

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5782404号
(P5782404)

(45) 発行日 平成27年9月24日(2015.9.24)

(24) 登録日 平成27年7月24日(2015.7.24)

(51) Int.Cl.

G 0 6 T 7/00 (2006.01)

F 1

G O 6 T 7/00

Q

請求項の数 5 (全 26 頁)

(21) 出願番号 特願2012-87667 (P2012-87667)
 (22) 出願日 平成24年4月6日 (2012.4.6)
 (65) 公開番号 特開2012-226744 (P2012-226744A)
 (43) 公開日 平成24年11月15日 (2012.11.15)
 審査請求日 平成27年3月26日 (2015.3.26)
 (31) 優先権主張番号 13/089, 460
 (32) 優先日 平成23年4月19日 (2011.4.19)
 (33) 優先権主張国 米国 (US)

早期審査対象出願

(73) 特許権者 596170170
 ゼロックス コーポレイション
 XEROX CORPORATION
 アメリカ合衆国、コネチカット州 068
 56、ノーウォーク、ピーオーボックス
 4505、グローバー・アヴェニュー 4
 5
 (74) 代理人 100079049
 弁理士 中島 淳
 (74) 代理人 100084995
 弁理士 加藤 和詳
 (72) 発明者 ルカ・マルケソッティ
 フランス国 グレノーブル 38000
 リュ・サン・ジャック 15

最終頁に続く

(54) 【発明の名称】 画質評価

(57) 【特許請求の範囲】

【請求項 1】

画像の画質を予測する方法であって、
 入力画像に関して、
 前記入力画像に関する意味内容の情報に基づいて第一の記述子を生成する工程と、
 前記入力画像から抽出される美的特徴に基づいて第二の記述子を生成する工程と、
 前記第一の記述子及び前記第二の記述子に基づいて品質値を前記入力画像に割り当てる
 ように訓練されたカテゴライザを用いて、前記第一の記述子及び前記第二の記述子に基づ
 いて前記品質値を前記入力画像に割り当てる工程と、
 前記入力画像の品質値を出力する工程と、
 を含む方法。

【請求項 2】

前記第一の記述子を生成する工程、前記第二の記述子を生成する工程、及び前記品質値
 を割り当てる工程は、コンピュータプロセッサを用いて行われる、請求項 1 に記載の方法
 。

【請求項 3】

前記第一の記述子は、前記入力画像と前記入力画像の画像データに関連する文字情報と
 のうちの少なくとも 1 つから生じる、請求項 1 に記載の方法。

【請求項 4】

前記文字情報は、前記入力画像に関連する文字標識と前記入力画像に組み込まれたメタ

データとのうちの少なくとも 1 つを含む、請求項 3 に記載の方法。

【請求項 5】

前記第一の記述子は、複数のコンテンツカテゴリ毎に値を含む、請求項 1 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本明細書に開示されている例示的な実施形態の様態は、写真画像の品質を評価するシステム及び方法に関し、美的特徴だけでなくコンテンツ特徴にも基づいて画質を予測する、自動システムにおける特定な用途を見出すものである。

10

【背景技術】

【0002】

プロやアマチュアの写真家により作られるデジタル写真画像は、ますます増えてきている。公衆のウェブサイトを通じてこれらの画像にアクセスすることができ、ウェブサイトの閲覧者がその品質やその他の特徴を評価することができる。

【0003】

画像品質評価の分野では、知覚画質を自動的に予測できる品質測定法を考案するために相当な努力がなされてきた。1つの目的は人間の好みと良い相関を有するデジタル画像から記述子を抽出することであった。過去には不規則な又は計画されたノイズ（例えば、ごま塩ノイズ、JPEGアーティファクト、リングング）及びブラー等の特定の信号レベルの劣化の有無が、写真画像の品質を規定するうえで一般に使用してきた。しかし、現在では高解像度デジタルセンサが容易に使用でき、写真家はこれらの劣化を克服できる。最近の画像品質評価では、低いレベルの画像品質よりも、高いレベルの特徴を評価することに焦点があてられてきている。以下を参照されたい。2006年ECCV(3)でのR. Datta et al.による「*Study ing aesthetics in photographic images using a computational approach*」の288～301ページ（以下「Datta2006」とする）、2007年MULTIMEDIA '07: Proc. 15th Intern'l Conf. on MultimediaでのR. Datta et al.による「*Learning the consensus on visual quality for next-generation image management*」の533～536ページ（以下「Datta2007」とする）、2008年10月15th IEEE Intern'l Conf. on image ProcessingでのR. Datta et al.による「*Algorithmic inferring of aesthetics and emotion in natural images: An exposition,*」の105～108ページ。

20

【発明の概要】

【発明が解決しようとする課題】

【0004】

画質に関する特徴は、画像のコンテンツには直接かかわりはなく、画像の知覚品質に影響を与える配色、構図、構成等の特定な視覚要素をとらえるために考案されているために、よく美的特徴と呼ばれる。

40

【0005】

訓練データとして使用可能な、ソーシャルネットワーク、写真共有ウェブサイト等を通じてアクセスできる注釈付の画像データが増えてきているが、ハイレベルの品質評価に関する課題は依然として残っている。第一にこれらのデータは大抵、固有のノイズと共に注釈を受けられているが、人間の好みを扱う場合、全員が一致することはあまりない。

【0006】

第二の課題は人間の好みをとらえるための特徴を考案することである。現時点で使用されている特徴は、人間の知覚と密接にかかわりあっているとは限らない。すなわち、それ

50

らは閲覧者が画質を評価するのに使用する全ての視覚情報をとらえるのに十分なほど強力ではない。

【0007】

自動の画像品質評価を改善できるシステム及び方法の必要性が残されている。

【課題を解決するための手段】

【0008】

例示的な実施形態の一様態によると、画像の画質を予測する方法は入力画像に関して、画像に関する意味内容情報に基づく第一の記述子を生成する工程と、画像から抽出される美的特徴に基づく第二の記述子を生成する工程とを含む。第一の記述子及び第二の記述子に基づいて品質値を画像に割り当てるために訓練されたカテゴライザを用いて、第一の記述子及び第二の記述子に基づいて品質値を画像に割り当て、画質値を出力する。10

