

(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2017-21606

(P2017-21606A)

(43) 公開日 平成29年1月26日(2017.1.26)

(51) Int.Cl.		F I				テーマコード (参考)
<b>G 0 6 F</b>	<b>17/30</b>	<b>(2006.01)</b>	<b>G 0 6 F</b>	<b>17/30</b>	<b>3 5 0 C</b>	<b>5 L 0 9 6</b>
<b>G 0 6 T</b>	<b>7/00</b>	<b>(2017.01)</b>	<b>G 0 6 F</b>	<b>17/30</b>	<b>1 7 0 D</b>	
			<b>G 0 6 T</b>	<b>7/00</b>	<b>3 0 0 F</b>	

審査請求 未請求 請求項の数 10 O L (全 33 頁)

(21) 出願番号	特願2015-139166 (P2015-139166)	(71) 出願人	899000068
(22) 出願日	平成27年7月10日 (2015.7.10)		学校法人早稲田大学
			東京都新宿区戸塚町1丁目104番地
		(74) 代理人	100080089
			弁理士 牛木 護
		(74) 代理人	100121153
			弁理士 守屋 嘉高
		(74) 代理人	100161665
			弁理士 高橋 知之
		(74) 代理人	100133639
			弁理士 矢野 卓哉
		(72) 発明者	松山 泰男
			東京都新宿区戸塚町1丁目104番地 学
			校法人早稲田大学内

最終頁に続く

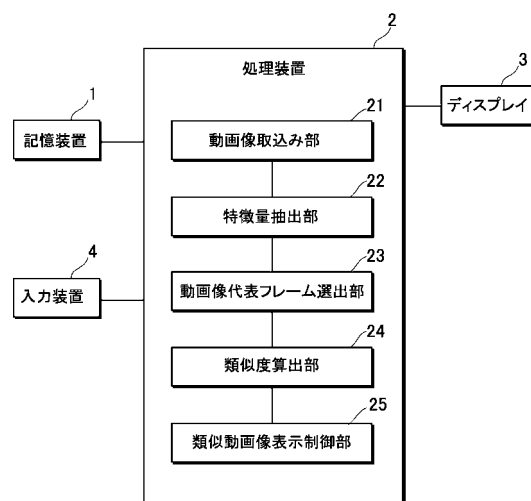
(54) 【発明の名称】 動画像検索方法、動画像検索装置及びそのプログラム

## (57) 【要約】

【課題】動画像そのものをクエリとする検索やグループ化を、従来よりも計算コストを低く抑えて高い精度で容易に行なう。

【解決手段】特徴量抽出部22は、クエリ動画像と対象動画像の特徴量をそれぞれ抽出する。動画像代表フレーム選出部23が、クエリ動画像と対象動画像について、時間区分k平均法、時間分割k平均法、時間拘束ペアウィズ・ニアレスト・ネイバー法、又は時間区分ペアウィズ・ニアレスト・ネイバー法の何れかを用いて、前記特徴量の関係から時系列で代表フレームを取得する。さらに類似度算出部24が、クエリ動画像と対象動画像との間の代表フレーム間類似度から、クエリ動画像と対象動画像との間の類似度を求めて、その結果を基にクエリ動画像に類似する動画像を検索する。

【選択図】図1



## 【特許請求の範囲】

## 【請求項 1】

クエリ動画像に対して対象動画像の類似度を算出して、前記クエリ動画像に類似する動画像を検索する動画像検索装置による動画像検索方法であって、

クエリ動画像のフレームから、フレームごとに第 1 の特徴量を抽出し、

対象動画像のフレームから、フレームごとに第 2 の特徴量を抽出し、

前記クエリ動画像について、抽出された第 1 の特徴量に対して、時間区分 k 平均法、時間分割 k 平均法、時間拘束ペアワイズ・ニアレスト・ネイバー法、又は時間区分ペアワイズ・ニアレスト・ネイバー法の何れかを用いて時系列で第 1 の代表フレームを取得し、

前記対象動画像について、抽出された第 2 の特徴量に対して、時間区分 k 平均法、時間分割 k 平均法、時間拘束ペアワイズ・ニアレスト・ネイバー法、又は時間区分ペアワイズ・ニアレスト・ネイバー法の何れかを用いて時系列で第 2 の代表フレームを取得し、

前記クエリ動画像における前記第 1 の代表フレームと、前記対象動画像における前記第 2 の代表フレームとの間の類似度から、前記クエリ動画像と前記対象動画像との間の類似度を求めて、前記クエリ動画像に類似する前記動画像を検索する、  
ことを特徴とする動画像検索方法。

10

## 【請求項 2】

前記第 1 の特徴量、及び前記第 2 の特徴量としてフレームシグネチャを用いることを特徴とする請求項 1 に記載の動画像検索方法。

## 【請求項 3】

前記第 1 の代表フレームの取得と前記第 2 の代表フレームの取得を同一の方法を用いて行なうことを特徴とする請求項 1 又は 2 に記載の動画像検索方法。

20

## 【請求項 4】

前記類似する動画像の検索では、

前記クエリ動画像における i 番目の前記第 1 の代表フレームが表わすフレーム数  $E^A_i$  と、前記対象動画像における j 番目の前記第 2 の代表フレームが表わすフレーム数  $E^B_j$  と、前記クエリ動画像における i 番目の前記第 1 の代表フレームと前記対象動画像における j 番目の前記第 2 の代表フレームとの間の前記代表フレーム間類似度  $s(i, j)$  と、をそれぞれ取得し、

ギャップペナルティ  $g$  を設定し、

30

前記クエリ動画像における前記第 1 の代表フレームの数が  $m$  であり、前記対象動画像における前記第 2 の代表フレームの数が  $n$  であるときに、 $(m+1) \times (n+1)$  の行列の各成分  $f(i, j)$  について、

$i = 0$  の行の各成分  $f(0, j)$  を、次の数 1 に基づいて算出し、

## 【数 1】

$$(0, -gE_1^B, -g\sum_{j=1}^2 E_j^B, -g\sum_{j=1}^3 E_j^B, \dots).$$

$j = 0$  の列の各成分  $f(i, 0)$  を、次の数 2 に基づいて算出し、

## 【数 2】

40

$$(0, -gE_1^A, -g\sum_{i=1}^2 E_i^A, -g\sum_{i=1}^3 E_i^A, \dots).$$

前記フレーム数  $E^A_i$  ,  $E^B_j$  に応じて算出される重みを  $r(i, j)$  としたときに、 $i = 0$  の行と  $j = 0$  の列以外の残りの行列成分  $f(i, j)$  を、次の数 3 に基づいて算出し、

【数 3】

$$f(i, j) = \max \{ f(i-1, j) - gE_i^A, f(i-1, j-1) + r(i, j)s(i, j), f(i, j-1) - gE_j^B \}$$

前記行列の最終行最終列の数値を利用して、前記クエリ動画像と前記対象動画像との間の前記類似度を算出することを特徴とする請求項 1 ~ 3 の何れか 1 項に記載の動画像検索方法。

【請求項 5】

前記類似する動画像の検索では、

10

前記クエリ動画像における  $i$  番目の前記第 1 の代表フレームが表わすフレーム数  $E_i^A$  と、前記対象動画像における  $j$  番目の前記第 2 の代表フレームが表わすフレーム数  $E_j^B$  と、前記クエリ動画像における  $i$  番目の前記第 1 の代表フレームと前記対象動画像における  $j$  番目の前記第 2 の代表フレームとの間の前記代表フレーム間類似度  $s(i, j)$  と、をそれぞれ取得し、

ギャップペナルティ  $g$  を設定し、

前記クエリ動画像における前記第 1 の代表フレームの数が  $m$  であり、前記対象動画像における前記第 2 の代表フレームの数が  $n$  であるときに、 $(m+1) \times (n+1)$  の行列の各成分  $f(i, j)$  について、

$i = 0$  の行の各成分  $f(0, j)$  と、 $j = 0$  の列の各成分  $f(i, 0)$  を、全て 0 に設定し、

20

前記フレーム数  $E_i^A$ 、 $E_j^B$  に応じて算出される重みを  $r(i, j)$  としたときに、 $i = 0$  の行と  $j = 0$  の列以外の残りの行列成分  $f(i, j)$  を、次の数 4 に基づいて算出し、

【数 4】

$$f(i, j) = \max \{ 0, f(i, -1, j) - gE_i^A, f(i-1, j-1) + r(i, j)s(i, j), f(i, j-1) - gE_j^B \}$$

前記行列の各成分の中の最大値を利用して、前記クエリ動画像と前記対象動画像との間の前記類似度を算出することを特徴とする請求項 1 ~ 3 の何れか 1 項に記載の動画像検索方法。

30

【請求項 6】

クエリ動画像に対して検索対象となる対象動画像の類似度を算出して、前記クエリ動画像に類似する動画像を検索する動画像検索装置であって、

前記クエリ動画像と前記対象動画像のフレームから、前記クエリ動画像に係るフレームごとの第 1 の特徴量と、前記対象動画像に係るフレームごとの第 2 の特徴量を抽出する特徴量抽出部と、

前記クエリ動画像と前記対象動画像について、抽出された第 1 の特徴量、及び第 2 の特徴量に対して、時間区分  $k$  平均法、時間分割  $k$  平均法、時間拘束ペアワイズ・ニアレスト・ネイバー法、又は時間区分ペアワイズ・ニアレスト・ネイバー法の何れかを用いて時系列で、第 1 の特徴量に係る第 1 の代表フレームと、第 2 の特徴量に係る第 2 の代表フレームを取得する代表フレーム選出部と、

40

前記クエリ動画像における前記第 1 の代表フレームと、前記対象動画像における前記第 2 の代表フレームとの間の類似度から、前記クエリ動画像と前記対象動画像との間の類似度を求めて、前記クエリ動画像に類似する前記動画像を検索する動画像検索部と、を備えたことを特徴とする動画像検索装置。

【請求項 7】

前記特徴量抽出部は、前記第 1 の特徴量、及び前記第 2 の特徴量としてフレームシグネチャを用いることを特徴とする請求項 6 に記載の動画像検索装置。

50

## 【請求項 8】

前記動画像検索部は、

前記クエリ動画像における  $i$  番目の前記第 1 の代表フレームが表わすフレーム数  $E^A_i$  と、前記対象動画像における  $j$  番目の前記第 2 の代表フレームが表わすフレーム数  $E^B_j$  と、前記クエリ動画像における  $i$  番目の前記第 1 の代表フレームと前記対象動画像における  $j$  番目の前記第 2 の代表フレームとの間の前記代表フレーム間類似度  $s(i, j)$  と、をそれぞれ取得し、

ギャップペナルティ  $g$  を設定し、

前記クエリ動画像における前記第 1 の代表フレームの数が  $m$  であり、前記対象動画像における前記第 2 の代表フレームの数が  $n$  であるときに、 $(m+1) \times (n+1)$  の行列の各成分  $f(i, j)$  について、

$i = 0$  の行の各成分  $f(0, j)$  を、次の数 5 に基づいて算出し、

## 【数 5】

$$(0, -gE_1^B, -g\sum_{j=1}^2 E_j^B, -g\sum_{j=1}^3 E_j^B, \dots).$$

$j = 0$  の列の各成分  $f(i, 0)$  を、次の数 6 に基づいて算出し、

## 【数 6】

$$(0, -gE_1^A, -g\sum_{i=1}^2 E_i^A, -g\sum_{i=1}^3 E_i^A, \dots).$$

前記フレーム数  $E^A_i$ 、 $E^B_j$  に応じて算出される重みを  $r(i, j)$  としたときに、 $i = 0$  の行と  $j = 0$  の列以外の残りの行列成分  $f(i, j)$  を、次の数 7 に基づいて算出し、

## 【数 7】

$$f(i, j) = \max \{ f(i-1, j) - gE_i^A, f(i-1, j-1) + r(i, j)s(i, j), f(i, j-1) - gE_j^B \}$$

前記行列の最終行最終列の数値を利用して、前記クエリ動画像と前記対象動画像との間の前記類似度を算出する構成としたことを特徴とする請求項 6 又は 7 に記載の動画像検索装置。

## 【請求項 9】

前記動画像検索部は、

前記クエリ動画像における  $i$  番目の前記第 1 の代表フレームが表わすフレーム数  $E^A_i$  と、前記対象動画像における  $j$  番目の前記第 2 の代表フレームが表わすフレーム数  $E^B_j$  と、前記クエリ動画像における  $i$  番目の前記第 1 の代表フレームと前記対象動画像における  $j$  番目の前記第 2 の代表フレームとの間の前記代表フレーム間類似度  $s(i, j)$  と、をそれぞれ取得し、

ギャップペナルティ  $g$  を設定し、

前記クエリ動画像における前記第 1 の代表フレームの数が  $m$  であり、前記対象動画像における前記第 2 の代表フレームの数が  $n$  であるときに、 $(m+1) \times (n+1)$  の行列の各成分  $f(i, j)$  について、

$i = 0$  の行の各成分  $f(0, j)$  と、 $j = 0$  の列の各成分  $f(i, 0)$  を、全て 0 に設定し、

前記フレーム数  $E^A_i$ 、 $E^B_j$  に応じて算出される重みを  $r(i, j)$  としたときに、 $i = 0$  の行と  $j = 0$  の列以外の残りの行列成分  $f(i, j)$  を、次の数 8 に基づいて算出し、

10

20

30

40

【数 8】

$$f(i,j) = \max \{ 0, f(i,-1,j) - gE_i^A, f(i-1,j-1) + r(i,j)s(i,j), f(i,j-1) - gE_j^B \}$$

前記行列の各成分の中の最大値を利用して、前記クエリ動画像と前記対象動画像との間の前記類似度を算出する構成としたことを特徴とする請求項 6 又は 7 に記載の動画像検索装置。

【請求項 10】

コンピュータに、請求項 1 ～ 5 の何れか 1 項に記載の動画像検索処理を実行させるための動画像検索プログラム。 10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、動画像検索方法、動画像検索装置及びそのプログラムに関する。

【背景技術】

【0002】

インターネット上には、莫大な量の動画像データが蓄積されている。これらの動画像の検索に関してよく使われる方法として、テキストベースの検索手法が知られている。このテキストベースの検索手法は、動画像のそれぞれに適切なテキスト情報を対応させ、テキスト検索を行なうことで所望の動画像を検索するものである。しかしながら、このような手法では、以下のような問題がある。 20

