

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5507386号
(P5507386)

(45) 発行日 平成26年5月28日(2014.5.28)

(24) 登録日 平成26年3月28日(2014.3.28)

(51) Int.Cl.

G06F 17/30 (2006.01)
H04N 5/91 (2006.01)

F 1

G06F 17/30 380B
G06F 17/30 170B
G06F 17/30 210D
H04N 5/91 N

請求項の数 6 (全 16 頁)

(21) 出願番号 特願2010-181562 (P2010-181562)
 (22) 出願日 平成22年8月16日 (2010.8.16)
 (65) 公開番号 特開2011-44140 (P2011-44140A)
 (43) 公開日 平成23年3月3日 (2011.3.3)
 審査請求日 平成25年8月12日 (2013.8.12)
 (31) 優先権主張番号 12/544,566
 (32) 優先日 平成21年8月20日 (2009.8.20)
 (33) 優先権主張国 米国(US)

早期審査対象出願

(73) 特許権者 596170170
 ゼロックス コーポレイション
 XEROX CORPORATION
 アメリカ合衆国、コネチカット州 06856、ノーウォーク、ピーオーボックス 4505、グローバー・アヴェニュー 45
 (74) 代理人 110001210
 特許業務法人 YKI 国際特許事務所
 (72) 発明者 マルコ ブレッサン
 フランス グルノーブル リュ エミール オジエ 2

審査官 鹿野 博嗣

最終頁に続く

(54) 【発明の名称】画像セットからの動画コンテンツの生成

(57) 【特許請求の範囲】

【請求項 1】

画像セットをクラスタ化し、各々が、不規則な一群の画像から成る複数の場面を生成する手順であって、前記クラスタ化は、前記画像セットの画像の対の間の対比較に基づく手順と、

場面内の画像を動画の一貫性に応じて順序付けることで、規則的な順序の画像から成る規則的場面を生成する手順と、

前記複数の場面から成る、前記規則的な順序の画像に基づいて順序付けられた画像として動画コンテンツを生成する手順と、から成り、

少なくとも、上述のクラスタ化する手順と順序付けの手順とがデジタルプロセッサによつて実行され、

前記場面内の画像を動画の一貫性に応じて順序付ける手順が、

画像の対の間での対比較に応じて場面内の前記複数の画像の最短経路を推定する手順から成ることを特徴とする画像処理方法。

【請求項 2】

請求項 1 に記載の画像処理方法であつて、さらに、

前記場面内の画像のタイムスタンプの集計に基づいて、各場面の場面日付を割り当てる手順を含み

前記動画コンテンツを生成する手順は、前記場面日付に基づいて前記動画コンテンツ内の場面を順序付ける手順を含むことを特徴とする画像処理方法。

10

20

【請求項 3】

画像セットをクラスタ化し、各々が、不規則な一群の画像から成る複数の場面を生成する手順と、

場面内の画像を動画の一貫性に応じて順序付けることで、規則的な順序の画像から成る規則的場面を生成する手順と、

前記複数の場面から成る、前記規則的な順序の画像に基づいて順序付けられた画像として動画コンテンツを生成する手順と、

各画像の表示時間を選択する手順と、

前記動画コンテンツに基づいて動画フレームを含む動画を生成する手順であって、前記動画コンテンツの各画像は、前記動画が所定のフレームレートで再生される場合、前記画像が選択された表示時間の間、表示されるのに有効な1つ以上の動画フレームを定義する手順と、から成り、

少なくとも、上述のクラスタ化する手順と順序付けの手順とがデジタルプロセッサによって実行され、

前記場面内の画像を動画の一貫性に応じて順序付ける手順が、

画像の対の間での対比較に応じて場面内の前記複数の画像の最短経路を推定する手順から成ることを特徴とする画像処理方法。

【請求項 4】

画像セットをクラスタ化して複数の場面を生成する手順と、動画の一貫性の基準に応じて各場面内で画像を順序付ける手順と、から成る画像処理方法であって、前記動画の一貫性の基準は、前記場面内の前記複数の画像の最短経路の推定を含み、前記推定は、画像の対比較基準を用いる画像処理方法をデジタルプロセッサが実行することが可能な命令を保存した記録媒体。

【請求項 5】

画像処理方法であって、

最短距離推定を含む動画の一貫性の基準に基づいて画像セットを自動的に順序付ける手順であって、当該自動的に順序付ける手順は、画像セットをクラスタ化して複数の場面を定義するクラスタを生成する手順と、前記動画の一貫性の基準に基づいて、各場面内の画像を自動的に順序付ける手順と、から成る手順と、

順序付けられた画像セットによって定義され、所定のフレームレートで再生されるフレームを含む動画として、前記画像を再生する手順と、から成り、

前記自動的に順序付ける手順が、

画像の対の間での対比較に応じて場面内の前記複数の画像の最短経路を推定する手順から成ることを特徴とする画像処理方法。

【請求項 6】

画像処理方法であって、

動画の一貫性の基準に基づいて画像セットを自動的に順序付ける手順と、

順序付けられた画像セットによって定義され、所定のフレームレートで再生されるフレームを含む動画として前記画像を再生する手順と、

一定エントロピー基準を用いて前記所定のフレームレートを選択する手順と、から成り、

前記自動的に順序付ける手順が、

画像の対の間での対比較に応じて場面内の前記複数の画像の最短経路を推定する手順から成る画像処理方法。

【発明の詳細な説明】**【技術分野】****【0001】**

以下の内容は情報処理技術、情報検索技術、画像技術、映像技術、通信技術、及び関連

10

20

30

40

50

技術に関するものである。

【背景技術】

【0002】

デジタル画像を生成する安価な光学スキャナやデジタルカメラ、大きな画像データベースの保存に十分な記憶容量を備えた大容量記憶媒体の継続的開発、高速デジタルデータネットワーク及び画像の配布を容易にするインターネット接続などの技術が組み合わされた結果、デジタル画像を取得し、それを配布することが、あらゆる場面で行われるようになっている。

【先行技術文献】

【特許文献】

10

【0003】

【特許文献1】米国特許第6789088号明細書

【特許文献2】米国特許第6396963号明細書

【特許文献3】米国特許第7451160号明細書

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかし、効率的に画像を閲覧し、選択する技術の進展は、画像の生成、保存、配布技術に対して遅れを取っている。一般的な方法では、画像クエリを受け取ったサーチエンジンが一括りの画像（場合によっては効率を高めるために、例えば「サムネイル」画像として小サイズ化されたもの）を検索し、取得した画像をモザイク状又は格子状の配列で表示する。モザイクに含まれる画像の数は、例えば表示される画像のサイズや数、表示技術の制限、人間の視覚系の制約等といった様々な要因の折り合いによって決まるが、通常は、1つのモザイクに約20枚の画像が含まれる。この表示方法では、ユーザは何千といった多数の画像を効率的に閲覧して選択することができない。

