(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2010-109592
(P2010-109592A)
(43)公開日　平成22年5月13日(2010.5.13)

(51)Int.Cl.			F I			テーマコード（参考）
H04N	5/91	(2006.01)	H04N	5/91	Z	5B057
G06T	1/00	(2006.01)	G06T	1/00	340A	5C053

審査請求 未請求　請求項の数 6　ＯＬ　（全 17 頁）

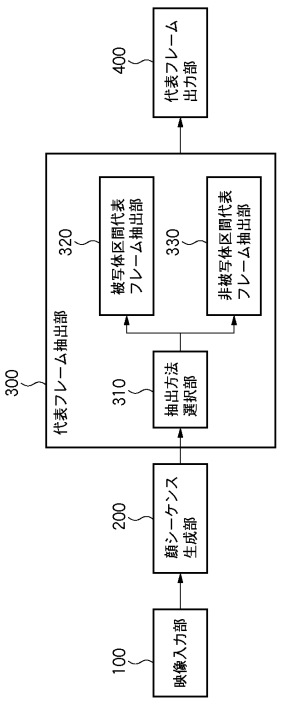
(21)出願番号　特願2008-278607 (P2008-278607)
(22)出願日　平成20年10月29日 (2008.10.29)

(71)出願人　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100076428
弁理士　大塚　康徳
(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)【要約】　（修正有）
【課題】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より適切な代表フレーム画像を抽出する。
【解決手段】顔検出部は、動画像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。顔追跡部は、顔検出部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時間区間を出力する。代表フレーム抽出部は、所定の時間区間のから当該区間の内容（映像）を良く表すフレームを１枚あるいは複数枚抽出する。抽出方法選択部、顔シーケンス生成部の出力に基づいて代表フレーム画像の評価ルール（抽出の基準）を変更する。被写体区間代表フレーム抽出部は、顔シーケンスに含まれるフレーム画像から、人物の内容を良く表す任意数の代表フレーム画像を抽出する。
【選択図】図２



【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画像と類似する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を対象として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画像シーケンスとして、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、
前記動画像データ内の各時刻において前記記憶手段に記憶された１以上の画像シーケンスを含むか否かに基づいて、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、前記動画像データ内の各時刻において前記記憶手段に記憶された１以上の画像シーケンスを含むか否かに基づいて異なる評価ルールで代表フレーム画像を抽出する抽出手段と、
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、少なくとも前記所定の画像パターンに類似する画像を含むフレーム画像を含む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当該時間区間における代表フレーム画像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】
前記抽出手段は、前記所定の画像パターンに類似する画像を含まないフレーム画像を含む時間区間の各々に含まれるフレーム画像間の動きベクトル分布を導出し、導出したベクトル分布に基づいてフレーム画像の各々の評価値を算出し、当該評価値が最大または最小となるフレーム画像を当該時間区間における代表フレーム画像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４】
前記所定の画像パターンは人物の顔画像であ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、
動画像データを入力する入力工程と、
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画像を含むフレーム画像を検出する検出工程と、
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画像と類似する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を対象として検出する追跡工程と、
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画像シーケンスとして、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工程と、
前記動画像データ内の各時刻において前記記憶部に記憶された１以上の画像シーケンスを含むか否かに基づいて、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、
前記複数の時間区間の各々について、前記動画像データ内の各時刻において前記記憶部に記憶された１以上の画像シーケンスを含むか否かに基づいて異なる評価ルールで代表フレーム画像を抽出する抽出工程と、
を備える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、個人でも大量の動画を撮影するようになってきている。一般に、映像データはデータ量が膨大であるため、当該映像データの利用者は、内容の概略を知りたい場合や所望のシーンを探す場合、映像を早回しや巻き戻しを行っている。そこで、動画の内容を短時間で把握するために、動画の内容を良く示すフレームを選択して提示する代表フレーム抽出技術が提案されている。

【０００３】

例えば、特許文献１には一台のカメラにより中断せずに撮影して得られる一連の画像をショットとし、ショットから先頭や末尾、中心など再生時刻に基づいてキーフレームを選択する技術が開示されている。そして、キーフレームの類似性に基づいて複数のショットをまとめて１つのシーンとして結合し、各シーンから所定枚数のキーフレームを選択するよう構成している。また、登場人物を含むフレームがキーフレームとして選択されるように、顔領域を含むフレームを優先的に選択するよう構成している。