(a) 動画像とテキスト情報の対応付け(ラベル付け)は手作業で行なう必要があり、コストがかかる。

(b) 汎用性が乏しい不適切なラベルが付けられることにより、検索の精度や効率が低下する可能性がある。

【0003】

上記問題点を解決する手法として、画像の内容(コンテンツ)そのものをクエリとして用いるコンテンツベースの検索が期待される。

【0004】

コンテンツベースの検索方法として、画像全体の輝度の変化と複雑度、エッジ量の複雑度、輝度分布を用いた簡易構図情報など画像全体に関する情報、またはそれらの組み合わせで、クエリ画像と検索対象画像との類似度を計算する手法が提案されている(例えば特許文献 1 参照)。また、複数の検索対象画像をそれぞれ所定の大きさの要素画像に分割し、分割された要素画像から代表的な要素画像を選択し、この代表的な要素画像を用いて描画した検索参照画像による画像の検索を行なう手法も提案されている(例えば特許文献 2 参照)。 30

【0005】

しかしながらこれらの提案手法は、静止画像をクエリとして用いて静止画像の検索を行なうものであり、動画像をクエリとして検索する手法については考慮されていない。 40

【先行技術文献】

【特許文献】

【0006】

【特許文献 1】特開 2001 - 56820 号公報

【特許文献 2】特開 2002 - 140331 号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

上記したように、従来技術においてはテキストベース又は静止画像をクエリとして用いる手法について開示があるのみであり、動画像をクエリとする、動画像のコンテンツに基 50

づいた動画画像検索について開示されたものではなく、検討されていなかった。本発明は、動画画像そのものをクエリとする検索を、計算コストを低く抑えつつ精度良く行なうことが可能な動画画像検索方法、動画画像検索装置及びそのプログラ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】

【0008】

本発明は、クエリ動画画像に対して対象動画画像の類似度を算出して、前記クエリ動画画像に類似する動画画像を検索する動画画像検索装置による動画画像検索方法であって、クエリ動画画像のフレームから、フレームごとに第1の特徴量を抽出し、対象動画画像のフレームから、フレームごとに第2の特徴量を抽出し、前記クエリ動画画像について、抽出された第1の特徴量に対して、時間区分k平均法、時間分割k平均法、時間拘束ペアワイズ・ニアレスト・ネイバー法、又は時間区分ペアワイズ・ニアレスト・ネイバー法の何れかを用いて時系列で第1の代表フレームを取得し、前記対象動画画像について、抽出された第2の特徴量に対して、時間区分k平均法、時間分割k平均法、時間拘束ペアワイズ・ニアレスト・ネイバー法、又は時間区分ペアワイズ・ニアレスト・ネイバー法の何れかを用いて時系列で第2の代表フレームを取得し、前記クエリ動画画像における前記第1の代表フレームと、前記対象動画画像における前記第2の代表フレームとの間の類似度から、前記クエリ動画画像と前記対象動画画像との間の類似度を求めて、前記クエリ動画画像に類似する前記動画画像を検索することを特徴とする。

10

【0009】

また本発明は、クエリ動画画像に対して検索対象となる対象動画画像の類似度を算出して、前記クエリ動画画像に類似する動画画像を検索する動画画像検索装置であって、前記クエリ動画画像と前記対象動画画像のフレームから、前記クエリ動画画像に係るフレームごとの第1の特徴量と、前記対象動画画像に係るフレームごとの第2の特徴量を抽出する特徴量抽出部と、前記クエリ動画画像と前記対象動画画像について、抽出された第1の特徴量、及び第2の特徴量に対して、時間区分k平均法、時間分割k平均法、時間拘束ペアワイズ・ニアレスト・ネイバー法、又は時間区分ペアワイズ・ニアレスト・ネイバー法の何れかを用いて時系列で、第1の特徴量に係る第1の代表フレームと、第2の特徴量に係る第2の代表フレームを取得する代表フレーム選出部と、前記クエリ動画画像における前記第1の代表フレームと、前記対象動画画像における前記第2の代表フレームとの間の類似度から、前記クエリ動画画像と前記対象動画画像との間の類似度を求めて、前記クエリ動画画像に類似する前記動画画像を検索する動画画像検索部と、を備えたことを特徴とする。

20

30

【発明の効果】

【0010】

上述した本発明の動画画像検索方法、動画画像検索装置およびそのプログラムにより、動画画像をクエリとする検索を、計算コストを低く抑えつつ精度良く行なうことができる。

【図面の簡単な説明】

【0011】

【図1】本発明の実施の形態による、動画画像検索装置のシステム構成を示すブロック図である。

【図2】本発明の実施の形態による動画画像検索装置の動作フローの説明図である。

40

【図3】本発明の実施の形態で適用する時系列情報を考慮したクラスタリングと、従来技術の時系列情報を考慮していないクラスタリングとの違いを示す説明図である。

【図4】本発明の実施の形態において、時間窓を用いたaffinity propagation法(TB-AP, Time-Bound Affinity Propagation)による代表フレーム選出アルゴリズムを示すフローチャートである。

【図5】本発明の実施の形態と従来技術において、メッセージ交換の違いを説明する説明図である。

【図6】本発明の実施の形態において、メッセージ交換のリンク行列を示す説明図である。

【図7】本発明の実施の形態において、TB-APにおける時間窓と、改良したk-means法やha

50

rmonic competition法における時間ブロックを、それぞれ時間軸上で示した説明図である。

【図 8】本発明の実施の形態における時間区分 k 平均法 (TP k-means) の説明図である。

【図 9 A】本発明の実施の形態における時間分割 k 平均法 (Time-Split k-means : TS k-means) について、全フレームを時系列に k 等分して割り振り、各クラスターで重心を求めた状態を示す説明図である。

【図 9 B】本発明の実施の形態における時間分割 k 平均法 (Time-Split k-means : TS k-means) について、各フレームで現在属するクラスターの重心と、時系列的に隣接する前後のクラスターの重心との間の距離を計算した状態を示す説明図である。

【図 9 C】本発明の実施の形態における時間分割 k 平均法 (Time-Split k-means : TS k-means) について、再クラスタリングを行なった状態を示す説明図である。

【図 9 D】本発明の実施の形態における時間分割 k 平均法 (Time-Split k-means : TS k-means) について、時系列が不連続なフレームの間で、クラスターを分割した状態を示す説明図である。

【図 9 E】本発明の実施の形態における時間分割 k 平均法 (Time-Split k-means : TS k-means) について、クラスターを分割した後に、各クラスターで重心を算出した状態を示す説明図である。

【図 10】本発明の実施の形態におけるペアワイズ・ニアレスト・ネイバー法 (PNN) におけるデータポイントの併合を示す説明図である。

【図 11】本発明の実施の形態における時間区分ペアワイズ・ニアレスト・ネイバー法 (TP-PNN) の説明図である。

【図 12】本発明の実施の形態における時間拘束ペアワイズ・ニアレスト・ネイバー法 (TB-PNN) の説明図である。

【図 13】本発明の実施の形態において、M 距離による類似度算出アルゴリズムを示すフローチャートである。

【図 14】実験で使用した元の動画像データの生成過程を表わす説明図である。

【図 15】実験で使用した動画像データの構造を示す説明図である。

【図 16】5 つのガウス型クラスターのそれぞれを、時間方向で 2 つに分けて 10 集団としたプロットであり、時間性を反映した代表点 (典型静止画) の出現と計算速度を調べるために使用した図である。

【図 17】動画像  $v_A$  と動画像  $v_B$  の各典型静止画の距離を行列として表した図である。

【図 18】実験で得られた大域アライメント用テーブルの一例を示す図である。

【図 19】実験で得られた局所アライメント用テーブルの一例を示す図である。

【図 20】大域アライメント用テーブルに対応して、計算を省略した個所を黒く塗り潰した図である。

【図 21】実験で得られた動画像検索における適合率 - 再現率曲線を示すグラフである。

【図 22 A】他の実験例として、異なる代表フレーム選出アルゴリズムを用いた実験で得られた、動画像検索における適合率 - 再現率曲線を示すグラフであって、クラスタリング時に冗長代表フレーム削除の事後処理を行なうことなく、そのまま類似度の算出を行なった場合の 5 つのクラスタリング手法における適合率 - 再現率曲線を表す。

【図 22 B】他の実験例として、異なる代表フレーム選出アルゴリズムを用いた実験で得られた、動画像検索における適合率 - 再現率曲線を示すグラフであって、クラスタリング時に冗長代表フレーム削除の事後処理を行なったときの 5 つのクラスタリング手法における適合率 - 再現率曲線を表す。

【図 23 A】他の実験例として、特徴量として CSD を用いた実験で得られた、動画像検索における適合率 - 再現率曲線を示すグラフである。

【図 23 B】他の実験例として、特徴量として Frame Signature を用いた実験で得られた、動画像検索における適合率 - 再現率曲線を示すグラフである。

【図 24】本発明の実施の形態による、動画検索装置の処理装置としてコンピュータを用いた場合のハードウェア構成図の例である。

10

20

30

40

50

## 【発明を実施するための形態】

## 【0012】

以下、添付図面に基づき、本発明における動画像検索方法と、動画像検索装置について説明する。

## 【0013】

図1は、本実施の形態における動画像検索装置のシステム構成の一例を模式的に示すものである。同図において、1は動画像の集合を記憶保存するデータベースとして用いるハードディスクドライブなどの記憶装置、2は類似画像の検索を行なうコンピュータなどの処理装置、3は検索に用いる動画像や検索対象の動画像を表示したり、検索結果を表示するディスプレイ、4はマウスやキーボード等の入力装置である。記憶装置1は処理装置2に読み出し可能な状態に接続される。処理装置2は、必要に応じて記憶装置1に蓄積された動画像を表示装置であるディスプレイ3に表示できるようになっている。なお、記憶装置1は処理装置2に内蔵されていても、また、処理装置2に通信手段を介して接続されるサーバ等のコンピュータ上に構築されていてもよい。また、クエリ画像と検索対象画像とを別々の記憶装置に記憶させておくこともでき、どのような形態であるかは特に限定されない。なお、処理装置2は入力装置4と接続されており、ユーザーによる入力を適宜受け付けることができる。

## 【0014】

処理装置2は、クエリ動画像及び対象動画像の処理を行なうため、動画像取込み部21と、特徴量抽出部22と、動画像代表フレーム選出部23とを備えている。動画像取込み部21は、記憶装置1に保存された動画像の中から、入力装置4を介してユーザーが選択したクエリ動画像、あるいは検索の対象となる対象動画像を取込むものである。特徴量抽出部22は、動画像取込み部21で取込んだクエリ動画像、あるいは対象動画像の特徴量を抽出する。動画像代表フレーム選出部23は、特徴量抽出部22で抽出した、クエリ動画像あるいは対象動画像の特徴量に基づき、クエリ動画像あるいは対象動画像内における一乃至複数の代表フレームをそれぞれ選出する。なお、上記動画像取込み部21、特徴量抽出部22、及び動画像代表フレーム選出部23は、クエリ動画像用と対象動画像用とに分けて構成しても構わない。

## 【0015】

処理装置2はその他に、クエリ動画像に類似する対象動画像を検索するため、類似度算出部24と、類似動画像表示制御部25とを備えている。類似度算出部24は、クエリ動画像内における各代表フレームの特徴量と対象動画像内における各代表フレームの特徴量とをそれぞれ比較し、類似度を算出するものである。類似動画像表示制御部25は、類似度算出部24で算出した類似度の高い順に、対象動画像をディスプレイ3に一乃至複数表示することができる。

## 【0016】

処理装置2としてコンピュータを用いる場合、類似動画像検索を行なうためのアプリケーションソフトウェアをインストールすることで、処理を行なうための各機能の一部または全部を実現することができ、処理装置2は、インストールされたアプリケーションソフトウェアに従った処理を実行することができる。

## 【0017】

図2は、本実施の形態における処理装置2の処理概要を示したものである。図中、左経路はクエリ動画像に対する処理フローを表し、右経路は対象動画像に対する処理フローを表す。図2の処理により、最終的にクエリ動画像に対する対象動画像の類似度が算出される。

## 【0018】

図2の左経路のクエリ動画像の処理フローを説明する。S110で、記憶装置1のデータベースからクエリ動画像を取り込む。次にS120で、クエリ動画像の特徴量を抽出する。特徴量としては、MPEG-7で提案されている画像特徴量の一つである色構造記述子( Color Structure Descriptor ; CSD ) や、ISO/IECで規定されているVideo Signature ( Multi

10

20

30

40

50



media content description interface, International Standard of ISO/IEC 15938-3, Amendment 4 (2010))を用いることができる。次にS 1 3 0で、典型静止画としてクエリ動画像の代表フレームを決める。このとき、動画像の経過時間の要素を考慮した手法を用いて代表フレームは決められる。クエリ動画像のフレーム数にもよるが、通常は複数の代表フレームが選択され、選択された代表フレームに対応するそれぞれの典型静止画が持つ特徴量の集合がクエリ動画像の特徴量とされる。

#### 【0019】

図2の右経路の対象動画像の処理フローもクエリ動画像の処理フローと同様の手順で行なわれる。すなわち、S 1 4 0で、記憶装置1のデータベースから対象動画像を取り込む。次にS 1 5 0で、対象動画像の特徴量を抽出する。特徴量として前記したCSDやVideo S  
ignatureを用いることができる。次にS 1 6 0で、対象動画像の典型静止画となる代表フレームを決める。このとき、前記したように、動画像の経過時間の要素を考慮した手法を用いて代表フレームは決められる。対象動画像のフレーム数にもよるが、通常は複数の代表フレームが選択され、選択された代表フレームに対応するそれぞれの典型静止画が持つ特徴量の集合が対象動画像の特徴量とされる。図2の左経路の処理と右経路の処理は、どちらを先に行なっても、また、並列的に行なっても構わない。

10

#### 【0020】

なお、動画像代表フレーム選出部23は、動画像の経過時間の要素を考慮した技術として、時間拘束アフィニティ伝播法(Time-Bound Affinity propagation: TB-AP)、時間区分k平均法(Time-Partition k-means: TP-k-means)、時間分割k平均法(Time-Split k  
-means: TS-k-means)、時間拘束ペアワイズ・ニアレスト・ネイバー法(Time-Bound pairwise nearest neighbor: TB-PNN)、あるいは時間区分ペアワイズ・ニアレスト・ネイバー法(Time-Partition pairwise nearest neighbor: TP-PNN)を用いることができる。これらについては後程詳しく説明する。

20

#### 【0021】

また、TP-k-means法の代わりに、その一般化である時間ブロックを導入したharmonic competitionを適用することも可能である。harmonic competition法については、「Matsuyama, Yasuo, "Harmonic Competition: A Self-Organizing Multiple Criteria Optimization," IEEE Transactions on Neural Networks, Vol.7, No.3, pp.652-668, MAY 1996」などに記載されている。