【0009】

第一の記述子は、画像及び画像の画像データに関連する文字情報のうちの少なくとも1つから生じ得る。

【0010】

文字情報は画像及び画像に組み込まれたメタデータに関連する文字標識のうちの少なくとも一方を含むことができる。

【0011】

第一の記述子は複数のコンテンツカテゴリ毎に値を含むことができる。

【0012】

この方法は入力画像からローカル特徴のセット抽出する工程と、ローカル特徴の分布を示す画像表現を生成する工程と、及び画像表現に基づいて複数のコンテンツカテゴリ毎に値を割り当てる工程とを含むことができる。20

【0013】

第二の記述子を生成する工程は画像のパッチから複数の美的特徴を抽出することを含むことができる。

【0014】

第二の記述子は、指向勾配のヒストグラム（HOG）特徴、スケール不变特徴変換（SIFT）特徴、及びガボール特徴のうちの少なくとも1つを含む一般特徴をさらに含むことができる。30

【0015】

ガボール特徴は、ガウスエンベロープにより変調された特定な周波数及び方向性を持つ正弦平面でよい。

【0016】

品質値を割り当てる工程は、随意に重み付けされた第一の記述子及び第二の記述子の組み合わせである組み合わせ記述子を生成することを含むことができる。

【0017】

組み合わせ記述子を以下の形をとる集計関数と組み合わせてもよい。

【数1】

$$x_t = [(1 - \alpha)x_a, \alpha x_c] \quad (4) \quad \text{span style="float: right;">40$$

但し

x_c は第一の記述子を示す

x_a は第二の記述子を示す

$0 < \alpha < 1$

【0018】

品質値を割り当てる工程は、の分類子のセットを用いて第二の記述子をコンテンツカテゴリごとに1つ分類して、コンテンツカテゴリ毎に各分類子スコアを生成し、重み付けした分類子スコアの組み合わせに基づく品質値を生成することを含むことができ、分類子スコア毎の重みは第一の記述子の各特徴の値から生じる。50

【 0 0 1 9 】

方法は訓練画像のセットを用いてカテゴライザを訓練することをさらに含むことができ、各訓練画像は関連する品質情報を有し、訓練には訓練画像毎に、

- a) 品質情報から品質値を生成することと、
- b) コンテンツ特徴の記述子を生成することであって、
- i) 画像に関連するコンテンツ情報、及び
- ii) 画像の画像データに基づいて意味的カテゴライザにより出力されるカテゴリ情報のうちの少なくとも一方に基づいて生成すること、
- c) 画像から抽出される美的特徴に基づいて美的特徴の記述子を生成することと、訓練セット内の各画像の品質値、コンテンツ特徴の記述子、及び美的特徴の記述子上のカテゴリライザを訓練することが含まれる。

【 0 0 2 0 】

品質情報には、画像毎に閲覧者の入力による品質に関連する序列又は画像の評価のうちの少なくとも 1 つが含まれ得る。

【 0 0 2 1 】

方法は情報検索システムにおける画像の検索のためのパラメータとして品質値を使用する工程をさらに含むことができる。

【 0 0 2 2 】

コンピュータプログラム製品は、コンピュータで実行したとき上記の方法を行う命令を符号化する非一時的記憶媒体を含むことができる。

10

【 0 0 2 3 】

上記の方法を行うシステムは、方法を行う命令を格納する非一時的メモリ及び命令を実行するコンピュータプロセッサを含むことができる。

【 0 0 2 4 】

別の様態では、画像に関する画像品質評価を生成するシステムは入力画像に関するコンテンツ情報に基づいて第一の記述子を生成するコンテンツ特徴エクストラクタと、画像から抽出される美的特徴に基づいて第二の記述子を生成する美的特徴エクストラクタと、第一の記述子及び第二の記述子に基づいて品質値を画像に割り当てるために訓練されたカテゴリライザとを格納するメモリを含む。コンピュータプロセッサは、エクストラクタとカテゴリライザを実装するメモリと通信する。

30

【 0 0 2 5 】

カテゴリライザは、少なくとも 1 つ分類子と、前期融合及び後期融合のうちの一方を行う融合モジュールと、を含むことができる。

【 0 0 2 6 】

融合モジュールは、組み合わせ記述子を生成することを含む前期融合を行うことができ、この組み合わせ記述子とは、第一の記述子と第二の記述子の随意的に重み付けされた組み合わせであり分類子に入力される。

【 0 0 2 7 】

少なくとも 1 つの分類子は複数のコンテンツカテゴリのうちの 1 つ毎に 1 つの分類子を含む複数の分類子を含むことができる。融合モジュールは、第二の記述子に基づいて複数の分類子により出力される複数の分類子スコアの重み付けされた組み合わせに基づいて品質値を生成することを含む後期融合を行い、それぞれの分類子スコアに関する重みは第一の記述子の各特徴値から生じる。

40

【 0 0 2 8 】

別の様態では、画像の画質を予測する方法が訓練画像のセットを受け取る工程を含み、セットの各画像は関連する画像の意味内容を示す文字情報と、ユーザにより提供される画質情報に基づく画質値とを有する。カテゴリライザは、画像から抽出される画像の標識、その画質値、及び特徴に基づいて訓練される。文字情報を有するが画質値のない入力画像に関して、この方法は画像の画像データから特徴を抽出する工程と、カテゴリライザを用いて文字情報及び抽出される特徴に基づいて品質値を画像に割り当てる工程とを含んでいる。

50

画質値が出力値である。

【図面の簡単な説明】

【0029】

【図1】訓練中及び使用中における例示的なカテゴライザへの入力の概略図である。

【図2】例示的な実施形態の一様態による画像品質評価に関するシステムの機能ブロック図である。

【図3】例示的な実施形態の別の様態による画像品質評価の方法を示すフローチャートである。

【図4】例示的な方法及び比較の方法により生成された結果を示すグラフである。

【図5】後期融合法における異なる の値を選択する影響を示すグラフである。 10

【図6】異なる数の方向を用いたガボール特徴の使用の精度への影響を示すグラフである。
。

【発明を実施するための形態】

【0030】

例示的な実施形態は画像品質評価のためのコンピュータで実行するシステム及び方法に関する。例示的なシステム及び方法は画像品質評価における特徴の組み合わせ及び異種の情報の統合を組み込む。画像の全特性を示す低レベルの画像特徴、及び画像領域の特性を示す高レベルの特徴の2種類の特徴が考えられる。2つの異なる種類の特徴に基づいて分類を組み合わせるために、2つの分類子融合法が提案される。第一に、組み合わせは低レベルの特徴とハイレベルの特徴の随意的に重み付けされた融合として実現される。第二に、組み合わせはコンテンツのクラス特定のスコアの後期融合における重みを付けて実現される。 20

