよい。

【0038】

また、顔の前を何かが横切った場合や、フラッシュなどの影響などの場合に、顔シーケンスが過分割されることを防ぐため、所定の時間間隔をあけたフレームを使って顔追跡を行っても良い。また、時間的に近接する二つの顔シーケンスの顔特徴の相関を求め、相関が高い場合は区間を1つに結合しても良い。すなわち、結合する前側の区間の開始から後ろ側の区間の終了時までを結合した区間の区間とし、付帯情報もあわせて統合する。代表パターンは簡単には片方の顔シーケンスのものをいれれば良い。

【0039】

顔シーケンスの類似度判定および結合は全ての前後の顔シーケンスについて順次行われ、類似する複数の顔シーケンスが1つの顔シーケンスに統合される。ただし、顔シーケンスに対応する映像区間が所定の時間以上離れている組は顔シーケンスの結合の候補としては用いない。また、映像中に人物が複数登場する場合には、複数の顔シーケンスで映像区間が重なる場合が生じるが、このような場合には、それぞれの顔シーケンスに対応する人物が別の人物と見なせるので顔シーケンスの結合の候補としては用いない。

【0040】

図7は、顔シーケンスの情報を記録した付帯情報の一例を示す図である。当該データには、検出された顔シーケンスの各々についての情報が記録されている。

【0041】

なお、ここでは、第1カラムはインデックスである「シーケンス番号」、第2カラムは動画像データ先頭からの秒数を示す「開始時刻」、および、第3カラムは顔シーケンスの継続時間を示す「長さ」が記録されるものとして説明する。なお、顔検出部が人物毎に検出可能であり、同一人物であるか、または誰であるかまで判別できる場合は人物IDを併せて記録しても良い。

【0042】

・代表フレーム画像抽出処理

代表フレーム抽出部2030は、所定の動画区間からその区間の内容を良く表すフレームを1ないし複数抽出する。代表フレーム抽出部2030は、顔シーケンス情報メモリ2031、代表フレーム抽出区間決定部2032、代表フレーム抽出位置決定部2033により構成される。顔シーケンス情報メモリ2031は、顔シーケンス生成部2020で生成した顔シーケンスデータを格納する部分である。

【0043】

図8は、顔シーケンスと代表フレーム画像を抽出する対象となる時間区間との関係を示す図である。図8に示される時間範囲においては3つの顔シーケンスA～Cが検出され、顔シーケンスAと顔シーケンスB、顔シーケンスBと顔シーケンスC、とは重複期間があ

10

20

30

40

50

る。このような、状態においては、代表フレーム画像を抽出する対象とする時間区間として、5つの時間区間1～5が設定される。つまり、登場人物の組み合わせが一定である時間区間を、代表フレーム画像抽出の対象となる時間区間として選定する。以下、この時間区間の選定処理について詳述する。

#### 【0044】

図9は、代表フレーム画像抽出の対象となる時間区間を選定するフローチャートである。

#### 【0045】

ステップS8001により処理を開始し、まず、ステップS8002では、顔シーケンスの構成が変化する時点(時刻)を求める。図8に示した3つの顔シーケンスに対して、これを求めた結果が、時間軸に直交する点線部分である。つまり、1以上の顔シーケンスの各々の開始時刻および終了時刻が分割タイミングとなる。この時間区間は、顔シーケンスの構成(登場人物の組み合わせなど)が一定の時間区間となり、これを以降では”構成一定区間”と呼ぶことにする。図8では、構成一定区間が5つ選定される。

#### 【0046】

ステップS8003では、構成一定区間の各々に対して、区間内のフレーム画像について所定の評価ルール(抽出の基準)に基づいて評価値を計算する。この計算方式には、いろいろなものが考えられるが、ここでは例として、顔画像そのものの評価を行う顔画像評価部を用いるものとする。これ以外にも、フレーム内の顔の位置や、フレーム内の彩度、色相の分布などを用いる方法があり、またこれらを混合して使用方法なども考えられる。また、前後のフレームとの画像全体の相関を求め、動きの激しくない箇所では評価値が高くなるよう調整しても良い。以下、具体的な算出方法の一例を述べる。なお、以下の説明においては、代表フレーム画像として適しているものほど評価値が大きくなるような評価計算式であるとする。