30

#### 【0022】

こうして、クエリ動画像と対象動画像について、それぞれの代表フレーム(典型静止画)が選出されると、S 1 7 0の類似度算出の処理が行なわれる。ここでの典型静止画の数は各動画像に依存するので、クエリ動画像と対象動画像との類似度は、ギャップの挿入が可能な重み付きM距離によって算出される。なお、M距離については後程詳しく説明する。S 1 8 0において、S 1 7 0で算出した類似度スコアを基に、類似度の高い対象動画像を検索してディスプレイ3に表示する。

#### 【0023】

上述の処理フローで、S 1 2 0, S 1 5 0における特徴量の抽出、S 1 3 0, S 1 6 0における代表フレームの選出、及びS 1 7 0における類似度の算出の3つの処理について以下に説明する。

40

#### 【0024】

##### ・フレームに対する特徴量の抽出

本実施の形態では、始めに動画像に含まれる各フレームの特徴量の抽出が実行される。ここで抽出された特徴量は、時系列を考慮したベクトル $\{x_t\}_{t=1}^n$ として表現される。例えば、類似画像検索で用いるアプリケーションソフトウェアによる処理が色情報を持つ動画像を扱えるように、当該ベクトルを先に述べたMPEG-7のCSDによって表現し、ベクトルをすべて正規化されたCSDヒストグラムとすることができる。

#### 【0025】

ただし、動画像の各フレームの特徴量はCSDを用いたものに限らず、当該特徴量はベク

50

トル量であればどのようなものでも構わない。前述したCSDヒストグラムの特徴量は以下の条件を適用して正規化している。

・ 正規化の条件

( i ) 各ベクトルの成分は非負とする。

( i i ) ベクトル成分全部の和は、1.0となるようにする。

これは、次の手順で得られる。

( a ) 動画像の各フレームについて、その特徴を表すベクトル量に変えておく ( 特徴量抽出 ) 。

( b ) 類似度の比較法およびそれをソフトウェアとして実現した時の汎用性を確保できるように、各ベクトルを正規化できる手法を採用する。

10

【 0 0 2 6 】

上記手順 ( b ) は、実質的に各ベクトルの最小値と総和を使うことで常に可能になる。また、CSDは手順 ( a ) における特徴ベクトルの単なる一例で、画像フレーム全体だけを対象にするさらに簡単なものがヒストグラムである。

【 0 0 2 7 】

CSDは簡単な特徴量ではあるが、良好な結果を得ることができて好ましい。また、CSDの代わりに、計算量が多くなるものの、別な手法による特徴量を用いることも可能である。例えば、動画像の各フレームの特徴量は、SIFT ( Scale-Invariant Feature Transform ) や、SURF ( Speeded Up Robust Features ) や、FREAK ( Fast Retina Keypoint ) や、bag of visual wordsなどを用いることができる。また、先に述べたVideo Signatureを用いることもできる。これらCSD以外の特徴量や、CSDを含む特徴量を組み合わせることも可能である。

20

【 0 0 2 8 】

例えば、前述した { CSD , FREAK , bag of visual words , SIFT , Video Signature... } などを個別の正規化済み特徴量の集団として、個別の正規化済み特徴量をNormalizedFeatureVector ( i ) と表記すると、動画像フレームの特徴量として、以下の数 1 に定義される合成特徴量ベクトルCompositeFeatureVectorを利用することもできる。ただし、正規化を行なう必要がない特徴量を用いる場合、NormalizedFeatureVectorは入力された特徴量をそのまま出力する。

【 0 0 2 9 】

30

【 数 1 】

$$\text{CompositeFeatureVector} = \sum_{i=1}^N w_i \times \text{NormalizedFeatureVector}(i)$$

【 0 0 3 0 】

ただし、合成特徴量ベクトルCompositeFeatureVectorは上式において正規化条件：  $w_i \geq 0$ 、 $\sum_{i=1}^N w_i = 1$  を満たすものとする。上記数 1 の式は、合成特徴量ベクトルCompositeFeatureVectorを得る一例に過ぎず、例えば特徴量同士を掛け算したものも、合成特徴量ベクトルCompositeFeatureVectorとして利用できる。したがって、合成特徴量ベクトルCompositeFeatureVectorは、次の数 2 の一般式で得られる。これは、幾つかの個別の特徴量FeatureVector ( i ) を組み合わせて最終的に正規化したものが、合成特徴量ベクトルCompositeFeatureVectorになるという表現を、数式化したものである。

40

【 0 0 3 1 】

【 数 2 】

$$\text{CompositeFeatureVector} = f(\text{FeatureVector}(1), \dots, \text{FeatureVector}(N))$$

【 0 0 3 2 】

・ 時系列を保つ代表フレームの選出

各フレームに対する特徴量を抽出した後、代表フレームの選出が行なわれる。ここでは、処理装置 2 が動画像内の各フレームを抽出した特徴量に基づいてクラスター化し、各ク

50

ラスターを代表するフレームを選出する。各クラスターを代表するフレームを典型静止画と呼ぶが、本実施の形態では、典型静止画の選出には時系列情報が加味されている。

【0033】

図3は、本実施の形態で用いる時系列情報を考慮した時間拘束アフィニティ伝播法(TB-AP)と通常のアフィニティ伝播法(AP)におけるクラスタリングの違いを説明する図である。図中上段にある時系列データは、特徴量として抽出した時間  $t = 1$  から時間  $t = 10$  までの  $n = 10$  個の時系列の特徴ベクトル  $\{x_t\}_{t=1}^{n=10}$  を表している。ここでの各ボックスは各時間  $t$  におけるフレームを表し、例えば時間  $t = 1$  のフレームには1の数字を、時間  $t = 5$  のフレームには5の数字が付してある。また、図で同じグラデーションで描いたフレームは、互いに類似しているフレーム(例えば動画の中の1シーン等)を表し、同じクラスに属するものとする。図示の例では、 $\{1, 2, 10\}$  のクラス、 $\{3, 4, 5\}$  のクラス、 $\{6, 7, 8, 9\}$  のクラスと、類似フレームのセットが3種類存在する。

10

【0034】

これらのフレームを特徴量別にクラスター化する場合、通常のアフィニティ伝播法(AP)を用いた場合には、図中左側下段に示すように、典型静止画として例えばフレーム  $\{2, 4, 7\}$  が選出される。しかし、これは時系列情報が考慮されておらず、動画像を検索するラベルとして適切ではない。時系列が正しく考慮されることにより、図中右側下段に示すように、典型静止画は  $\{2, 4, 7, 10\}$  と選出される。動画像の経過時間を考慮する時間拘束アフィニティ伝播法(TB-AP)では、そのような時系列を考慮したクラスタリングが可能となる。つまり図3の例では、時間拘束アフィニティ伝播法を適用すると、典型静止画が正しく  $\{(1, \underline{2}, 0), (1, \underline{4}, 1), (1, \underline{7}, 2), (0, \underline{10}, 0)\}$  と選出される。

20

【0035】

なお、これらの典型静止画のフレーム数は動画像の長さやコンテンツに応じて可変する。したがって、各々の典型静止画は、時間順序におけるその左右に、責任のある(responsible)フレームを含んでいる。

【0036】

以下、時系列を保つ典型静止画を選出するためのアルゴリズムの一つであるTB-APのアルゴリズムについて、図4のフローチャートを参照しながら説明する。

30

【0037】

S210: Similarityの計算

ここでは、時間  $t = 1$  から時間  $t = n$  までの動画像フレームについて、特徴量として抽出された  $n$  個の時系列の特徴ベクトル  $\{x_t\}_{t=1}^n$  が与えられる。前述のように、各々の特徴ベクトル  $\{x_t\}_{t=1}^n$  は、合計が1となる正規化されたCSDヒストグラムである。時間  $t = i$  の特徴ベクトル  $x_i$  から、時間  $t = j$  の特徴ベクトル  $x_j$  への関係性を表わすSimilarity測度  $s(x_i, x_j)$  を、 $s(i, j)$  として定義する。そして、特徴ベクトル  $x_i$  が特徴ベクトル  $x_j$  よりも時間  $t = k$  の特徴ベクトル  $x_k$  に類似している時かつその時に限り、 $s(k, i) > s(k, j)$  を満たすような性質を持つものとする。

40

【0038】

本実施の形態では、以下の数3の式で計算したSimilarity測度  $s(i, j)$  を用いることとする。

【0039】

【数3】

$$s(i, j) = \bar{D} - \|x_i - y_j\|.$$

【0040】

ただし、 $\text{bar}(D)$  (以下、数3の右辺第1項のように、記号Dにオーバーラインが付いたものを、本文中では $\text{bar}(D)$ と表記する)は、例えば入力装置4によりユーザーが設定可能な定数パラメータである。

50

## 【 0 0 4 1 】

Similarityは、「値が大きいほど関係性も大きく、値が小さいほど関係性も小さい」というルールに則った値の設定であれば、数3の式に限らずどのようなものでも構わない。また、 $\text{bar}(D)$ の選び方はTB-APの結果に影響しないので、ここでは0としておくが、当該パラメータは動画像間を比較する上では重要な要素となり、0以外の値をとりうる。

## 【 0 0 4 2 】

S 2 2 0 : 行列 A を初期化

ここでは、初期値を零行列にした行列  $A = [a_{ij}]$  を用意する ( $a_{ij} = 0$ )。この行列 A は、availabilityを表す行列である。availabilityは、典型静止画候補のデータポイント j からデータポイント i へ送られるメッセージで、データポイント i がデータポイント j の典型静止画として選ぶことがどれだけ適切かを表す値である。

10

## 【 0 0 4 3 】

S 2 3 0 : 行列 A を用いて行列 R を算出 (または更新)

ここでは先ず、スライドする時間窓の長さを、その中心が対角成分に位置するように、 $W = 2 - 1$  に設定した後、時間窓内での時系列を考慮して、特徴ベクトル  $x_i, x_j, x_k$  を選択する。

## 【 0 0 4 4 】

次に、responsibilityを表す行列 R の成分  $r_{ij}$  を、先に説明したavailability行列 A を利用して計算更新する。responsibilityは、データポイント i から典型静止画候補のデータポイント j へ送られるメッセージで、データポイント i がデータポイント j の典型静止画としてどのくらい適切かを表す値である。このresponsibility行列 R の計算と、後述するavailability行列 A の計算では、データポイント i, j を次の数4の範囲に設定して算出する。つまりTB-APでは、メッセージ交換を行なう範囲に数4に示す制限を設けている。

20

## 【 0 0 4 5 】

## 【 数 4 】

$$i: 1 \leq i \leq n,$$

$$j: 1 \leq j \leq n, \text{ constrained by } i - \omega < j < i + \omega.$$

30

## 【 0 0 4 6 】

responsibility行列 R は通常対称的であり、行列 A を用いて数5に示すように更新される。

## 【 0 0 4 7 】

## 【 数 5 】

$$\rho_{ij} := s(i, j) - \max_{k: k \neq j} \{a_{ik} + s(i, k)\},$$

$$r_{ij} := (1 - \lambda)\rho_{ij} + \lambda r_{ij}.$$

## 【 0 0 4 8 】

40

数5において、 $r_{ij}$  は、データポイント i とデータポイント j との間のresponsibilityの伝播値である。また、 $(0, 1)$  は設計パラメータとしてのダンピング係数であり、これを用いて後述する S 2 5 0 での収束を遅らせ、安定した結果を得ることができる。行列 R も初期値は零行列とされるが、例えばダンピング係数が0であれば、responsibilityの伝播値  $r_{ij}$  がそのまま行列 R の値として更新される。

## 【 0 0 4 9 】

S 2 4 0 : 行列 A を算出

ここでは、データポイント i, j を数4で示した範囲に制限し、次の数6を用いてavailability行列 A の成分  $a_{ij}$  を更新する。

## 【 0 0 5 0 】

50

【数 6】

$$\alpha_{ii} := \sum_{k \neq i} \max\{r_{ki}, 0\},$$

$$\alpha_{ij} := \min\{0, r_{jj} + \sum_{k: k \neq i, k \neq j} \max\{r_{kj}, 0\}\}, (i \neq j),$$

$$a_{ij} := (1 - \lambda)\alpha_{ij} + \lambda a_{ij} \text{ (including } i = j\text{)}.$$

【0051】

10

数 6 において、 $a_{ij}$  は、データポイント  $i$  とデータポイント  $j$  との間のavailabilityの伝播値である。

【0052】

S 2 5 0：行列 A と行列 R の総和が収束したか否かを判定

以上のように算出した行列 A の成分  $a_{ij}$  と行列 R の成分  $r_{ij}$  とを加算し、その値  $a_{ij} + r_{ij}$  が収束の基準を満たすまで、S 2 3 0 と S 2 4 0 の手順を繰り返す。一方、 $a_{ij} + r_{ij}$  の値が収束の基準を満たすと、次の S 2 6 0 の手順に移行する。

【0053】

S 2 6 0：代表フレーム選出

ここでは、以下の数 7 に定義されるインデックスが、典型静止画のインデックスとして採用される。

20

【0054】

【数 7】

$$j: \arg \max_{i: i - \omega < j < i + \omega \text{ for } 1 \leq i \leq n} \{a_{ij} + r_{ij}\}$$

【0055】

つまり、 $a_{ij} + r_{ij}$  の値が最大となるデータポイントでの静止画が、典型静止画として採用される。例えば、データポイント  $i$  は、 $a_{ij} + r_{ij}$  の値が最大となるデータポイント  $j$  を中心とするクラスターに属する。最終的に代表フレームとして選出される典型静止画の集合は、こうしたインデックスを収集して見つけ出される。

30

【0056】

上記一連の S 2 3 0 ~ S 2 6 0 のフローの中で、時間窓を設定する上で必要な  $\omega$  は、行列 A や行列 R を算出する際のメッセージ交換の範囲を制限するパラメータとなる。この制限下におけるメッセージ交換の様子を、図 5 に示す。同図において、上段の「AP」は、従来技術となるアフィニティ伝播法によるメッセージ交換の範囲を示している。また下段の「TB-AP」は、本実施の形態における時間拘束アフィニティ伝播法によるメッセージ交換の範囲を示している。ここでは、前記図 3 で説明したものと同一フレームが時系列に並べられているものとする。