【0031】

データから直接人間の好みを自動予測することが有効であり得る事例が数多く存在する。例えば、画像検索システムでは、従来のコンテンツベースの検索方法の結果を例示的なシステムを用いて改良又はまとめることができる。また、システムを直接カメラに配置して所与のシーンに写し出されたものに対してリアルタイムで提案を行うことも可能である。
。

【0032】

図1は例示的な実施形態の様態を概略的に示す。図2は画像品質評価を画像12に割り当てる例示的なシステム10を示す。図3は画像品質評価を画像に割り当てる例示的な方法を示し、この画像は平均的な閲覧者の知覚画質を一致する知覚画質を提供することを目的とする。図1に示す通り、実行時間では、例示的なシステム10はコンテンツ特徴14と美的特徴16の両方を含むことができ、分類される画像12に関する特徴のセットを生成する。組合せカテゴライザ18は、画像品質評価20を入力画像12に割り当てるために訓練されている。評価は、コンテンツ特徴（ハイレベル特徴）14と美的特徴（低レベル特徴）16の関数である品質値20の形をとる。 30

【0033】

従って、画像品質評価20は一連の人間の閲覧者により知覚される画質を予測する、画像12の品質の自動評価である。本システムにより出力される画像品質評価20は、例えば、0は最も低い画質、すなわち非常に悪い画像を示し、1は最も高い画像、すなわち非常に良い画像を示す0から1の段階等の、所定の最大値と最小値の範囲内の実数の値であるスコアでよい。つまり星評価（例えば、1つ星は最も低くK星は最も高い、Kは5～10の所定の最大値）のような量子化されたスコアである。別の実施形態では、値20は序列（例えば、1最も高い画質に対応し、その他の画像は高い値を有するに従って低い序列に対応する）又は可能な言語による表現（「低品質」、「平均品質」、「良品質」、及び「優良品質」等の）の有限集合から選択される画質の言語による表現でよい。 40

【0034】

コンテンツ特徴14とはカテゴリにより画像の目で見える知覚コンテンツを示す。意味内容のカテゴリのセットを規定することができ、それぞれいくつかの入力画像12内に存 50

在することが期待され得る概念を示す。一例として、あるカテゴリのセットは以下の概念を表すことができる。動物、風景、人々、都市、花、対象等。少なくとも 5 つのカテゴリ等のあらゆる数のカテゴリを提供できることを理解されたい。例えば、10 のコンテンツカテゴリまでよく、またいくつかの実施形態では 20 まであり、又は 100 コンテンツカテゴリまでよく、それより多くてもよい。画像 12 には、1 つ以上のコンテンツ特徴 14 に基づいてカテゴリ毎に値を含む特徴ベクトルの形をとるコンテンツ特徴の記述子 22 を割り当てることができる。画像のコンテンツ上の意味、画像に組み込まれるメタデータ及び / 又は画像に関連する GPS 又はその他の位置情報を示す文字標識等の画像 12 に関連する文字情報 23 からコンテンツ特徴 14 を抽出することができる。あるいは又はさらに、以下に記載する通りコンテンツ特徴を画像データ自体、すなわち画像画素から、例えば、意味的カテゴリライザを用いて抽出することもできる。コンテンツ特徴の記述子を形成する値は、量子化することができる、例えば、二値 (0 又は 1) 又は [0, 1] の範囲等の所定の範囲内の実数の値でよい。

【0035】

美的特徴 16 は画像の画素に基づいて抽出することができる特徴である。全ての好適な美的特徴の組み合わせを選択することができ、それによりブラー、粒状感、コントラスト、彩度、配色、三分割法の順守、独自性、補色の使用、被写界深度、美しい形状の外観、画像サイズ、画像アスペクト比、赤目又は他のアーティファクトの存在等の知覚画質問題の範囲を評価する。

【0036】

カテゴリライザ 18 は、100 から 100 万の訓練画像等の訓練画像 30、32、34 のセットを用いて訓練される。各訓練画像 30、32、34 は画素の配列を有する画像データ 36 を含む。訓練画像はそれぞれ人間の批評家により割り当てられるスコア（又は複数のスコア）等の品質情報 38 で標識される。訓練画像 30、32、34 を、批評家により割り当てられるカテゴリ等のコンテンツ情報 40（情報 23 に類似する）で標識することができる。品質情報 38 及びコンテンツ情報 40 のうちの 1 つ以上に基づいて訓練画像の削減したセット 42 を訓練のために選択することができる。

【0037】

コンテンツ特徴 50 及び美的特徴 52 等の特徴と正規化された品質値 54 が画像毎に生成される。コンテンツと美的特徴の記述子 56、58（分類される画像 12 の特徴 14、16 と同じ形で）がカテゴリライザ 18 に入力される。カテゴリライザ 18 は、入力された特徴の記述子 56、58 及び訓練画像 30、32、34 等のそれぞれの関連する品質値 54 に基づいて、品質値 20 を新しい画像 12 に割り当てるなどを覚える。

【0038】

次に図 2 を参照すると、画質システム 10 が動作する典型的な環境が示されている。システム 10 は図示されたサーバコンピュータ等の 1 つ以上のコンピュータ装置を備え、メインメモリ 60 及びデータメモリ 62 として図示されているメモリ、コンピュータの CPU 等のプロセッサ 64、及び他の装置と通信するための 1 つ以上のネットワークインターフェース（I/O）66 を含み、全てデータ / コミュニケーションバス 68 により通信可能に接続されている。コンピュータは、デスクトップ、ラップトップ等の PC、パームトップコンピュータ、携帯情報端末（PDA）、デジタルカメラ、サーバコンピュータ、携帯電話、タブレットコンピュータ、ポケットベル、又は例示的な方法を行うための命令を実行できるその他のコンピュータ装置でよい。