#### 【0047】

代表フレーム抽出区間決定部2032の中に配置される顔状態判定部(不図示)は、主要顔判定部(不図示)で評価値が高い被写体について、被写体を良く表す顔画像であるか否かを示す評価値を算出する。例えば、顔の向き、目の開閉具合、表情、照明による影、顔の一部が他のオブジェクトで隠れていないかに着目し評価地を導出する。顔の向きや表情に関しては顔画像パターンに対して向きや表情の教師データを与えて前述したニューラルネットワークの重み、閾値を求めることで実現できる。

#### 【0048】

なお、顔の状態を正確に判定するためには、顔画像の中に目、口、鼻などの顔の各パーツが存在することが重要であると考えられる。すなわち、顔が横方向や斜めを向いているものよりも、正面を向いているものの方が顔の特徴を正確に表現している。したがって、顔状態判定部は顔シーケンス中の各顔画像の顔の向きを検出する構成をもつ。例えば、前述したニューラル・ネットワークによる顔判別器と同じ構成の複数の顔判別器を備える。但し、各顔判別器の判別のためのパラメータを顔の向きごとにサンプル学習によりチューニングし設定しておく。そして、複数の顔判別器のうち、もっとも出力の高い、すなわち尤度の高い顔判別器に対応した顔の向きを出力し、正面を向いた場合に高い評価値を与える。

#### 【0049】

また、例えば、顔画像から目、口、鼻などのパーツを個別に探索し、それぞれの存在の有無を解析結果として出力するようにしてもよい。また、目が開いているか、閉じているかを判定し解析結果を出力するようにしてもよい。また、顔に対する照明状態がよく全体的に肌部分が明るく撮影されている場合には部分的に陰がある場合よりも高い評価値を与えてもよい。影や隠れについては、たとえば参考文献3にあるようなEigenfaceと呼ばれる手法で顔パターンをモデル化し、モデルパラメータを使って近似された顔画像と元画の顔画像との近似差を評価することで影や隠れ領域を求めることが出来る。モデルパラメータの一部には顔に当たった照明成分が含まれるので、その成分の強さから照明の

10

20

30

40

50



方向と強さを求めることが出来る。

参考文献 3 : M. Turk and A. Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuroscience 3 (1): 71-86, 1991

【 0 0 5 0 】

ステップ S 8 0 0 4 では、各構成一定区間から、一番評価値が高いフレームを代表フレームとして選択して、これを記録する。これを以降では " 代表フレーム候補位置 " と呼ぶ。この処理により、各構成一定区間について、代表フレーム候補があるものについては、その代表フレーム候補を抽出したこととなる。

【 0 0 5 1 】

上述の代表フレーム画像抽出の対象となる時間区間の選定および各時間区間の代表フレーム候補を抽出した後、以下の処理により代表フレーム画像が抽出される。

【 0 0 5 2 】

代表フレーム抽出位置決定部 2 0 3 3 は、代表フレーム抽出区間決定部 2 0 3 2 で抽出された情報を基に、最終的な代表フレームを決定する。ここでは、被写体である登場人物が複数の場合には、その登場人物が一緒に写っているという情報を重視する。これは家庭内のイベントや出来事を主な撮影対象とする場合などに有効である。

【 0 0 5 3 】

この場合、複数の人物が同時に写っている時間区間から得られた代表フレーム候補を、最終的な代表フレームに入れることが重要となる。そこで、図 8 における時間区間 2 と時間区間 4 の代表フレーム候補を、最終的な代表フレーム候補とする。また、それ以外の構成一定区間でも、評価値が高い代表フレーム候補がある場合には、併せて最終的な代表フレームに追加する。また、同じ登場人物の構成一致区間が存在する場合には、代表フレーム候補の評価値の高いほうを選択する。なお、代表フレーム画像として適しているものほど評価値が小さくなるような評価計算式である場合には最小の評価値のフレーム画像を選択する。