20

【0005】

加えて、現行のモザイク表示では通常、画像を互いに関連付けることは行われない。一つのモザイクにおいて画像は普通、タイムスタンプ、撮影者、画像サイズ、画像名、画像の種類などのメタデータに記されている個々の基準の順に分類される。クエリ結果の表示にモザイク表示が用いられる場合、画像は普通はそのクエリへの個々の関連度を示す基準の順に分類される。いずれの場合にも、（例えば）20枚の画像からなる第一のモザイク上の画像が、間に挟まれた何百もの画像によって隔てられている、後のモザイクの画像とほぼ同一である可能性がある。これでは多数の類似した画像の中から「最高」の画像を選択するのは容易ではない。

30

【0006】

メタデータ情報は、画像をグループ化するモザイク表示技術を増進するために用いることができる。この方法は通常、手動又は半手動で行われ、どのメタデータが扱うタスクと関連しているかをユーザが指示しなければならない。さらに、もし異なる装置（例えば異なるデジタルカメラ）によって生成された画像を集めることが目的の場合、異なる装置によって生成されたタイムスタンプが互いに同期していないという更なる問題が起こり得る。

40

【0007】

一例として、子供の誕生日パーティーについて考える。そこでは、参加している親の数人がデジタルカメラを持参し、そのパーティーの間中、写真を撮るものとする。そして、異なるデジタルカメラを用いて得られたこれらの写真を魅力的で感じの良いプレゼンテーションとして見せたいという希望があるとする。そのための一つの方法はタイムスタンプによって画像を並び替えるというものである。しかし、その結果は見た目にあまり魅力的ではないかも知れない。なぜならば、ほぼ同一時刻に異なるカメラによって得られた異なる画像は、相違するカメラによる非常に異なった観点から撮られた可能性があるからである。その結果は、ある画像から次の画像への観点が一貫せず「飛躍」している。加えて、

50

そのパーティーにおいてもし一つのカメラのデジタル時計が他のデジタルカメラに対して例えば10分だけずれているならば、その時間的にずれたカメラによって得られた写真は同期しないことになる。従って、例えば、そのプレゼンテーションでは子供が誕生日ケーキのろうそくを吹き消し、次いで誕生日の食事へと移るところが示されるかもしれないが、時間的にずれたカメラによって得られた、その子供がろうそくを吹き消す写真が食事の写真の中にちりばめられることになる。

【0008】

上記の例は写真を選択する際の難しさを示す実例である。例えば、写真を検討している人が、ある瞬間ににおいてある視点から撮られた、子供がろうそくを吹き消している画像が特に魅力的であると思うかもしれない。その検討者は、その瞬間の、他のさらに良い画像が無いかどうかを知りたいと思うかもしれない。残念ながら、タイムスタンプによって並んだプレゼンテーションの一貫性がないため、得られた画像はほぼ同一時刻の他の視点で得られた非常に多くの他の画像の中に分散する可能性が高く、検討者にとって魅力があると感じられた写真に類似した写真を比較することが難しくなる。たとえ画像がさらにカメラの識別番号（このメタデータが利用可能であると仮定して）によって並び替えられたとしても、検討者はやはり、魅力的と思える写真を撮ったカメラの視点に類似した視点を有する、異なるカメラによって撮られたよく似た写真を見逃してしまうかもしれない。

10

【0009】

画像選択に限界があることに加え、現在の画像配布形態には幾つかの関連する欠点がある。例えば、オンラインの画像シェアサイトでは、写真を提示するために、通常、モザイク又はスライドショーが用いられている。モザイクの使用については既に論じたが、その欠点は画像シェアの場合においても存在する。

20

【0010】

スライドショーでは通常、予め選択された画像間の時間間隔で以て画像が連続的に表示される。スライドショーを用いることでユーザの写真の閲覧可能な速度を高めることができるが、各画像の閲覧時間が短くなってしまう。スライドショーの速度が速すぎると流れが「飛躍」する可能性があり、閲覧者をいらつかせることになる。もし速度が遅すぎなるば、スライドショーの全閲覧時間が長くなりすぎる可能性がある。スライドショーを作成することもまた、制作者にとって多大な時間を必要とし得る。なぜならば、制作者は通常、スライドショーの流れにおける画像の順番を手動で選択しなければならないからである。あるいは、自動的に順序が選択される場合（例えばタイムスタンプに基づいて）には、スライドショーにおける表示の「飛躍」が大きくなってしまう可能性がある。

30

【0011】

これらの方法もまた、同一のイベントにおける異なる撮影者による画像を組み合わせるための有用な方法を提供することができない。また、撮影者、カメラの識別番号、タイムスタンプ等のメタデータに基づくもの以外の有効な画像検索方法を提供することができない。こういった、装置によって生成されるメタデータは、実際に画像化された対象との、すなわち、撮影された出来事や人物等との関連が弱い。

【課題を解決するための手段】

【0012】

ここにおける実例として開示されるいくつかの実施形態では、画像処理方法は、画像セットをクラスタ化して、各々が、不規則な一群の画像から成る複数の場面を生成する手順と、場面内の画像を動画の一貫性に応じて順序付けることで、規則的な順序の画像から成る規則的場面を生成する手順と、前記複数の場面から成る、前記規則的な順序の画像に基づいて並べられた画像として動画コンテンツを生成する手順と、から成る。

40

【0013】

ここにおける実例として開示されるいくつかの実施形態では、記録媒体が、画像セットをクラスタ化して複数の場面を生成する手順と、動画の一貫性の基準に応じて各場面内で画像を順序付ける手順と、から成る画像処理方法を実行することが可能な命令を保存する。

50

【0014】

ここにおける実例として開示されるいくつかの実施形態では、画像処理方法が、自動的に、動画の一貫性の基準に基づいて画像セットを順序付ける手順と、該画像を、順序付けられた画像セットによって規定され、予め定められたフレームレートで再生されるフレームから成る動画として再生する手順と、から成る。