【0039】

デジタルプロセッサ 64 をシングルコアプロセッサ、デュアルコアプロセッサ（又はより一般的にマルチコアプロセッサにより）、デジタルプロセッサ及び協働数値演算プロセッサ、デジタルコントローラ等によりさまざまに組み込むことができる。

【0040】

メモリ又は一連のメモリ 60、62 は、ランダム・アクセス・メモリ（RAM）、リード・オンリー・メモリ（ROM）、磁気ディスク又はテープ、光ディスク、フラッシュメ

10

20

30

40

50

モリ、又はホログラムメモリ、又はそれらの組み合わせ等のあらゆる種類の非一時的コンピュータ可読媒体でよい。下記に示す通り、メモリ60はサーバコンピュータの動作及び画質を評価する例示的な方法を行うための命令を格納することができる。メモリ62は、例示的な方法により処理される画像12及び処理されたデータ20を格納する。

【0041】

ネットワークインターフェース66は、コンピュータがコンピュータネットワーク70、例えばローカル・エリア・ネットワーク(ＬＡＮ)、インターネット、電話回線、有線接続、又はそれらの組み合わせ等のワイド・エリア・ネットワーク(ＷＡΝ)等の有線又は無線接続を介して、他の装置と通信することを可能にし、変調器/復調器(ＭＯＤＥＭ)を含むことができる。10

【0042】

分類される画像12は、PC、ラップトップ、カメラ、携帯電話等の汎用又は専用コンピュータ装置等の画像のソース80から、又はフラッシュドライブ、ディスク、ポータブル・ハード・ドライブ、カメラ・メモリ・スティック等の非一時的メモリ格納装置からシステム10に入力される。例示的な実施形態では、ソース80は画像12及びサーバコンピュータにホストされたウェブポータルに画像をアップロードするためのウェブブラウザ84を格納するメモリ82を含むクライアントのコンピュータ装置であり、プロセッサ86により実行される。メモリ82及びプロセッサ86をメモリ60、62及びプロセッサ64と同じように設定することができる。1つ以上のインターフェース88、90により、コンピュータ80はシステム10及びコンピュータモニタ、LCDスクリーン等のディスプレイ装置92、及びタッチスクリーン、キーボード、キーパッド、カーソルコントロール・装置又はそれらの組み合わせ等の1つ以上のユーザ入力装置94と通信することができる。20

【0043】

訓練画像32、34、36等の訓練画像のセットは、データベース(又は複数のデータベース)100内に格納される。データベース100をメモリ62又は、システム10に(例えば、ネットワーク70を介して)アクセス可能なメモリ内に格納することができる。。

【0044】

システム10は、画像12、32、34、36を、JPEG、TIFF、GIF、JBIG、BMP又はその他の画像に使用される共通のファイル形式等のあらゆる便宜なファイル形式で受け取ることができ、処理をする前にそれらを随意的に別の好適な形式に変換することができる。画像は、文字/図及び写真等を含む各写真、ビデオ画像、三次元画像、組合せ画像でよい。一般に、各入力デジタル画像は、画像を形成する画素の配列に関する画像データを含む。画像データは、RGB等の色分解のセットごとにグレースケール値等の色の値を含むことができる、又は、画像データを異なる色で表すことができる異なる色空間で表すことができる。一般に、「グレースケール」とは、あらゆる単一の色分解の光学濃度の値のことを指すが、表示される(RGB、L*a*b*、YCbCr、等)。例示的なシステム及び方法はモノクロ画像(単一色分解)及び多色画像(複数の色分解)に適用できる。本明細書に使用される通り、メタデータ、HTML標識、関連するファイル内の情報等として、画像データ添付される文字情報23、40は「画像データ」には含まれない。3040

【0045】

システム10、図3に示す例示的な方法を行うための、メインメモリ60に格納されたソフトウェア命令を含む。これらの命令はプロセッサ64により実行され、次の連続構成要素として示されている。品質値エクストラクタ102、コンテンツ特徴エクストラクタ104、美的特徴エクストラクタ106、及び組合せカテゴライザ18。組合せカテゴライザ18は、1つ以上の特徴の分類子110及び融合モジュール112を含むことができる。

【0046】

1020304050

簡単に言うと、品質値エクストラクタ 102 は訓練画像 32、34、36 毎に品質値 54 を抽出する。品質値は単純に人間の批評家により提供され、情報画像に関する品質情報 38 における情報から生じる数字の実数の値でよい。例えば、批評家は 1 ~ 10 の整数を各画像 32、34、36 に割り当て、10 を最も高い品質又は序列（1 は最も高い）とすることができる。いくつかの訓練画像 32、34、36 は数人の批評家により評価されたものでもよく、その場合、品質値 54 平均は又は批評家スコア 38 の別の関数でよい。カテゴライザ 18 が訓練されたら、構成要素 102 はもはや必要ないことを理解されたい。

【 0047 】

コンテンツ特徴エクストラクタ 104 は、画像 12、32、34、36 に関するコンテンツ特徴 14、50 を抽出し、画像の意味内容を表現するコンテンツ特徴 14 又は 50 に基づいて各画像に関するコンテンツ特徴の記述子 22、56 を生成する。ある実施形態では、コンテンツ特徴エクストラクタ 104 は意味的カテゴライザ 114 を含む、又はアクセスする。この意味的カテゴライザ 114 は画像 12、32、34、36 の画像データからのコンテンツ特徴 14、50 のいくつかの又は全てを抽出する。従って、例示的なコンテンツ特徴の記述子 22、56 は、美的特徴のいずれにも基づいていない。

【 0048 】

美的特徴エクストラクタ 106 は、画像 12、32、34、36 に関する美的特徴 16、52 を抽出し、画像に関する特徴 16、52 から各美的特徴の記述子 24、58 を生成し、この美的特徴の記述子 24、58 が画像の美的外観を少なくとも部分的に予測する。従って、例示的な美的特徴記述子 24、58 はコンテンツ特徴のどれにも基づいていない。