【０００４】

特許文献２では、動画に映っている人物の構成が変化した場合に、動画のシーンを分割すると共に、人物の構成を示すインデックスをシーンのそれぞれに付与しシーンを検索する技術が開示されている。さらに、特許文献３では、顔に特化した検出を行う技術が開示されている。具体的には、検出した顔の識別を行うことで映像中に登場する人物の顔を区別し、顔が検出された区間に対して、顔の向き、サイズ、顔の数などに基づいて代表フレームを選択している。

【０００５】

さらに、特許文献４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。具体的には、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インデックスを生成している。

【特許文献１】特開２００２－２２３４１２号公報

【特許文献２】特開２００５－１０１９０６号公報

【特許文献３】特開２００１－１６７１１０号公報

【特許文献４】特許第３３１２１０５号

【発明の開示】

【発明が解決しようとする課題】

【０００６】

しかしながら、特許文献１に記載の技術においては、顔を含むキーフレームを選択した場合であっても、顔の領域が小さかったり横を向いていたりするなど、利用者にとって良い状態のフレーム画像ではない場合があった。また、デジタルカメラでの撮影対象は子供の成長など家族を被写体とする場合が多い。その場合、人物に着目して代表フレームを抽出する特許文献２，３に記載の技術においては、家族の顔が写ったフレームばかりが並ぶことになってしまっていた。つまり、人物・顔を検出できた動画区間に着目して代表フレームを選択しているため、人物・顔が検出できなかった風景や印象に残る被写体を含むフレームは代表フレームとして選択されることが無かった。さらに、特許文献４に記載の技術においては、フレームごとに評価値を求めているため、ホームビデオで動画の内容把握

を目的とした場合には多数の類似したフレームがインデックスとなってしまい冗長になってしまっていた。

【０００７】

つまり、特定の被写体（例えば顔）に着目して代表フレームを選ぶと、”だれが”映っているかはわかるが、“何処で”映したという情報が欠けてしまっていた。そのため、ホームビデオなどで撮影したパーソナルコンテンツなどにおいては、必ずしも適切な代表フレームが抽出できていないという問題があった。

【０００８】

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】

【０００９】

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画像と類似する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を対象として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画像シーケンスとして、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、前記動画像データ内の各時刻において前記記憶手段に記憶された１以上の画像シーケンスを含むか否かに基づいて、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、前記動画像データ内の各時刻において前記記憶手段に記憶された１以上の画像シーケンスを含むか否かに基づいて異なる評価ルールで代表フレーム画像を抽出する抽出手段と、を備える。

【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の制御方法は以下の構成を備える。すなわち、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、動画像データを入力する入力工程と、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画像を含むフレーム画像を検出する検出工程と、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画像と類似する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を対象として検出する追跡工程と、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画像シーケンスとして、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工程と、前記動画像データ内の各時刻において前記記憶部に記憶された１以上の画像シーケンスを含むか否かに基づいて、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、前記複数の時間区間の各々について、前記動画像データ内の各時刻において前記記憶部に記憶された１以上の画像シーケンスを含むか否かに基づいて異なる評価ルールで代表フレーム画像を抽出する抽出工程と、を備える。

【発明の効果】

【００１１】

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】

【００１２】

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。

【００１３】

（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、顔画像が含まれている時間区間か否かに基づいて、情報処理装置における代表フレーム画像の評価ルール（基準）を変更する。それにより、顔画像が含まれている時間区間からは人物の画像を代表フレーム画像として選択し、顔画像が含まれていない時間区間からは風景の画像を代表フレーム画像として選択する方法について説明する。
【００１４】
＜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。
【００１５】
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。
【００１６】
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を経由してカムコーダ１０１２から読み取った動画像データを記憶する。
【００１７】
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。
【００１８】
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。なお、各部の詳細動作については後述する。
【００１９】
１００は映像入力部であり、ＩＥＥＥ１３９４Ｉ／Ｆ１０１０を介してカムコーダ１０１２から動画像データを入力する。なお、映像入力部１００は、動画像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、動画像データには、時系列に並んだ複数のフレーム画像が格納されている。
【００２０】
２００は顔シーケンス生成部であり、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。
【００２１】
顔シーケンス生成部２００は、画像メモリ２１０、顔検出部２２０、顔追跡部２３０、顔シーケンス記憶部より構成される。画像メモリ２１０は、映像入力部１００から出力された動画像データをフレームごとに一時的にＲＡＭ１００６へ記憶する。顔検出部２２０は、動画像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。顔追跡部２３０は、顔検出部２２０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時間区間を出力する。