【 0 0 5 4 】

このようにして、動画像データに複数の登場人物が写っている時間区間が存在する場合には、必ずその時間区間の代表フレーム画像を抽出することが可能となる。

【 0 0 5 5 】

なお、一般に一覧表示を目的とする場合には代表フレーム画像は少ない枚数が好ましく、検索に使用するインデックス情報を目的とする場合には代表フレーム画像は多くの枚数が存在したほうが良い。このように、目的によって必要な代表フレーム画像の数は異なるため、記憶部に格納する際に評価値を合わせて格納しておき、ユーザから指定された枚数の代表フレーム画像を残しておくことで必要な代表フレームの枚数を評価値の高いものから順に選ぶことが出来る。

【 0 0 5 6 】

また、代表フレーム画像の抽出結果を記憶部に格納する際のデータ形式は、システムが解釈可能である限り任意のデータ形式が使用できる。例えば、テキスト形式や、独自の D T D ( Document Type Definition ) を定義した X M L ( Extensible Markup Language ) 形式等で表現するようにしてもよい。

【 0 0 5 7 】

図 1 0 は、代表フレーム画像の一覧の出力例を示す図である。抽出した代表フレーム画像の一覧は、ディスプレイ 1 0 0 8 に表示するよう構成してもよいし、プリンタ 1 0 0 9 によって印刷出力するよう構成しても良い。

【 0 0 5 8 】

以上説明したとおり第 1 実施形態に係る情報処理装置によれば、動画像データから顔シーケンスを検出する。そして、検出された複数の顔シーケンスの時間区間に基づいて、登場人物の組み合わせなどが一定の時間区間 ( 構成一定区間 ) を選定する。構成一定区間の各々について代表フレーム画像を抽出することにより、より適切な代表フレーム画像を抽出することが出来る。

10

20

30

40

50

## 【 0 0 5 9 】

特に、代表フレーム画像を、時間軸上で複数の画像シーケンスが重なり合う構成一定区間から優先して抽出することにより、より適切かつより少ない枚数の代表フレーム画像を抽出することも出来る。

## 【 0 0 6 0 】

(第2実施形態)

第2実施形態では、フレーム画像の構図を指定するための1以上の”構図テンプレート”を利用して代表フレーム画像を抽出する手法について説明する。

## 【 0 0 6 1 】

つまり、代表フレーム抽出の対象となる時間区間の選定を行う際に、主要な被写体の判定や適切な撮影構図の判定を利用することで、適切な代表フレームを保有する可能性のある代表フレーム抽出区間を選択することが可能としている。なお、構成や処理の内容は、代表フレーム抽出区間決定部2032以外の部分については、第1実施形態と同様であるので説明は省略する。

10

## 【 0 0 6 2 】

図11は、代表フレーム画像抽出の対象となる時間区間を選定するフローチャートである。

## 【 0 0 6 3 】

ステップS10001により処理を開始し、まず、ステップS10002では、顔シーケンスの構成が変化する時点(時刻)を求める。図8に示した3つの顔シーケンスに対して、これを求めた結果が、時間軸に直交する点線部分である。この時間区間は、顔シーケンスの構成(登場人物の組み合わせなど)が一定の時間区間となり、これを以降では”構成一定区間”と呼ぶことにする。図8では、構成一定区間が5つ選定される。