【 0049 】

コンテンツ特徴の記述子 56、美的特徴の記述子 58 及び品質値 54 を用いて訓練された組合せカテゴライザ 18 は、コンテンツ及び美的特徴の記述子 22、24（従って、いずれの関連する品質情報に基づいていない）に基づいて新しい画像 12（関連する品質情報を持たない）分類することができる。融合モジュール 112 は二つ以上の分類子 110 から結果を融合する、又は例えば、後期融合法又は前期融合法をそれぞれ用いて分類子 110 を入力する前に記述子 22、24 を融合する。ある実施形態では、融合モジュール 112 は、分類子 110 を入力する前に、特徴の記述子 22 及び 24（又は訓練中の 56 及び 58）を单一の記述子に融合させる前期融合モジュールである。別の実施形態では、融合モジュール 112 は、各分類子 110 によりカテゴリ毎に 1 つ出力されるスコアを受け取り、分類子 110 により出力される、重み付けされたスコアの関数として全体の品質スコア 20 を生成する後期融合モジュールである。

【 0050 】

様々なソフトウェアの構成要素 102、104、106、18、114 が、独立したモジュールとして記載されているが、これらのうちの 1 つ以上を分離して追加モジュールする、又は組み合わせることができることを理解されたい。構成要素のうちのいくつかは、前のモジュールの入力上で機能することができる。いくつかのケースでは、それから抽出される画像データ又は情報を、さらに処理するために前のモジュールに戻すことができる。

【 0051 】

本明細書で使用される用語「ソフトウェア」は、コンピュータ又はその他のデジタルシステムを設定して、ソフトウェアの目的とする作業を行うために、コンピュータ又はその他のデジタルシステムにより実行できる、あらゆる命令の集まり又はセットを包含することを意図する。本明細書に使用される用語「ソフトウェア」は、ローカル又はリモートの記憶媒体に格納されるそのような命令を包含することを意図する。

【 0052 】

図 3 は、図 2 のシステムを用いて行われる画像品質評価の方法を示す。方法は S100 から始まる。

10

20

30

40

50

【0053】

S102で、訓練画像32、34、36のセットが提供される。訓練画像32、34、36はコンテンツ情報品質スコアを含む文字情報38、40と関連することができる。

【0054】

S104で、訓練画像毎に品質値54が、例えば品質情報38（エクストラクタ102により）から抽出される。ある実施形態では、（平均の）ユーザスコアが量子化され得る。例えば、閾値スコア を選択し、画像 < 閾値の場合、不良（例えば、-1）と標識し、その他の画像、すなわちそれらの画像 閾値 th の場合、良（例えば、+1）と標識することにより二値化される。別の実施形態では、訓練画像を随意的にフィルタ（S106）にかけて、例えば、第一の閾品質値を超えた画像、及び第二の高い閾値より低い画像を取り除くことにより、訓練画像42の削減したセットを供給する。これにより、ボーダーラインの画像を取り除く。S104では、残った画像が、第一の閾値の以下の場合、不良（例えば、-1）で標識し、その他の画像、すなわち、それらの画像が第二の閾値以上の場合、良（例えば、+1）で標識する。 |

【0055】

S108で、（随意的にフィルタにかけられた）訓練画像32、34、36のセットからコンテンツ特徴50が（エクストラクタ104により）抽出される。例えば、i) 手動で適用された標識40と、ii) 画像データに基づいて意味的カテゴリライザ114により出力される、コンテンツの値のうちの1つ以上から、これらの特徴を抽出することができる。

【0056】

S110で、抽出されるコンテンツ特徴50に基づいてコンテンツ特徴の記述子56が画像32、34、36ごとに（エクストラクタ104により）生成される。

【0057】

S112で、訓練画像32、34、36から美的特徴52が（エクストラクタ106により）抽出される。

【0058】

S114で、抽出された美的特徴52に基づいて美的特徴の記述子58が訓練画像32、34、36ごとに（エクストラクタ106により）生成される。

【0059】

S116で、S104で抽出された品質値54及び各コンテンツ及びS110及びS114で生成された美的特徴の記述子56、58に基づいて分類子110が訓練される。特に、訓練された分類子110のパラメータはメモリ60内に格納される。

【0060】

S118で、あらゆる品質情報38を持たない新しい画像12（テスト画像）がシステム10に入力される。

【0061】

S120で、コンテンツ特徴14が新しい画像12から（エクストラクタ104により）抽出される。

【0062】

S122で、S120で（エクストラクタ104により）抽出されたコンテンツ特徴14に基づいてコンテンツ特徴の記述子22が新しい画像12に対して生成される。

【0063】

S124で、美的特徴16が新しい画像12から（エクストラクタ106により）抽出される。

【0064】

S126で、S124で抽出された美的特徴16に基づいて美的特徴の記述子24が新しい画像12に対して（エクストラクタ106により）生成される。

【0065】

S128で、新しい画像12に関する美的記述子24及びコンテンツ記述子22に基づ

10

20

30

40

50

いて品質値 20 が訓練されたカテゴライザ 18 を用いて生成される。上記に記載した通り、品質値 20 は、例えば、画像 12 を品質に関して良、又は不良として指定する量子化された二進値（例えば、+1、-1）でよい、又は、例えば、0 から 1 又は 1 から 10 の範囲の実数の値でよい。

【0066】

S130 で、品質値 20 が、例えば、サーバコンピュータ 10 上に常駐するメモリ格納装置 62 に出力される又は I/O 装置 66 を介して、外部メモリ、クライアント装置 80、又はプリンタ、ディスプレイ装置 92、又はその他の外部の出力装置等の外部の装置に出力される。

【0067】

S132 で、処理を実行する別のコンピュータへ品質スコア 20 を入力することができる。方法は S134 で完了する。

【0068】

一度カテゴライザ 18 訓練されたら、新しい画像に関してステップ S104～S116 を繰り返す必要はないことを理解されたい。また、全てのステップが示された順番で行われる必要もない。例えば、ステップ S108 と S112 は入れ替える又は同時に使うことができる。

【0069】

図 3 に示される方法は、コンピュータ上で実行されるコンピュータプログラム製品内で実行することができる。コンピュータプログラム製品は、ディスク、ハードドライブ等の方法を実行する制御プログラムが記録されている非一時的コンピュータ可読記憶媒体を含むことができる。