【図面の簡単な説明】**【0015】**

【図1】イメージのセットから動画を生成するシステムの一例。

【図2】実際に構築された図1のシステムの方法を用いて生成された動画コンテンツの選択された定量的計量を図示するもの。

10

【図3】実際に構築された図1のシステムの方法を用いて生成された動画コンテンツの選択された定量的計量を図示するもの。

【発明を実施するための形態】**【0016】**

以下では、画像セットが動画コンテンツとして表示されるインターフェースを用いて画像を閲覧し、任意に選択する方法及び装置が開示される。本明細書で用いられている「動画コンテンツ」なる用語は、効果的な動画話法の原則に関して並べられた画像セットのことを指す。これらの原則には、コンテンツを一又は複数の「場面」に分割することが含まれる。各場面は、視覚的に一貫性がある画像の小セットである。互いに類似した画像セットは視覚的に一貫性がある例の一つであり、類似した画像をグループ化することで場面を求めることができる。この類似性は、いかなる選択された類似性基準によっても与えることが可能であり、その類似性基準は、任意であるが、画像に関連付けられたメタデータを含んでもよい。

20

【0017】

効果的な動画話法の原則には更に、各場面内で画像を一貫性があるように順序付けることも含まれる。これは、動画コンテンツが一つの場面内で滑らかに進むべきであるという期待を反映するものである。例えば、一つの場面の中で、人物は、ある場所から別の場所へ滑らかに動くことが期待されており、一つの場面の中において一つの場所から別の場所へ突然「テレポート」することは期待されていない。

【0018】

30

一方、場面と場面の間の一貫性は、通常、効果的な動画話法の原則ではない。例えば、ある人が特定の地点、例えば屋内で立っている状態で一つの場面が終了し、次の場面がその人が全く異なるどこか、例えば屋外で立っている状態で始まるのは珍しいことではない。場面間の一貫性は通常は効果的な動画話法の原則ではないが、場面間のある関係を任意に組み込んでも良い。例えば、幾つかの実施形態では、複数の場面は、各場面の画像の平均タイムスタンプに基づいて順序付けられて上映される。場面の順序付けに他の基準を用いることもありえる。

【0019】

ここで、効果的な動画話法の原則に関して順序付けられた画像セットを動画コンテンツとして示すことには特定の利点があることがわかる。一つの場面内で一貫性を有するようにして、プレゼンテーションがスムーズで、美的に心地よいものとなることが保証される。見た目に一貫性があることはまた、順序がばらばらである場合に対して、一定の時間内でユーザがナビゲーションできる画像の量を増加させる。これは、情報がオンラインで処理される場合（ひとまとまりの文章を読む場合と同様に）、例えば一単語のような新たな少しの情報のそれぞれを処理するために必要な時間は、現在の前後関係に基づくその単語の予測可能性と結びついているからである。さらに、画像を選択するという目的において、一つの場面内で一貫性をもって画像を示すことで、一貫性を決定するのに用いられる類似基準に基づいて互いに類似した画像をグループ化することに対する傾向を与える。結果として、ユーザは大雑把な意味で類似している一群の画像（例えば、全てが一人の子供がろうそくを吹き消している場面のもの）を自然に特定することができる。次いで、

40

50

その場面の類似した画像の中から最高の画像を選択すべく、場面の一貫性によって共にグループ化される類似した画像をさらに詳細に比較するために、例えば多くの動画再生エンジンで利用することが可能な一枚ずつの順次表示モードを用いて、その場面の個々の画像を検討することができる。

【0020】

一方、ここで、画像セットに対して全体的に一貫性を課すように試みることは実際的でなく、また望ましくないことがわかる。例えば、本明細書で用いられている誕生日パーティーの例において、もしそのパーティーが（食事中に）ダイニングルームから（子供用のゲームのために）居間に移動するならば、ダイニングルームで撮られた写真と居間で撮られた写真との間で一貫性を課すように試みることは望ましくないであろう。このような差は、ヨーロッパをあちこち回る一週間の休暇旅行中に、さまざまな国、異なる大聖堂や建物などにおいて撮られた写真を見せるといった他の利用形態においては、さらに大きなものとなるかもしれない。ここで、このような画像のグループ間の大きな差は、場面の概念を動画技術から画像表示技術にアレンジすることで容易に調整可能であることがわかる。10

【0021】

効果的な動画話法の原則に関して並べられた動画コンテンツとして画像セットを表示する開示された方法の別の利点は、複数の画像のプレゼンテーションを、M P E G、D i v X、X v i d、F F m p e g、a v i、w m v など、標準の動画コーデック用いて保存することが可能であるということと、複数の画像のプレゼンテーションを、MPlayer (<http://www.mplayerhq.hu/>より入手可能)、VLC media player (<http://www.videolan.org/vlc/>より入手可能)、Windows (登録商標) Media Player™ (Microsoft Corporation, Redmond, WA, USAより入手可能)、Quicktime (Apple Corporation, Cupertino, CA, USAより入手可能)などの汎用の動画又はマルチメディアプレーヤーを用いて動画として表示することが可能であることである。これにより、画像を高速で表示（動画のフレームレートの上限まで、通常は一秒あたり 24 ~ 30 フレーム程度）することが可能となり、且つ、個々の画像を選択するための一枚ずつの順次表示モードのような便利な組み込み機能を利用することが可能となる。当然のこととして、本明細書においては、動画は予め定められたフレームレートで再生されるフレームから成る。20

【0022】

また、当然のこととして、動画コンテンツが効果的な動画話法の原則に関して順序付けられた画像セットを指している場合、その動画コンテンツは従来の動画保存形式で保存される必要はなく、さらに、必ずしも動画又はマルチメディアプレーヤーエンジンを用いて動画として表示される必要もない。例えば、効果的な動画話法の原則に関して順序付けられた画像セットから成る動画コンテンツが、O D P 又はP P T などの従来のプレゼンテーション形式で以てプレゼンテーションスライドとして保存され、OpenOffice.org Impress (<http://openoffice.org/>より入手可能) やMicrosoft PowerPoint (Microsoft Corporation, Redmond, WA, USAより入手可能)などの汎用スライドプレゼンテーションアプリケーションを用いて表示されることが考えられる。また、動画コンテンツを非標準の形式を用いて保存し、その動画コンテンツを専用のソフトウェアを用いて再生することも考えられる。30

【0023】

さらに、当然のこととして、「動画コンテンツ」という用語は、効果的な動画話法の原則に関して順序付けられた画像セットが、動画のフレームレートで以て動画として表示されることを求めるものではない。実際、たとえ画像が従来の動画形式で保存されていて汎用の動画又はマルチメディアプレーヤーエンジンを用いて再生されたとしても、それらの画像は、再生レートが動画フレームレートよりもずっと遅くなるように複製することができる。例えば、もしその動画が一秒あたり 30 フレームの動画フレームレートで再生され、画像を一秒あたり一枚の速度で表示したいとすると、動画が一秒あたり 30 フレームで再生され、画像が一秒間表示されるように、各画像は連続 30 枚のフレームを適切に規定することができる。40

