

明 細 書

発明の名称： 類似性検出装置及び指向性近傍検出方法

技術分野

[0001] 本発明は、特徴検出装置及び指向性近傍検出方法に関し、特に、多次元連続データ集合から所望のデータを検索する特徴検出装置及び指向性近傍検出方法に関する。

背景技術

[0002] 近傍検索技術は、様々な応用範囲で用いられる重要な技術であり、例えば、画像やWebページ等のデータの類似性を考慮し、膨大なデータベースの中から特徴が似たデータを検出する等のアプリケーションに用いられる。非常に単純な方法として、それぞれのデータの特徴を距離の定義された空間の一点（例えばユークリッド空間の一点）で表現し、クエリ点からの全てのデータへの距離を計算することで、クエリに近いデータを検出することができる。しかし、このような距離を直接計算する方法は、非常にコストがかかり、データ数の増大とともに計算負荷が非線形に増加してしまうため、様々な方法が提案されている。

[0003] 特許文献1（特開2004-021430号公報）には、ユーザの操作を煩雑にすることなく、特定の被写体を含む画像を類似画像としての確に検索し得る画像検索装置が開示されている。その画像検索装置は、検索の対象となる画像を複数含む検索対象画像群の中から、検索キーである基準画像に類似する画像を抽出する画像検索装置であって、基準画像及び検索対象画像群に含まれる各画像をそれぞれ複数の領域に分割する手段と、基準画像及び検索対象画像群に含まれる各画像の各領域から少なくとも1つの特徴量を抽出する手段と、検索対象画像群に含まれる各画像について領域の一部を選択する手段であって、検索対象画像群から画像を順次選択し、基準画像の各領域から抽出された特徴量と、当該選択した画像の各領域から抽出された特徴量と、の類似性に基づいて、所定数の当該選択された画像の領域を選択する手

段と、前記検索対象画像群に含まれる各画像について選択された一部の領域の特徴量に基づいて、検索対象画像群から基準画像に類似する画像を抽出する手段と、を含むことを特徴としている。

[0004] 特許文献2（特開2005-070927号公報）には、画像自体の一致度に拘泥されず、描画形状が類似している画像を検出することができるような画像特徴を取得する画像特徴取得方法が開示されている。その画像特徴取得方法は、二次元画像の画像特徴を取得する画像特徴取得方法であって、上記二次元画像を所定の大きさにサイズ変更し、該画像がカラー画像であれば一色の濃淡の階調で形成するスケール画像に変換し、各スケール画像に対して、二次元画像の左右のいずれか一方に設定した起点から水平方向へ連なる画素列を、該画素列の終点と次の画素列の起点とを鉛直方向へ順番に連結して水平成分の一次元ラスタ画像を作成し、二次元画像の上下のいずれか一方に設定した起点から鉛直方向へ連なる画素列を、該画素列の終点と次の画素列の起点とを水平方向へ順番に連結して鉛直成分の一次元ラスタ画像を作成し、これらの一次元ラスタ画像に対して適宜の変換処理を実行する。

[0005] 特許文献3（特開平10-326286号公報）には、類似検索の精度が向上でき、また重要な類似データが検索結果からぬけおちる可能性が低い類似検索装置が開示されている。その類似検索装置は、複数の対象に対して各々作成され、対象を特徴づける複数の属性をベクトル構成要素とするベクトルデータを複数蓄積するベクトルデータベースと、指定された類似検索対象に対するベクトルデータを作成する対象ベクトルデータ作成部と、複数の検索条件を生成する検索条件集合生成部と、この検索条件集合生成部で生成された個々の検索条件ごとに、上記ベクトルデータベースに蓄積された複数のベクトルデータの中から、上記検索条件を満足し、かつ上記対象ベクトルデータに類似するベクトルデータを検索する類似検索エンジンと、上記類似検索エンジンにより検索された結果を個々の検索条件ごとに表示する検索結果表示部とを備えたことを特徴としている。

[0006] 一般に、高次元の空間における近傍検索は、低次元の場合に比べ、更に問

題が困難になる。そのため、大量な高次元データに対して、厳密に距離を計算して近傍データを求めるのではなく、近似的に、もしくは確率的に近距離のデータを求める近似的近傍検索方法が提案されている。その一つの代表例は、LSH (Locality Sensitive Hashing) (非特許文献1 参照) である。LSHは、任意の2点に対して、その2点の距離が近いほど高い確率で衝突する(同じ値をもつ)ハッシュ関数を利用する方法であり、クエリ入力に対する近傍検出演算にかかる時間を削減することができる。ここで、LSHを用いると、あるクエリ q に対してデータ p が衝突する確率は、その距離 $d(p, q)$ にのみ依存するため、クエリ q を中心とする1つの円の円周上にある複数のデータは、全て等確率でクエリ q に衝突することになる。

先行技術文献

特許文献

[0007] 特許文献1：特開2004-021430号公報

特許文献2：特開2005-070927号公報

特許文献3：特開平10-326286号公報

非特許文献

[0008] 非特許文献1：Datar, M., Immorlica, N., Indyk, P., and Mirrokni, V. 2004. Locality sensitive hashing scheme based on p -stable distributions. In Proceedings of the ACM Symposium on Computational Geometry

発明の概要

[0009] 第1の問題点は、2つのデータの特徴の近さの指標として、方向に代表される詳細な位置関係が考えられていないことである。具体的には、ある方向に関する距離を他の方向に関する距離より重要視する場合等、方向の違いを

区別して近傍検索を行うことができない。その理由は、従来の近傍検索はその距離のみに依存する手法が用いられているからである。

[0010] 第2の問題点は、ある観点に対する類似度のみに関心がある場合、従来手法では、余計な計算時間がかかってしまうことである。その理由は、従来手法では、方向の考慮がないため、全方向での近傍検出によって候補を抽出した後に、関心がある方向に沿ったデータのみを取り出すという2段階の処理が必要であるからである。方向によって関心の高さの違いが大きい（ある方向に沿った距離の重みと、他の方向の距離の重みの差が非常に大きい時）ほど、近傍検出した候補の数が膨大になるため、計算時間の削減効果はあまり期待できなくなってしまう。

[0011] 本発明の目的は、特徴の違いが、距離として定義された空間内での近傍検出において、任意の類似度判定基準（任意の方向と、任意の距離の重要度）を指定して、近傍データを高速に検出する類似性検出装置及び指向性近傍検出方法を提供することにある。

[0012] 本発明による類似性検出装置は、乱数発生部と初期化部とデータ登録部と検索部とデータ処理部とを備えている。その乱数発生部は、入力装置を介して入力された複数の方向パラメータと複数の強度パラメータとに基づいて複数の乱数情報を算出する。その複数の方向パラメータは、ユークリッド空間上の方向を示している。その初期化部は、その複数の乱数情報に基づいて複数のキー計算関数を算出する。そのデータ登録部は、入力装置を介して入力された複数の検索対象データに基づいて、その複数のキー計算関数に対応する複数のテーブルを算出し、その複数のテーブルをテーブル保持装置に記録する。その複数の検索対象データは、それぞれ、そのユークリッド空間上の点を示している。その複数の検索対象データのうちの任意の2つの検索対象データがそれぞれ示している2つの点の距離は、その2つの検索対象データの類似度を示している。その複数のテーブルのうちの任意のキー計算関数に対応するテーブルは、複数のキーを複数のデータリストに対応付け、その複数のデータリストのうちの任意のキーに対応するデータリストに属するデー

タがその任意のキー計算関数に代入されることにより算出される値がその任意のキーに等しくなるように算出される。その検索部は、その複数のテーブルを参照して、入力装置を介して入力された検索条件が示しているクエリに基づいて候補データリストを算出する。その候補データリストは、その複数のキー計算関数に対応する複数の検索データリストを含んでいる。その複数の検索データリストのうちのその任意のキー計算関数に対応する検索データリストは、その複数のデータリストのうちの、そのクエリがそのキー計算関数に代入されることにより算出されるクエリ値に対応するデータリストを示している。そのデータ処理部は、その候補データリストに属する複数の検索データからその検索条件が示している条件を満足する検索結果データを算出し、その検索結果データを出力装置に出力する。

[0013] 本発明による指向性近傍検出方法では、入力装置を介して入力された複数の方向パラメータと複数の強度パラメータとに基づいて複数の乱数情報を算出する。また、その複数の乱数情報に基づいて複数のキー計算関数を算出する。また、入力装置を介して入力された複数の検索対象データに基づいて、その複数のキー計算関数に対応する複数のテーブルを算出する。また、その複数のテーブルをテーブル保持装置に記録する。また、その複数のテーブルを参照して、入力装置を介して入力された検索条件が示しているクエリに基づいて候補データリストを算出する。また、その候補データリストに属する複数の検索データからその検索条件が示している条件を満足する検索結果データを算出する。また、その検索結果データを出力装置に出力する。その複数の方向パラメータは、ユークリッド空間上の方向を示している。その複数の検索対象データは、それぞれ、そのユークリッド空間上の点を示している。その複数の検索対象データのうちの任意の2つの検索対象データがそれぞれ示している2つの点の距離は、その2つの検索対象データの類似度を示している。その複数のテーブルのうちの任意のキー計算関数に対応するテーブルは、複数のキーを複数のデータリストに対応付け、その複数のデータリストのうちの任意のキーに対応するデータリストに属するデータがその任意の

キー計算関数に代入されることにより算出される値がその任意のキーに等しくなるように算出される。その候補データリストは、その複数のキー計算関数に対応する複数の検索データリストを含んでいる。その複数の検索データリストのうちのその任意のキー計算関数に対応する検索データリストは、その複数のデータリストのうちの、そのクエリがそのキー計算関数に代入されることにより算出されるクエリ値に対応するデータリストを示している。

[0014] 第1の効果は、膨大なデータの中で、あるクエリ点からある興味方向に沿った自由度の高い近傍点の検出が行えることである。その理由は、任意の2点が、任意の興味方向に沿った任意の距離重み（重要度）で測った距離が短いほど高い確率で同じエントリに登録されるテーブルを具備するデータ管理機能を提供することで、指向性近傍検出機能を実現するからである。