【0070】

あるいは、電波通信及び赤外線通信の間に生成される音波又は光波等の伝送媒体を用いたデータ信号として制御プログラムが組み込まれた伝送搬送波等の一時的媒体内で、この方法を実行することもできる。

【0071】

次に、システム及び方法の様々な様態をより詳細に説明する。

コンテンツ特徴の抽出

【0072】

例示的な実施形態は、画像 12 に割り当てられた画質値 20 が、画像の意味内容に部分的に依存することを前提とする。従って、下記の例で証明される通り、画像 12 の主題を示すコンテンツ特徴 14 を用いることにより、その品質の評価を改善することができる。画像の意味内容 14 は、手書きの注釈 23、その画像と別の画像の類似性に基づいて文字情報を別の画像から画像に組み込む自動モデルにより作られる文字標識、メタデータ、GPS 又はその他の位置情報、分類スコア等のうちの 1 つ以上から生じ、それらうちのいくつかは画像のコンテンツベースの記述子を生成するために用いられる。

【0073】

画像 12、32、34、36 は、 N_c 個のセットの異なる概念又はコンテンツカテゴリを用いて示される。各画像 $X(i)$ に特徴ベクトル $x_c(i)_{22 \times 56}$ の形をとる第一の記述子を次のように割り当てることができる。 $|x_c| = N_c$ （すなわち、カテゴリ毎に 1 特徴）、ここで各次元は特定の概念が画像 $X(i)$ 内に表されている可能性を示す。

【数 2】

$$x_c(i) = [w_1, \dots, w_j, \dots, w_{N_c}] \quad (1)$$

但し各特徴の可能性（重み w ）は二進値又は実数の値でよい。

第一の記述子 $x_c(i)$ において、全ての重みを正規化することができ、その合計は 1 となる。

【0074】

例えば、コンテンツ情報 23、40 が画像のテキストの記述を含む人間により割り当て

10

20

30

40

50

られるカテゴリ又は標識上の情報の場合、コンテンツ特徴エクストラクタ 104 は、コンテンツ特徴記述子 22、56 を生成する際に、この情報を用いることができる。

【0075】

一実施形態では、このテキストコンテンツ情報は、閲覧者カテゴリの所定のセットから選ばれる。例えば、ユーザが見る画像は、セット内の 1 つ以上の（しかし一般には全部ではない）閲覧者カテゴリを選択に限定される。閲覧者カテゴリをシステム 10 により用いられるコンテンツカテゴリと連携させることができ、そのため各閲覧者カテゴリはコンテンツカテゴリのうちの各 1 つに対応する（いくつかのケースでは、二つ以上の閲覧者カテゴリを同じコンテンツカテゴリに連携させることができる）。従って、例えば、閲覧者が「風景」の標識だけを画像に添付した場合、1 の特徴値 w_j が対応する「風景」コンテンツ特徴に割り当てられ、(0、1、0、0、0、0、0) 等の第一の記述子を生成することができる。コンテンツカテゴリは、例えば、動物、風景、人々、都市、花、対象等である。
10

【0076】

標識 23、40 が自由形式の場合、すなわち、どのカテゴリにも限定されていない場合、システム 10 は構文解析を行って文字情報 23、40 を分析してコンテンツカテゴリとして認識されるテキストを識別することができる。例えば、「これは赤いバラです」という文を例にとると、システムは、バラ（名詞）を抽出し画像をコンテンツカテゴリ「花」に 1 の特徴の重み w_j とともに割り当てる。
20

【0077】

追加的に又は代替的に、コンテンツ特徴記述子 22、56 は、画像 12、32、34、36 の画像データ（画素）から抽出される特徴に基づく。この場合、コンテンツ特徴の記述子の生成子は、1 つ以上の視覚コンテンツ分類子を含むことができる意味的カテゴリ 114 にアクセスすることができる。カテゴリ 114 は、画像を单一の最も可能性の高いカテゴリ又は確率的に全てのカテゴリに割り当てることができる。従って、第一の記述子は、テキストコンテンツ及び / 又はテキストコンテンツ及び画像データコンテンツの両方の関数である画像データベースの重み又は一連の重みを含むことができる。
20

【0078】

例示的な意味内容の特徴の記述子 22、56 を生成するために使用することができる、画像の視覚コンテンツを分類するための方法は、例えば、米国公開広報第 2007005356 号、第 20070258648 号、第 20080069456 号、第 20080317358 号、第 20090144033 号、第 20100040285 号、第 2010092084 号、第 20100098343 号、第 20100318477 号、米国特許出願第 12/512,209 号、第 12/693,795 号、第 12/960,018 号、及び Perronnin, F., Dance, C による 2007 年 CVPR での「Fisher Kernels on Visual Vocabularies for image Categorization」に記載されている。
30

【0079】

例えば、意味的カテゴリ 114 はパッチエクストラクタを含み、画像 12、32、34、36 の形状、質感、色等のパッチの特徴に関連するコンテンツが抽出及び分析される。パッチは、画像の区分化すること、特に関心のあるポイントを検出すること、規則的な格子を考慮すること、又は単に画像パッチを不規則にサンプリングすることにより得ることができる。各パッチから抽出される低レベルの特徴（ベクトル等）を融合して画像の印として機能する特徴ベクトルを形成することができる。別の方法では、画像の特徴ベクトルをクラスタに割り当てることができる。例えば、K 平均法等を用いて訓練画像から抽出された低レベルの特徴をクラスタリングすることにより視覚の語彙を前もって得る。次いで、各パッチベクトルが最も近いクラスタに割り当てられ、割り当てのヒストグラムを生成することができる。他の方法では、確率論的枠組みが用いられる。例えば、混合ガウス分布モデル（GMM）等の基本的な発生モデルが存在することを前提とし、そこから全てのベクトルが放射される。この場合、期待値最大化（EM）アルゴリズムを用いて視覚
40