20

## 【 0 0 6 4 】

ステップS10003では、構成一定区間内の各フレームを、予め指定されている構図テンプレートと照合して合致するものがあるかどうかの構図判定を行う。

## 【 0 0 6 5 】

図12は、構図テンプレートの例を示す図である。構図テンプレートは、フレーム画像の構図の典型的なパターンを示すものである。構図テンプレート1201(テンプレート番号1)は、一人の人(X)が一定時間以上、ある程度の大きさの顔が写っている構図を示している。一方、構図テンプレート1202(テンプレート番号2)は、二人の人(XとY)が一定時間以上、両方とも、顔がある程度の大きさで、かつ、同じぐらいの大きさで写っているものを示している。これらは、あくまでも例であり、これら以外にも、特定のパターンの動作や移動などの構図テンプレートもあってよい。

30

## 【 0 0 6 6 】

このように、構図テンプレートを利用することで、個々のフレーム画像に代表フレーム画像として適切なものが含まれるかどうかを調べる。これにより、後述の処理と合わせて、主要な被写体の顔の選択も同時に行っている。この構図テンプレートと比較して、顔の位置、大きさ、撮影された時間などを基準として、どれだけ満たしているかを数値化し、これを適合値とする。

40

## 【 0 0 6 7 】

ステップS10004では、各構成一定区間に対して、この適合値が一番大きい構図テンプレートを選び、その適合値がある閾値を超えた場合、それが該当する構成テンプレートとみなし、その構成テンプレート番号を記録する。

## 【 0 0 6 8 】

図13は、図7の顔シーケンスに対して代表フレーム抽出区間決定処理を適用した結果を示す図である。

## 【 0 0 6 9 】

なお、ここでは、第1カラムは図8における”構成一定区間”の番号であり、第2カラムは適合すると判定された”テンプレート番号”である。ここで、構成一定区間5に対し

50

ては、構図テンプレートでの適合値が閾値を超えなかったため、構図テンプレート番号は「なし」になっている。なお、図13の第5カラムは、各構成一定区間の登場人物であり、これは各区間にある顔シーケンスの組み合わせに相当する。

**【0070】**

ステップS10005では、構図テンプレート番号が決定されている構成一定区間に対して、その区間内の動画のフレームについて評価値を計算する。計算方式には、いろいろなものが考えられるが、ここでは例として、前述した顔画像そのものの評価を行う顔画像評価部を用いるものとする。しかし、これ以外にも、フレーム内の顔の位置や、フレーム内の彩度、色相の分布、それから前述した構図テンプレートの適合値などを用いる方法があり、またこれらを混合して使用方法なども考えられる。また、前後のフレームとの画像全体の相関を求め、動きの激しくない箇所で評価値が高くなるよう調整しても良い。

10

**【0071】**

ステップS10006では、構図テンプレート番号がある構成一定区間に対して、最も評価値が高いフレームを代表フレームとして選択し記録する。前述の図13における第3カラムは動画内の選択されたフレームの位置（動画先頭からの秒数）で、これを代表フレーム候補位置と呼ぶ。また、第4カラムは、その代表フレーム候補の評価値である。これにより、各構成一定区間について、代表フレーム候補があるものについては、その代表フレーム候補を抽出したこととなる。

**【0072】**

上述の代表フレーム候補を抽出した後、以下の処理により代表フレーム画像が抽出される。

20

**【0073】**

代表フレーム抽出位置決定部2033は、図13のように記録された情報を基に、最終的な代表フレームを決定する。ここでは、被写体である登場人物が複数の場合には、その登場人物と一緒に写っているという情報を重視する。これは家庭内のイベントや出来事を主な撮影対象とする場合などに有効である。

この場合、複数の人物が同時に写っている時間区間から得られた代表フレーム候補を、最終的な代表フレームに入れることが重要となる。そこで、図13のような状況においては、時間区間2と時間区間4の代表フレーム候補を、最終的な代表フレーム候補とする。また、それ以外の構成一定区間でも、評価値が高い代表フレーム候補がある場合には、併せて最終的な代表フレームに追加する。また、同じ登場人物の構成一致区間が存在する場合には、代表フレーム候補の評価値の高いほうを選択する。