【００２２】

３００は代表フレーム抽出部であり、所定の時間区間のから当該区間の内容（映像）を良く表すフレームを１枚あるいは複数枚抽出する。代表フレーム抽出部３００は、抽出方法選択部３１０、被写体区間代表フレーム抽出部３２０、非被写体区間代表フレーム抽出部３３０より構成される。

【００２３】

抽出方法選択部３１０は、顔シーケンス生成部２００の出力に基づいて代表フレーム画像の評価ルール（抽出の基準）を変更する。被写体区間代表フレーム抽出部３２０は、顔シーケンスに含まれるフレーム画像から、人物の内容を良く表す任意数の代表フレーム画像を抽出する。また、非被写体区間代表フレーム抽出部３３０は、いずれの顔シーケンスにも属さない時間区間のフレーム画像から、風景や印象に残るオブジェクトを良く表す任意数の代表フレームを抽出する。

【００２４】

４００は代表フレーム出力部であり、抽出した代表フレームを例えばディスプレイ１００８へ表示したりプリンタ１００９により印刷したりする。

【００２５】

＜装置の動作＞

図４は、第１実施形態に係る情報処理装置の動作フローチャートである。

【００２６】

ステップＳ１００では、映像入力部１００は所望の動画像データをフレームごとに画像メモリ２１０に読み込む。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ＭＰＥＧ、ＪＰＥＧ等の方式により圧縮符号化されている場合は、画像データを対応する復号方式にしたがって復号し、ＲＧＢ各画素により構成される画像データを生成する。

【００２７】

ステップＳ２００では、顔検出部２２０は、動画像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、動画像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。

【００２８】

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、それ以外は非顔であると判別する。

【００２９】

図９は、ニューラル・ネットワーク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、様々な大きさの顔の検出に対応するため、図９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、それぞ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。

【００３０】

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

【００３１】
ステップＳ３００では、顔追跡部２３０は、顔検出部２２０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する。そして、ステップＳ４００では、被写体が出現する区間を顔シーケンスとして出力する。すなわち、所定フレーム間隔で検出された顔画像のそれぞれについて後続するフレームで追跡を行い、連続した顔画像の集まりを画像シーケンス（顔シーケンス）として決定し出力する。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。

【００３２】
図５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。

【００３３】
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、顔検出部によって同一人物であるか、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。

【００３４】
図８は、顔シーケンスの生成の処理フローを示す図である。また、図７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。以下、顔シーケンス生成の詳細について説明する。

【００３５】
ステップＳ３０１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２２０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。

【００３６】
ステップＳ３０２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。すなわち、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。

【００３７】
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、例えばＲＧＢ各々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。

【００３８】
ステップＳ３０３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を判定し、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断しステップＳ３０４に進む。また、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。

【００３９】
ステップＳ３０４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Ｓ３０１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

ケンスを取得する。

【００４０】

なお、ここでは、顔検出部２２０で人物顔パターンが検出されたフレーム画像に後続するフレーム画像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、例えば動画像に含まれる各フレーム画像から動きベクトルを導出し、導出した動きベクトルに基づいて顔パターンの追跡を行うよう構成してもよい。

【００４１】

また、顔の前を何かが横切ったりフラッシュなどの影響によって顔シーケンスが過分割されることを防ぐため、所定の時間間隔だけ離れたフレームを使って顔追跡を行っても良い。また、時間的に隣接する２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２つの顔シーケンスを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１つの区間とし、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用いれば良い。

【００４２】

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、類似の顔シーケンスが統合される。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。

【００４３】

・代表フレーム画像の抽出

以下では、ステップＳ４００で出力された１以上の顔シーケンスの情報に基づいて、動画像データから代表フレーム画像を抽出する方法について説明する。

【００４４】

ステップＳ５００では、動画像データに含まれる各時間区間が、顔シーケンスが含まれる期間（以下、被写体区間と呼ぶ）であるか顔シーケンスが含まれない期間（以下、非被写体区間と呼ぶ）であるかを判定する。

【００４５】

図１１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１１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とは重複期間がある。このような、状態においては、被写体区間として区間Ｂと区間Ｃとが設定され、非被写体区間として区間Ａと区間Ｃとが設定される。

【００４６】

抽出方法選択部３１０は、顔シーケンス生成部２００の出力に基づいて代表フレーム画像の評価ルール（抽出の基準）を切り替える。具体的には、何れかの顔シーケンス内（すなわち被写体区間）の時間区間に対しては被写体区間代表フレーム抽出部３２０によって代表フレーム画像を抽出させ、そうでない場合（すなわち非被写体区間）に対しては非被写体区間代表フレーム抽出部３３０よって代表フレーム画像を抽出させる。つまり、図１１の状況において、抽出方法選択部３１０は、区間Ａおよび区間Ｃに対する動画像データを非被写体区間代表フレーム抽出部３３０が処理するよう制御する。一方、区間Ｂおよび区間Ｄに対する動画像データを被写体区間代表フレーム抽出部３２０が処理するように制御する。