[0015] 第2の効果は、指向性近傍検索処理の高速化である。その理由は、事前に所望の指向性を登録した複数のテーブルにデータを登録しておくことで、クエリ点からの指向性近傍検索に当たって、クエリ点と同じエントリに登録されたデータのみを扱えば良く、全てのデータとの距離を計算する必要がないためオンライン処理時間が大幅に短縮化されるからである。

図面の簡単な説明

[0016] [図1]図1は、本発明による類似性検出装置を示すブロック図である。

[図2]図2は、テーブル管理装置を示すブロック図である。

[図3]図3は、初期化動作を示す流れ図である。

[図4]図4は、検索動作を示す流れ図である。

[図5]図5は、指向パラメータの例を示す表である。

[図6]図6は、指向パラメータに対応する衝突確率等値線を示すグラフである。

[図7]図7は、指向パラメータの例を示す表である。

[図8]図8は、複数のテーブルセットの例を示す表である。

[図9]図9は、複数の指向パラメータに対応する複数の衝突確率等値線を示すグラフである。

[図10]図 1 0 は、画像を示す図である。

[図11]図 1 1 は、複数の検索対象データ示すグラフである。

[図12]図 1 2 は、複数のテーブルセットの例を示す表である。

[図13]図 1 3 は、検索条件のクエリの例を示すグラフである。

[図14]図 1 4 は、検索結果データの例を示すグラフである。

[図15]図 1 5 は、複数の時系列データを示すグラフである。

[図16]図 1 6 は、複数の時系列データのうちの複数の検索対象データを示すグラフである。

[図17]図 1 7 は、複数の衝突確率等値線を示すグラフである。

[図18]図 1 8 は、他の複数の衝突確率等値線を示すグラフである。

[図19]図 1 9 は、衝突確率分布を示すグラフである。

[図20]図 2 0 は、他の衝突確率分布を示すグラフである。

[図21]図 2 1 は、特徴分布を示すグラフである。

[図22]図 2 2 は、移動体管理システムを示すブロック図である。

[図23]図 2 3 は、複数のテーブルセットを示す表である。

発明を実施するための形態

[0017] 図面を参照して、本発明による類似性検出装置の実施の形態を説明する。

その類似性検出装置 1 は、図 1 に示されているように、複数のコンピュータが互いに双方向に情報を伝送することができるよう接続されている。

[0018] その複数のコンピュータの各コンピュータは、図示されていないが、CPUと、記憶装置と、インターフェースとを備えている。

[0019] そのCPUは、そのコンピュータにインストールされているコンピュータプログラムを実行することにより、その記憶装置とそのインターフェースとを制御する。

[0020] その記憶装置は、そのコンピュータプログラムを記録し、そのCPUにより作成される情報を一時的に記録する。

[0021] そのインターフェースは、そのコンピュータに接続されている外部機器により生成される情報をそのCPUに出力したり、そのCPUにより生成され

た情報をその外部機器に出力したりする。

[0022] その外部機器としては、入力装置、出力装置、通信装置、リムーバルメモリドライブが例示される。

[0023] その入力装置は、ユーザに操作されることにより情報を作成し、その情報をそのCPUに出力する。その入力装置としては、キーボード、ポインティングデバイス、タッチパネルが例示される。

[0024] その出力装置は、そのCPUにより生成される情報をユーザに認識可能に出力する。その出力装置としては、ディスプレイ、音響装置、タッチパネルが例示される。

[0025] その通信装置は、通信ネットワークを介してそのCPUにより作成された情報を他のコンピュータに送信し、その通信ネットワークを介して他のコンピュータから受信された情報をそのCPUに出力する。その通信装置は、更に、そのコンピュータにインストールされるコンピュータプログラムを他のコンピュータからダウンロードすることに利用される。

[0026] そのリムーバルメモリドライブは、記録媒体が挿入されたときに、その記録媒体に記録されているデータを読み出すことに利用される。そのリムーバルメモリドライブは、更に、コンピュータプログラムが記録されている記録媒体が挿入されたときに、そのコンピュータプログラムをそのコンピュータにインストールするときに利用される。その記録媒体としては、磁気ディスク（フレキシブルディスク、ハードディスク）、光ディスク（CD、DVD）、光磁気ディスク、フラッシュメモリが例示される。

[0027] その複数のコンピュータは、パラメータ管理装置2と、乱数発生装置3と、テーブル管理装置5と、データ処理装置6とを含んでいる。

[0028] パラメータ管理装置2は、入力装置7を備えている。パラメータ管理装置2は、入力装置7を介してパラメータリスト8がパラメータ管理装置2に入力されるように、入力装置7を制御する。

[0029] 乱数発生装置3は、テーブル管理装置5から乱数情報を要求されたときに、テーブル管理装置5から出力された情報に基づいて乱数情報を算出する。

[0030] テーブル管理装置 5 は、テーブル保持装置 10 と、入力装置 11 とを備えている。テーブル管理装置 5 は、入力装置 11 を介して入力データ 12 がテーブル管理装置 5 に入力されるように、入力装置 11 を制御する。テーブル管理装置 5 は、入力データ 12 と乱数発生装置 3 により算出された乱数情報とに基づいて複数のテーブルを算出する。テーブル管理装置 5 は、その複数のテーブルがテーブル保持装置 10 に記録されるように、テーブル保持装置 10 を制御する。テーブル管理装置 5 は、更に、データ処理装置 6 からデータリストを要求されたときに、その複数のテーブルを参照して、データ処理装置 6 から出力された情報に基づいてデータリストを算出し、そのデータリストをデータ処理装置 6 に出力する。

[0031] データ処理装置 6 は、入力装置 14 と、出力装置 15 とを備えている。データ処理装置 6 は、入力装置 14 を介して検索条件 16 がデータ処理装置 6 に入力されるように、入力装置 14 を制御する。検索条件 16 は、クエリと条件とを示している。データ処理装置 6 は、更に、そのクエリをテーブル管理装置 5 に出力することにより、テーブル管理装置 5 にデータリストを要求する。データ処理装置 6 は、その条件とテーブル管理装置 5 により算出されたデータリストとに基づいて検索結果データを算出する。その検索結果データは、そのデータリストが示すデータのうちのその条件を満足するデータを示している。データ処理装置 6 は、更に、テーブル管理装置 5 により算出された検索結果データがユーザに認識可能に表現されるように、出力装置 15 を制御する。

[0032] 図 2 は、テーブル管理装置 5 を示している。テーブル管理装置 5 にインストールされるコンピュータプログラムは、テーブル管理装置 5 に複数の機能をそれぞれ実現させる複数のコンピュータプログラムから形成されている。

[0033] テーブル管理装置 5 にその複数の機能をそれぞれ実現させた場合、テーブル管理装置 5 は、乱数情報取得部 21 と、初期化部 22 と、データ登録部 23 と、検索部 24 とを含んでいる。

[0034] 乱数情報取得部 21 は、パラメータ管理装置 2 に入力されたパラメータリ

スト 8 をパラメータ管理装置 2 から収集する。パラメータリスト 8 は、テーブルパラメータと管理パラメータと指向パラメータとを示している。そのテーブルパラメータは、テーブルの面数 P とウィンドウ幅 W と基数 C とビット長 B とを示している。面数 P は、正の整数を示している。ウィンドウ幅 W は、正の実数を示している。基数 C は、2 以上の整数を示している。ビット長 B は、正の整数を示している。その管理パラメータは、次元数 D を示している。その指向パラメータは、集合 U を示している。

[0035] 集合 U は、指向パラメータ u_j から形成され、以下の式により表現される。

$$U = \{u_j \mid u_j \in R^D \times R_+\} \quad (j = 1, \dots, D)$$

ここで、集合 R^D は、 D 次元数ベクトル空間を示している。例えば、集合 R^2 は、2 次元数ベクトル空間を示している。

[0036] 集合 R_+ は、正の実数の集合を示している。すなわち、指向パラメータ u_j は、以下の式により表現される。

$$u_j = \langle v_j, \sigma_j \rangle$$

ここで、方向パラメータ v_j は、 D 次元のベクトルを示している。強度パラメータ σ_j は、正の実数であり、方向パラメータ v_j を重視する程度を示している。

[0037] 乱数情報取得部 21 は、パラメータリスト 8 のうちのウィンドウ幅 W とビット長 B と次元数 D と集合 U とを乱数発生装置 3 に出力することにより、乱数発生装置 3 に乱数情報を要求する。乱数情報取得部 21 は、乱数発生装置 3 により算出された乱数情報を乱数発生装置 3 から収集する。乱数情報取得部 21 は、その要求を P 回繰り返すことにより、複数 (P 個) の乱数情報を乱数発生装置 3 から収集し、乱数情報集合 Z を作成する。

[0038] 乱数情報集合 Z は、以下の式により表現される。

$$Z = \{Z_p\} \quad (p = 1, \dots, P)$$

ここで、集合 Z_p は、乱数発生装置 3 により算出された 1 つの乱数情報を示している。

[0039] 集合 Z_p は、以下の式により表現される。

$$Z_b = \{Z_b \mid Z_b = \langle \Phi^{(p)}_b, R^{(p)}_b \rangle \mid b = 1, \dots, B\}$$

ここで、乱数 $R^{(p)}_b$ は、一様分布 $U[0, W]$ に従う乱数を示している。

ランダムベクトル $\Phi^{(p)}_b$ は、 D 次元のベクトルを示している。

[0040] ランダムベクトル $\Phi^{(p)}_b$ は、以下の式により表現される。

$$\Phi^{(p)}_b = V A^{-1/2} A_b$$

[0041] 行列 V は、以下の式により表現される。

$$V = (v_1, \dots, v_D)$$

[0042] 対角行列 $A^{-1/2}$ は、以下の式により表現される。

$$A^{-1/2} = d \{a_g \mid 1/\sigma_1, \dots, 1/\sigma_D\}$$

[0043] ベクトル A_b は、 D 次元ベクトルであり、ベクトル A_b の d 成分 $A_{b,d}$ ($d = 1, \dots, D$) は、正規分布 $N(0, 1)$ に従う乱数を示している。

[0044] 初期化部 22 は、更に、パラメータ管理装置 2 により収集されたパラメータリスト 8 と乱数情報取得部 21 により収集された乱数情報とに基づいて複数 (P 個) のキー計算関数を算出する。その複数のキー計算関数のうちの任意の自然数 p ($p \leq P$) に対応するキー計算関数 $L^{(p)}(x)$ は、 D 次元ベクトル x の関数である。