30

**【0074】**

このようにして、動画データに複数の登場人物が写っている時間区間が存在する場合には、必ずその時間区間の代表フレーム画像を抽出することが可能となる。特に、予め指定した構図テンプレートにより示される構図のフレーム画像が抽出されるため、より適切な代表フレーム画像を抽出することが可能となる。

**【0075】**

（第3実施形態）

40

第3実施形態では、図13に示される情報に基づいて、より少ない枚数の代表フレーム画像を抽出する例について説明する。具体的には、同じ人物の組み合わせに対して代表フレーム画像を1つのみ抽出する手法について説明する。なお、構成や処理の内容は、代表フレーム抽出位置決定部2033以外の部分については、第2実施形態と同様であるので説明は省略する。

**【0076】**

図14は、代表フレーム抽出位置決定部2033の動作フローチャートである。

**【0077】**

ステップS13001により処理を開始し、まず、ステップS13002では、図13の結果に基づいて、代表フレーム候補に写っている被写体である登場人物の情報を取得す

50

る。ここでは、動画像データの全登場人物が A , B , C となっているという情報を得る。人物の特定については、前述したような、その顔画像そのもの類似度、個人の弁別や同定を行うような方法を用いても構わない。

【 0 0 7 8 】

ステップ S 1 3 0 0 3 では、ステップ S 1 3 0 0 2 で取得した登場人物を含む代表フレーム候補の組み合わせを導出する。

【 0 0 7 9 】

図 1 5 は、ステップ S 1 3 0 0 3 における処理の結果を示す図である。第 1 カラムは、代表フレームの抽出位置の組み合わせの番号（代表フレーム抽出位置候補番号）であり、図 1 3 から " A のみ "、" A と B "、" B のみ "、" B と C " の 4 つの組み合わせがあることがわかる。第 2 カラムは、その組み合わせの内容で、構成一致区間番号の集合になっている。

10

【 0 0 8 0 】

ステップ S 1 3 0 0 4 では、代表フレーム抽出位置の組み合わせの中で、構成一致区間の個数が少ないものを探索し、代表フレーム画像を抽出する候補として残留させる。前述の図 1 5 の第 3 カラムは、構成一致区間の個数であり、ここでは個数が一番少ない（ 2 個 ）代表フレーム抽出位置候補番号 3 と代表フレーム抽出位置候補番号 4 を候補として残留させる。

【 0 0 8 1 】

ステップ S 1 3 0 0 5 では、残留した代表フレーム抽出位置候補について、その中の代表フレーム候補の評価値の合計を計算し、代表フレーム抽出位置候補の評価値とする。前述の第 4 カラムが、その合計（評価値）である。なお、代表フレーム抽出位置候補のうち、構成一致区間の個数が最低でなかったものについては、計算する必要がないので、図 1 5 では「 x 」と表記されている。

20

【 0 0 8 2 】

ステップ S 1 3 0 0 6 では、代表フレーム候補に複数の被写体が存在し、かつ、代表フレーム抽出位置候補の評価値が最大のものを最終的な代表フレームとして選択する。前述の図 1 5 の第 5 カラムに、最終的に選択された代表フレームの位置が示されている。

【 0 0 8 3 】

このようにして、動画像データに複数の登場人物が写っている時間区間が存在する場合には、より適切な代表フレーム画像を抽出することが可能となる。

30

【 0 0 8 4 】

（他の実施形態）

以上、本発明の実施形態について詳述したが、本発明は、複数の機器から構成されるシステムに適用しても良いし、また、一つの機器からなる装置に適用しても良い。

【 0 0 8 5 】

なお、本発明は、前述した実施形態の機能を実現するプログラムを、システム或いは装置に直接或いは遠隔から供給し、そのシステム或いは装置が、供給されたプログラムコードを読み出して実行することによっても達成される。従って、本発明の機能処理をコンピュータで実現するために、コンピュータにインストールされるプログラムコード自体も本発明の技術的範囲に含まれる。