【0024】

図1を参照して、画像セット10を、効果的な動画話法の原則に関して順序付けられた画像セットから成る動画コンテンツに変換する、一例となるシステムについて説明する。画像セット10は、画像がタイムスタンプ情報又は画像の順序を定義するのに利用することもあり得る他のメタデータを有しているかもしれないものの、不規則な画像セットとして扱われる。

【0025】

動画コンテンツへの変換の最初の処理作業は、複数の画像を一又は複数の場面に分割することである。そうするために、画像セット10が、類似性基準（この場合、値が大きいほど類似性が高いことが示される）や、距離又は発散基準（この場合、値が小さいほど類似性が大きいことが示される）といった適切な比較基準によって測定されるような類似性を有する画像を特定するために、クラスタ化される。図示された実施形態では、このクラスタ化する手順は、画像対比較部12を用いて計算された画像間の対比較に基づいて行われる。比較基準は画像自体に由来する情報を利用してもよいし、画像に関連するメタデータを用いることもできるし、画像由来の情報及び関連するメタデータ情報の組み合わせを用いることもできる。例えば、一つの方法では、各画像は局所的特徴に基づく特徴ベクトルによって表され、比較はフィッシャーカーネルを採用する。（Conference on Computer Vision and Pattern Recognition 2007 (CVPR07)における、Perronnin et al.による "Fisher kernel on visual vocabularies for image categorization" を参照。）画像全体の比較を行うため、幾つかの実施形態では画像が一定の数の領域に分けられ、領域と対応する領域との比較が行われる。最終的な類似基準は、領域あたりの比較基準の和、平均、又は他の集計単位である。

10

20

【0026】

任意であるが、選択された比較基準には、撮像装置の識別番号、画像の題材の識別番号、又は他の使用可能なメタデータといったメタデータが含まれる。他の考えられる比較基準では、顔に焦点を合わせる。画像内で顔を識別する既知の技術を用いて、顔が描かれている画像の部分間の類似度を、適切な比較基準によって比較することができる。例えば、比較基準は顔認識基準であり得る。これにより、次のクラスタ化する手順が、画像に現れている個々の人物に関連した情報に基づいて行われる。

【0027】

30

比較部12の出力は、各種の形態や種類であり得る。その出力は、画像グラフ14として見ることができる。この画像グラフ14では、グラフの節点は画像セット10の画像であり、その節点は対比較基準の関数に相当する重みを有するエッジによって接続される。あるいは、比較部12の出力は、要素 d_{ij} が i 番目の画像と j 番目の画像を比較する比較基準である対類似行列として見ることもできる。（「類似」行列という用語は従来より用いられているが、比較を行う際には、類似基準あるいは距離又は発散基準を採用することができることに注意。）例えば、画像 $I_i, I_j [I_n]_{n=1, \dots, N}$ 、ここにおいて $[I_n]_{n=1, \dots, N}$ が画像セット10である、を考えると、画像間の比較基準は $d_{ij} = f(I_i, I_j)$ と記述することができる。グラフの観点からすると、要素 d_{ij} は画像の節点 $I_i, I_j [I_n]_{n=1, \dots, N}$ を接続するエッジの重みである。通常、類似行列は対称行列（すなわち、 $d_{ij} = d_{ji}$ が常に成立つ）又は非対称行列（すなわち、 $d_{ij} \neq d_{ji}$ があり得る）のいずれかである。もし比較基準 f が非対称であるならば、対称比較行列は $d_{ij} = f(I_i, I_j) + f(I_j, I_i)$ などの適切な対象化を用いることによって作成することができる。

40

【0028】

クラスタ化は、比較基準に基づき、クラスタ化モジュール16によって適切に行うことができ、スペクトルクラスタリング、カーネル関数に基づいたクラスタリング、確率的潜在意味解析 (probabilistic latent semantic analysis : PLSA) などの、適切なクラスタ化アルゴリズムをどれでも用いることができる。例えば、スペクトルクラスタリング技術は、より少ない次元でのクラスタリングのために次元縮小を実施するために、データの類似行列のスペクトルを利用する。スペクトルクラスタリングの方法の幾つかの定式では、

50

対称類似行列が固有値分解を用いて分解され、行列の積 $A \cdot D \cdot A^T$ を生成する。ここで、 D は、対角線要素が該対称類似行列の固有値である対角行列であり、 A は他の行列、そして上付きの「 T 」は転置処理を意味する。スペクトルクラスタリングの枠組みの中では、行列 A の列は画像をクラスタと関連付けるものと解釈される。

【0029】

クラスタ化は、例えば最大演算子を用いて、各画像が全くただ一つのクラスタにのみ割り当てられる強分割を採用することができる。又は、クラスタ化は、例えば確率の閾値を採用することにより、各画像が通常一又は複数のクラスタに割り当てられる弱分割を採用することもできる。

【0030】

対比較基準に基づいたクラスタ化技術が本明細書では具体例として説明されるが、K 平均法などのように、画像間での対類似に基づいて機能するものよりも、画像の特徴ベクトル又は他の表現に直接に基づいて機能するクラスタ化技術を含む、他のクラスタ化技術を採用することもできる。

【0031】

クラスタ化モジュール 16 の出力は、複数の場面 20 である。クラスタ化モジュール 16 の出力である各場面は、クラスタ化によって生成された、不規則な画像のクラスタから成る。もしクラスタ化において強分割が採用されるならば、各画像は一つのクラスタに割り当てられる。一方で、クラスタ化において弱分割が採用されるならば、ある画像が二以上のクラスタに割り振られることがあり得る。平均的には、弱分割は、強分割と比較すると、一つのクラスタにより多くの画像が含まれる（一つの画像が二つのクラスタに割り当てられることは「複製された」状態と見ることができる）。一つのクラスタにより多くの画像が含まれるならば、画像間が滑らかに移行する可能性が高くなる。従って、クラスタ内の動画の一貫性が実現される可能性が向上する。一方で、画像選択といった幾つかの用途では、異なる場面の間で画像を繰り返すのは好ましくないかもしれない。そのような用途においては、強分割が好適に用いられる。