[0045] キー計算関数 $L^{(p)}(x)$ は、 D 次元ベクトル x を用いて、以下の式により表現される。

$$L^{(p)}(x) = (f^{(p)}_1(x), \dots, f^{(p)}_B(x))$$

[0046] 基本関数 $f^{(p)}_b(x)$ は、以下の式により表現される。

[数1]

$$f^{(p)}_b(x) = \left[\frac{\Phi^{(p)}_b \cdot x + R^{(p)}_b}{W} \right] \bmod C$$

[0047] データ登録部 23 は、入力装置 11 を介してテーブル管理装置 5 に入力データ 12 が入力されるように、入力装置 11 を制御する。入力データ 12 は、データ集合 X を示している。

[0048] データ集合 X は、データ x_i から形成され、以下の式により表現される。

$$X = \{x_i \mid x_i \in R^D\} \quad (i = 1, \dots, N)$$

ここで、集合 R^D は、 D 次元数ベクトル空間を示している。自然数 N は、データ集合 X のデータ x_i の要素の総数を示し、1 より大きい自然数を示している。データ x_i は、 D 次元数ベクトルを示している。

- [0049] データ登録部 23 は、更に、入力データ 12 と初期化部 22 により算出された複数のキー計算関数とに基づいて、複数（ P 個）のテーブルを作成する。その複数のテーブルのうちの任意の自然数 p に対応するテーブルは、複数のエントリから形成されている。そのテーブルは、その複数のエントリの各エントリが 1 つのキーと 1 つのバリューとのペア（組）を示すことにより、複数のキーを複数のバリューに対応付けている。すなわち、その複数のキーのうちの任意のキーは、その複数のバリューのうちの 1 つのバリューに対応している。
- [0050] その複数のキーのうちの任意のキーは、初期化部 22 により算出された複数のキー計算関数のうちの任意の自然数 p に対応するキー計算関数 $L^{(p)}(x)$ に、データ集合 X のうちのいずれかのデータ x が代入されることにより算出された値を示している。更に、その複数のキーは、互いに異なり、データ集合 X に属する全てのデータ x がキー計算関数 $L^{(p)}(x)$ に代入されることによりそれぞれ算出される複数の値を含んでいる。
- [0051] その複数のバリューのうちの任意のキー L に対応するバリュー $Table^{(p)}[L]$ は、所定のデータの集合であるデータリストを示している。このとき、バリュー $Table^{(p)}[L]$ は、そのデータリストに属するデータのうちの任意のデータ x がキー計算関数 $L^{(p)}(x)$ に代入されることにより算出された値がキー L に一致するように、算出される。
- [0052] データ登録部 23 は、更に、その複数のテーブルがテーブル保持装置 10 に記録されるように、テーブル保持装置 10 を制御する。
- [0053] 検索部 24 は、データ処理装置 6 からクエリが出力されたときに、データ登録部 23 により作成された複数のテーブルを参照して、そのクエリに基づいてデータリストを算出する。すなわち、検索部 24 は、そのクエリ q に基

づいてその複数のテーブルに対応する複数の検索キーを算出する。その複数の検索キーのうちの自然数 p に対応する検索キー $L^{(p)}(q)$ は、初期化部 22 により算出された複数のキー計算関数のうちの自然数 p に対応するキー計算関数 $L^{(p)}(x)$ にクエリ q が代入されることにより算出された値を示している。

[0054] 検索部 24 は、データ登録部 23 により作成された複数のテーブルを参照して、その複数の検索キーに基づいてデータリスト $List<x>$ を算出する。データリスト $List<x>$ は、データ集合 X に属するデータのうちの所定のデータの集合を示している。データリスト $List<x>$ は、その複数のテーブルのうちの自然数 p に対応するテーブルがキー $L^{(p)}(q)$ に対応するエントリを含むときに、そのエントリが示すバリュー $Table^{(p)}[L]$ が示すデータリストに属するデータを含むように、算出される。

[0055] このとき、データ処理装置 6 は、検索条件 16 と検索部 24 により算出されたデータリスト $List<x>$ とに基づいて検索結果データ x_* を算出する。検索結果データ x_* は、データリスト $List<x>$ に属するデータのうちの検索条件 16 が示す条件が満足するデータの集合を示している。データ処理装置 6 は、更に、検索結果データ x_* が表示されるように、出力装置 15 を制御する。

[0056] 本発明による指向性近傍検出方法の実施の形態は、類似性検出装置 1 により実行され、初期化動作とデータ登録動作と検索動作とを備えている。

[0057] 図 3 は、その初期化動作を示している。ユーザは、まず、パラメータリスト 8 を用意し、入力装置 7 を操作することにより、パラメータリスト 8 をパラメータ管理装置 2 に入力する。パラメータ管理装置 2 は、入力装置 7 を介してパラメータリスト 8 が入力されると、パラメータリスト 8 を記憶装置に記録する。テーブル管理装置 5 は、パラメータ管理装置 2 に入力されたパラメータリスト 8 をパラメータ管理装置 2 から取得する（ステップ S1）。

[0058] パラメータリスト 8 は、テーブルパラメータと管理パラメータと指向パラメータとを示している。そのテーブルパラメータは、テーブルの作成に必要な

である情報を示し、すなわち、テーブルの面数 P とウィンドウ幅 W と基数 C とビット長 B とを示している。面数 P は、正の整数を示している。ウィンドウ幅 W は、正の実数を示している。基数 C は、2 以上の整数を示している。ビット長 B は、正の整数を示している。その管理パラメータは、その他管理運用に必要である情報を示し、すなわち、次元数 D を示している。その指向パラメータは、特徴検出の方向や距離の重み情報を含み、近傍検出における指向性を表現し、集合 U を示している。

[0059] 集合 U は、指向パラメータ u_j から形成され、以下の式により表現される。

$$U = \{u_j \mid u_j = \langle v_j, \sigma_j \rangle\} \quad (j = 1, \dots, D)$$

ここで、方向パラメータ v_j は、 D 次元のベクトルを示している。強度パラメータ σ_j は、正の実数であり、方向パラメータ v_j を重視する程度を示している。

[0060] テーブル管理装置 5 は、テーブル面番号 p に 1 を代入する（ステップ S 2）。テーブル管理装置 5 は、テーブル面番号 p が示す値が面数 P の大小関係を算出する（ステップ S 3）。

[0061] テーブル管理装置 5 は、テーブル面番号 p が示す値が面数 P より大きくないうちに（ステップ S 3、 $p \leq P$ ）、パラメータリスト 8 のうちのウィンドウ幅 W とビット長 B と次元数 D と集合 U とを乱数発生装置 3 に出力することにより、乱数発生装置 3 に乱数情報を要求する。乱数発生装置 3 は、テーブル管理装置 5 から乱数情報を要求されたときに、ウィンドウ幅 W とビット長 B と次元数 D と集合 U とに基づいて、集合 Z_p を示す乱数情報を算出する。

[0062] 集合 Z_p は、要素 Z_b から形成され、以下の式により表現される。

$$Z_p = \{Z_b \mid Z_b = \langle \Phi^{(p)}_b, R^{(p)}_b \rangle\} \quad (b = 1, \dots, B)$$

ここで、乱数 $R^{(p)}_b$ は、一様分布 $U[0, W)$ に従う乱数を示し、 B 回の独立な試行により生成される。ランダムベクトル $\Phi^{(p)}_b$ は、 D 次元のベクトルを示している。

[0063] ランダムベクトル $\Phi^{(p)}_b$ は、以下の式により表現される。

$$\Phi^{(p)}_b = V \Lambda^{-1/2} A_b$$

[0064] 行列 V は、以下の式により表現される。

$$V = (v_1, \dots, v_D)$$

[0065] 対角行列 $\Lambda^{-1/2}$ は、以下の式により表現される。

$$\Lambda^{-1/2} = \text{diag} \{1/\sigma_1, \dots, 1/\sigma_D\}$$

[0066] ベクトル A_b は、 D 次元ベクトルであり、ベクトル A_b の d 成分 $A_{b,d}$ ($d = 1, \dots, D$)は、正規分布 $N(0, 1)$ に従う乱数を示している。ベクトル A_b の全ての成分は、それぞれ、独立な試行により生成される。テーブル管理装置5は、乱数発生装置3により算出された乱数情報を乱数発生装置3から収集する（ステップS4）。

[0067] テーブル管理装置5は、その乱数情報が収集された後に、テーブル面番号 p を1だけインクリメントし、すなわち、テーブル面番号 p が示す値に1を加算することにより算出された和をテーブル面番号 p に代入する（ステップS5）。

[0068] テーブル管理装置5は、テーブル面番号 p をインクリメントした後に、テーブル面番号 p が示す値と面数 P との大小関係を算出する（ステップS3）。テーブル管理装置5は、テーブル面番号 p が示す値が面数 P より大きくないないときに（ステップS3、 $p \leq P$ ）、ステップS4～S5を繰り返して実行する。

[0069] テーブル管理装置5は、テーブル面番号 p が示す値が面数 P より大きいときに（ステップS3、 $p > P$ ）、乱数情報集合 Z とパラメータリスト8とに基づいて、複数（ P 個）のキー計算関数を算出する（ステップS6）。

[0070] ここで、乱数情報集合 Z は、ステップS4が繰り返し実行されることにより収集された複数（ P 個）の乱数情報を示している。その複数のキー計算関数のうちのテーブル面番号 p に対応するキー計算関数 $L^{(p)}(x)$ は、 D 次元数ベクトル x の関数である。

[0071] 乱数情報集合 Z は、以下の式により表現される。

$$Z = \{Z_p\} \quad (p = 1, \dots, P)$$

[0072] キー計算関数 $L^{(p)}(x)$ は、 D 次元数ベクトル x を用いて、以下の式によ

り表現される。

$$L^{(p)}(x) = (f^{(p)}_1(x), \dots, f^{(p)}_B(x))$$

ここで、基本関数 $f^{(p)}_b(x)$ は、前述の式により表現される（数 1 参照）。

[0073] テーブル管理装置 5 は、その複数（P 個）のキー計算関数が算出された後に、エントリが空の複数（P 個）のテーブルを作成する（ステップ S 7）。その複数のテーブルは、その複数のキー計算関数に対応している。テーブル管理装置 5 は、テーブル保持装置 10 を制御することにより、その複数（P 個）のテーブルをテーブル保持装置 10 に記録する。