40

【 0 0 8 6 】

その場合、プログラムの機能を有していれば、オブジェクトコード、インタプリタにより実行されるプログラム、OS に供給するスクリプトデータ等、プログラムの形態を問わない。

【 0 0 8 7 】

プログラムを供給するための記録媒体としては、例えば、フロッピー（登録商標）ディスク、ハードディスク、光ディスク（CD、DVD）、光磁気ディスク、磁気テープ、不揮発性のメモリカード、ROMなどがある。

【 0 0 8 8 】

50

また、コンピュータが、読み出したプログラムを実行することによって、前述した実施形態の機能が実現される。その他、そのプログラムの指示に基づき、コンピュータ上で稼動しているOSなどが、実際の処理の一部または全部を行い、その処理によっても前述した実施形態の機能が実現され得る。

【0089】

さらに、記録媒体から読み出されたプログラムが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれる。その後、そのプログラムの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によっても前述した実施形態の機能が実現される。

【図面の簡単な説明】

【0090】

【図1】第1実施形態に係る情報処理装置の内部構成図である。

【図2】第1実施形態に係る情報処理装置の機能ブロック図である。

【図3】各機能部内部の詳細機能ブロックを示す図である。

【図4】ニューラル・ネットワークの手法により画像中から顔を検出する様子を例示的に示す図である。

【図5】動画データから顔シーケンスを生成する様子を例示的に示す図である。

【図6】顔シーケンスを検出するためのフローチャートである。

【図7】顔シーケンスの情報を記録した付帯情報の一例を示す図である。