【0032】

任意であるが、クラスタ化モジュール 16 は、クラスタ化によって生成されたクラスタのいずれにもうまく合わない画像を削除する異常画像削除サブモジュール 22 を含んでもよい。このような異常フィルタリングは、例えば、他の画像と非常に異なっているために動画の一貫性を大幅に低下させるであろう画像を削除することにより、動画コンテンツの一貫性を向上させるのに有用である。クエリに基づいた画像検索アプリケーションでは、異常画像を削除することは検索結果における有効なフィルター効果を及ぼす。検索された画像が他の検索された画像と大きく異なっていると、無関係である可能性がより高いからである。幾つかの実施形態では、例えば、もしクラスタ内の画像の数が非常に少ない場合やクラスタ内における画像間の平均の又は集計した距離が大きすぎる場合には、異常画像削除サブモジュール 22 は画像のクラスタ全体を削除することもできる。幾つかの実施形態では、図示された任意の専用の異常画像削除サブモジュール 22 を動作させることなく、選択されたクラスタ分割自体において異常画像の削除を実施してもよい。例えば、閾値に基づいた弱分割が用いられる場合、クラスタ化の結果得られるクラスタに帰属する確率がどのクラスタの基準にも合わない「異常」画像は効果的に削除される。

【0033】

現実に成された実行では、クラスタ化は、事後分布において閾値 0.4 を採用した弱分割による PLSA が用いられた。これにより、複数の画像のうちの幾つかを複製し、場合によっては最終的な順序から幾つかの異常画像を削除することで、さらに同質の場面が得られることがわかった。現実に成された実行ではまた、小さすぎる場面、即ち含まれる画像が 10 枚未満の場面は削除された。さらに一般的には、小さなクラスタを削除するためのクラスタサイズ閾値は、画像セット 10 のサイズ、画像セット 10 の画像の種類、望ましい動画話法の構成などに基づいて適切に選択される。又は、クラスタ化において、不適な程度に小さな場面（すなわち、作成される動画コンテンツの表示を考えた場合、閲覧に

10

20

30

40

50

は不適な程度に短い場面)を防ぐために、閾値のサイズよりも大きなクラスタを出力するようにさせることもできる。上述したように、クラスタ内の画像間の総距離といった他の基準もまた、削除する場面を選択する基準として用いることができる。

【0034】

クラスタ化モジュール16によって実行されるクラスタ化処理の目的は、一つの画像から次の画像に移る際に、画像に描かれる人物や物体などが突然にその位置、大きさ、向きなどを変えることがない、一貫性のある画像の流れを形成するために結び付けることができそうな意味において「類似」している画像を合わせてグループ化することである。クラスタ化モジュール16によって実行されるクラスタ化は互いに類似した画像を場面20にグループ化する。ここで、比較の尺度は比較部12によって採用される比較基準である。結果として、一つの場面における画像は互いに類似しているものと期待され、従って、実質的に動画の一貫性を乱さない順番で配列することができる可能性が高い。10

【0035】

終盤が近づくにつれて、場面内順序付けモジュール26が動画の一貫性に関して場面内で画像を順序付けることにより、規則的な順序の画像から成る規則化場面のセット30を生成する。N枚の画像セット10からクラスタ化されたS < N枚の画像を有する一つの場面の考える。言い換えると、その場面は、N枚の画像セット10からクラスタ化されることによって選択された不規則なS枚の画像の小セットとみなすことができる。その画像のセットの特定の順序をI₁, I₂, ..., I_Sと定める。動画の一貫性のためには、その並びにおいて二つの連続する画像はできる限り類似していることが望ましい。従って、全体として、20

【数1】

$$\sum_{i=1}^{S-1} d_{i,i+1}$$

(ここで、d_{i, i+1}は類似基準である)の値を最大化する、又は

【数2】

$$\sum_{i=1}^{S-1} d_{i,i+1}$$

(ここで、d_{i, i+1}は相違又は発散基準である)の値を最小化することが望ましい。

【0036】

別の言い方をすると、画像のペア間での対比較に関して、一つの場面内において複数の画像を通る最短経路を推定することが望ましい。そのような最短経路推定は時に「巡回セールスマン問題」と呼ばれ、最短経路を推定する数多くの方法が知られている。とはいえる、完全な最短距離を決定することは、クラスタ化モジュール16によって生成された典型的な場面のような多数の対象のセットに関しては、困難または不可能である。従って、「最短経路推定」という語句及びそれに似た表現は、最短経路問題に対する大凡の又は推定の解を指しているのであり、通常、とりわけ場面内の画像の数が増えるにつれ、得ることが困難又は不可能となる、完全な解を指しているのではない。40

【0037】

幾つかの実際に構築されたシステムでは、最短経路推定は、Lin-Kernighan巡回セールスマン解決法を用いて行われる。(Applegate et al., "Concorde TSP Solver" (2006) (available at <http://www.tsp.gatech.edu/concorde/>)を参照。)任意ではあるが、例え50

ば、画像に伴うタイムスタンプのメタデータより利用できる時間情報のような追加の方法を用いることができる。Lin-Kernighan巡回セールスマン解決法では、経路における第一の節点（本願の場合には第一画像）を選択する必要がある。実際に構築されたシステムでは、経路の第一画像は、場面を構成する複数の画像のうち、最も早いタイムスタンプを有する画像であるように選択された。他に、例えば、場面を定義するクラスタの重心に最も近い画像を選択するといった、第一画像を選択するための基準もあり得る。半管理セッティングでは、ユーザは候補結果に基づいてこの第一画像を選出又は選択するであろう。さらに、Lin-Kernighan巡回セールスマン解決法は単に、最短経路を推定する実例にすぎず、最短経路推定を実行するのに、さらに一般的で実質的には、いかなる正確な又は大凡の最短経路を推定する又は決定する技術をも用いることが可能である。

10

【0038】

任意であるが、非一貫場面削除モジュール32は、十分な動画の一貫性を有していない場面を削除する。例えば、任意で設けられる非一貫場面削除モジュール32は、推定された最短距離に基づいて並べられた、場面内の連続する画像間の平均の又は集計した距離が閾値よりも大きい全ての場面を削除しても良い。又は、推定された最短距離に基づいて並べられた、場面内の連続する画像間の平均の又は集計した距離が閾値よりも大きい場合には、場面内順序付けモジュール26を（望ましくはより短い）推定最短距離を与えるために再適用しても良い。不所望なほど高い動画一貫性を有する規則的場面が場面内順序付けモジュール26によって出力された場合には、クラスタ化モジュール16を再適用することや、画像対比較部12を再適用することで対画像比較基準を再定義することもあり得る。クラスタの数は、例えば、個別に検討した場合に動画の一貫性が満足できないような場面を一つにまとめるといった、場面の質を改善するうえで影響を及ぼすパラメータの一つである。