40

【0057】

通常の「AP」では、全てのデータポイントに対して、メッセージの交換が行われるが、本実施の形態では、時間窓で制限された近傍のデータポイントに対して、メッセージの交換が行なわれる。具体的には、図 5 に示す「AP」の例では、データポイント  $i = 5$  を中心として、データポイント  $j = 1$  から  $j = 10$  の全範囲にデータ交換が行なわれる。一方、同図「TB-AP」の例では、パラメータ  $\omega = 4$  に設定して、時間窓  $2\omega - 1 = 7$  の範囲でメッセージ交換を行なう構成になっている。そのため、データポイント  $i = 5$  を中心として、データポイント  $j = 2$  から  $j = 8$  の範囲にメッセージ交換の範囲が制限される。

【0058】

図 6 は、本実施の形態の TB-AP を適用したメッセージ交換のリンク行列を示している。

50

ここでもパラメータ  $\alpha = 4$  に設定しており、データポイント  $i$  を 1 から 10 の何れかに選択したときに、どのデータポイント  $j$  との間でメッセージ交換が行なわれるのかを塗り潰して示している。

#### 【0059】

通常APの特性として、メッセージ交換を行なったデータポイントのみが典型静止画の候補として最終的に選出される。これは、本実施の形態のTB-APでも同じことがいえる。本実施の形態のTB-APでは、メッセージを交換する範囲を時系列を考慮して制限しているため、結果としてメッセージ交換を行わない時系列的に遠いデータポイント（例えば、図3に示す時間  $t = 10$ ）の典型静止画も、クラスタリングにより代表フレームの一つとして選出できる。

#### 【0060】

また、TB-APにおける典型静止画の採用に関して、連続する類似フレームについては時間窓を越えても同じ典型静止画をとることができる。具体的には、いま第  $i$  フレームに対する典型静止画  $e_i$  を決定するとする。このとき、第  $i$  フレームより前にある第  $k$  フレームから第  $i - 1$  フレームまでの特徴ベクトルが、閾値などを用いて類似していると判定され、そして第  $k$  フレームに対する典型静止画  $e_k$  と第  $i$  フレームに対する典型静止画  $e_i$  も類似していると判定される場合、新たな典型静止画  $e_i$  を生成せずに典型静止画  $e_k$  をそのまま用いることができる。これによって、本発明の性能を向上させることができる。

#### 【0061】

以上のように、本実施の形態では上記TB-APに基づく処理を類似動画像検索処理装置2が実行することにより、時系列を考慮した典型静止画の集合を、クエリ動画像および対象動画像よりそれぞれ抽出することができる。そしてこの抽出された代表フレームの特徴量を、当該動画像の特徴量として決定する。特に、本実施の形態で採用するTB-APでは、選択したデータポイントと共にスライドする時間窓を設けているので、メッセージ交換の回数が制限され、メッセージ伝達の複雑さは  $O(n)$  となる。一方、通常APでの複雑さは  $O(n^2)$  である。つまり、APの計算コストはデータポイントの数の2乗に比例するため、データポイントの数が多くなると計算時間を要するといった問題点がある。それに対して、本実施の形態では一般的に  $< n$  であるため、TB-APはAPよりも計算コストを低く抑えることができる。

#### 【0062】

なお、類似度に関するデータの正規化と距離測度に関して、上記Similarity測度  $s(i, j)$  は、アルゴリズムが収束すればどのようなものを用いてもよい。パラメータ  $\alpha$  を適切に設定すれば、数3におけるSimilarity測度  $s(i, j)$  の形式は、アルゴリズムの収束をもたらす。

#### 【0063】

$\bar{D}$  の基準となる選択肢は次の通りである：

(a) 全データ距離の平均値：これは、データのサイズが小さい場合にのみ許される。

(b) 各々の特徴ベクトル  $x_t$  が各成分の総和が1になる非負成分だけを有するように正規化されている場合、 $\bar{D}$  は  $\frac{1}{2} \left( \frac{d-2}{2d} \right)$ 、または  $\left( 1 - \frac{1}{d} \right)$  のいずれかである：

ただし、 $d$  はシンプレックスに存在する特徴ベクトル  $x_t$  の次元を表わす。 $\frac{1}{2}$  はシンプレックスのエッジであり、 $\left( \frac{d-2}{2d} \right)$  は内球の半径であり、 $\left( 1 - \frac{1}{d} \right)$  は外球の半径を表す。なお、 $\approx$  は近似記号を意味し、 $d$  が十分大きい場合に当該近似が有効となる。

また、以上の基準の他にシステムの用途に応じた最適な値を実験的に求めることもできる。

#### 【0064】

例えば、前述のCSDではHSV（色相 - 彩度 - 明度）色空間表現を用いており、 $H, S, V$  の各値の比率を、 $H : S : V = 12 : 8 : 8$  に量子化するのが好ましいとされる。この場合、各画像は  $12 \times 8 \times 8 = 768$  色で表現され、CSDによる特徴量は768次元のヒストグ

10

20

30

40

50

ラムとなるので、 $d = 768$ となり、外球の半径を表す値を $\text{bar}(D)$ として用いる。

【0065】

次に、TB-APの代替手法として、動画像の経過時間の要素を考慮したクラスタリング手法について説明する。始めに改良したk-means法(k平均法)やharmonic competition法について、図7を参照しながら説明する。図7左側には、TB-APにおける時間軸上でスライドする時間窓Wが示されており、図7右側には、改良したk-means法やharmonic competition法における時間軸上でスライドする時間窓Wが示されている。

【0066】

TB-APは、時間軸に沿って重なり合うようにスライドする時間窓W中の一連の画像フレームを計算対象とする。これに対して、代替技術である改良したk-meansやharmonic competitionでは、重なりの無い時間窓Wの画像フレームを計算対象としている。つまりTB-APでは、時間窓Wのずらし幅が時間窓Wのフレーム長さよりも小さく、一乃至複数フレームずつ時間窓Wがスライディングするのに対し、改良したk-meansやharmonic competitionでは、時間窓Wのずらし幅と時間窓Wのフレーム長さが一致し、時間窓Wの単位で時間窓Wがスライディングしている。このような手法を時間拘束(Time-Bound)に対して時間区分(Time-Partition)と記載し、前記改良したk-meansを時間区分k平均法(Time-Partition k-means: TP k-means)と適宜記載する。

【0067】

改良したk-means(TP k-means)やharmonic competitionを用いた場合、時間軸に沿って画像フレームを時間窓Wに分割し、分割された時間窓Wの中でクラスタリングがなされる。ここで得られる代表点ベクトルは、元の画像そのものに対応するものではなく、k-meansやharmonic competitionに基づく計算により生成された類似画像(人工画像)の特徴ベクトルである。したがって、その代表点ベクトルと最も類似した特徴を持つフレームを、典型静止画として選出するという、二段階の計算手順が必要となる。この二段階の計算により得られた典型静止画は、TB-APの典型静止画と近似している。

【0068】

以下にTP k-meansの処理フローについて、図8を参照して説明する。始めに時間軸に沿って時間窓Wで動画像を分割し、時間窓Wで区切られた複数のフレームに区分けする。区分けされた時間窓W内でフレームをk等分し、それぞれのクラスターにフレームを割り振る。続いて、それぞれのクラスターに属するフレームの特徴ベクトルを基にクラスターの重心を求める。重心は特徴ベクトルの算術平均により求めることができる。各フレームについて、フレームの特徴ベクトルと、時間窓内全ての重心との間の距離を計算し、その結果を基に再クラスタリングする。再クラスタリング後、再度クラスターの重心を求め、時間窓内全ての重心と各フレームとの間の距離の再計算と再クラスタリングを、処理が収束するまで行なう。各クラスターにおいて、収束後の重心に最も近いフレームを典型静止画として選出する。この処理を、時間窓を移動させつつ行なう。

【0069】

次に、k-meansを用いた他のクラスタリング法である時間分割k平均法(Time-Split k-means: TS k-means)について、図9A~図9Eを参照して説明する。なお、これらの各図において、1~25の連続する数字は、時系列な動画像フレームの時間tに相当する。前記TP k-meansでは、あらかじめ決められたサイズの時間窓Wで動画像を区切り、その時間窓Wの中でk-meansに基づくクラスタリングを行なっている。これに対して、このTS k-meansは、動画像の全フレームに対してk-meansに基づくクラスタリングを行なう手法である。ただし、時系列を考慮したフレームの割り振りとクラスターの再分割を行なう点が、一般的なk-meansと異なる。

【0070】

TS k-meansでは以下のような処理が行なわれる。始めに、動画像の全フレームを時系列に基づいてk等分し、それぞれのクラスターにフレームを割り振る。続いて、それぞれのクラスターに属するフレームの特徴ベクトルを基にクラスターの重心を求める。重心は特徴ベクトルの算術平均により求めることができる。例として図9Aでは、全部で25個の動

10

20

30

40

50

画像フレームを時系列に5等分し( $k = 5$ )、曲線で囲まれたそれぞれのクラスターで重心1~重心5を算出している。次に各フレームについて、フレームの特徴ベクトルと、現在属するクラスターの重心、及び、時系列の近い前後のクラスターの重心との間の距離を計算し、その結果を基に再クラスタリングする。これは図9Bに示すように、例えば時間 $t = 18$ のフレームについて、現在属するクラスターの重心4と、時系列的に隣接する前後のクラスターの重心3及び重心5との間の距離をそれぞれ計算する。その結果、このフレームは重心3との間の距離が最も近いので、再クラスタリングでは図9Cに示すように、重心4ではなく重心3のクラスターに割り振られる。

#### 【0071】

次に、再クラスタリングされた各クラスター内で時系列の不連続性を判断し、不連続がある場合にはクラスターを分割する。これは図9Dに示すように、例えば重心3と重心4のクラスターについて、これらのクラスターが時系列的に不連続なフレームの間で分割される。クラスターが分割された場合にはクラスターの数 $k$ より増えるが、再クラスタリングによりクラスターが消滅した場合には数が減ることもありうる。再クラスタリング及び分割後、再度クラスターの重心を求め、重心の属するクラスター及び、時系列が前後のクラスターに属する各フレームとの間の距離の再計算と再クラスタリングを、処理が収束するまで行なう。図9Eに示す例では、重心3と重心4のクラスターを分割した後に、7つのクラスターでそれぞれ重心1~重心7を算出し、時間 $t = 1 \sim 25$ の各フレームについて、上述した距離の計算と再クラスタリングを収束するまで繰り返す。そして、各クラスターにおいて収束後の重心に最も近いフレームを典型静止画として選出する。

#### 【0072】

続いて、改良したペアワイズ・ニアレスト・ネイバー法(pairwise nearest neighbor: PNN)を用いたクラスタリング手法について説明する。PNNはまず、与えられた全てのデータポイント間の距離を計算する。次に、距離の近いデータポイント同士から順に併合する。これを併合後のデータポイントが全て一定の距離以上になるまで繰り返していく。これにより、類似度の高いデータポイント群を一つの集合としてまとめたクラスター群を求めるアルゴリズムである。通常のパNNについては「W. Equitz, "A new vector quantization clustering algorithm," IEEE Trans ASSP, vol. 37, pp. 1568-1575, 1989」などに記載されている。通常のパNNでは、ペアの特徴量の平均値がそのペアを表すものとして用いられるが、本実施の形態で説明する改良したPNNでは、後述するようにペアの一方を残し、他方を除去する点が異なる。また、時系列を考慮した時間窓 $W$ の設定のやり方に応じて、Time-Bound PNN(TB-PNN)とTime-Partition PNN(TP-PNN)とがある。

#### 【0073】

図10は、改良したPNNの併合を説明する図である。図中、 $x_i$  や  $x_j$  は各々のフレームの特徴量に対応したデータポイント $x_i$ を表し、 $\mu$  は全データポイントの重心を表す。PNNでは、与えられた全データポイント間の距離を計算した後、それらのデータポイントの組み合わせの中で、最も距離の近いペアを見つける。図10に示す例では、 $a$ と $b$ のデータポイントが、最も距離の近いペアに相当する。次に、そのペアを併合する。本実施の形態に係る改良されたPNNにおける「併合」とは、全データポイントの重心に近いものをそのまま残し、遠いものを削除することをいう。図10に示す例では、 $b$ のデータポイントを削除して、 $a$ のデータポイントを残す。なお、通常のパNNでは、 $a$ のデータポイントと $b$ のデータポイントの重心(中点)を計算し、それを $a$ と $b$ の代表値としている。その後、残ったデータポイント間の距離が一定値( $\epsilon$ )以上になるまで繰り返し、残ったデータポイントに対応するフレームが典型静止画となる。それぞれの典型静止画は、その典型静止画に併合されたフレームを、それぞれのメンバーであるとして扱う。

#### 【0074】

図11は、上述のTP-PNNを説明する図である。TP k-meansと同様に、改良したTP-PNNでは、あらかじめ決められたサイズの時間窓 $W$ で動画像を区切り、その時間窓 $W$ の中でPNNに基づくクラスタリングを行なっている。このとき通常のパNNと異なる点は、まず時間窓 $W$ ごとに重心(各フレームの特徴量の算術平均)を求め、2つのデータポイント(フレ

10

20

30

40

50



ムの特徴量)を併合する際に、その重心に距離の近いデータポイントをそのまま残し、遠いデータポイントを削除することである。この処理を隣接するペアの距離が所定の値より大きくなるまで繰り返し、残ったデータポイントに対応するフレームを典型静止画とする。各典型静止画には、それまでに併合によって削除されたフレーム群が属する。TP-PNNでは、この処理を各時間窓(重複しない)に対して行なう。

#### 【0075】

これに対し、TB-PNNでは図12に示すように、TB-APと同様にスライドする時間窓の制限をかけ、各データポイントの時間窓の範囲のみとの距離計算を行なう。そして、時間窓により制限されたデータポイントの組み合わせのうち、距離の近いものから順にデータポイントの併合を行なう。ただし、併合を行なう際、TP-PNNと異なり、動画フレーム全体の重心に近いものをそのまま残し、遠いものを削除する。この処理を隣接するペアの距離が所定の値以上となるまで繰り返す。これもTP-PNNと同様に残ったデータポイントに対応するフレームを典型静止画とする。各典型静止画には、それまでに併合によって削除されたフレーム群が属する。

#### 【0076】

以上説明したように、動画像の経過時間の要素を考慮した時間的な制限を導入して1乃至複数の代表フレームを取得することで、通常のAPよりも検索精度を上げることができる。また、通常のk-means及びPNNの各手法に対しても同様の結果が期待できる。TB-APでは、時間窓の中の一連の画像フレームを計算対象として、そこから一段階で代表フレームを選出することが可能となり、計算手順を簡素化できる。他のTP-k-means、TS k-means、TB-PNN、TP-PNNの4手法は、TB-APに比べ計算量が少ないという利点がある。精度に関して、TB-APより若干劣る場合もあるが些少であり、いずれも高い精度を保つことが可能である。