【図8】顔シーケンスと代表フレーム画像を抽出する対象となる時間区間との関係を示す図である。

【図9】代表フレーム画像抽出の対象となる時間区間を選定するフローチャートである。

【図10】代表フレーム画像の一覧の出力例を示す図である。

【図11】代表フレーム画像抽出の対象となる時間区間を選定するフローチャートである。

【図12】構図テンプレートの例を示す図である。

【図13】図7の顔シーケンスに対して代表フレーム抽出区間決定処理を適用した結果を示す図である。

【図14】代表フレーム抽出位置決定部の動作フローチャートである。

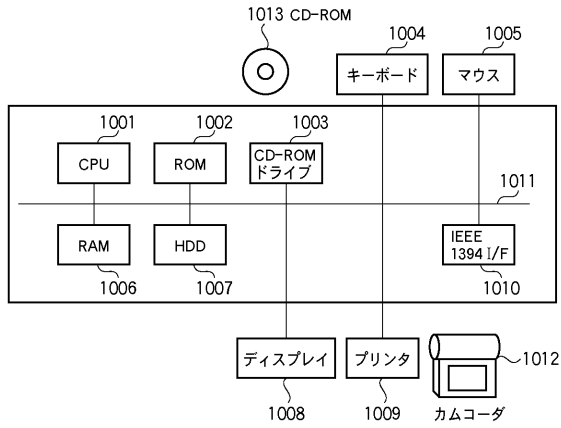
【図15】ステップS13003における処理の結果を示す図である。

10

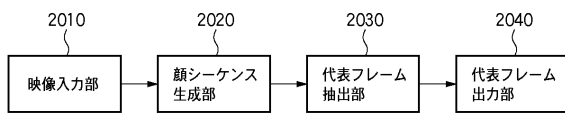
20

30

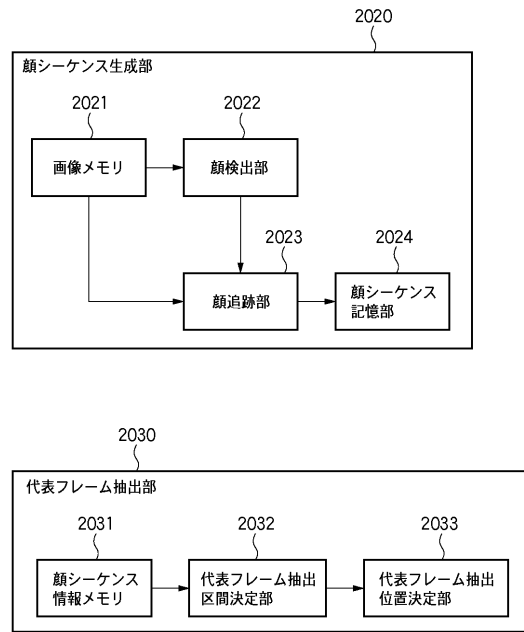
【図1】



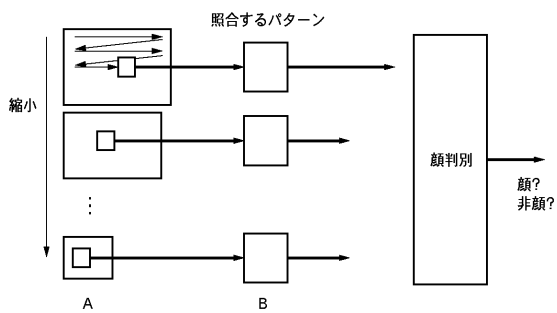
【図2】



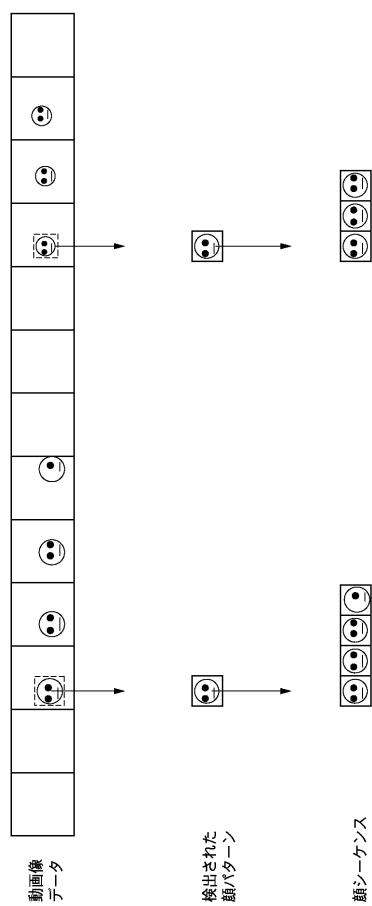
【図3】



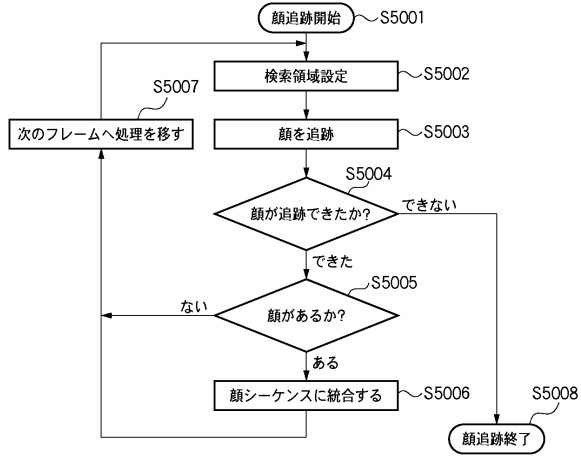
【図4】



【図5】



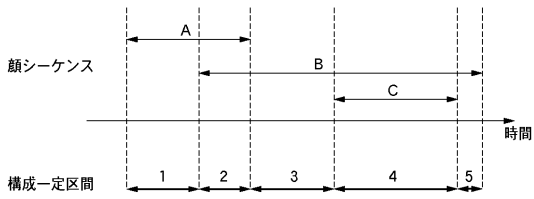
【図6】



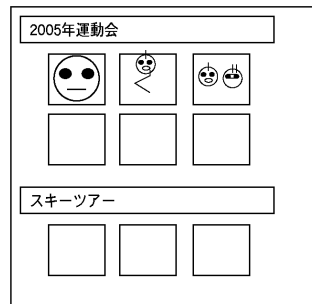
【図7】

シーケンス番号	開始時刻	長さ
1	43.50	2.83
2	55.34	12.58
3	58.35	14.33
4	96.24	8.73
5	105.13	6.24
6	119.15	19.52
7	131.45	2.30
8	158.83	13.5
...	...	...