20

【0039】

場面内規則性を持った場面のセット30は（これは本明細書では規則的場面のセット30とも称される）、各場面内で実質的な動画の一貫性をもたらす。幾つかの実施形態では、これは最終的な動画コンテンツの出力となることがある。例えば、各規則的場面は動画コンテンツの単位として扱うことができ、プレゼンテーションアプリケーションなどを用いて表示することができる。例示される実施例を含む他の実施形態では、規則的場面30はさらに、その規則的場面30の順序を選択するために処理される。幾つかの実施形態では、場面の数が、規則的場面30を手動で順序付けることがあり得る程度に十分に少ないかもしれない。とは言え、手動で場面を順序付けることは、クラスタ化によって多数の場面が生成される画像のセットの数が多い場合には実際的ではない。また、手動での順序付けは、場面の数が少ない場合であっても望ましくないかもしれない。

30

【0040】

従って、図示される実施形態では、規則的場面30の順序付けは、場面順序付けモジュール36によって自動的に実施される。もし規則的場面30の順序がある用途において重要ではないと考えられる場合には、場面順序付けモジュール36は、例えばクラスタ化の間に用いられたクラスタの任意のインデックス化に従って場面を順序付けるといった、任意の、好ましくは計算効率の良い順序付け方案を採用することができる。又は、場面順序付けモジュール36は、クラスタのサイズによって規則的場面30を順序付けてもよい。その結果、動画コンテンツは、サイズ（すなわち、表示の観点からした長さ）が小さくなる（もしくは大きくなる）順序で以て示される。この方法は、例えば、画像セット10がクエリに基づく画像検索システムによって検索された「トップN」の画像である場合に、適切かもしれない。このような適用では、最大の場面には最も関連のある画像が含まれる可能性が高い。さらに具体的な例として、もしクエリが「cats」という文字である場合には、最大の場面にはそのクエリ「cats」によって意図された可能性が高い猫科の動物の画像が含まれる可能性が高い。一方で実際には、より小さな場面にはアンドリュー・ロイド・ウェバーによるブロードウェーミュージカルの「Cats」などに関連した画像が含まれるかもしれない。

40

50

【0041】

別の実施形態では、場面順序付けモジュール36は、場面に含まれる画像のタイムスタンプの平均又は他の集計に基づいて場面を順序付ける。これは実際に構築されたシステムで採用されている方法である。この場面順序付け基準は誕生日パーティーなどのイベントで取得された画像から動画コンテンツを構築するのに適切かもしれない。この場合、閲覧者が大凡時間順にそのイベントの画像を見たい可能性がある。そのような適用では、平均タイムスタンプの昇順に場面を並べることが好ましい。平均集計関数の代わりに、場面内の最小タイムスタンプのような別の集計関数を用いても良い。

【0042】

場面順序付けモジュール36が output するのは、場面内規則性を持った規則的場面のセット40である。幾つかの実施形態では、これは最終的な動画コンテンツの出力となり得る。図示される実施例のような他の実施形態では、場面内規則性を持った規則的場面のセット40は更に、任意に、連続する場面間での移行を選択するために処理され、さらに任意に、場面内で連続する画像間での移行を選択するために処理され、さらに任意に、各画像の表示時間を決定するために処理される。この移行の選択は、任意に、手動で行うことができる。

10

【0043】

又は、図示される実施例に示されるように、移行選択モジュール44が自動的に連続する場面間の移行を選択し、自動的に場面内の画像間での移行を選択し、自動的に各画像の表示時間を選択する。幾つかの実施例では、これらの値は均一な初期値として選択される。例えば、ある方法では、各画像に関し、一定の表示時間（例えば0.2秒）を与え、一つの場面内で連続する画像間で急に移行をし、連続する場面間では0.5秒のフェードトゥブラックによる移行のような特定の移行を行う。さらに一般的には、移行は、例えば、急な（すなわち強移行）、フェードトゥブラック、モーフィングなどの、映画のようなあらゆる適切な移行を含む。

20

【0044】

幾つかの現実に成された実行では、一つの場面における最後の画像から直後の場面における最初の画像への移行である場面内移行には、初期設定として0.5秒のフェードトゥブラック移行が設定され、一場面内の連続する画像間の移行は急な（強）移行であり、各画像の表示時間はその場面内で一定エントロピー率をもたらすように選択される。一定エントロピー率をもたらす画像表示時間を選択するのは、有効な実際のコミュニケーションは、ノイズのある通信路で最も効率的に情報を送信する方法は一定の割合で送ることであるという情報理論原則に従うということを示唆する証拠に基づいている。（Genzel et al ., "Entropy Rate Constancy in Text", in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 199-206を参照。）スピーチの場合、長い時間にわたって一定エントロピー率を維持することは、より注意深く発音していることになり、冗長さはあまり見られない。（Aylett, "Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech", in Proc. ICPHS99, 1999を参照。）動画の分野において、これは、画像の移行は、次に来る情報の不明確性が増すにつれて遅くすべきであるという原則に言い換えることができる。もし類似する画像間の移行の方が、類似していない画像間の移行よりも生じ易いと仮定すると、一定エントロピー率をもたらす、連続する画像間の距離に比例する移行速度が適切に定義される。全ての任意の画像間の平均距離 m_d と標準偏差 s_d を定義し、さらに m 及び M がそれぞれ最小及び最大画像移行速度であると定義し、両者の距離が $d_{i, i+1}$ （本具体例において、 d は類似基準ではなく、距離又は発散基準である）で表される二つの画像 I_i 及び I_j 間の移行速度を考えると、距離が $d_{i, i+1}$ だけ離れた二つの連続した画像間の移行速度は、次のロジスティック関数 $T(d_{i, i+1})$ によって適切に定義される：

30

40

【数3】

$$T(d_{i,i+1}) = \frac{M + m \cdot e^{\frac{-k(d_{i,i+1} - md)}{sd}}}{1 + e^{\frac{-k(d_{i,i+1} - md)}{sd}}} \quad (1),$$

10

ここにおいて、パラメータ K は変化の速度を調節するものである。現実に成された実行では、M = 0 . 5 s e c、m = 0 . 0 5 s e c なる値が選択された。