ボキャブラリを放出することができる。いずれの場合でも、語彙内の各視覚単語は、一般的な低レベルの特徴のグループに対応する。視覚単語はそれぞれ、視覚（デジタルよりむしろ）対象の種類（例えば、ボール又は球、ロッド又はシャフト等）、特徴のある背景（例えば、星空、青空、草原等）等の中間レベルの画像特徴に（おおよそ）対応できる。分類される画像を考慮すると、確率論的モデルの場合は、抽出される各特徴ベクトルは前もって訓練された語彙又は全ての視覚単語の中からその最も近い視覚単語に確率的な方法で割り当てられる。この割り当てに基づいて、単一のコンテンツカテゴリを割り当てることができる、又は全てのコンテンツカテゴリに対する確率的な割り当てを行うことができる。

美的特徴の抽出

10

【0080】

美的特徴エクストラクタ106は、画像12、32、34、36毎に、直接画像上で評価される特徴から成るベクトル $x_a(i)_{24,58}$ の形をとる第二の特徴記述子を規定する。

【0081】

美的特徴は、低レベルの特徴、すなわち画像の全範囲の特性をとらえる特徴、及び画像内のローカル領域に焦点を合わせる高レベルの特徴を含むことができる。

【0082】

低レベルの特徴を抽出する技術には、配色、露光又はエッジ分布、明度、モーションブラー、ダイナミックレンジ、焦点外背景、白黒、明暗対照法、ロングシャッタースピード、三分割法、マクロ／クローズアップ写真、対称性とパターン、リード線、オフセンタ被写体、自然構成、視点変化等のうちの1つ以上が含まれ得る。画像内のローカル領域の特徴を示す高レベルの記述子は、K平均法の区分化、特徴抽出法によりその領域を抽出し、幾何学的コンテキストを用いることができる。理由は優勢領域又は写真の主題の位置を見ることにより構図及び構成の特性をとらえるためである。当技術分野で多くの特徴機構が実行及びテストされているが、画像分類の文字に使用される典型的な記述子はこのコンテキストには用いられていない。

20

【0083】

例示的な低レベルの美的特徴うちの少なくともいくつかをData2006に記載された方法に従って演算することができる。Data2006の特徴には、f1：光の使用の特徴を示す平均画素明度、f2：マルチカラー画像モノクロから区別する相対的配色、セピア又は単に低コントラスト画像、f3平均彩度、f4平均色相、f5、f6及びf7三分割法特徴、(f2、f3及びf4に類似するが三分割法に従う画像に対して画像が人間に好みに合うかどうか評価するため画像の領域を考慮している。) f8及びf9：画像と他の画像を比較して主題がより親しみがあるかどうか評価し、親しみのない（すなわちオリジナル）画像に対して高い値が得られる親しみの測定、f10～f21：Daubechiesのウェーブレット変換を用いて演算される粒状感の特徴、f22：画像の2つの寸法の合計である画像サイズ、f23：画像の2つの寸法の比であるアスペクト比、f24～f47+：画素をクラスタリングして類似色のパッチ及び補色パッチの配列にすることを記載した区分化に関連する特徴、f54、f55：被写界深度に関連する特徴、及びf56：完全な月のような形状に対して高い値が得られ、凹又は複雑な形状に関しては低い値が得られるスコアによる凸形状に関連する特徴、が含まれる。さらに、これらの各特徴の非線形の累乗を用いてこの一連の特徴を増大させることができる。すなわちそれらを二乗、三乗、及び平方根して、各画像を示す $D = 224$ の次元特徴ベクトルを得ることができる。以上Data2007、及び／又はY.Ko、X.Tang、and F.Jing.の2006年6月「The design of High-level features for photo quality assessment Vol.」のページ419～426（以後本明細書ではKe2006と呼ぶ）に記載されている特徴を参照。Keの特徴は、エッジの空間分布、ブラー、その配色ヒストグラムとプロ画像及びスナップ写真の配色の間の距離、簡単、コントラスト、及び明度の測定と

30

40

50

しての色相カウントである。

【0084】

画像カテゴリのうちの少なくともいくつかは、特徴値を正規化して [0 、 1] の範囲にすることができる、高い値のスコアは高い知覚品質と関連する傾向にある。使用される特徴のうちのいくつかは、他の特徴と実質的に重複してもよく、例示的な実施形態では、カテゴリライザ 18 の重みを特徴に割り当てるため、それにより方法が損なわれることはない。

【0085】

指向勾配 (H O G) の特徴のヒストグラム、スケール不变特徴変換 (S I F T) 特徴 (又は、その他の勾配ベースの特徴) 、及びガボール特徴のうちの 1 つ以上等のより、高レベルの記述子の特徴を有する美的特徴の記述子 24 、 58 では、これらの美的特徴を組み合わせることができる。 H O G 、 S I F T 及びガボール特徴により画像の構造及びそれらの質感の特性をとらえることができる。これら三つの特徴は美的特徴よりは一般的であるが、それらは美的分類に関して上記の美的記述子の特徴を補うことができる。

【0086】

例えば、ガボールフィルタは質感の区分化、顔認識、対象検知等に有用である。ガボールフィルタは、人間の視覚システムと類似する周波数及び方向の表現を有する線形フィルタである。空間領域では、 2 D ガボールフィルタは、 x 軸及び y 軸にそれぞれ沿って変動する、 x 軸及び y 軸にそれぞれ沿って空間センタ周波数を有する正弦平面波により変調されたガウスカーネル関数である。ガボールフィルタは自己相似であり、したがって 1 つのマザーウェーブレットから全てのフィルタを拡張及び回転することにより生成することができる。各ガボールフィルタは方向性を有し、従って複数の方向性のそれぞれを用いて例示的な記述子に関する特徴を提供することができる。

【0087】

ガボールフィルタはガウスエンベロープ (関数) により変調された、特定の周波数及び方向性の正弦平面とみなすことができ、下記のように表される。 :

【数 3】

$$h(x, y) = e^{\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}} e^{-j2\pi(u_0x + v_0y)} \quad (2)$$

30

及びその周波数応答 :

【数 4】

$$H(u, v) = 2\pi\sigma_x\sigma_y e^{-2\pi^2[(u-u_0)^2\sigma_x^2 + (v-v_0)^2\sigma_y^2]} \quad (3)$$