#### 【0077】

上記した時系列を保つ代表フレームの選出に際して、時間窓の幅が小さすぎる場合や、動画像中でほとんど変化の無いシーンが続く場合などで、連続する類似のフレームにより必要以上に代表フレームが選出される可能性がある。このような場合には、時間順序が連続する代表フレーム同士の差を求め、その差が閾値以下となるような場合は、後続の代表フレームを消去し、グループをまとめることで、冗長な代表フレームを削除することができる。

#### 【0078】

・動画像間の解析による類似度の算出

次に、異なる典型静止画の集合を有する異なる動画像同士の比較を行なうときに用いる動画像間の解析による類似度の算出手法について説明する。コンテンツに基づいた動画像間の検索をするためには、クエリの動画像とターゲットの動画像の類似度を比較する必要がある。ここでは、以下に説明する2つの方法の何れかで行なう。本実施の形態では比較対象を動画像とするが、代表フレームのようなexemplarを持つ時系列データであれば、これに限るものではない。

#### 【0079】

(1) 大域アライメントを求めるM距離を適用した類似度の比較

本実施の形態では、類似度を比較するための尺度としてM距離(M-distance)と呼ぶ距離を用いる。大域アライメント用のM距離は、文字列の類似度を測るためのレーベンシュタイン距離(L-distance)や、生命情報学において大域アライメントを求めるためのNeedleman - Wunschアルゴリズムを一般化したものである。したがって、この大域アライメント用のM距離は、レーベンシュタイン距離や、Needleman - Wunschアルゴリズムのそれぞれを特例として含んでいる。

#### 【0080】

以下、大域アライメント用のM距離を利用したアルゴリズムについて、図13のフローチャートを参照しながら詳しく説明する。

#### 【0081】

S 3 1 0 : クエリ動画像と対象動画像の各典型静止画を用意

クエリ動画像に相当する動画像  $v_A$  に対して、典型静止画の集合  $\{e^A_i\}$  と、当該典型静止画が所有するフレーム数の集合  $\{E^A_i\}$  ( $i = 1, 2, \dots$ ) が与えられる。また、動画像  $v_A$  と比較される対象動画像に相当する動画像  $v_B$  に対しても、同様の集合  $\{e^B_j\}$ ,  $\{E^B_j\}$  ( $j = 1, 2, \dots$ ) が与えられる。ここでは、数 3 における Similarity 測度  $s(i, j)$  が、代表フレーム間類似度として選択される。

【 0 0 8 2 】

S 3 2 0 : テーブルを作成し、次の動的計画法の手順に従って経路を元に辿る。

動画像  $v_A$  の典型静止画の数を  $m$  とし、動画像  $v_B$  の典型静止画の数を  $n$  としたときに、 $(m + 1) \times (n + 1)$  の行列を大域アライメントのテーブルとして作成する。行列の各成分  $f(i, j)$  は、動的計画法により次の手順で算出される。 10

【 0 0 8 3 】

まず、S 3 2 1 では、設計パラメータとして設定したギャップペナルティ  $g$  が選択される。このギャップペナルティ  $g$  を導入することで、クエリ動画像の典型静止画の数  $m$  と、対象動画像の典型静止画の数  $n$  が異なっても、これらの時系列を考慮した典型静止画同士の類似性を、正しく判定できるようになる。

【 0 0 8 4 】

次の S 3 2 2 では、 $\{i = 0\}$  の行と、 $\{j = 0\}$  の列について、ギャップ  $g$  にフレームの総数を掛け合わせて、段階的に値が下降するように、テーブルを構成する行列の各成分を計算する。具体的には、 $\{i = 0\}$  の行は、次の数 8 に基づいて行列成分  $f(0, j)$  を算出する。 20

【 0 0 8 5 】

【 数 8 】

$$(0, -gE_1^B, -g\sum_{j=1}^2 E_j^B, -g\sum_{j=1}^3 E_j^B, \dots).$$

【 0 0 8 6 】

また、 $\{j = 0\}$  の列は、次の数 9 に基づいて行列成分  $f(i, 0)$  を算出する。

【 0 0 8 7 】

【 数 9 】

$$(0, -gE_1^A, -g\sum_{i=1}^2 E_i^A, -g\sum_{i=1}^3 E_i^A, \dots).$$

【 0 0 8 8 】

さらに S 3 2 3 では、 $(i, j) = (1, 1)$  の位置から開始し、 $\{i = 0\}$  の行と  $\{j = 0\}$  の列以外の残りの行列成分  $f(i, j)$  を、次の数 10 に基づいて算出する。

【 0 0 8 9 】

【 数 10 】

$$f(i, j) = \max$$

$$\{f(i-1, j) - gE_i^A, f(i-1, j-1) + r(i, j)s(i, j), f(i, j-1) - gE_j^B\}$$

【 0 0 9 0 】

その後、最大値を与える成分に向かって矢印を描く。この計算は、全ての行列成分が算出されるまで続けられる。ここで  $r(i, j)$  は、典型静止画の所有度を反映する重みを表し、その値は例えば、次の数 11 に基づいて算出される。

【 0 0 9 1 】

10

20

30

40

【数 1 1】

$$r(i, j) = (E_i^A + E_j^B) / 2$$

【0 0 9 2】

数 1 0 に示すように、ここでの行列成分  $f(i, j)$  は、行または列の数が大きくなる順に向けて逐次的に生成される。そして、 $i$  行  $j$  列の行列成分  $f(i, j)$  は、 $(i - 1)$  行  $j$  列を經由し  $i$  行  $j$  列に至る経路、 $i$  行  $(j - 1)$  列を經由し  $i$  行  $j$  列に至る経路、 $(i - 1)$  行  $(j - 1)$  列を經由し  $i$  行  $j$  列に至る経路のうち、コストの最も大きい経路が最適経路として選択され、そのときの数値が当該行列の成分として逐次的に決められていく。

10

【0 0 9 3】

S 3 3 0 : 大域アライメントの決定

テーブルを構成する全ての行列成分の算出が終了した後、行列  $f(i, j)$  の最終行最終列の成分から矢印を元に辿ることで、大域アライメントの経路を得ることができる。

【0 0 9 4】

S 3 4 0 : 類似度の算出

動画画像  $v_A$  と動画画像  $v_B$  の類似度  $u(A, B)$  を、次の数 1 2 に基づいて算出する。

【0 0 9 5】

【数 1 2】

20

$$u(A, B) = h(f_{\text{last}}) / w(\sum_i E_i^A, \sum_i E_i^B)$$

【0 0 9 6】

ただし  $h(x)$  は正の増加関数とし、 $f_{\text{last}}$  は当該行列の最終行最終列の数値である。また、 $w$  は平均化関数とする。当該平均化関数は例えば、 $\sum_i E_i^A$  と  $\sum_i E_i^B$  との算術平均とすることができる。

【0 0 9 7】

(2) 局所アライメントを求める M 距離を適用した類似度の比較

比較すべき動画画像が互いに大きく相違している場合、上述した大域アライメントは人間の感覚からずれたものとなる。このような場合には、局所アライメントが利用できる。局所アライメントは、配列が全体としては似ておらず、部分的類似を見つけない場合に有効である。以下に記述するアルゴリズムは、上記の大域的アライメントによる方法を、局所アライメントに適応させたもので、類似度の尺度として局所アライメント用の M 距離を用いる。この局所アライメント用の M 距離は、生命情報学における局所アライメントを求めるための Smith-Waterman アルゴリズムの一般化となっている。したがって、本実施の形態で適用する局所アライメント用の M 距離は、Smith-Waterman アルゴリズムを特例として含んでいる。

30

【0 0 9 8】

以下、局所アライメント用の M 距離を利用したアルゴリズムについて、図 8 のフローチャートを参照しながら詳しく説明する。ただし、既に説明した大域アライメント用の M 距離と重複する手順が多いため、違いを記述するに留める。

40

【0 0 9 9】

S 3 1 0 : クエリ動画画像と対象動画画像の各典型静止画を用意

これは、大域アライメント用の M 距離と同じ処理手順である。

【0 1 0 0】

S 3 2 0 : テーブルを作成し、次の動的計画法の手順に従って経路を元に辿る。

動画画像  $v_A$  の典型静止画の数を  $m$  とし、動画画像  $v_B$  の典型静止画の数を  $n$  としたときに、 $(m + 1) \times (n + 1)$  の行列を局所アライメントのテーブルとして作成する。行列の各成分は、動的計画法により次の手順で算出される。

50

【 0 1 0 1 】

先ず、S 3 2 1では、設計パラメータとして設定したギャップペナルティ  $g$  が選択される。これは、大域アライメント用の  $M$  距離と同じ処理手順である。

【 0 1 0 2 】

次の S 3 2 2では、 $\{i = 0\}$  の行と、 $\{j = 0\}$  の列について、各成分をすべて 0 に設定する。

【 0 1 0 3 】

さらに S 3 2 3では、 $(i, j) = (1, 1)$  の位置から開始し、 $\{i = 0\}$  の行と  $\{j = 0\}$  の列以外の残りの行列成分  $f(i, j)$  を、次の数 1 3に基づいて算出する。

【 0 1 0 4 】

【 数 1 3 】

10

$$f(i, j) = \max$$

$$\{0, f(i, -1, j) - gE_i^A, f(i - 1, j - 1) + r(i, j)s(i, j), f(i, j - 1) - gE_j^B\}$$

【 0 1 0 5 】

その後、0 ではない最大値を与える成分に向かって矢印を描く。数 1 3の右辺では数 1 0と異なり、max 記号の中の最左成分を 0 としている。これにより、スコア値が負になった場合に値を 0 にリセットし、アライメントを新しく始めることができる。

【 0 1 0 6 】

S 3 3 0：局所アライメントの決定

テーブルを構成する全ての行列成分の算出が終了した後、行列  $f(i, j)$  の中の最大値  $f_{\max}$  の成分から矢印を元に辿ることで、局所アライメントの経路を得ることができる。

20

【 0 1 0 7 】

S 3 4 0：類似度の算出

大域アライメント用の  $M$  距離と同じ処理手順を用い、動画像  $v_A$  と動画像  $v_B$  の類似度  $u(A, B)$  を、数 1 2に基づいて算出する。

【 0 1 0 8 】

以上述べたように、フレームに対する特徴量の抽出、時系列を保つ代表フレームの選出、及び、動画像間の解析による類似度の算出の手順を、類似動画像検索処理装置 2 が処理実行して、動画像  $v_A$  に対する個々の動画像  $v_B$  の類似度  $u(A, B)$  を算出する。これにより、クエリ動画像に類似する類似動画像を対象動画像の中から検索することができる。

30

【 0 1 0 9 】

次に、上記動画像検索装置を用いた実験の詳細と結果を述べる。

【 0 1 1 0 】

先ず、実験データについて、検索のエンドユーザーは人間であるので、類似度の最終判断は人間のもつ感性に強く依存する。したがって、類似度の判断は、題目にできるだけ依存しないことが望ましい。しかし、単なるセットでは、プレーンな機器でも判断するのが容易になり過ぎるので、実験では次の条件を満足するデータセットを用意した。

40

【 0 1 1 1 】

- ・元の動画像データは、ラベル付けされていないものとする。
- ・検索における適合率および再現率を、自動的に判断できるものとする。
- ・図 3 の右下に示すような時間に依存する性質を判別できるように、各動画像は、シーンの時間変化を捉えた情報を有するものとする。

【 0 1 1 2 】

図 1 4 は、元の動画像データの生成過程を表したものである。グループ A の動画像は「鳥瞰図」を表し、グループ B の動画像は「鳥」を表す。実験において、これらのグループ A, B より 20 の動画クラスが作られる（図 1 5 を参照）。各クラス 1 ~ 2 0 はそれぞれ 21

50

個の動画像を構成するので、互いに異なる合計で420個のテスト動画を使用した。

#### 【0113】

##### ・事前実験

APアルゴリズムの収束はデータに依存するので、数5と数6におけるパラメータの範囲を確認する必要がある。試行データの集合として、5集団のガウス型クラスターを生成し、さらに、それぞれの時間方向について2集団に分け、10集団とした。図16は、これらのデータを図示したものである。

#### 【0114】

図16に示す試行データの集合に関して、本実施の形態で提案するTB-APと、通常のAPに対して実験を行なった。主な目的は収束のためのパラメータの範囲と、時間窓パラメータの効果とを確認することにある。繰り返しの実験により、パラメータ  $[0.3, 1.0]$  の選択はデータの特徴に依存する。ただし  $\omega$  は正の微少量であり、図16の実験では  $\omega = 0.02$  とした。

10

#### 【0115】

表1は、収束の傾向と計算時間を示したものである。収束は、初期値で正規化された  $a(i, k)$  および  $r(i, k)$  の総和により、それが  $10^{-5}$  未満であるか否かによりチェックされた。表1では、時間窓  $W$  のパラメータ  $\omega$  を変化させたときの典型静止画の数「exemplars」、反復回数「iterations」、メッセージ伝達数「message passing」をそれぞれ示している。表1より、例えば  $\omega = 10$  のとき、本実施の形態のTB-APは通常のAPのメッセージ伝達と比べて6倍高速であることがわかる。

20

#### 【0116】

##### 【表1】

$\omega$	exemplars	iterations	message passing
1(no AP)	100	428	42800
5	20	1528	1344640
10	14	968	1752080
20	14	937	3298240
50	11	988	7360600
100 = $n$	11	1063	10630000

30

#### 【0117】

$\omega = 1$  の場合は、全フレームが典型静止画であることを意味する。 $\omega = n = 100$  は、時間依存性のある典型静止画抽出を含まない通常のAPの場合に相当する。これは収束傾向を見るため比較例として求めた。 $\omega = 5, 10, 20, 50$  とした場合を見ると、メッセージ伝達数に単調増加が認められ、これは時間窓のサイズに起因するためCPUタイムに直接関わる。しかし、通常のAPのメッセージ伝達数に対して、TB-APのメッセージ伝達数の比率は小さいまま保たれているので、これらは望ましい範囲にあるといえる。

#### 【0118】

準備した図15の動画像セット(クラス1~20)に対して、信頼できるフレームを伴う典型静止画を見つけ出すのにTB-APを適用した。これは、各動画像に「数値的な注釈」を付与するのに重要な手順であり、図15に示す整理されていない動画像の集合を、数値的なタグという観点から構造化することと等価である。こうしたタグは、オンラインとオフラインの何れかで算出できる。