[0074] そのデータ登録動作は、その初期化動作が実行された後に実行される。ユーザは、まず、入力データ 12 を用意し、入力装置 11 を操作することにより、入力データ 12 をテーブル管理装置 5 に入力する。入力データ 12 は、データ集合 X を示している。

[0075] データ集合 X は、データ x_i から形成され、以下の式により表現される。

$$X = \{x_i \mid x_i \in R^D\} \quad (i = 1, \dots, N)$$

ここで、集合 R^D は、D 次元数ベクトル空間を示している。自然数 N は、データ集合 X の要素であるデータ x_i の総数を示し、1 より大きい自然数を示している。データ x_i は、D 次元数ベクトルを示している。

[0076] テーブル管理装置 5 は、入力装置 11 を介して入力データ 12 が入力されると、記憶装置を制御することにより、入力データ 12 をその記憶装置に記録する。

[0077] テーブル管理装置 5 は、入力装置 11 を介して入力データ 12 が入力されると、その初期化動作により作成された複数（P 個）のテーブルに対応する複数（P 個）のテーブル作成動作を実行する。その複数のテーブル作成動作のうちのテーブル面番号 p に対応するテーブル作成動作は、データ集合 X に属する全てのデータに対応する複数（N 個）のエントリ作成動作から形成されている。

[0078] その複数（N 個）のエントリ作成動作のうちのあるデータ x_i に対応するエ

ントリ作成動作では、テーブル管理装置 5 は、その初期化動作により作成された複数（ P 個）のキー計算関数のうちのテーブル面番号 p に対応するキー計算関数 $L^{(p)}(x)$ にデータ x_i を代入することによりキー $L^{(p)}(x_i)$ を算出する。テーブル管理装置 5 は、その初期化動作により作成された複数（ P 個）のテーブルのうちのテーブル面番号 p に対応するテーブルに、キー $L^{(p)}(x_i)$ に対応するエントリが存在しているかどうかを判別する。

[0079] テーブル管理装置 5 は、そのテーブルにキー $L^{(p)}(x_i)$ に対応するエントリが存在するときに、そのエントリのバリュー $Table^{(p)}[L^{(p)}(x_i)]$ が示すデータリストにデータ x_i を追加する。テーブル管理装置 5 は、そのテーブルにキー $L^{(p)}(x_i)$ に対応するエントリが存在しないときに、キー $L^{(p)}(x_i)$ に対応するエントリをそのテーブルに追加する。このとき、そのエントリのバリュー $Table^{(p)}[L^{(p)}(x_i)]$ は、データ x_i のみを要素とするデータリストを示している。

[0080] テーブル管理装置 5 は、その複数（ N 個）のエントリ作成動作の全てを実行することにより、その複数（ P 個）のテーブルのうちのテーブル面番号 p に対応するテーブルを作成する。テーブル管理装置 5 は、その複数（ P 個）のテーブル作成動作の全てを実行することにより、その複数（ P 個）のテーブルの全てを作成する。テーブル管理装置 5 は、テーブル保持装置 10 を制御することにより、その複数のテーブルをテーブル保持装置 10 に記録する。

[0081] このように作成されたテーブルでは、任意の 2 つのデータに対応する 2 つのキーがそれぞれ示す任意の 2 点が、指向パラメータ u_j が示す方向に沿った任意の距離重み（重要度）で測った距離が短いほど高い確率で同じエントリに登録される。

[0082] 図 4 は、その検索動作を示している。ユーザは、まず、検索条件 16 を用意し、入力装置 14 を操作することにより、検索条件 16 をデータ処理装置 6 に入力する。検索条件 16 は、クエリ q と条件とを示している。その条件としては、「クエリ q の最近傍のデータを 1 つ取得する」「クエリ q に近い

ものからK個のデータを取得する」が例示される。データ処理装置6は、入力装置14を介して検索条件16が入力されると、記憶装置を制御することにより、検索条件16をその記憶装置に記録する（ステップS11）。データ処理装置6は、検索条件16が入力されると、更に、クエリqをテーブル管理装置5に出力することにより、テーブル管理装置5にデータリストを要求する（ステップS12）。

[0083] テーブル管理装置5は、データ処理装置6からクエリqが出力されると、テーブル面番号pに1を代入し、空のデータリスト $List<x>=\phi$ を用意する（ステップS13）。テーブル管理装置5は、テーブル面番号pが示す値が面数Pの大小関係を算出する（ステップS14）。テーブル管理装置5は、テーブル面番号pが示す値が面数Pより大きくないときに（ステップS14、 $p \leq P$ ）、クエリqに基づいて検索キー $L^{(p)}(q)$ を算出する（ステップS15）。

[0084] 検索キー $L^{(p)}(q)$ は、その初期化動作により算出された複数のキー計算関数のうちのテーブル面番号pに対応するキー計算関数 $L^{(p)}(x)$ にクエリqが代入されることにより算出される値を示している。

[0085] テーブル管理装置5は、そのデータ登録動作により作成された複数のテーブルのうちのテーブル面番号pに対応するテーブルを参照して、検索キー $L^{(p)}(q)$ に対応するバリュー $Table^{(p)}[L^{(p)}(q)]$ を取得する（ステップS16）。テーブル管理装置5は、バリュー $Table^{(p)}[L^{(p)}(q)]$ が示すデータリストに属するデータをデータリスト $List<x>$ に追加する（ステップS17）。

[0086] ここで、テーブル管理装置5は、データリスト $List<x>$ とバリュー $Table^{(p)}[L^{(p)}(q)]$ とに重複して属するデータがあった場合に、バリュー $Table^{(p)}[L^{(p)}(q)]$ が示すデータのうちのデータリスト $List<x>$ に登録されていないデータのみをデータリスト $List<x>$ に追加する。

[0087] テーブル管理装置5は、ステップS17が実行された後に、テーブル面番

号 p を 1 だけインクリメントし、すなわち、テーブル面番号 p が示す値に 1 を加算することにより算出された和をテーブル面番号 p に代入する（ステップ S 1 8）。テーブル管理装置 5 は、テーブル面番号 p をインクリメントした後に、テーブル面番号 p が示す値と面数 P との大小関係を算出する（ステップ S 1 4）。テーブル管理装置 5 は、テーブル面番号 p が示す値が面数 P より大きくないときに（ステップ S 1 4、 $p \leq P$ ）、ステップ S 1 5～S 1 8 を繰り返して実行する。

[0088] テーブル管理装置 5 は、テーブル面番号 p が示す値が面数 P より大きいときに（ステップ S 1 4、 $p > P$ ）、データリスト $List<x>$ をデータ処理装置 6 に出力する。こうして出来上がったデータリスト $List<x>$ は、クエリ q の特定方向を適当な重みで重要視して距離計算した時に近距離となるデータの候補となっている。

[0089] データ処理装置 6 は、テーブル管理装置 5 によりデータリスト $List<x>$ が出力されると、検索条件 1 6 が示す条件とデータリスト $List<x>$ とに基づいて検索結果データ x_* を算出する（ステップ S 1 9）。

[0090] 検索結果データ x_* は、データリスト $List<x>$ に属するデータのうちのその条件が満足するデータの集合を示している。もし、その条件が「クエリ q に近いものから K 個のデータを取得する」を示し、データリスト $List<x>$ が K 個以下のデータしか含まない場合には、検索結果データ x_* は、データリスト $List<x>$ の全てのデータを示している。なお、データリスト $List<x>$ からクエリ q に近いデータを選択する計算としては、普通のユークリッド距離を用いても良いし、指向パラメータで指定した重要度を考慮して距離を計算しても良い。

[0091] データ処理装置 6 は、検索結果データ x_* が算出された後に、出力装置 1 5 を制御することにより、検索結果データ x_* を出力装置 1 5 に表示させる（ステップ S 2 0）。

[0092] このような指向性近傍検出方法によれば、ユーザが興味の方角やその度合い等を示す関心情報を考慮して指向パラメータを設定しておくことで、類似

性検出装置 1 は、高次元データにおいても、その関心に沿った中で、関心中心から近いデータを高速に優先的選択する指向性近傍検出が可能である。

[0093] 既存手法では、同じ距離（類似度）にあるデータは同じ確率で選択するため、ある検索中心から同心球上の点は、全て同じ確率で検出される。そのため、ある興味方向にある距離 r のデータだけを取り出すには、一度半径 r の球内にあるデータを全て選択し、そこからある方向のみのものを優先的に取り出す必要がある。このような手法では、指向性が強くなるほど（ある方向の重要度の重みが他の方向の重要度の重みに比べて大きくなるほど）、不必要なデータを大量に取得しなければならず、その分計算時間が増大する。

[0094] それに対して、本発明による技術では、衝突確率等値線が等方的な球ではなく、衝突確率等値線が興味方向に長軸をもつ楕円体として制御できるため、任意の方向に任意の重みで指向性近傍検出を行うことが可能となり大幅な時間削減が望める。

[0095] なお、類似性検出装置 1 は、データ処理装置 6 が、乱数情報取得部 2 1 と初期化部 2 2 とデータ登録部 2 3 と検索部 2 4 とがそれぞれ実現される複数のコンピュータに置換されることができる。更に、類似性検出装置 1 は、パラメータ管理装置 2 と乱数発生装置 3 とテーブル管理装置 5 とデータ処理装置 6 とが実現される 1 つのコンピュータに置換されることもできる。更に、パラメータ管理装置 2 と乱数発生装置 3 とデータ処理装置 6 と乱数情報取得部 2 1 と初期化部 2 2 とデータ登録部 2 3 と検索部 2 4 とのいずれかが実現される複数のコンピュータから形成されることもできる。このように置換された類似性検出装置も、類似性検出装置 1 と同様にして、任意の方向に任意の重みで指向性近傍検出を行うことが可能となる。

[0096] 次に、本発明による指向性近傍検出方法が二次元（ $D = 2$ ）ユークリッド空間に適用される例を考える。例えば、図 5 に示されているように、集合 U が複数の指向パラメータ 3 6 を複数の方向パラメータ 3 7 と複数の強度パラメータ 3 8 とに対応付けている例を考える。