【００４７】

・被写体区間からの代表フレーム抽出（Ｓ６００）

被写体区間代表フレーム抽出部３２０は、顔シーケンスを含む一連の時間区間内を対象にして任意数の代表フレームを抽出するものであり、主要顔判定部３２１と、顔状態判定

部３２２と、被写体区間代表フレーム判定部３２３から構成される。

【００４８】

主要顔判定部３２１は撮影者が意図した主要な被写体であるかという観点から評価値を求める。例えば、単一の顔シーケンスから構成される区間においては、撮影者が顔がフレーム内に入るよう操作しているために発生する顔のフレーム内での動きパターン、出現時間などから主要顔の評価値を算出する。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。

【００４９】

図１０は、格子状のブロックに分割したフレーム画像内における顔中心の軌跡をプロットした図である。顔画像のフレーム内での動きパターンは、顔の中心部（顔中心）の軌跡を求め、図１０の各ブロック内に位置した時間の総和とブロック外へ移動した時の方向別の回数を入力として導出される。このようにして多数の動きパターンを用意し、また、主要顔であるか否かを教師データとして前述したニューラルネットワークの重み、閾値を求めておくことで、評価値を算出する。

【００５０】

なお、複数の顔シーケンスから構成される区間においては、各々の顔シーケンスについて主要被写体評価を行なう。その後、相対的な評価値が所定の閾値より低い顔シーケンスを対象外とし、顔の相対的なサイズや接地箇所などから撮影者と被写体の相対的な距離を推測する。また、時間的重なりが長い場合には重なり区間の評価値を相対的に高くすると良い。

【００５１】

顔状態判定部３２２は、主要顔判定部３２１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。

【００５２】

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３２２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。

【００５３】

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

【００５４】

被写体区間代表フレーム判定部３２３は、１以上の顔シーケンスが含まれる時間区間の動画像データから代表フレーム画像を抽出する。例えば、主要顔判定部３２１により主要であると判定され、かつ、顔状態判定部が出力する評価値と入力された動画区間の長さとが所定の閾値を超えたフレーム画像を抽出する。１つの代表フレーム画像を抽出する場合には、評価値が最大となったフレーム画像を抽出する。また、複数の代表フレームを出力する場合には、局所最大や動画区間を分割し、分割した各区間内で最大となるフレームを出力すればよい。

【００５５】

また、代表フレーム間の間隔や、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。区間の長さが所定の値より短い場合や評価値が所定の値に満たない場合には必ずしも代表フレームを選択する必要はない。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。

【００５６】

・非被写体区間からの代表フレーム抽出（Ｓ７００）

非被写体区間代表フレーム抽出部３３０は、風景判定部３３１と、注目点判定部３３２と、非被写体区間代表フレーム判定部３３３とから構成される。

【００５７】

風景判定部３３１は、風景を撮影している可能性が高いフレームに対して高い評価値を与える。例えば、所定時間以上のパンを行なっている場合は　撮影場所の特徴的な風景を撮影している場合が極めて高い。また、ズーム情報（光学ズーム・デジタルズーム）が利用できる場合はワイド側である場合にテレ側より風景を撮影している場合が高いので、高い評価値を出力する。

【００５８】

より具体的には、パンの検出はフレーム画像からオプティカルフローなどによって動きベクトル分布の傾向から判定する。このとき背景の動きベクトルの流れに対して中心付近で異なる場合は何らかの被写体を撮影している可能性が高い。このような場合は高い評価値を与えない。以下では、動きベクトルを使った被写体判定の具体例について説明する。

【００５９】

まず、フレーム画像内の各点でオプティカルフローによって動きベクトルを求める。次にフレーム外周付近の点の動きベクトルを用いてパンの最中かを判定する。時間方向に移動平均を取ることで安定した検出が可能となる。パン状態でない場合には動きベクトルから撮影者が風景を撮っているのか判別できないので被写体領域の検出は行なわない。パン状態の場合はフレーム外周付近の点の動きベクトルを用いて中央付近の動きベクトルを一次補間によって求める。