40

#### 【0119】

こうして、TB-APにより得られた、比較対象とする二つのクラスの動画像(動画像  $v_A$  と動画像  $v_B$ ) の各典型静止画について、それらのSimilarity測度  $s(i, j)$  から得られる距離行列の一例を図17に示す。同図において、動画像  $v_A$  は4つの典型静止画を有し、動画像  $v_B$  は3つの典型静止画を有するので、当該行列は  $3 \times 4$  行列である。表中の数字は、動画像  $v_A$  と動画像  $v_B$  との典型静止画(代表フレーム)間の距離行列の各成分を表し、例えば動画像  $v_A$  の代表フレーム2と、動画像  $v_B$  の代表フレーム3との距離は

50

0.26となる。当該距離は各代表フレーム間の類似度に対応して数値化したもので、ここでは距離が小さいほど代表フレーム同士が類似したものとなり、反対に距離が大きいほど代表フレーム同士が互いに相違したものとなる。

【0120】

図18は、上述した大域アライメント用のM距離を適用して作成したテーブルの一例である。この大域アライメント用のテーブルを作成する上で、アルゴリズムのパラメータは、 $\text{bar}(D) = 1 / (1 - (1 / d))$ 、 $d = 3 \times 256 = 768$ 、 $g = 0.05$ にそれぞれ設定した。同図において、テーブルの右下に記載される数字26.60が最終成分の数値 $f_{last}$ であり、そこから元に辿る最適経路を大域アライメントとして網掛けで示している。その後、動画像 $v_A$ と動画像 $v_B$ の類似度 $u(A, B)$ を求めるために、算出された数値 $f_{last}$ は平均化関数で正規化される。例えば、ギャップペナルティー $g$ とフレーム数の集合との積の総和で正規化する場合、図18において、その総和は、 $0.05 \times (8 + 12 + 10 + 7) + 0.05 \times (12 + 7 + 10) = 1.85 + 1.45 = 3.30$ であるので、正規化された類似度 $u(A, B)$ は $26.60 / (1.85 + 1.45) = 8.06$ と算出される。また適合パターンは、当該行列の最終行最終列より矢印を辿って探すことができる。こうして、動画像 $v_A$ と動画像 $v_B$ との間で、大域アライメント用の重み付きM距離による適切な類似度 $u(A, B)$ の算出が可能となる。

10

【0121】

図19は、上述した局所アライメント用のM距離を適用して作成したテーブルの一例である。アルゴリズムのパラメータは、 $\text{bar}(D) = (1 - (1 / d))$ 、 $d = 3 \times 256 = 768$ 、 $g = 0.05$ と設定した。局所アライメント用のテーブルでは、大域アライメント用のテーブルでスコアがマイナスになる成分、すなわち数8や数9で算出した成分が、すべて0にリセットされることが確認できる。ここでは最大値 $f_{max}$ が26.95であり、その最大値 $f_{max}$ の成分から元に辿る最適経路を、局所アライメントとして網掛けで示している。こうして、比較する動画像が大きく異なるものであっても、動画像 $v_A$ と動画像 $v_B$ との間で、局所アライメント用の重み付きM距離による適切な類似度 $u(A, B)$ の算出が可能となる。

20

【0122】

なお、大域アライメントにおける好ましい変形例として、上述したM距離による類似度の算出は、その計算量を削減するために、図20に示すように、シーケンス全体の割合に対して任意の幅 $r$ だけを計算に用いることもできる。これは、M距離をコンテンツベースの動画像検索の類似度算出に特化させることで高速化した算出手法である。図20において、黒く塗りつぶされた部分は、計算を省略した個所を示す。またこの図は、大域アライメントによる図18のテーブルに対応する。この場合、計算の範囲は以下の数14によって定められる。

30

【0123】

【数14】

$$\begin{cases} 0 \leq i \leq n \\ 0 \leq j \leq m \\ j - (n \times r) < i < j - (n \times r) \end{cases}$$

40

【0124】

上記数14において、 $m$ は動画像 $v_A$ の代表フレーム数であり、 $n$ は動画像 $v_B$ の代表フレーム数である。幅 $r$ は、類似度比較における計算範囲の絞り具合を表すパラメータであって、計算範囲を数14のように絞ることで、通常の全ての行列成分を算出する計算と比べて、 $100 \times (1 - r)^2$  %の計算量削減が可能となる。図20の例では、計算範囲を絞ったにもかかわらず、計算範囲を絞っていない図18の計算結果と同じ結果を算出していることが確認できる。

【0125】

50

次に、前述した時間窓のパラメータと、類似度  $u(A, B)$  を算出した後の再現率や適合率との関係について説明する。ここでは、全ての動画画像が類似度の高い順に分類された後、再現率と適合率が次の数 15 のように定義される。

【0126】

【数 15】

$$\text{recall} = |\text{correct videos found}| / N_{\text{same\_class}}$$

$$\text{precision} = |\text{correct videos found}| / |\text{top rank videos to be checked}|$$

【0127】

10

再現率 (recall) : = 検索によって見つかった正解動画画像の数 / 同じクラスに属する動画画像数

適合率 (precision) : = 検索によって見つかった正解動画画像の数 / 検索された上位動画画像数

【0128】

再現率とは、検索要求を満たす全動画画像に対しての、検索要求を満たす検索結果の割合を意味し、検索の網羅性を表す指標である。一方、適合率とは、全検索結果に対しての、検索要求を満たす検索結果の割合を表し、検索の正確さを表す指標である。一般に再現率や適合率が高いほど検索の性能が高いとみなされる。

【0129】

20

図 21 は、実験で得られた適合率 - 再現率曲線の結果を示している。同図において、縦軸は適合率を表し、横軸は再現率を表す。曲線がグラフ中右上に行くほど、検索性能が高い。ここでは、TB-AP を適用した時間窓パラメータが「 $\tau = 5$ 」と「 $\tau = 10$ 」の各実験を載せている。図中枠内に記載されるように、数 12 で示した関数  $h(x)$  を定義し、そこで使用する係数  $a$  を  $a = 1$  に設定した。また比較として、通常の AP を適用した「Plain AP」の実験結果も併せて掲載した。

【0130】

これらの実験結果から、再現率として 20% までを要求する関心領域 (図中「ROI」で囲んだ領域) において、本実施の形態の TB-AP による適合率は非常に高く、処理能力は満足のいくものであることが読み取れる。一方、通常の AP は時間軸を考慮していないため、本発明の TB-AP と比べてその性能は劣ることが見てとれる。

30

【0131】

次に、代表フレームを選出する際に用いたクラスタリング手法の違いによる検索結果の違いについて示す。図 22 A および図 22 B は、前述した 5 つのクラスタリング手法における適合率 - 再現率曲線の実験結果を示している。

【0132】

実験データとして、互いに異なる 2100 個の動画画像を用意した。次にこれらを 21 個ずつに分け、合計 100 個の動画セットにする。各セットにおいて、21 個の動画画像のうちの一つをクエリとし、このクエリの 10% 以上のフレームを残りの 20 個の動画画像の中にそれぞれ一様乱数によって得られた位置に挿入した。これによりクエリを除く 2000 個のテスト動画を被検索対象群とした。

40

【0133】

上記の実験データに対して、フレームに対する特徴量の抽出、時系列を保つ代表フレームの選出、及び動画画像間の解析による類似度の算出を行ない、クエリに関連する動画画像の検索を行なった。特徴量は CSD を用い、類似度は局所アライメントを基に求めた。

【0134】

図 22 A は、クラスタリング時に冗長代表フレームの削除の事後処理を行なうことなく、そのまま類似度の算出を行なった場合の 5 つのクラスタリング手法における適合率 - 再現率曲線を表す。5 つの手法とも概ね良好な結果が得られており、差はほとんど見られない。

50

## 【 0 1 3 5 】

図 2 2 B は、クラスタリング時に冗長代表フレームの削除の事後処理を行なったときの 5 つのクラスタリング手法における適合率 - 再現率曲線を表す。このとき閾値 = 0.05 としている。代表フレームの数が概ね半分に減ることで、事後処理の無いものに比べれば適合率が下がっているが差は小さい。

## 【 0 1 3 6 】

上記 5 つのクラスタリング手法では、TB-AP に比べて他の TP k-means、TS k-means、TP-PNN、及び TB-PNN は計算量が少ないため、処理速度を大幅に早くできる。また、局所最適化を避けるためには、初期値の設定で工夫する必要があるが、TB-AP、TP-PNN、及び TB-PNN については、そのアルゴリズムから不適切な局所最適化は生じない。

10

## 【 0 1 3 7 】

次に、別の実験例として、前記実験例において用いた CSD のかわりに、Video Signature に係る特徴量を本発明に適用した実験例を以下に示す。Video Signature は ISO/IEC 標準の中で仕様が公開されているものであり (Multimedia content description interface, International Standard of ISO/IEC 15938-3, Amendment 4 (2010))、コンテンツの同一性を判断するための一意に識別可能な特徴量を規定している。コンテンツの複製 (コピー) を高精度に検知することができるため、インターネット上の不正コピーや不正流通の検知に用いることができるものと期待されている。

## 【 0 1 3 8 】

Video Signature の規格では、特徴量のデータは、フレーム単位の特徴を記述する Frame Signature、Frame Confidence 及び Word、並びに連続する複数フレームから構成される区間単位の特徴を記述する BagOfWords から構成されている。本実験例では、この中の Frame Signature (フレームシグネチャ) に着目して計算を行なっている。なお、これは Frame Signature に限定するという意味ではなく、Video Signature の規格にある他の特徴量を用いたり、また、先に述べたように複数の特徴量を合成特徴量ベクトルとして利用することもできる。

20

## 【 0 1 3 9 】

Frame Signature は、特徴量として、映像の各フレームを部分領域に分割し、部分領域間の輝度の差分を三値化し、380 次元のベクトルとして表わすものである。このとき、フレーム内の中央領域の方を周辺領域よりも密に部分領域としてサンプリングするようにしている。本実験例では、各フレームを  $32 \times 32$  の部分領域に分割し、YCbCr の色彩空間の中の Y の輝度の平均値を各部分領域の輝度情報として用いた。

30

## 【 0 1 4 0 】

先のクラスタリングの実験と同様に、実験データとして、まず互いに異なる 2100 個の動画画像を用意した。次にこれらを 21 個ずつに分け、合計 100 個の動画セットにする。各セットにおいて、21 個の動画画像のうちの一つをクエリとし、このクエリの 10% 以上のフレームを残りの 20 個の動画画像の中にそれぞれ一様乱数によって得られた位置に挿入した。これによりクエリを除く 2000 個のテスト動画を被検索対象群とした。また、単純にクエリの一部を挿入するのではなく、挿入する動画画像に以下のような 5 つの処理を施して検索しにくくしている。

40

## 1) フレームレートを増加 (フレームの削除) (frame)

挿入する動画画像の 2 または 3 または 6 フレーム毎 (一様乱数により決定) に 1 フレームを削除することにより、フレームレートの増加を施している。

## 2) グレースケール化 (gray)

挿入する動画画像の各フレーム画像に対して、

$$Y = 0.11448 \times R + 0.58661 \times G + 0.29891 \times B$$

の式によりグレースケール変換を施している。ここで、R、G、B (赤、緑、青) は各フレーム画像を表現する色彩空間を表し、値はそれぞれ 0 ~ 255 の整数値を取る。

## 3) 輝度を低下 (br)

挿入する動画画像の各フレーム画像の R、G、B の値に輝度の低下率を表す変数 P を乗ず

50



る。Pの値は0.6～0.9の中で一様乱数を取る。

#### 4) サイズ変化 (縮小及び拡大) (resize)

挿入する動画像の各フレーム画像サイズを、0.5～2.0倍の範囲から一様乱数を取り変更を施している。

#### 5) JPEG (Joint Photographic Experts Group) 変換 (jpeg)

JPEG変換は、画像の周波数成分のうち一部の情報を削減することで圧縮を施している。

実験ではこの圧縮率を高めることで、挿入する動画像の品質を低下させた。本実験中では画像処理ライブラリのOpenCVを使用しており、この中でJPEGの品質を決定する変数CV\_IMWRITE\_JPEG\_QUALITYに対して、20～80の中から一様乱数を取って設定することで、圧縮を行なった。この変数CV\_IMWRITE\_JPEG\_QUALITYは0～100の値を取り、100のとき画像の品質が最高となる。

10

#### 【0141】

上記の実験データに対して、フレームに対する特徴量の抽出、時系列を保つ代表フレームの選出、及び、動画像間の解析による類似度の算出を行ない、クエリに関連する動画像の検索を行なった。また、比較例として、特徴量としてCSDを用いた実験も行なった。

#### 【0142】

図23Aおよび図23Bは、実験で得られた適合率 - 再現率曲線の結果を示している。図23AはCSDを用いた場合、図23BはFrame Signatureを用いた場合である。典型静止画の取得方法はTP-PNNを用い、各パラメータの値はギャップペナルティ  $g = 0.2$ 、CSDを用いた場合  $\gamma = 0.1$ 、 $\text{bar}(D) = 0.05$ 、距離計算はユークリッド距離、Frame Signatureを用いた場合  $\gamma = 320$ 、 $\text{bar}(D) = 15$ 、距離計算は絶対値距離を用いた。このとき  $\gamma$  はTP-PNNで併合する条件の閾値である。各パラメータについては、適正な比較のため、圧縮率 (代表フレームの取得数) が同等となるように  $\gamma$  を設定し、また、それぞれの方法で結果が最良となるように  $\text{bar}(D)$  の値を選択している。

20

#### 【0143】

図23Aに示すように、輝度が落ちている場合は、CSDによる検索は難しくなる。グレースケール化された場合にはほとんど検索できない。また、JPEG変換によっても精度が低下した。ただし、フレームレートの変換とサイズ変換に関しては高い精度を保っていた。一方、図23Bに示すように、Video Frame Signatureを用いた場合、CSDに比べ実験を行なったすべての変換に対して精度低下の差異が少なくなることが確認できた。

30

#### 【0144】

以上述べたように、Frame Signatureを採用することで、CSDを用いた場合と比較して、輝度変化などの変更に強いという利点がある。また、特徴ベクトルのデータ量がCSDと比べて小さいため、比較時の距離計算を高速に行なうことが可能となる利点がある。

#### 【0145】

以上、本実施の形態における動画像検索装置のシステム構成とアルゴリズムを提示したが、動画像は大きな容量サイズであり、各典型静止画とそれらの責任のあるフレームの集合は、通常オフラインで算出される。そのため、大きなデータ集合の構造情報に対する数値ラベルとして、こうした集合を利用できる。また本実施の形態では、各々の典型静止画が責任のあるフレームを含んでいることを考慮して、レーベンシュタイン距離を含む新規な類似度の算出方法を、M距離として提案している。このM距離は、選出した典型静止画に対して、大域アライメントと局所アライメントの両方で利用が可能である。