[0097] 複数の指向パラメータ 3 6 は、2 つの指向パラメータ u_1 , u_2 から形成され

ている。複数の方向パラメータ 37 は、 $(1/\sqrt{2}, 1/\sqrt{2})$ を示す方向パラメータ v_1 と $(-1/\sqrt{2}, 1/\sqrt{2})$ を示す方向パラメータ v_2 とから形成されている。複数の強度パラメータ 38 は、2 を示す強度パラメータ σ_1 と 1 を示す強度パラメータ σ_2 とから形成されている。指向パラメータ u_1 は、方向パラメータ v_1 を強度パラメータ σ_1 に対応付けている。指向パラメータ u_2 は、方向パラメータ v_2 を強度パラメータ σ_2 に対応付けている。

[0098] すなわち、集合 U は、以下の式により表現される。

$$U = \{ \langle (1/\sqrt{2}, 1/\sqrt{2}), 2 \rangle, \langle (-1/\sqrt{2}, 1/\sqrt{2}), 1 \rangle \}$$

[0099] このとき、方向パラメータ v_1 は、図 6 に示されているように、 x 軸から反時計回りに 45 度傾いた方向を示している。方向パラメータ v_2 は、方向パラメータ v_1 が示す方向と直交する方向を示している。集合 U は、強度パラメータ σ_2 が示す値の 2 倍の値を強度パラメータ σ_1 が示すことから、方向パラメータ v_1 の方向を方向パラメータ v_2 の方向に比べて 2 倍重視することを意味する。

[0100] 入力データ 12 は、集合 X に属するデータの分布が任意とする。検索条件 16 の条件としては、「クエリ q の入力に対して、クエリ q との距離が最も近いデータである」ことを示す。

[0101] このとき、衝突確率等値線 31 は、方向パラメータ v_1 の方向に長軸をもつ楕円に形成される。すなわち、衝突確率等値線 31 の上のデータは、全て同じ確率で検出される。類似性検出装置 1 は、方向パラメータ v_2 の方向に比べ、方向パラメータ v_1 の方向に 2 倍遠い場所を同一視して、集合 X に属するデータから検索結果データを検索することができる。すなわち、類似性検出装置 1 は、検出確率を制御することができるため、方向パラメータ v_1 の方向に強度パラメータ σ_1 の重みで指向性近傍検出を行うことが可能となる。

[0102] 本発明による類似性検出装置の第 2 の実施の形態は、既述の実施の形態に記載されたパラメータリスト 8 が示す指向パラメータが他の指向パラメータ

に置換されている。

[0103] その指向パラメータは、集合 U_k から形成される集合を示し、その集合は、以下の式により表現される。

$$\{U_k\} \quad (k = 1, \dots, K)$$

[0104] 集合 U_k は、以下の式により表現される。

$$U_k = \{u_j \mid u_j \in R^D \times R_+\} \quad (j = 1, \dots, D)$$

このとき、集合 U_k は、互いに異なる。すなわち、方向パラメータ v 又は強度パラメータ σ のどれかが異なっている。

[0105] 例えば、簡単のため $D = 2$ の2次元の場合を考えると、集合 U_k は、以下の式により表現される。

$$U_k = \{\langle v_x, \sigma_x 2^p \rangle, \langle v_y, \sigma_y 2^p \rangle\}$$

ここで、方向パラメータ v_x は、 x 軸方向の単位ベクトルを示している。方向パラメータ v_y は、 y 軸方向の単位ベクトルを示している。変数 p は、0以上の整数を示している。強度パラメータ σ_x は、 x 軸方向の基本重みを示している。強度パラメータ σ_y は、 y 軸方向の基本重みを示している。

[0106] 例えば、指向パラメータ U としては、図7に示されているように、 $p = 0, 1, 2, 3$ の4通りとし、 $(\sigma_x, \sigma_y) = (1, 3), (3, 1)$ の2通りとすると、 $K = 2 \times 4 = 8$ 通りの指向パラメータ集合 U_k が指定される。

[0107] このとき、その類似性検出装置は、その指向パラメータに基づいて、その指向パラメータが示す集合 U に属する複数の集合に対応する複数のキー計算関数集合を作成する。その複数のキー計算関数集合のうちの集合 U_k に対応するキー計算関数集合は、複数のキー計算関数を示している。その複数のキー計算関数は、既述の実施の形態における類似性検出装置1により集合 U に基づいて算出される複数のキー計算関数と同様にして、集合 U_k に基づいて算出される。

[0108] その類似性検出装置は、入力データ12とその複数のキー計算関数集合とに基づいて、図8に示されているように、その複数のキー計算関数集合52に対応する複数のテーブルセット51を作成する。複数のテーブルセット5

1の任意のデータセットは、複数のデータから形成されている集合を示している。

[0109] その複数のデータは、既述の実施の形態における複数のデータと同様にして、作成されている。すなわち、その複数のデータのうちの任意のデータ面番号 p に対応するデータは、複数のキーを複数のバリューストに対応付けている。

[0110] その複数のキーのうちの任意のキーは、その複数のキー計算関数のうちのデータ面番号 p に対応するキー計算関数 $L^{(p)}(x)$ に、データ集合 X のうちのいずれかのデータ x が代入されることにより算出された値を示している。

[0111] その複数のバリューストのうちの任意のキー L に対応するバリュースト $e(L)$ は、所定のデータの集合であるデータリストを示している。このとき、バリュースト $e(L)$ は、そのデータリストに属するデータのうちの任意のデータ x がキー計算関数 $L^{(p)}(x)$ に代入されることにより算出された値がキー L に一致するように、算出される。

[0112] 図9は、集合 U_k が指定する指向性をそれぞれ示す複数の衝突確率等値線を示している。その複数の衝突確率等値線は、それぞれ、方向パラメータの方向に長軸をもつ楕円に形成され、互いに異なっている。

[0113] その複数の衝突確率等値線のうちの集合 U_k に対応する衝突確率等値線の上のデータは、図6に示される衝突確率等値線31と同様にして、複数のデータのデータ51のうちの集合 U_k に対応するデータセットが示す複数のデータをを用いて、全て同じ確率で検出される。その類似性検出装置は、既述の最良の実施の形態で述べた方法と同様に、それぞれの指向パラメータに対応するデータを指定された面数だけ作成したデータを作成・管理する。

[0114] ユーザは、その検索動作の際に、検索条件16と異なる他の検索条件をその類似性検出装置に入力する。

[0115] その検索条件は、指向性パラメータと検索条件16が示すクエリと条件とを示している。その指向性パラメータは、方向と強度のペアの集合を示して

いる。その類似性検出装置は、複数のテーブルセット 51 からその指向性パラメータに一番近い指向パラメータを示す集合 U_k に対応するテーブルセットを選択する。その類似性検出装置は、そのテーブルセットに属する複数のテーブルを用いて複数の検索データリストを算出し、データリスト $List < x >$ を算出する。

[0116] その複数の検索データリストのうちのあるテーブルセットに対応する検索データリストは、そのテーブルセットの複数のバリューのうちのクエリのキーに対応するバリューが示すデータリストに一致している。データリスト $List < x >$ は、その複数のテーブルセットに対応する複数の検索データリストを含んでいる。

[0117] このような第2の実施の形態における類似性検出装置は、予め複数のテーブルセットを作成し、検索に合わせて利用するテーブルを選択する機能が設けられていることにより、事前に指向性パラメータが不明な場合でも、検索動作時に「この点のこちらの方向にこれくらいの強さで」という指向性パラメータを合わせて入力することで、指向性近傍検索が可能となる。

[0118] なお、ここでは、8通りのみの指向性パラメータを持つテーブルを用意する例を示したが、一般には、もっと多くても良く、様々な方向や強度の組み合わせを用意しておくほど様々な検索要求に対して精度良い出力を返すことができる。そのテーブルセットの個数は、計算資源等とのトレードオフで決めれば良い。

[0119] また、検索の指向性パラメータの指定として、この例では、検索者が指向性パラメータ U の書式（方向と強度のペアの集合）で入力するとしたが、これは、より直感的な入力を定義しデータ処理装置において指向性パラメータの書式に変換しても良い。

[0120] より直感的な入力の例として、関心のない軸方向の単位長さと同じ視する関心がある方向にそった長さを指定することで、検索の指向性を指定することも可能である（例えば x 軸から 45° 傾いた方向に沿った長さ 1 は、 x 軸から -45° 傾いた方向に沿った長さ 3 に相当する場合は、 x 軸から -45

傾いた方向は45°。傾いた方向より実距離で3倍遠くても同じ確率で検出することになる)。

[0121] 更に、本例では、検索における指向性情報に対して、最適なラールを一選択して結果を出力する例を記載したが、複数のラールを組み合わせた指向性近傍検索結果を出力しても良い。具体的には、例えば、検索に指定された指向性情報と、ラールの持つ指向性情報との距離を用いて、その重みで、各ラールで行った近傍検索結果を表示しても良い。

実施例 1

[0122] 本発明による指向性近傍検出方法における第2の実施例について詳細に説明する。本実施例では、ある特徴に関して類似する画像を高速検出する手法について説明する。

[0123] 画像を示す画像データは、一般に、D次元特徴ベクトルで表現することが可能である。様々な手法が考えられるが、ここでは、一例として、図10に示されているように、画像61が適当なメッシュで複数の画像62-1〜62-16に切られており、各画像62-1で画素値を定義することによって表現される画像データを考える。

[0124] ここで、それぞれの画像における画素値をラスタ・スキャン (Raster scan) して、画素値を並べることで特徴ベクトルを作成するとする。なお、ラスタ・スキャンでは、左上からはじめて、左から右に順に値を読んでいる、右端までいったら一段おいて同じことを繰り返す。

[0125] 例えば、4×4の画像に区切られた任意の画像の特徴ベクトルuは、16次元ベクトルで表現される。例えば、16個の単位ベクトル $e_1 \sim e_{16}$ と16個の画素値とを用いて、以下の式により表現される。

[数2]

$$\sum_{i=1}^{16} c_i e_i = n$$

ここで、単位ベクトル e_i は、i番目の成分のみが1を示し、あとの成分が

0を示すベクトルである。

[0126] ここで、ユーザは、ある特徴方向に関してのみ興味があるとする。一番簡単には、ユーザは、画像61のうちのある一部分である興味領域63のみに関心があり、それ以外の類似性を無視するような場合が考えられる。その場合、ユーザは、特徴ベクトル u の第1成分と第2成分と第5成分と第6成分とのみを重視し、他の画素を重視しないことになる。