【0045】

移行選択モジュール 44 の出力は、定義された画像のプレゼンテーション時間だけでなく、場面間及び場面内で連続する画像間の移行が定義された規則的場面の規則的セットを含んだ動画コンテンツ 50 である。動画コンテンツ 50 は任意に、様々な方法で、更に処理を行うことなく保存することができる。例えば一つの方法では、N 枚から成る画像 10 が識別インデックス i (ここで 1 ≤ i ≤ N)と共に保存される。そして、動画コンテンツ 50 は次の様に保存することができる。(1) 規則的経路ベクトル P = [p₁, p₂, …, p_L]。ここで L は動画コンテンツ 50 内の画像の枚数 (異なる場面の複製された画像を含む) であり、ベクトル要素 p_i は動画コンテンツ 50 の再生中に表示される 1 番目の画像の識別インデックスを含む。(2) 移行ベクトル T = [t₁, t₂, …, t_L]。ここで、ベクトル要素 t_i は動画コンテンツ 50 の再生中に表示される i 番目の画像のプレゼンテーション時間を含む情報と、1 番目の画像の表示から (1 + 1) 番目の画像の表示への移行に関する移行情報とを包含している。

20

【0046】

任意であるが、動画コンテンツ 50 を様々な方法でさらに処理し、パッケージングしても良い。例えば、幾つかの実施形態では、動画コンテンツ 50 は、各画像が一つのスライドとして表されたスライドプレゼンテーションファイルとして保存される。OpenOffice.org Impress や Microsoft PowerPoint などの既存のスライドプレゼンテーションプログラムは、組み込みの「スライドショー」プレゼンテーションモードを備えており、移行及び画像 (即ちスライド) の表示時間を、各移行及び各スライドのそれぞれに関して個別に設定することが可能である。スライドプレゼンテーションとしてパッケージされた動画コンテンツ 50 は ODP や PPT といった従来のプレゼンテーション形式で保存することができ、OpenOffice.org Impress や Microsoft PowerPoint などの汎用スライドプレゼンテーションアプリケーションを用いて表示することができる。

30

【0047】

図示された実施例では、動画コンテンツ 50 は、予め定められたフレームレートで再生されるフレームから成る動画として動画生成モジュール 52 によってパッケージングされる。各画像のプレゼンテーション時間を動画のフレームレートに適応させるために、画像を有効な回数、複製することができる。例えば、もある画像の表示時間が t_d であり、動画フレームの持続時間が t_f (t_d < t_f であると仮定する) であるとすると、画像を t_d / t_f に最も近い整数に等しい回数だけ複製することにより、その画像が表示される。より具体的な例として、もし t_d = 0 . 5 s e c 及び t_f = 0 . 0 4 s e c (これは 25 フレーム / 秒のフレームレートに相当する) であるならば、t_d / t_f = 12 . 5 となり、その画像は 12 又は 13 (丸め方による) フレームのいずれかに適切に複製される。画像 10 の解像度及び動画の解像度に応じて、動画生成モジュール 52 は適切な動画の解像度を有するフレームを生成するために、画像に対し、ダウンサンプル、アップサンプル、又はその他の処理を行っても良い。カラーパレットやアスペクト比などの他の画像の特性もまた、画像を動画に適合させるために適切に調整することができる。

40

50

【0048】

移行は次のように動画生成モジュール52によって適切に実行される。急な（強）移行は次の画像を表示するフレームに直ちに進むことによって実行され、一方、フェードトゥブラックのような他の移行は連続的に「より黒い」、又は所望の移行をシミュレートしたり生じさせる、一連のフレームを用いて適切に実行される。

【0049】

動画生成モジュール52はMPEG、DivX、Xvid、FFmpeg、avi、wmvなどの選択された動画コーデックを用いて生成されたフレームから成る動画の形式を整えたりコード化したりする。こういった形式を整えることやコード化は、コーデックによって定義される画像圧縮などの処理を含んでいても良い。選択された動画コーデックに沿ってコード化された動画は動画ファイル記憶部54に保存される。動画ファイル記憶部54には、コンピュータのハードディスク、光ディスク、静電記憶装置、又は他の記憶媒体、デジタル動画レコーダのハードディスク、光ディスク、又は他の記憶媒体、等が含まれる。10

【0050】

保存された動画はMPlayer、VLC media player、Windows Media PlayerTM、Quicktimeなどの汎用動画又はマルチメディアプレーヤーを用いて動画を再生することで適切に表示される。適切に再生するために、選択された汎用動画又はマルチメディアプレーヤーエンジンには、動画が保存されたコーデックを「読む」ことができるデコーダがインストールされているべきである。幾つかの実施形態では、保存作業は割愛しても良い。例えば、動画はインターネット上で又は他のデジタルネットワーク上でストリーミングされても良い。20

【0051】

処理要素12、16、22、26、32、36、44、52、56は、関連する処理を実行するために適切にプログラムされたデジタルプロセッサを含む一又は複数のデジタル素子によって適切に具現化される。図示された実施例では、コンピュータCはこれらの要素を具現化するために構成されたデジタルプロセッサとメモリ（要素は非図示）を含んでおり、さらに、動画コンテンツを動画、スライドショープレゼンテーションなどとして表示するディスプレイDを含む。コンピュータの代わりに、他のデジタル処理装置を用いることができる。別の実例では、処理要素12、16、22、26、32、36、44、52、56は、画像セット10を取得するデジタルカメラによって実現され得る。これによつて、デジタルカメラが動画コンテンツ50を生成し、それをデジタルカメラの画面に表示し、カメラのメモリーカードにその動画を保存し、その動画をコンピュータや他の装置にダウンロードする、といったことができる。他の実施形態では、様々な処理要素が、二つ以上の異なるデジタル装置によって具現化されても良い。例えば、動画生成要素12、16、22、26、32、36、44、52はコンピュータによって具現化される一方、マルチメディアプレーヤーエンジン56は、デジタル動画デコーダー、DVD、又は動画再生機能を備える他の装置と動作可能なように接続されたテレビ受信器によって具現化されてもよい。30

【0052】

他の実施形態では、一又は複数の記録媒体が、処理要素12、16、22、26、32、36、44、52、56を具現化するためにデジタル処理装置によって実行可能な命令を保存する。この一又は複数の記録媒体には、例えば、光ディスクや他の光記録媒体、ハードディスクドライブや他の磁気記録媒体、フラッシュメモリや他の静電記録媒体、インターネットサーバに関連した記録装置、などが含まれる。40

【0053】

図2、3を参照して、ヨーロッパの修道院を歩いて回る間にデジタルカメラで撮った画像に適用される図1のシステムの、現実に成された実行によって生成された動画コンテンツのための幾つかの計量が示される。図2は二次元グリッドを描いているが、その軸は、事後分布において閾値0.4を採用した弱分割によるPLSAクラスタ化アルゴリズムを採用したクラスタ化モジュール16によって生成された動画コンテンツの場面に対応して50