【００６０】

次に、一次補間によって求めた動きベクトルと中心付近の各点でオプティカルフローによって得られた動きベクトルとの差分をもとめ、差分ベクトルが所定の式一以上の長さとなる点の包領域を被写体の領域と判定する。このとき、被写体の面積を使って時間方向での移動平均を求め、所定の面積を占める場合には被写体を撮影していると判断する。

【００６１】

注目点判定部３３２は、認識対象ではないが一般に印象に残るオブジェクトを撮影している場合に高い評価値を与える。大きくパンをしている場合には風景判定部３３１で説明した動きベクトルを使った被写体判定による方法があり、この場合には面積に基づいて評価値を出力する。以下では、パン状態でない場合の具体例について説明する。

【００６２】

まず、前後するフレーム差分によって画像を複数の領域に分割し、注目点である可能性が高い領域に対して高い評価値を出力する。例えば、それぞれの領域の位置、エッジ、彩度、色相の分布、動きパターンなどの特徴量を教師データとして前述したニューラルネットで学習し、当該領域が注目点である可能性が高い場合に高い評価値を出力する。また、

撮影時のズーム情報が利用できる場合はワイド側である場合にテレ側より高い評価値を出力するとよい。

【００６３】

非被写体区間代表フレーム判定部３３３は、何れの顔シーケンスにも含まれない時間区間の動画像データから代表フレーム画像を抽出する。例えば、風景判定部３３１が出力する評価値と入力された動画区間の長さとが所定の閾値を超えたフレーム画像を抽出する。１つの代表フレーム画像を抽出する場合には、評価値が最大となったフレーム画像を抽出する。

【００６４】

なお、評価値が所定の閾値より高い区間内では、動きベクトルを時間的に積分した値が所定の閾値を超えるたびに代表フレーム画像を順次追加出力しても良い。これによって風景全体を代表フレーム画像として出力することが出来る。

【００６５】

さらに、注目点判定部３３２が出力する評価値と入力された動画区間の長さとが所定の閾値を超えたフレーム画像を抽出する。１つの代表フレーム画像を抽出する場合には、評価値が最大となったフレーム画像を抽出する。

【００６６】

複数の代表フレームを出力する場合には、局所最大や動画区間を分割し、分割した各区間内で最大となるフレームを出力すればよい。また、代表フレーム画像間の時間間隔や、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。なお、区間の長さが所定の値より短い場合や評価値が所定の値に満たない場合には必ずしも代表フレーム画像を抽出する必要はない。また、風景判定部３３１と注目点判定部３３２のいずれかから得られる評価値のみを用いるよう構成しても良い。

【００６７】

図６は、代表フレーム画像の位置情報を出力したデータの一例を示す図である。第１カラムは動画先頭からの秒数を示し、第２カラムは図５で示した”シーケンス番号”を示し、第３カラムは”評価値”を示す。なお、第２カラムにおいて番号が無いものは非被写体区間から抽出した代表フレームであることを示している。なお、第３カラムに記述する評価値として最終評価にいたる途中結果を併記しても良い。たとえば非被写体区間であれば、風景判定部３３１における評価値と注目点判定部３３２における評価値を併記することで、利用者が所望とする代表フレーム画像に絞り込むことが可能となる。

【００６８】

１つの時間区間に対し１個の代表フレーム画像のみを格納してもよいし複数個格納しても良い。また、前述したように、ある時間区間について所定の閾値を超える評価値のフレーム画像が無い場合などは、代表フレームを格納しないよう構成してもよい。

【００６９】

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。

【００７０】

また、本実施形態は例えば特許文献１に開示される技術と組み合わせることも出来る。すなわち、本実施形態による方法で抽出した代表フレームをキーフレームとして用い、類似性などに基づいて更に絞込んで提示することが可能である。

【００７１】

また、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ＤＴＤ（Document Type Definition）を定義したＸＭＬ（Extensible Markup Language）形式等で表現するようにしてもよ

い。

【００７２】

図１３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１００８に表示するよう構成してもよいし、プリンタ１００９によって印刷出力するよう構成しても良い。

【００７３】

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が含まれる時間区間と顔シーケンスが含まれない時間区間の各々について代表フレーム画像を抽出することが出来る。それにより、顔シーケンスが含まれない時間区間からは例えば風景のフレーム画像が抽出されることになる。このような、フレーム画像を代表フレーム画像として抽出することにより、利用者は、例えば、”どこで”撮影した動画像データであるかを端的に知ることが可能となる。

【００７４】

また、顔シーケンスごとに代表フレーム抽出を行なうのではなく、顔シーケンスが生成された時間区間の動画像データを対象に代表フレーム画像の抽出を行なうよう構成してもよい。そのように構成することで、複数の人物のうちどちらが主要な人物かを評価して、その評価に基づいた代表フレーム抽出を行なうことができる。