[0127] ここで、重視する重みを10倍とすると（非興味方向に沿った距離の1に対して興味方向に沿った距離の10を同一視すると）、指向性パラメータ $U = \{u_i = \langle v_i, \sigma_i \rangle \mid i = 1, \dots, 16\}$ を $v_i = e_i, \sigma_i = 10$ （ $i = 1, 2, 5, 6$ ）, 1（その他）と設定する。この指向性パラメータ U を用いて、既述の最良の実施の形態に示した指向性近傍検出を行うことで、あるクエリ画像 q の入力に対して、第1、2、5、6番目の画素の類似性を重視した類似画像検索が可能である。

[0128] 更に、本発明による指向性近傍検出方法は、一般に、画素を適当な割合で組み合わせを特徴方向として類似画像の検出を行うことも可能である。例えば、第1、2、5、6番目の画素値の比率の類似性を重視する場合を考える。

[0129] 興味方向として、第1、2、5、6番目の画素を同じ重みで足し合わせた方向パラメータ v_1 を、以下の式により定義する。

$$v_1 = (e_1 + e_2 + e_5 + e_6) / 2$$

[0130] 方向パラメータ v_1 は、グラム・シュミットの直交化法等を用いれば、方向パラメータ v_1 と $\{e_i \mid i = 2, \dots, 16\}$ とから、互いに直交する正規直交座標系 $\{v_1, v_2, \dots, v_{16}\}$ を構成することができる。

[0131] 先の例と同様に方向パラメータ v_1 の方向を他の方向に比べて10倍重視する場合、指向性パラメータ U を $U = \{u_i = \langle v_i, \sigma_i \rangle \mid (i = 1, \dots, 16) \text{ を } \sigma_i = 10 \text{ (} i = 1 \text{)}, 1 \text{ (その他) と設定することにより、第1, 2, 5, 6番目の画素値の比率が同じ画像を類似画像として検出することができる。これは、例えば画素値として色の濃さを0~1の実数値で表した}$

ときに、対象部分の濃さの比率が近ければ、対象部分以外の画素値の違いが多少大きくても類似画像として検出されることになる。

実施例 2

[0132] 次いで、本発明による指向性近傍検出方法における第2の実施例について詳細に説明する。本実施例では、多次元関数の局所的特徴に着目した類似関数検索を行う。典型的な例として、多次元ベクトルの集合で表現される時系列データ $\{v(t_i)\}$ $i = 1, 2, \dots$ を考える。

[0133] ここで、 $v(t_i)$ は、時刻 t_i における多次元ベクトルを示している。その時系列データとしては、ある地点で、時刻 t_i に測定された加速度センサーのそれぞれ x 軸方向、 y 軸方向、 z 軸方向の加速度の値を3次元ベクトルとして表現したものが例示される。その場合、3次元ベクトル空間上における時刻をパラメータにした点の集合が時系列データとなる。

[0134] 本実施例は、大量な時系列データ（例えば、大量の地点での加速度データの時系列データ）の中で、あるクエリとなる時系列データに対して、その部分的な領域のみで類似する時系列データを高速検索する。

[0135] 図11は、本実施例で検索対象となる複数の時系列データ $71-1 \sim 71-T$ を示している。ユーザは、まず、複数の時系列データ $71-1 \sim 71-T$ を本実施例における類似性検出装置に入力する。その類似性検出装置は、第2の実施の形態における類似性検出装置と同様にして、複数の指向パラメータに対応する複数のテーブルに時系列データの各点を順次登録する。

[0136] ここで、その複数の指向パラメータは、簡単に、 x 軸方向に長軸をもち y 軸方向に短軸をもつ指向パラメータと、その長軸から90度回転した方向に指向性をもつ他の指向パラメータとの2種類とする。指向パラメータの強度は、一般に、複数指定可能だが、ここでは、例えば3:1の比率で設定するものとする。

[0137] その類似性検出装置は、図12に示されているように、その複数の指向パラメータ 74 に対応する複数のテーブルセット 75 を作成する。ここで、複数のテーブルセット 75 の各テーブルセットに属するテーブルの数は、 K 個

とする。その類似性検出装置は、既述の実施の形態における類似性検出装置と同様にして、全てのテーブルに、時系列データの各点に対するキーを計算してバリューに登録する。

[0138] 検索のときは、ユーザは、図13に示されているように、時系列データ76を入力し、時系列データ76のうちの興味領域（検索対象領域）を示す検索範囲パラメータ77をクエリとして指定する。ここで、検索範囲パラメータ77は、時系列データ76の興味領域が指定できれば良い。例えば、座標値の範囲を示すデータでも良いし、各データのパラメータ（例えば、時刻 t 等）の範囲を示すデータでも良い。また、検索範囲パラメータ77は、時系列データ76のなかから興味領域の範囲内のデータのみ抜き出した部分時系列データそのものでも良い。

[0139] その類似性検出装置は、既述の実施の形態における類似性検出装置と同様にして、複数のテーブルセット75のうちのユーザにより選択されたテーブルセットに属する複数のテーブルを参照して、図14に示されているような時系列データ78を抽出し、時系列データ78がユーザ認識可能に表現されるように出力装置15を制御する。時系列データ78は、その複数のテーブルに対応する複数の検索データリストに含まれるデータから形成されている。その複数の検索データリストのうちのあるテーブルに対応する検索データリストは、そのテーブルに対応するキー計算関数にそのクエリが代入されることにより算出されるクエリ値に対応するバリューが示すデータリストを示している。

[0140] 本発明による指向性近傍検出方法における他の実施例について詳細に説明する。本実施例では、範囲限定類似関数検索を行うための、入力される時系列データと任意の時系列データの、その範囲を限定した類似度を判定する機能について、以下に詳細に説明する。簡単のため、ここでは、図15に模式的に示される2次元空間における時系列データ81に対する時系列データ82の類似度を計算することを考える。ここで、検索範囲の限定として、両者の時系列データのうちのユーザにより指定された時間間隔 $[t_{min} : t_m$

a ×」に属する部分のみの局所的な特徴を比較することを考える。

[0141] 時系列データ 81 は、その時間間隔に属するデータとして、 $\{S_i\} : i = 1, \dots, 9$ を有している。時系列データ 82 は、その時間間隔に属するデータとして、 $\{R_i\} : i = 1, \dots, 6$ を有している。このとき、図 16 に示されているように、時系列データ 81 から範囲限定された部分時系列データ 83 は、 $\{S_i\} : i = 1, \dots, 9$ から形成され、時系列データ 82 から範囲限定された部分時系列データ 84 は、 $\{R_i\} : i = 1, \dots, 6$ から形成される。

[0142] ここで、ユーザは、更に、部分時系列データ 83 の点群 $\{S_i\} : i = 1, \dots, 9$ から中心的な参照点 85 をクエリとして抽出する。参照点 85 は、部分時系列データ 83 を最も良く代表する点であり、例えば全ての点の平均でも良いし、平均に一番近い点でも良いし、最大・最小値の平均（この例では S_1 と S_9 の平均）等でも良い。

[0143] 次に、その類似性検出装置は、参照点 85 と、部分時系列データ 83 及び部分時系列データ 84 の各点が複数のテールセット 75 の各々に属する複数のテールにおいて同じエントリに登録される確率（衝突確率）を計算し、複数の指向パラメータ 74 に対応する複数の衝突確率分布を算出する。その複数の衝突確率分布は、それぞれ、参照点 85 から部分時系列データ 83 と部分時系列データ 84 とに属する各データまでの距離の関数を示している。

[0144] 図 17 は、複数の指向パラメータのうちの第 1 指向パラメータに対応する衝突確率等値線を示している。その衝突確率等値線 86 は、 $\times$ 軸方向に長軸をもち y 軸方向に短軸を有している。図 18 は、その複数の指向パラメータのうちの第 2 指向パラメータに対応する衝突確率等値線を示している。その衝突確率等値線 87 は、 $\times$ 軸方向に長軸をもち y 軸方向に短軸を有している。

[0145] 図 19 は、衝突確率分布を示している。その衝突確率分布 88 は、部分時系列データ 83 と部分時系列データ 84 とに属する各データが参照点 85 と

その第1指向パラメータに基づいて算出されたテーブルで同じエントリに登録される衝突確率を示している。

[0146] 図20は、他の衝突確率分布を示している。その衝突確率分布89は、部分時系列データ83と部分時系列データ84とに属する各データが参照点85とその第2指向パラメータに基づいて算出されたテーブルで同じエントリに登録される衝突確率を示している。

[0147] これらの値は、各指向性パラメータの全テーブル数Kに対して、実際に両者が衝突していた（同じエントリに登録されていた）数Lを測定し、 L/K で衝突確率を計算することで近似される。衝突確率分布88と衝突確率分布89とは、時系列データの各点と参照点85の衝突確率が、両者の場所の位置関係と指向性パラメータの示す方向や強度によって異なる値を取ることを示している。衝突確率分布88と衝突確率分布89とは、この衝突確率分布は、比較する二つの部分時系列データが近いほど同じ形をとるため、この形の違いを比較することで任意の二つの部分時系列データの類似性を評価することが可能である。

[0148] 図21は、特徴分布を示している。その特徴分布90は、衝突確率分布88と衝突確率分布89とでの各部分時系列データの衝突確率の差を示し、各部分時系列データの特徴分布90の形は、類似度を部分時系列データの類似度として評価されることができる。この類似度の計算には、滑らかに補間した関数間の内積を計算する方法、もしくは、Earth Movers Distanceのような距離評価尺度を使う方法等が考えられるが、これに限るものではなく、関数の形（ある距離が与えられた時の関数値）の類似度が定量的に評価できるものであれば良い。

[0149] 上記の手法によって、クエリで指定された時系列データに対して、指定された検索範囲に限定した時に、一番類似性の高い時系列データを抽出することが可能である。具体的には、クエリで指定された部分時系列データから参照点を計算し、複数のテーブルセット75に属する全てのテーブルにおける参照点と衝突するデータの中で検索範囲内のもののみを抽出（本例では時間