40

#### 【0146】

また、上述した5つのクラスタリング手法による検索結果の違いを調べた実験例では、クエリ動画像と対象動画像に対して同じクラスタリング手法を用いて代表フレームを選出している。このようにすることで、同一の規範で代表フレームを選出することができ、類似度比較の際に、異なる規範を用いて代表フレームを選出することによる不整合の発生を防ぐことができる。ただし、必ずしも同じクラスタリング手法をとる必要があるとは限らず、検索する動画像のデータサイズやクラスタリングの処理速度、検索精度等を考慮して、クエリ動画像と対象動画像とで異なるクラスタリング手法を用いて代表フレームを選出

50

しても構わない。

【0147】

なお、上述した処理装置2による動画像検索処理は、FPGA (field-programmable gate array) 等を用いた電子回路により実行することができるが、その機能の全部又は一部をコンピュータにより実行することができる。図24は、処理装置2としてコンピュータ100を用いる場合のハードウェア構成の一例である。コンピュータ100には、各種制御を行なうCPU (Central Processing Unit) 102、主記憶装置103、補助記憶装置104、通信インターフェース105、及びドライブ装置106などがバス101を介して相互に接続されている。

【0148】

CPU 102は、補助記憶装置104に格納されているプログラムを読み出し、その指示に従って本発明に係る動画検索処理を実行することができる。主記憶装置103はランダムアクセスメモリ等からなる記憶装置であり、補助記憶装置104に格納されているプログラムやデータ等の一部を一時的に記憶し、また、CPU 102による処理結果を一時的に記憶するものである。補助記憶装置104は、例えばハードディスクドライブやSSD (Solid State Drive) 等からなる記憶装置であり、本発明に係る動画検索処理を実行するためのプログラムや処理に用いるデータを記憶する。通信インターフェース105は、ネットワークを介して外部のサーバや外部の記憶装置と通信を行なうためのものである。コンピュータ100は通信インターフェース105を介して、外部のサーバからプログラムをダウンロードしたり、外部の記憶装置に対してデータの読み出しや書き込みを行なったりすることができる。ドライブ装置106は、記録媒体107に対してデータの読み出しや書き込みを行うために用いるものである。記録媒体107はプログラムやデータを記録するものであり、CD (Compact Disk) やDVD (Digital Versatile Disk) 等を用いたROM (Read only Memory) 等がある。また、記録媒体107としては、その他USB (Universal Serial Bus) メモリ等の電子デバイスメモリを用いることも可能である。なお、これらのハードウェアは一例として示したものであり、必要に応じて取捨選択可能であり、また、他の要素を加えても構わない。

【0149】

以上を纏めると、本実施の形態では、クエリ動画像に対して検索対象となる対象動画像の類似度を算出して、クエリ動画像に類似する動画像を検索する動画像検索装置のシステムを用いた動画像検索方法を提案する。

【0150】

特にこの方法では、処理装置2に組み込んだ特徴量抽出部22が、クエリ動画像のフレームから、フレームごとに第1の特徴量を抽出し、対象動画像のフレームから、フレームごとに第2の特徴量を抽出する。次に、処理装置2に組み込んだ動画像代表フレーム選出部23が、クエリ動画像について、特徴量抽出部22で抽出された第1の特徴量に対して、TP k-means、TS k-means、TB-PNN、又はTP-PNNの何れかを用いて、時系列で第1の代表フレームを取得し、対象動画像について、特徴量抽出部22で抽出された第2の特徴量に対して、TP k-means、TS k-means、TB-PNN、又はTP-PNNの何れかを用いて、時系列で第2の代表フレームを取得する。そして、処理装置2に組み込まれた動画像検索部としての類似度算出部24や類似動画像表示制御部25が、クエリ動画像における第1の代表フレームと、対象動画像における第2の代表フレームとの間の代表フレーム間類似度から、クエリ動画像と対象動画像との間の類似度を求めて、クエリ動画像に類似する動画像を検索する。

【0151】

このような方法を採用すると、クエリ動画像と対象動画像に対して、動画像の経過時間の要素を考慮した時間的な制限を導入して、1乃至複数の第1の代表フレームと第2の代表フレームをそれぞれ取得することが可能になる。そして、この時系列の第1の代表フレームと第2の代表フレームを利用して、クエリ動画像と対象動画像との間で類似性を判定することで、対象動画像の中からクエリ動画像に類似する動画像を、通常のAPよりも精度

を上げて検索できる。また、k-means、及びPNNの各手法においても同様の結果が期待できる。また、TP k-means、TS k-means、TB-PNN、TP-PNNの各手法は、TB-APに比べ計算量が少なく、結果的に動画画像そのものをクエリとする検索やグループ化を、計算コストを低く抑えて高い精度で容易に行なうことが可能になる。

#### 【0152】

上記の方法では、特徴量抽出部22が第1の特徴量及び第2の特徴量として、Video Signatureに係る特徴量として規定されるFrame Signatureを用いるのが好ましい。これにより、CSDを用いた場合と比較して、輝度変化などの変更があっても高い検索精度を維持できる。また、データ量がCSDと比べて小さいため、比較時の距離計算を高速に行なうことができる。

10

#### 【0153】

上記クエリ動画画像に類似する動画画像の検索では、クエリ画像におけるi番目の第1の代表フレームが表わすフレーム数 $E^A_i$ と、対象動画画像におけるj番目の第2の代表フレームが表わすフレーム数 $E^B_j$ を取得し、クエリ画像におけるi番目の第1の代表フレームと対象動画画像におけるj番目の第2の代表フレームとの間のSimilarity測度 $s(i, j)$ を、前記代表フレーム間類似度として取得する。

#### 【0154】

そして、ギャップペナルティ $g$ を設定した上で、クエリ動画画像における第1の代表フレームの数が $m$ であり、対象動画画像における第2の代表フレームの数が $n$ であるときに、 $(m+1) \times (n+1)$ の行列の各成分 $f(i, j)$ について、 $i=0$ の行の各成分 $f(0, j)$ を数8に基づいて算出し、 $j=0$ の列の各成分 $f(i, 0)$ を、数9に基づいて算出する。さらに、フレーム数 $E^A_i$ 、 $E^B_j$ に応じて算出される重みを $r(i, j)$ としたときに、 $i=0$ の行と $j=0$ の列以外の残りの行列成分 $f(i, j)$ を、数10に基づいて算出する。こうして、行列の最終行最終列の数値 $f_{last}$ を算出したら、その数値 $f_{last}$ を利用してクエリ動画画像と対象動画画像との間の類似度 $u(A, B)$ を、例えば数11で算出し、クエリ動画画像に類似する動画画像を検索する。

20

#### 【0155】

このような方法を採用すれば、クエリ動画画像と対象動画画像との間で第1の代表フレームと第2の代表フレームの数が異なっても、ギャップペナルティ $g$ を導入することで、双方の代表フレーム同士の類似性とその時系列の類似性を、正しく判定して検索を行なうことができる。また、こうして得られた行列 $f(i, j)$ の各成分から、大域アライメントの経路を得ることができる。しかも、ここで得られる動画画像データに適合した類似度の算出は、レーベンシュタイン距離や、Needleman - Wunschアルゴリズムのそれぞれを単なる特例として含んでおり、生命情報配列の類似度を測る手法にも適用できる。

30

#### 【0156】

また別な手法として、上記クエリ動画画像に類似する動画画像の検索では、フレーム数 $E^A_i$ 、 $E^B_j$ と、Similarity測度 $s(i, j)$ をそれぞれ取得し、ギャップペナルティ $g$ を設定した上で、 $(m+1) \times (n+1)$ の行列の各成分 $f(i, j)$ について、 $i=0$ の行の各成分 $f(0, j)$ と、 $j=0$ の列の各成分 $f(i, 0)$ を、全て0に設定する。そして、フレーム数 $E^A_i$ 、 $E^B_j$ に応じて算出される重みを $r(i, j)$ としたときに、 $i=0$ の行と $j=0$ の列以外の残りの行列成分 $f(i, j)$ を、数13に基づいて算出する。こうして、行列の各要素の中で最大値 $f_{max}$ を算出したら、その数値 $f_{max}$ を利用してクエリ動画画像と対象動画画像との間の類似度 $u(A, B)$ を、例えば数12で $f_{last}$ を $f_{max}$ に代えて算出し、クエリ動画画像に類似する動画画像を検索してもよい。

40

#### 【0157】

この場合も、クエリ動画画像と対象動画画像との間で第1の代表フレームと第2の代表フレームの数が異なっても、ギャップペナルティ $g$ を導入することで、双方の代表フレーム同士の類似性とその時系列の類似性を、正しく判定して検索を行なうことができる。また、こうして得られた行列 $f(i, j)$ の各成分から、局所アライメントの経路を得ることができる。しかも、ここで得られる動画画像データに適合した類似度の算出は、Smith-

50

Watermanアルゴリズムを単なる特例として含んでおり、生命情報配列の類似度を測る手法にも適用できる。

【 0 1 5 8 】

そして、上述した作用効果は、上記の処理を実行することができる動画像検索装置や、上記の処理をコンピュータに実行させるための動画像検索プログラムにおいても同様に実現する。

【 0 1 5 9 】

以上、本発明の実施の形態について説明したが、本発明はこうした実施の形態に何等限定されるものではなく、本発明の要旨を逸脱しない範囲内において、種々なる形態で実施し得ることは勿論である。例えば上述した各パラメータ値の設定は、入力装置 4 を介して適宜変更することも可能である。