【００７５】

（変形例）

第１実施形態では、被写体パターンとして人物の顔を検出し、顔シーケンスを生成するようにした。しかし、撮影対象が把握できるものであれば、その他のオブジェクトを被写体のパターンとして取り扱っても良い。

【００７６】

例えば、走る電車を軌道に近い地上から撮影する場合は、電車を検出・追跡し、電車が撮影されている区間か否かで代表フレームを抽出する基準を換えれば良い。なお、電車撮影区間から代表フレームを抽出するには、オーディオトラックのレベルとパン操作から先頭車両の前部が映った所に高い評価値をつけるよう構成すると好適である。

【００７７】

図１２は、顔シーケンスと電車シーケンスとを含む動画像データから代表フレーム画像を抽出する際の区間設定の一例を示す図である。図に示されるように　顔と電車など複数の種別の被写体種別を検出する場合は、ショットの境界では時間区間を区切るよう制御するとよい。

【００７８】

このように、被写体種別として異なる種類の画像を利用することにより、動画像データからより適切な代表フレーム画像を抽出可能となる。なお、被写体種別として、たとえば、”家族の人の顔”と”家族以外の人の顔”などの種別を用いても良い。

【００７９】

（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。

【００８０】

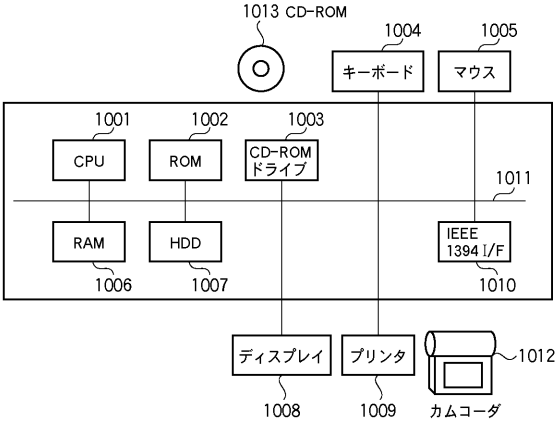
なお、本発明は、前述した実施形態の機能を実現するプログラムを、システム或いは装置に直接或いは遠隔から供給し、そのシステム或いは装置が、供給されたプログラムコードを読み出して実行することによっても達成される。従って、本発明の機能処理をコンピュータで実現するために、コンピュータにインストールされるプログラムコード自体も本発明の技術的範囲に含まれる。

【００８１】

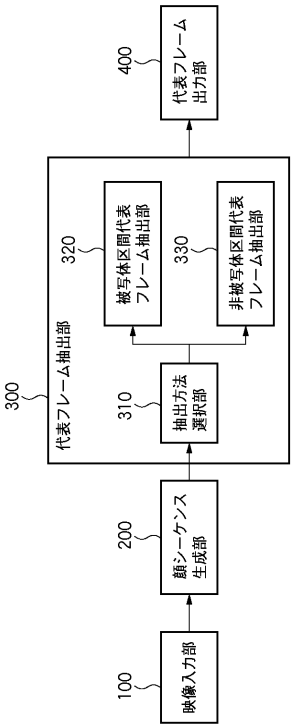
その場合、プログラムの機能を有していれば、オブジェクトコード、インタプリタにより実行されるプログラム、ＯＳに供給するスクリプトデータ等、プログラムの形態を問わない。

【００８２】
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。
【００８３】
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。
【００８４】
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】
【００８５】
【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】第１実施形態に係る情報処理装置の動作フローチャートである。
【図５】顔シーケンスの情報を記録したデータの一例を示す図である。
【図６】代表フレーム画像の位置情報を出力したデータの一例を示す図である。
【図７】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。
【図８】顔シーケンスの生成の処理フローを示す図である。
【図９】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１０】格子状のブロックに分割したフレーム画像内における顔中心の軌跡をプロットした図である。
【図１１】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１２】顔シーケンスと電車シーケンスとを含む動画像データから代表フレーム画像を抽出する例を示す図である。
【図１３】代表フレーム画像の一覧の出力例を示す図である。