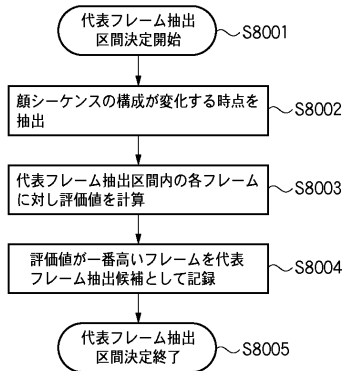
【図8】



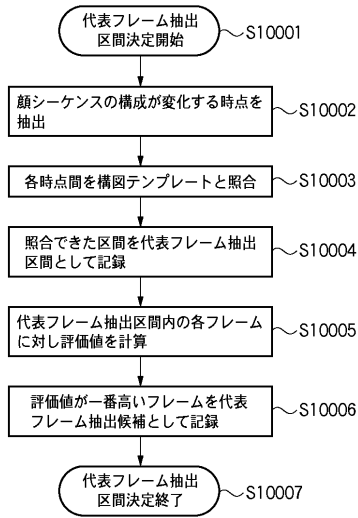
【図10】



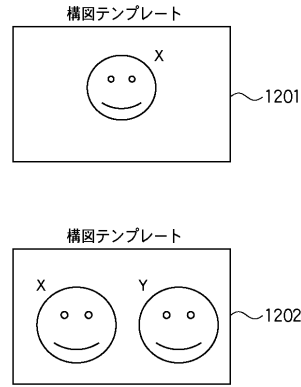
【図9】



【図 1 1】



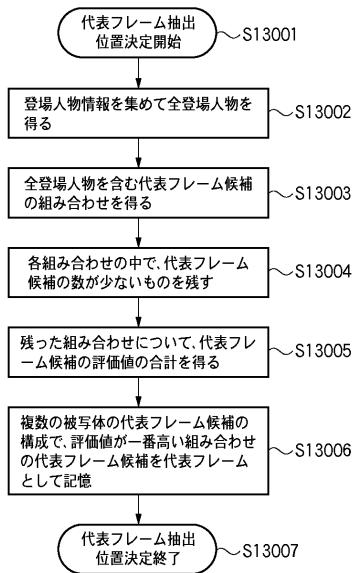
【図 1 2】



【図 1 3】

構成一定区間 番号	構図テンプレート 番号	代表フレーム 候補位置	代表フレーム 候補評価値	登場人物
1	1	2.1	0.42	A
2	2	4.2	0.82	A, B
3	1	5.4	0.55	B
4	2	9.3	0.90	B, C
5	なし	なし	なし	なし

【図 1 4】



【図 1 5】

代表フレーム抽出 位置候補番号	構成一定区間 番号	構成一定区間 個数	代表フレーム候補 評価値合計	代表フレーム 候補位置
1	1, 2, 4	3	X	X
2	1, 3, 4	3	X	X
3	1, 4	2	1.32	X
4	2, 4	2	1.72	4.2, 9.3



---

フロントページの続き

- (72)発明者 相馬 英智  
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
- (72)発明者 東條 洋  
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
- (72)発明者 八代 哲  
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

審査官 畑中 高行

- (56)参考文献 特開2005-101906(JP,A)  
特開2005-227957(JP,A)  
特開2008-211485(JP,A)  
特開2003-348528(JP,A)  
国際公開第2006/028116(WO,A1)  
特開2008-271587(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N5/76-5/956  
G11B27/00-27/06  
G11B27/10-27/34  
G11B20/10-20/16