が指定された時間範囲のものを抽出)し、各時系列データに対応する衝突確率分布を作成し、クエリで指定された部分時系列データの衝突確率分布との類似性を評価すれば良い。

[0150] ここで、本例では検索範囲の指定を連続した一つの区間としたが、一般に複数の区間で行うことも、それぞれの区間で同様の手法を適用し、各区間における類似性を合わせて評価することで容易に実現可能である。

[0151] 次に本実施例における効果を説明する。本実施例によれば、実際のクエリと範囲の入力に対して、検索対象の全ての時系列データの中から最も類似する時系列データの抽出処理を大幅に高速化することが可能である。

[0152] なぜなら、通常的手法では、検索範囲のみを取り出した全ての部分時系列データとクエリの部分時系列データの比較をオンラインで行わなくてはならないため、検索対象データ数が増加すると計算時間も増加し、応答速度が遅くなってしまう。

[0153] しかし、本実施例によれば、検索対象の時系列データは予めオフラインでテーブルに登録しておくことが可能であり、オンライン処理で実際に評価が必要なのは、参照点と衝突するデータによって構成される時系列データ及びその対応する衝突確率分布のみであるからである。特に検索対象時系列データの数が増大の時に、対象データ数の大幅な削減が期待できる。

[0154] なお、本実施例では、時刻をパラメータとして扱ったが、例えば、時刻自体を適当に定数倍したり、時刻に適当な変換を施したりして時系列データの値と比較できる形で座標値の一つとして取り組んで検索を行うことも可能である。

実施例 3

[0155] 次に本発明の第三の実施例について詳細に説明する。第三の実施例では、本発明による類似性検出装置は、図22に示されているように、移動体管理システム100に適用されている。

[0156] 移動体管理システム100は、ネットワーク103を介して、複数のコンピュータが互いに双方向に情報を伝送することができるよう接続されてい

る。ネットワーク 103 としては、インターネット、携帯電話網等が例示される。その複数のコンピュータは、移動体管理装置 101 と複数の移動体端末 102-1 ~ 102-3 とを含んでいる。複数の移動体端末 102-1 ~ 102-3 は、それぞれ、ユーザにより携帯され、現在位置情報を取得できるデバイスである。そのデバイスとしては、スマートフォンや GPS ロガー等が例示される。

[0157] 移動体管理装置 101 は、既述の実施の形態における類似性検出装置を備え、複数の移動体端末 102-1 ~ 102-3 の移動方向を管理し、似た方向に進んでいる移動体端末をグルーピングする機能を提供する。例えば、移動体管理装置 101 は、複数の移動体端末 102-1 ~ 102-3 の各移動体端末に関してその現在位置と適当な時間 τ だけ前にいた位置のペアを管理する。ここで、簡単のため扱う空間を 2 次元空間とし、検出したい主たる移動方向を x 軸方向とし、 y 軸方向に比べ x 軸方向に適当に指向性を持たせるように強度を設定した指向性パラメータを設定する。強度の例として例えば、 x 軸方向へは v_τ の定数倍とし、 y 軸方向は例えば x 軸方向の強度の 0.1 倍等とすることで、 x 軸方向の強度は検出したい速度上限を規定し、 y 軸方向は方向のぶれの許容範囲を規定することができる。

[0158] 複数の移動体端末 102-1 ~ 102-3 は、ある時刻 t にそれぞれ位置 1010、位置 1020、位置 1030 に存在し、時刻 $t + \tau$ にそれぞれ位置 1011、位置 1021、位置 1031 に移動したとする。複数の移動体端末 102-1 ~ 102-3 は、GPS によって現在位置を取得し、定期的にネットワーク 103 を通して移動体管理装置 101 に位置情報をアップロードする。

[0159] 移動体管理装置 101 は、既述の実施の形態における類似性検出装置と同様にして、図 23 に示されているように、複数の指向性パラメータ 111 に対応する複数のテーブルセット 112 を作成する。移動体管理装置 101 は、複数のテーブルセット 112 を用いて複数の移動体端末 102-1 ~ 102-3 の位置を管理し、位置 1010、位置 1011、位置 1020、位置

1021, 位置1030, 位置1031に対して各テーブルのキーを計算し、そのキーに対応するバリューに登録する。移動体管理装置101は、このテーブルセットを用いて、複数の移動体端末102-1~102-3のうちのx軸方向に進む移動体端末をグループ化し、そのグループの特性に関する情報を含む移動体グループ情報105を出力する。

[0160] 次に、移動体管理装置101がこの移動体グループ情報105を作成する手法を説明する。もし、移動体端末がx軸方向に想定速度程度の速度で移動していれば、時刻 t と時刻 $t + \tau$ における位置座標は、同じキーを持つ可能性が高いが、直交するy軸方向に移動している移動体端末は、異なるキーを持つ可能性が高いため、両者を区別することができる。

[0161] 本実施例では、移動体端末102-1と移動体端末102-2の位置情報である位置1010, 位置1011, 位置1020, 位置1021は、同じエントリに登録され、移動体端末102-3の位置情報である位置1030及び位置1031は異なるエントリに登録される可能性が高い。そのため、x軸方向に同じような速度で移動している移動体端末102-1及び移動体端末102-2を同じグループとしてグルーピングすることが可能となる。

[0162] 本例では、1種類の指向性パラメータのみの場合を示したが、一般には、複数の指向性パラメータを指定することで、複数の関心方向や、複数の移動速度上限を設定することが可能であり、様々な移動体端末のグルーピングポリシーを作成することが可能である。

[0163] また、ある移動体端末に注目し、その移動体端末が登録されているテーブルエントリをモニターすることで、この移動体端末と並走する他の移動体端末を検出することもできる。例えば、検索対象移動体端末の入力に対して、過去もしくは現在に、この検索対象移動体端末と同じグループに属する移動体端末を並走移動体端末として出力することができる。

[0164] その際に、必要ならばグルーピングポリシーを合わせて設定することで、そのポリシーに合わせてグルーピングを修正することも可能である。例えば、グルーピングポリシーとして、「検索対象移動体端末と過去10分以上半

径 100 m 以内に存在しながら並走した移動体端末」のように設定すれば、本発明によって出力された移動体端末グループの中から条件を満足させるもののみをフィルタリングして出力すれば良い。

[0165] 更に、移動体端末の数が大量であった時は、各テーブルの各エントリに登録された移動体端末のうち、時刻 t と時刻 $t + \tau$ における位置のどちらも同じエントリに登録されているものの数を数え、その数の多い順に適当な数だけエントリに登録された移動体端末の集合を抽出する。

[0166] これは、全移動体端末の中から、同じ方向に沿って進む移動体端末をその合計数が多いグループからサンプリングすることを意味する。それらのサンプリングされた移動体端末をその移動方向や速度からグルーピングすることで、システム全体の移動体端末移動パターンの主たる傾向（多くの移動体端末がどの方向にどれくらいの速度で移動しているか、等）を取り出して、その移動体端末グループのサイズや方向・速度等の移動体端末グループ情報を作成し、移動体端末の移動状態を分類・表示ことが可能である。

[0167] 次に、本実施例の効果について述べる。移動体管理システム 100 は、基本的に、移動体端末の位置アップデートに対して管理するテーブルを更新し、上記の方法でテーブルから関心の高い移動体端末をサンプリングしてグルーピングを行うため、全ての移動体端末の移動に関する相互関係を計算してグルーピングする必要がない。そのため移動体端末数が増加してもその計算負荷の増加を抑えることができ、計算時間の大幅な削減が望める。

[0168] なお、本発明で示した全ての実施の形態及び実施例においては、テーブルのキーの計算方法として、既述の実施の形態におけるキー計算関数 $L^{(p)}(x)$ と基本関数 $f^{(p)}_0(x)$ とランダムベクトル $\Phi^{(p)}_0$ とを用いたが、指向性を指定するパラメータによって、望む方向・強度に対応する非対称性が導入されたものであればこれに限るものではない。

[0169] なお、本発明で示した全ての実施の形態及び実施例では、テーブルにおけるデータの衝突は、厳密に同じエントリに登録されるとして説明したが、バリエーションとしてテーブルのエントリ距離を定義し、近くのエントリ（例

例えば、キー計算関数 $L^{(p)}(x)$ であらわされるキーの一ビットだけ異なる「隣のエントリ」等）までに登録されているデータは衝突すると衝突範囲を拡大しても良い。

[0170] <備考>

以上、本発明の実施の形態及び実施例を詳述してきたが、実際には、上記の実施の形態及び実施例に限られるものではなく、本発明の要旨を逸脱しない範囲の変更があっても本発明に含まれる。

[0171] <付記>

上記の実施の形態及び実施例の一部又は全部は、以下の付記のように記載することも可能である。但し、実際には、以下の記載例に限定されない。

[0172] [付記1]

入力装置を介して入力された複数の方向パラメータと複数の強度パラメータとに基づいて複数の乱数情報を算出する乱数発生部と、

前記複数の乱数情報に基づいて複数のキー計算関数を算出する初期化部と、

入力装置を介して入力された複数の検索対象データに基づいて、前記複数のキー計算関数に対応する複数のテーブルを算出し、前記複数のテーブルをテーブル保持装置に記録するデータ登録部と、

前記複数のテーブルを参照して、入力装置を介して入力された検索条件が示すクエリに基づいて候補データリストを算出する検索部と、

前記候補データリストに属する複数の検索データから前記検索条件が示す条件を満足する検索結果データを算出し、前記検索結果データを出力装置に出力するデータ処理部と

を具備し、

前記データ登録部は、前記複数のテーブルのうちの任意のキー計算関数に対応するテーブルを算出する際、複数のキーを複数のデータリストに対応付け、前記複数のデータリストのうちの任意のキーに対応するデータリストに属するデータを前記任意のキー計算関数に代入することにより算出される値

が前記任意のキーに等しくなるように該テーブルを算出し、

前記候補データリストは、前記複数のキー計算関数に対応する複数の検索データリストを含み、

前記複数の検索データリストのうちの前記任意のキー計算関数に対応する検索データリストは、前記複数のデータリストのうちの、前記クエリが前記キー計算関数に代入されることにより算出されるクエリ値に対応するデータリストを示す

類似性検出装置。

[0173] [付記 2]