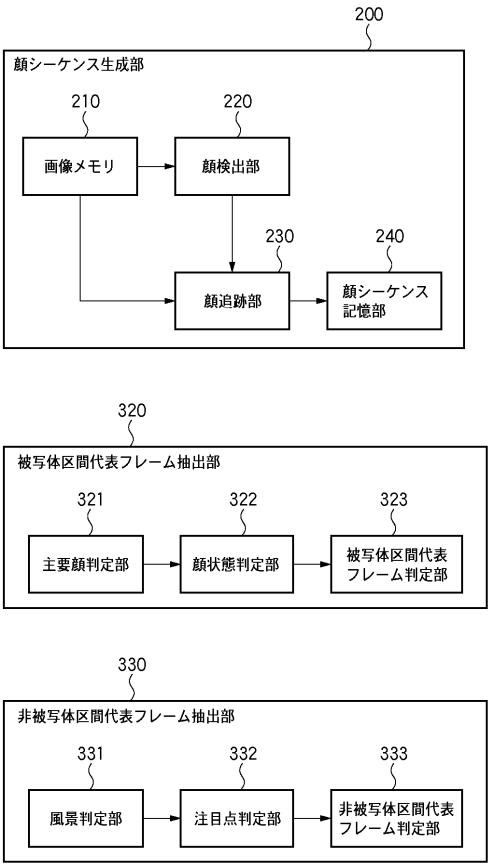
【図１】


1013 CD-ROM
1004
キーボード
1005
マウス
1001
CPU
1002
ROM
1003
CD-ROM
ドライブ
1011
RAM
HDD
IEEE
1394 I/F
1006
1007
1010
ディスプレイ
プリンタ
1012
1008
1009
カムコーダ

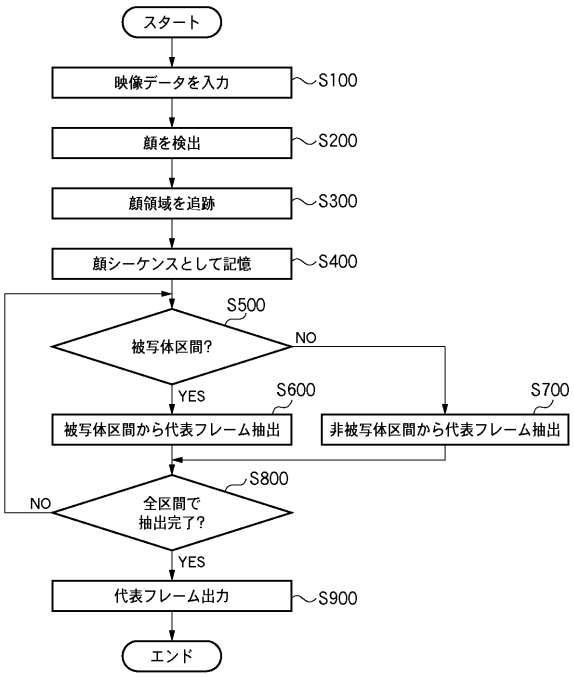
【図２】


400
代表フレーム
出力部
300
代表フレーム抽出部
320
被写体区間代表
フレーム抽出部
330
非被写体区間代表
フレーム抽出部
310
抽出方法
選択部
200
顔シーケンス
生成部
100
映像入力部

【図３】


200
顔シーケンス生成部
210
画像メモリ
220
顔検出部
230
顔追跡部
240
顔シーケンス
記憶部
320
被写体区間代表フレーム抽出部
321
主要顔判定部
322
顔状態判定部
323
被写体区間代表
フレーム判定部
330
非被写体区間代表フレーム抽出部
331
風景判定部
332
注目点判定部
333
非被写体区間代表
フレーム判定部

【図４】


スタート
映像データを入力
S100
顔を検出
S200
顔領域を追跡
S300
顔シーケンスとして記憶
S400
S500
被写体区間?
NO
YES
S600
被写体区間から代表フレーム抽出
S700
非被写体区間から代表フレーム抽出
S800
全区間で
抽出完了?
NO
YES
代表フレーム出力
S900
エンド