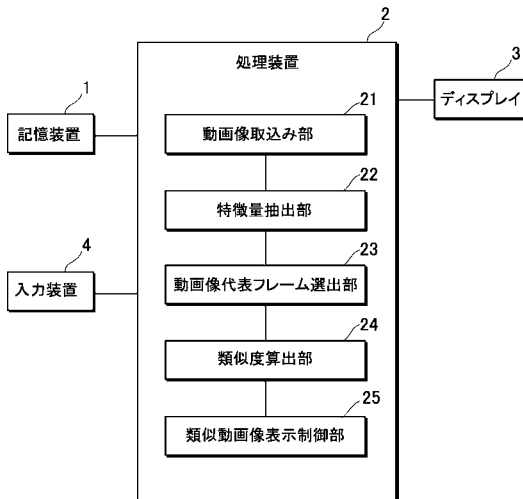
10

【 符号の説明 】

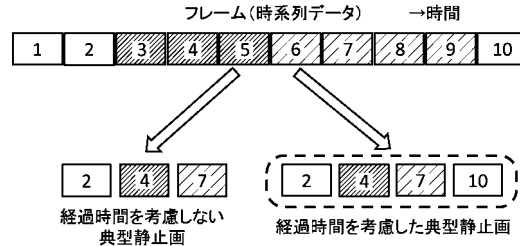
【 0 1 6 0 】

- 2 2 特徴量抽出部
- 2 3 動画像代表フレーム選出部
- 2 4 類似度算出部（動画像検索部）
- 2 5 類似動画像表示制御部（動画像検索部）

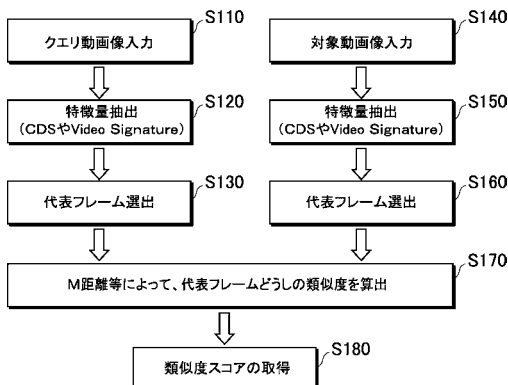
【 図 1 】



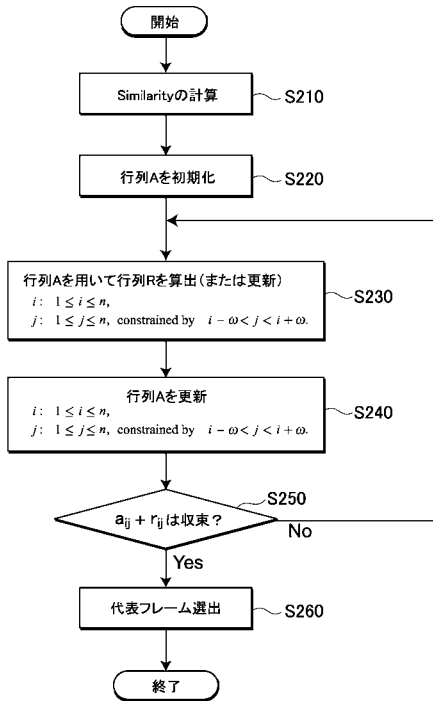
【 図 3 】



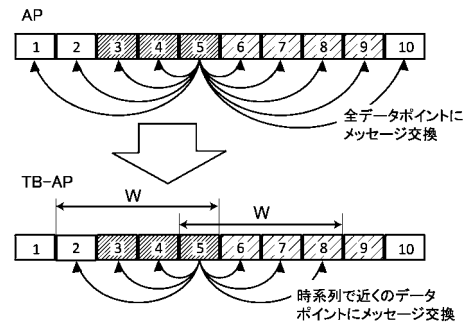
【 図 2 】



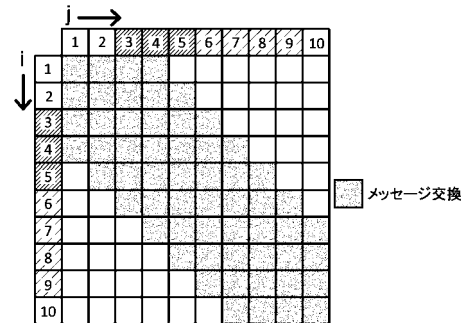
【図 4】



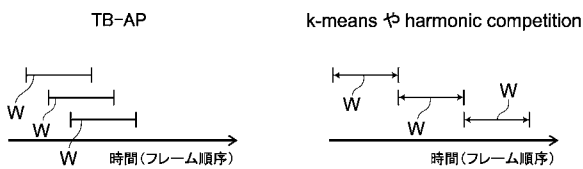
【図 5】



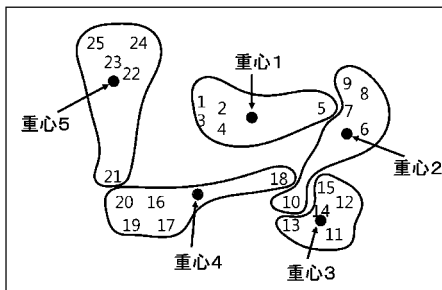
【図 6】



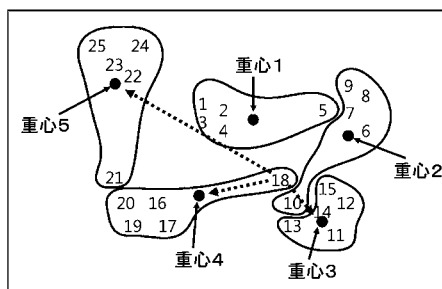
【図 7】



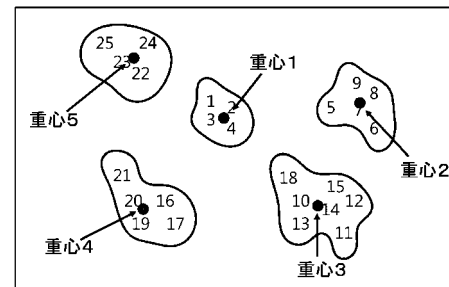
【図 9 A】



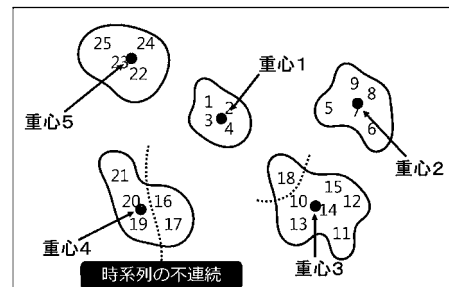
【図 9 B】



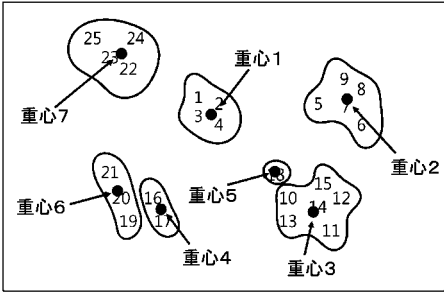
【図 9 C】



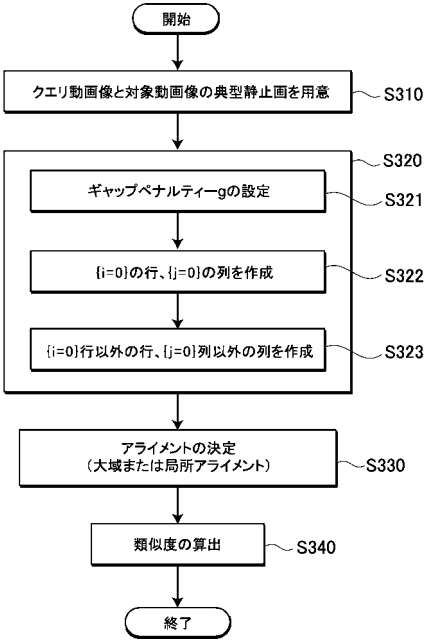
【図 9 D】



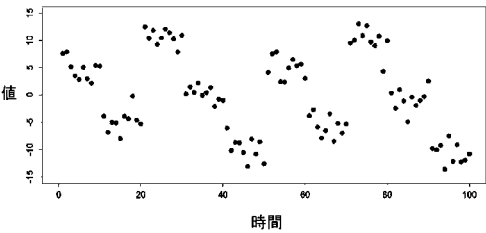
【図 9 E】



【図 1 3】



【図 1 6】



【図 1 9】

		動画像V <sub>B</sub>			
↓ i	→ j	典型静止画1 E <sup>B</sup> <sub>1</sub> = 8	典型静止画2 E <sup>B</sup> <sub>2</sub> = 12	典型静止画3 E <sup>B</sup> <sub>3</sub> = 10	典型静止画4 E <sup>B</sup> <sub>4</sub> = 7
		0	0	0	0
典型静止画1 E <sup>A</sup> <sub>1</sub> = 12	↑	0	9.54	10.53	10.03
典型静止画2 E <sup>A</sup> <sub>2</sub> = 7	↑	0	9.19	17.81	18.28
典型静止画3 E <sup>A</sup> <sub>3</sub> = 10	↑	0	8.83	18.68	26.95

【図 2 0】

		動画像V <sub>B</sub>			
↓ i	→ j	典型静止画1 E <sup>B</sup> <sub>1</sub> = 8	典型静止画2 E <sup>B</sup> <sub>2</sub> = 12	典型静止画3 E <sup>B</sup> <sub>3</sub> = 10	典型静止画4 E <sup>B</sup> <sub>4</sub> = 7
		0.0	-0.4	-1.0	-1.85
典型静止画1 E <sup>A</sup> <sub>1</sub> = 12	↑	-0.6	9.54	10.13	9.63
典型静止画2 E <sup>A</sup> <sub>2</sub> = 7	↑	-0.95	9.19	17.81	17.88
典型静止画3 E <sup>A</sup> <sub>3</sub> = 10	↑	-1.45	8.69	18.68	26.95

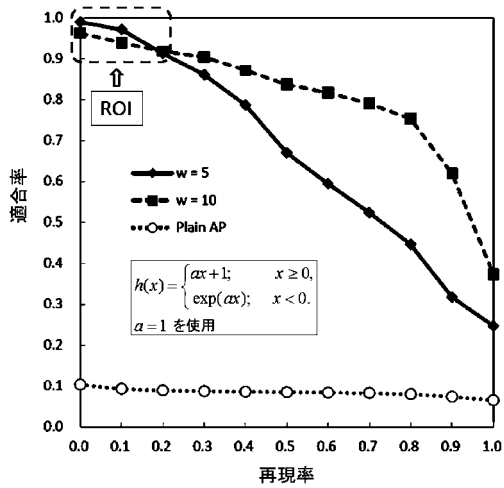
【図 1 7】

		動画像V <sub>A</sub>			
↓ i	→ j	典型静止画1 E <sup>A</sup> <sub>1</sub> = 8	典型静止画2 E <sup>A</sup> <sub>2</sub> = 12	典型静止画3 E <sup>A</sup> <sub>3</sub> = 10	典型静止画4 E <sup>A</sup> <sub>4</sub> = 7
		0.0	-0.4	-1.0	-1.85
典型静止画1 E <sup>B</sup> <sub>1</sub> = 12	↑	-0.6	9.54	10.13	9.63
典型静止画2 E <sup>B</sup> <sub>2</sub> = 7	↑	-0.95	9.19	17.81	17.88
典型静止画3 E <sup>B</sup> <sub>3</sub> = 10	↑	-1.45	8.69	18.68	26.95

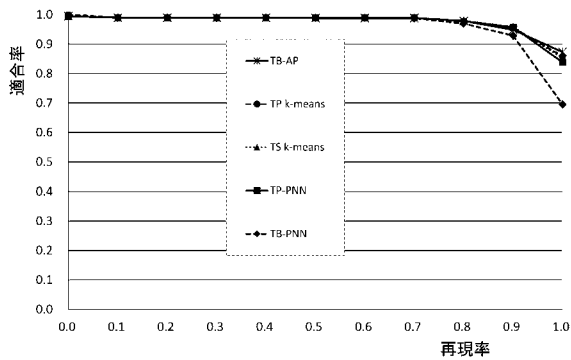
【図 1 8】

		動画像V <sub>B</sub>			
↓ i	→ j	典型静止画1 E <sup>B</sup> <sub>1</sub> = 8	典型静止画2 E <sup>B</sup> <sub>2</sub> = 12	典型静止画3 E <sup>B</sup> <sub>3</sub> = 10	典型静止画4 E <sup>B</sup> <sub>4</sub> = 7
		0.0	-0.4	-1.0	-1.85
典型静止画1 E <sup>A</sup> <sub>1</sub> = 12	↑	-0.6	9.54	10.13	9.63
典型静止画2 E <sup>A</sup> <sub>2</sub> = 7	↑	-0.95	9.19	17.81	17.88
典型静止画3 E <sup>A</sup> <sub>3</sub> = 10	↑	-1.45	8.69	18.68	26.95

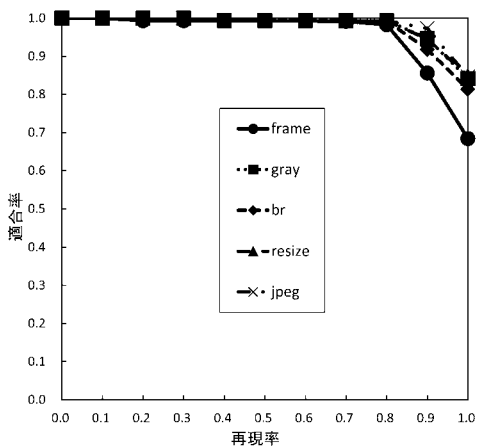
【図 2 1】



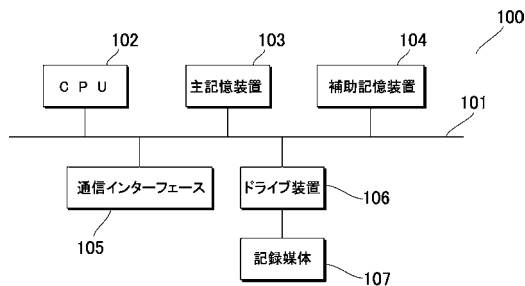
【図 2 2 A】



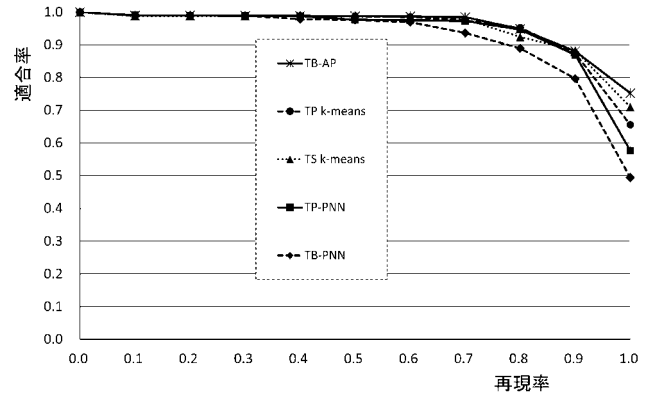
【図 2 3 B】



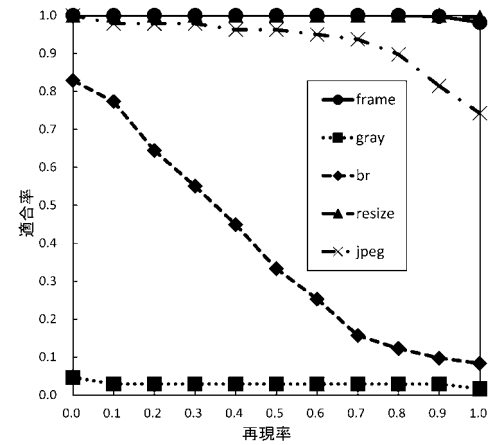
【図 2 4】



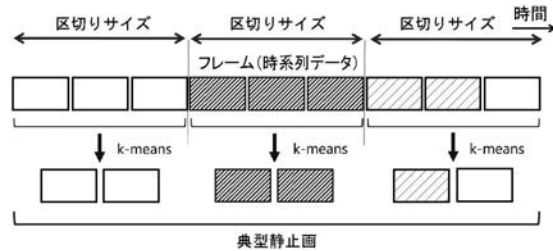
【図 2 2 B】



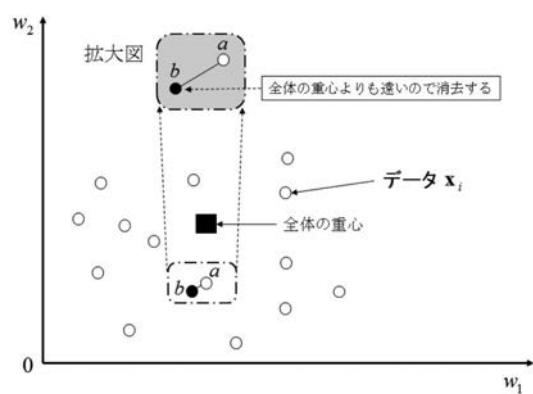
【図 2 3 A】



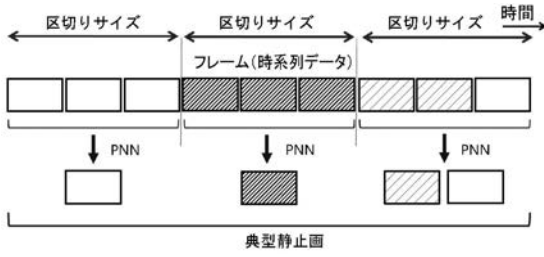
【図 8】



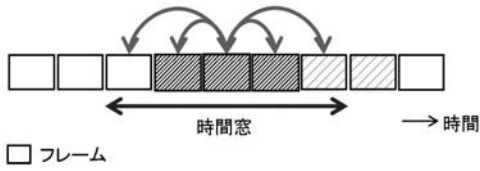
【図 1 0】



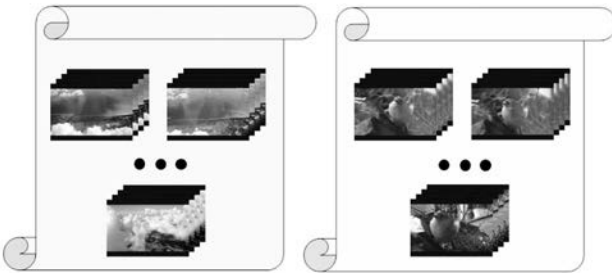
【図 1 1】



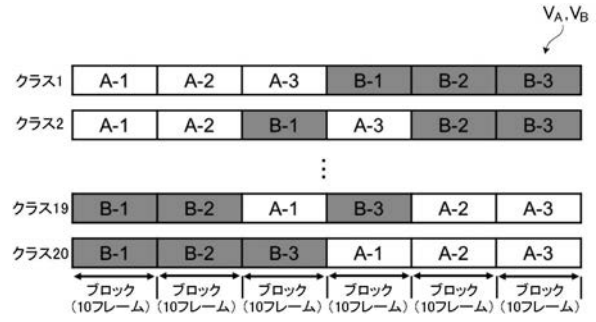
【図 1 2】



【図 1 4】



【図 1 5】





---

フロントページの続き

- (72)発明者 森脇 雅史  
東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内
- (72)発明者 横手 良太  
東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内
- (72)発明者 堀江 輝樹  
東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内
- (72)発明者 鹿野 晶滉  
東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内
- (72)発明者 岩瀬 弘道  
東京都新宿区戸塚町1丁目104番地 学校法人早稲田大学内
- Fターム(参考) 5L096 JA03 JA11 JA18 KA09