付記 1 に記載の類似性検出装置であって、

前記データ登録部は、前記複数の検索対象データに基づいて複数のテーブルセットを更に算出し、

前記検索部は、前記複数のテーブルセットのうちの前記検索条件が示すテーブルセットが前記複数のテーブルを示すときに、前記複数のテーブルを参照して前記候補データリストを算出する

類似性検出装置。

[0174] [付記 3]

付記 1 乃至 2 のいずれかに記載の類似性検出装置であって、

前記複数の方向パラメータは、前記複数の検索対象データのうちの類似する 2 つの状態変化類似データが前記任意のキー計算関数に代入されることによりそれぞれ算出される 2 つの値が等しくなるように、又は、前記 2 つの値の差が所定の値より小さくなるように、設定される

類似性検出装置。

[0175] [付記 4]

付記 1 乃至 3 のいずれかに記載の類似性検出装置であって、

前記複数の検索対象データは、それぞれ画像を示す

類似性検出装置。

[0176] [付記 5]

付記 1 乃至 4 のいずれかに記載の類似性検出装置であって、
前記データ登録部は、前記複数の検索対象データが更新されたときに、前記複数のテーブルを更新する
類似性検出装置。

[0177] [付記 6]

入力装置を介して入力された複数の方向パラメータと複数の強度パラメータとに基づいて複数の乱数情報を算出するステップと、
前記複数の乱数情報に基づいて複数のキー計算関数を算出するステップと、
、

入力装置を介して入力された複数の検索対象データに基づいて、前記複数のキー計算関数に対応する複数のテーブルを算出するステップと、

前記複数のテーブルをテーブル保持装置に記録するステップと、

前記複数のテーブルを参照して、入力装置を介して入力された検索条件が示すクエリに基づいて候補データリストを算出するステップと、

前記候補データリストに属する複数の検索データから前記検索条件が示す条件を満足する検索結果データを算出するステップと、

前記検索結果データを出力装置に出力するステップと、

前記複数のテーブルのうちの任意のキー計算関数に対応するテーブルを算出する際、複数のキーを複数のデータリストに対応付け、前記複数のデータリストのうちの任意のキーに対応するデータリストに属するデータを前記任意のキー計算関数に代入することにより算出される値が前記任意のキーに等しくなるように該テーブルを算出するステップと

を具備し、

前記候補データリストは、前記複数のキー計算関数に対応する複数の検索データリストを含み、

前記複数の検索データリストのうちの前記任意のキー計算関数に対応する検索データリストは、前記複数のデータリストのうちの、前記クエリが前記キー計算関数に代入されることにより算出されるクエリ値に対応するデータ

リストを示す

指向性近傍検出方法。

[0178] [付記 7]

付記 6 に記載の指向性近傍検出方法であって、

前記複数の検索対象データに基づいて複数のテーブルセットを更に算出するステップと、

前記複数のテーブルセットのうちの前記検索条件が示すテーブルセットが前記複数のテーブルを示すときに、前記複数のテーブルを参照して前記候補データリストを算出するステップと

を更に具備する

指向性近傍検出方法。

[0179] [付記 8]

付記 6 乃至 7 のいずれかに記載の指向性近傍検出方法であって、

前記複数の方向パラメータは、前記複数の検索対象データのうちの類似する 2 つの状態変化類似データが前記任意のキー計算関数に代入されることによりそれぞれ算出される 2 つの値が等しくなるように、又は、前記 2 つの値の差が所定の値より小さくなるように、設定される

指向性近傍検出方法。

[0180] [付記 9]

付記 6 乃至 8 のいずれかに記載の指向性近傍検出方法であって、

前記複数の検索対象データは、それぞれ画像を示す

指向性近傍検出方法。

[0181] [付記 10]

付記 6 乃至 9 のいずれかに記載の指向性近傍検出方法であって、

前記複数の検索対象データが更新されたときに、前記複数のテーブルを更新するステップ

を更に具備する

指向性近傍検出方法。

[0182] [付記 1 1]

大量の検索対象データに対して、任意の類似度判定基準で類似するデータを高速に検出する類似性検出装置であり、

前述の類似性検出装置は、前述の検索対象データを登録・管理するための、類似度判定基準を設定するパラメータによって作成される乱数情報と関連付けられたテーブルを保持し、

任意の参照点からの近傍検索要求に対して、前述のテーブルを利用し、前述の設定された類似度判定基準によって類似性判断を行い、近傍と判断されるデータを出力する

ことを特徴とする類似性検出装置。

[0183] [付記 1 2]

付記 1 1 に記載の類似性検出装置において、複数の異なる類似度判定基準を有し、それぞれの類似度判定基準と関連付けられた複数のテーブルを有し、検索対象データを全てのテーブルで並列に管理する

ことを特徴とする類似性検出装置。

[0184] [付記 1 3]

付記 1 1 に記載の類似度判定基準として、興味の方角と、その方角の重要度パラメータの組の集合とする

ことを特徴とする類似性検出装置。

[0185] [付記 1 4]

付記 1 1 に記載の検索対象データとして、任意の 2 点の距離がユークリッド空間上の距離として定義されている

ことを特徴とする類似性検出装置。

[0186] [付記 1 5]

付記 1 2 に記載の類似性検出装置において、検索においてその検索中心とともに検索類似度判定基準を入力し、複数のテーブルにおける近傍検出結果と、各テーブルに指定された類似度判定基準パラメータと検索類似度判定基準との関連性を用いて出力する近傍を決定する

ことを特徴とする類似性検出装置。

[0187] [付記 1 6]

付記 1 1 又は付記 1 2 に記載の検索対象データとして、予めその値が静的に固定されたデータだけでなく、動的にその値を更新するデータであり、値の更新に合わせて前述のテーブルの登録状態を更新する

ことを特徴とする類似性検出装置。

[0188] [付記 1 7]

付記 1 1 又は付記 1 2 に記載の類似性検出装置として、前述のテーブルを用いた類似性判断結果を利用して、各データの状態変化が類似するものをグルーピングする

ことを特徴とする類似性検出装置。

[0189] [付記 1 8]

付記 1 1 に記載のテーブルは、キーとバリューのペアで構成され、ある任意のデータの登録に当たって、前記乱数情報を用いてデータを登録すべきキーを計算し、対応するバリューに当たるデータのリストに追加登録することでデータを管理する

ことを特徴とする類似性検出装置。

[0190] [付記 1 9]

付記 1 1 に記載の類似性判断として、同一もしくは差異の小さいキーを持つデータ同士を類似性が高いと扱う

ことを特徴とする類似性検出装置。

[0191] なお、本出願は、日本出願番号 2 0 1 1 - 2 1 9 5 4 7 に基づく優先権を主張するものであり、日本出願番号 2 0 1 1 - 2 1 9 5 4 7 における開示内容は引用により本出願に組み込まれる。

請求の範囲

- [請求項1] 検索対象データに対して、任意の類似度判定基準で類似するデータを検出する類似性検出装置であって、
- 類似度判定基準を設定するパラメータによって作成される乱数情報と関連付けられたテーブルを保持する手段と、
- 検索対象データを前記テーブルに登録し管理する手段と、
- 任意の参照点からの近傍検索要求に対して、前記テーブルを利用し、前記設定された類似度判定基準によって類似性判断を行い、近傍と判断されるデータを出力する手段と
- を具備する
- 類似性検出装置。
- [請求項2] 請求項1に記載の類似性検出装置であって、
- 複数の異なる類似度判定基準を設定する手段と、
- 前記複数の異なる類似度判定基準のそれぞれと関連付けられた複数のテーブルを保持する手段と、
- 前記検索対象データを前記複数のテーブルで並列に管理する手段と
- を更に具備する
- 類似性検出装置。
- [請求項3] 請求項1又は2に記載の類似性検出装置であって、
- 興味の方向と、該方向の重要度パラメータとの組の集合を、前記類似度判定基準とする手段
- を更に具備する
- 類似性検出装置。
- [請求項4] 請求項1乃至3のいずれか一項に記載の類似性検出装置であって、
- 任意の2点の距離がユークリッド空間上の距離として定義されているデータを、前記検索対象データとする手段
- を更に具備する
- 類似性検出装置。

- [請求項5] 請求項 1 乃至 4 のいずれか一項に記載の類似性検出装置であって、
 検索の中心とともに検索類似度判定基準を入力し、前記複数のテーブルにおける近傍検出結果と、前記複数のテーブルの各々に指定された類似度判定基準パラメータと検索類似度判定基準との関連性を用いて出力する近傍を決定する手段
 を更に具備する
 類似性検出装置。
- [請求項6] 請求項 1 乃至 5 のいずれか一項に記載の類似性検出装置であって、
 予め値が静的に固定されたデータ、及び動的に値を更新するデータのいずれも、前記検索対象データとして使用する手段と、
 前記検索対象データとして使用されるデータの値の更新に合わせて前記テーブルの登録状態を更新する手段と
 を更に具備する
 類似性検出装置。
- [請求項7] 請求項 1 乃至 6 のいずれか一項に記載の類似性検出装置であって、
 前記テーブルを用いた類似性判断結果を利用して、状態変化が類似するデータをグルーピングする手段
 を更に具備する
 類似性検出装置。
- [請求項8] 請求項 1 乃至 7 のいずれか一項に記載の類似性検出装置であって、
 前記テーブルをキーとバリューとの組で構成し、ある任意のデータの登録に当たって、前記乱数情報を用いてデータを登録すべきキーを計算し、対応するバリューに当たるデータのリストに追加登録することでデータを管理する手段
 を更に具備する
 類似性検出装置。
- [請求項9] 請求項 1 乃至 8 のいずれか一項に記載の類似性検出装置であって、
 同一もしくは差異の小さいキーを持つデータ同士を類似性が高いと

判断する手段
を更に具備する
類似性検出装置。

[請求項10] 類似性検出装置により実施され、検索対象データに対して、任意の類似度判定基準で類似するデータを検出するための指向性近傍検出方法であって、

類似度判定基準を設定するパラメータによって作成される乱数情報と関連付けられたテーブルを保持することと、

検索対象データを前記テーブルに登録し管理することと、

任意の参照点からの近傍検索要求に対して、前記テーブルを利用し、前記設定された類似度判定基準によって類似性判断を行い、近傍と判断されるデータを出力することと

を含む

指向性近傍検出方法。