いる。各グリッド要素(i, j)におけるグレースケール強度は場面*i*の画像と場面*j*の画像との間の距離の総計を指している。予想された通り、クラスタ化によって、図2において左上から右下に走る暗い斜めの線によって示されているように、 $i = j$ となるグリッド要素が最大の類似性を持つ。図3は26枚の画像から成る場面の一連の画像間の距離 $d_{i, i+1}$ (本具体例において、dは類似基準ではなく、距離又は発散基準である)をプロットしたものである。ここで、場面内順序付けモジュール26は異なる三つの最短経路推定技術を用いるように構成された。(1)画像が画像のタイムスタンプメタデータによって順序付けられる時間順序(この方法は時間的にデジタルカメラによって相次いで得られた画像が類似している可能性が高いということを仮定している)、(2)第一画像が選択され、その第一画像に最も類似した第二画像が次に選択され、その第二画像に最も類似した画像が次に選択され...、という手順を不規則な場面内の全ての画像が選択されるまで行うという欲張り選択法、そして、(3)Lin-Kernighan巡回セールスマントリアル解法である。

【0054】

引き続いて図3を参照すると、時間順序方法はひどく発散した結果となった。これは、ある場合には根底にある仮定が非常に良いかもしれない(例えば、同一人物の写真を何枚か続けて撮る)し、別の場合には根底にある仮定が非常に貧弱であるかもしれない(例えば、人物の写真に次いでステンドグラスの窓の写真が来る)からである。欲張り法は、初期は画像選択の結果が非常に良かった。なぜならば、初期の段階では、場面の中で、利用できる多数の画像の中から類似した画像を見つけるのは容易だからである。しかし、欲張り法による選択処理が進むにつれ、残っている利用可能な画像の数は着々と少なくなつて行き、結果的に、動画コンテンツが場面内で大きくなると共に、連続した画像間の距離 $d_{i, i+1}$ は次第に大きくなる。Lin-Kernighan巡回セールスマントリアル解法は場面全体の経路の最適化を伴うため、効果的に、場面全体における連続した画像間の良好な「平均」距離 $d_{i, i+1}$ を与える。

【0055】

図3は、26枚の画像のみから成る比較的短い場面に関する結果を示す。通常、所定のクラスタ化アルゴリズムにおいては、場面内の画像の枚数が増えるにつれて、得られる動画の一貫性が高まると期待される。この予想では、全ての画像が同一のソースから得られる。結果の質は、画像セット10が複数のソースから生成されるので、結果の質が高まると期待される。クラスタ化と場面の長さはトレードオフの関係にある。例えば、非常に類似した画像のみをクラスタ化する積極的分割基準を用いると、通常、短い規則的な場面をもたらす非常に類似した画像の小さなクラスタが出力される。こういった短い規則的な場面は、短いにもかかわらず、場面内の画像の類似性が高いため、動画の一貫性が高いかもしれない。一方で、あまり類似していない画像をグループ化する積極性の低い分割基準を用いると、類似性が低い、より大きなクラスタが出力される。結果として得られる規則的な場面は、場面内での画像間の相違性が高いため、より長いにもかかわらず、動画の一貫性はさらに低いかもしれない。動画の並びのダイナミック性は場面の長さと関係がある。従って、クラスタの数のパラメータを、出力される動画の所望のダイナミック性を調節するために用いることができる。

【符号の説明】

【0056】

10 画像セット、12 画像対比部、14 画像グラフ、16 クラスタ化モジュール、20 場面セット、22 異常画像削除サブモジュール、26 場面内順序付けモジュール、30 場面内規則性を持った場面のセット、32 非一貫場面削除モジュール、36 場面順序付けモジュール、40 場面内規則性を持った規則的場面のセット、44 移行選択モジュール、50 動画コンテンツ、52 動画生成モジュール、54 動画ファイル記憶部、56 マルチメディアプレーヤーエンジン、C コンピュータ、D ディスプレイ。

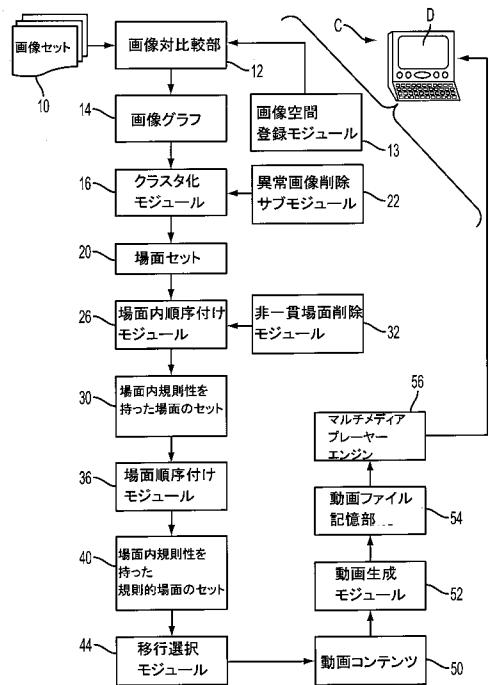
10

20

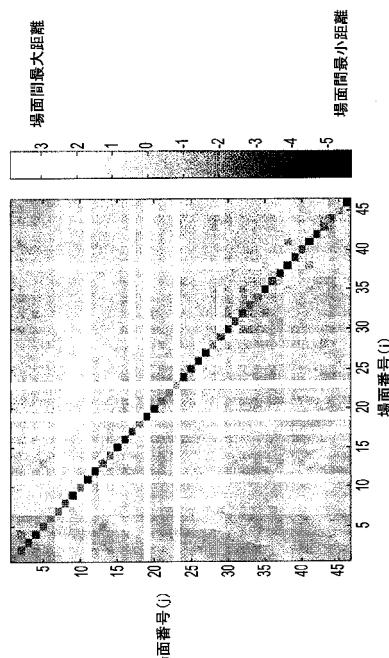
30

40

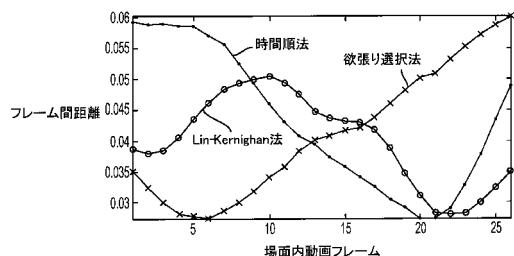
【図1】



【図2】



【図3】



フロントページの続き

(56)参考文献 米国特許出願公開第2006/0259863(US,A1)
特開平10-023408(JP,A)
特開2000-067057(JP,A)

(58)調査した分野(Int.Cl., DB名)

G 06 F 17 / 30
H 04 N 5 / 91