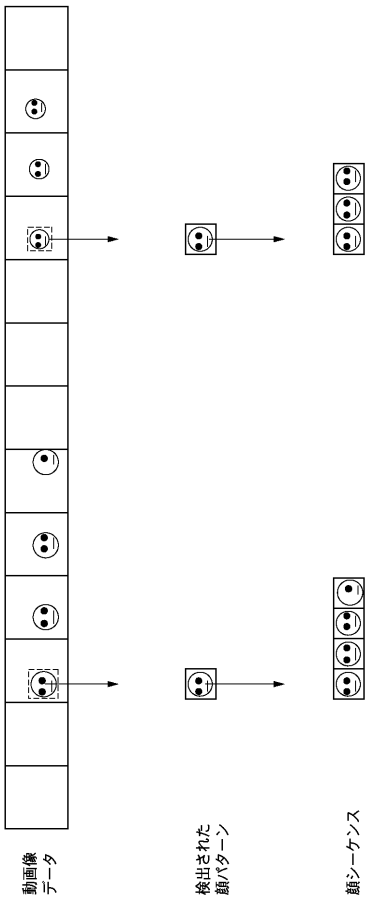
【図５】

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
…	…	…

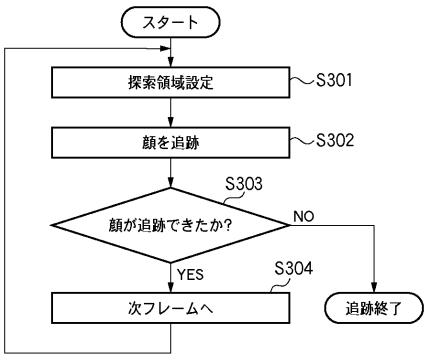
【図６】

代表フレーム	シーケンス番号	評価値
24.3	--	0.9
44.25	1	0.5
60.38	2	0.4
60.38	3	0.8
87.53	--	0.95
99.52	4	0.7
108.35	5	0.85
130.45	6	0.65
135.65	6	0.9
165.19	8	0.5
…	…	…

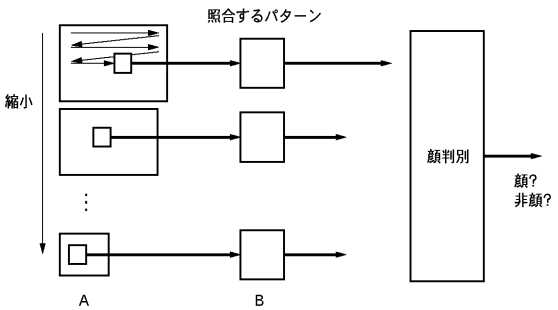
【図７】



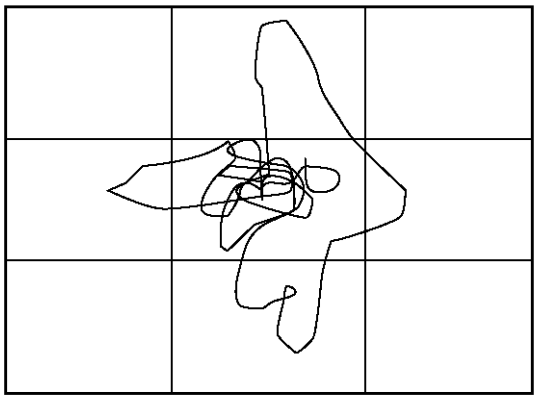
【図８】



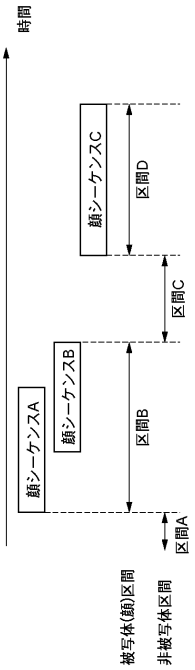
【図９】



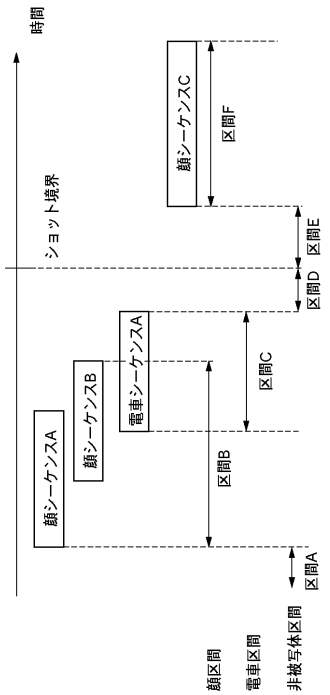
【図１０】



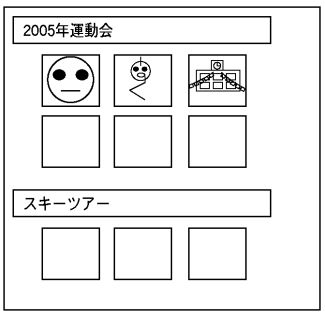
【図１１】



【図１２】



【図１３】



フロントページの続き

(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
(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　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

Ｆターム(参考) 5B057 BA02 CA08 CA12 CA16 CH18 DA08 DC34
5C053 GA11 GB19 GB36 GB37 HA29 JA22 LA01 LA15