【0088】

これはガウス関数を周波数領域内の (u_0 、 v_0) により変換することと同等である。従って、ガボール関数を周波数内的位置 (u_0 、 v_0) に、すなわち、オリジナルから

【数 5】

$$\sqrt{u_0^2 + v_0^2}$$

40

の距離で

【数 6】

$$\tan -1 \frac{u_0}{v_0}$$

の方向にシフトしたガウス関数とみなすことができる。上記の方程式 2 及び 3 において、 (u_0 、 v_0) はガボールフィルタ空間センタ周波数である。パラメータ (x 、 y)

50

はガウスエンベロープの x 方向と y 方向に沿った標準偏差である。 σ_x 、 σ_y はフィルタの帯域幅を決定し、 j は指数関数の複雑度を表す。

【0089】

入力画像データからガボールフィルタを生成する様々なソフトウェア・アプリケーション（例えば、Matlabで実行されている）が市販されており、それらは例示的な実施形態において、例えば、 σ_x 、 σ_y のデフォルト値を用いて例示的な記述子に関する1つ以上の特徴を生成している。値 u_0 、 v_0 は美的特徴として使用することができる。異なるガボールフィルタ方向性により、異なる画像の端を検知する。

【0090】

1999年のInternational Conference on Computer Vision (ICCV)で「Object Recognition From Local Scale-Invariant Features」の中で Loweにより記載された通り、例示的な勾配ベースの特徴はSIFT記述子であり、各パッチに関して演算される。SIFT記述子は、例えば、空間的位置の44のグリッド上の8つの配向面で演算され、128次元ベクトル（つまり、この実施形態では特徴ベクトルごとに128の特徴）を与えるガウス導関数等の近隣の画像のマルチ画像表現である。SIFT特徴を用いたある説明例では、5階級の規則的な格子上（16画素毎に）の 32×32 画素パッチから特徴を抽出して128次元のSIFT記述子を提供する。特徴の数は隨意的に減少する、例えば主成分分析（PCA）を用いて例えば64次元まで減少する。

訓練に使用する品質値の抽出

【0091】

例示的な実施形態では、訓練画像32、34、36の品質値54は、写真共有ウェブサイトから得た画像に関する品質情報38（例えば、品質スコア、序列）から生じる。ソーシャルネットワークには、大抵テキストラベル（「好き」、「嫌い」）又は数値の段階（評価）を含む注釈ポリシがある。例えば、写真ネットのウェブサイト上では、ユーザは1（見苦しい）から7（美しい）の間のスコアで画像を評価することができる。このウェブサイトのユーザはサイトの運営管理者から次のようなガイドラインを提供されている。「7に近い評価の理由：a）見た目が良い、b）魅力的／注意を引く、c）面白い構図、d）色使いが良い、e）（フォトジャーナリズム場合）ドラマがある、ユーモアがある、インパクトがある、f）（スポーツの場合）絶頂期、アスリートの頑張り」。評価が、「好き」等の選択可能なテキスト回答の形をとれば、その評価をスコアに換算することができる。

【0092】

いくつかの実施形態では、訓練に使用される画像32、34、36のうちの少なくともいくつか又は全てが、少なくとも5又は少なくとも10スコア／序列38を有する画像等のように複数のスコア／序列を有する。中間、中央、又はそれらのスコアのモード等の平均を、品質値54としてとることができる。いくつかの実施形態では、異常値、例えば中間から1つ又は2つより多い標準偏差分異なるスコア38を品質値54の演算から除外することができる。

【0093】

標識からの素のスコア3を用いてカテゴライザ18を訓練することができる、又はスコアを二値化あるいは量子化（例えば、3つ以上の量子化レベルに）することができる。ある実施形態では、画像(i)について、この画像 $q_{av}(i)$ に関して得られるスコア38の平均を画像毎に算出する。2つの閾値 $\mu_1 = \mu + \sigma / 2$ 及び $\mu_2 = \mu - \sigma / 2$ が規定され、 μ は平均スコア等のユーザスコアの範囲のほぼ中間の値をとることができ、 σ は、不確定スコアを有するいくつかの画像を除去して選択した数字（又は他の因子）であり、どの画像の割合が中間、中間範囲なのかを決定する好適な数字である。これにより基本的に3つの範囲が作られる。次いで各画像は、 $q_{av}(i) < \mu_1$ の場合、ラベル「良」(+1)で注記され、 $q_{av}(i) > \mu_2$ の場合、「不良」(-1)で注記される。従って $\mu_2 < q_{av}(i) < \mu_1$ の中間範囲内の画像32、34、36は、カテゴライザを訓練

10

20

30

40

50

する目的のため除去され、無視される。の値は訓練セットのサイズに依存する。が大きすぎる場合、適切な数の良と不良の画像を有することはできない。小さすぎる場合、セット内の不確定画像が多くなりすぎる可能性があり、分類子システムを正確に訓練することができない。例として、可能性のあるスコアの範囲が1～7の場合、 μ は約3～6、例えば約5でよく、は、1～2等の0.5から3でよい。この場合、 < 4 又は良に分類される画像がない可能性がある。ユーザにより割り当てられるスコアの範囲に基づいて閾値を設定するために用いられるスコア（この例では5）を変えることができ、「良」と「不良」で標識されるスコア $q_{av}(i)$

が、ほぼ同数になるように選択できることを理解されたい。

カテゴライザ

10

【0094】

画像品質評価に関して、例示的なカテゴライザ18は複数のソース（画像、手動の注釈、及び文字標識）から生じ得る異種の特徴（視覚及び文字）の組み合わせを、入力として受け取ることができる。具体的には、通常コンテンツベースの分類に関する用いられる特徴14は、コンテンツに依存する画像品質評価を提供する品質の分析のために特に考案された、別の特徴16と組み合わされる。画像12の主題に関する情報の使用により、より正確な画像の品質の予測が可能となる。具体的には、分類の枠組みの中のコンテンツ特徴と美的特徴の融合に関して、2つのソリューション（前期融合及び後期融合）が考えられる。

【0095】

20

カテゴライザ分類子110は、スペースロジスティック回帰、単純ベイズ、線形判別分析、サポートベクタマシン（SVM）、線形回帰、又はその他の好適なマシン学習法等のあらゆる好適な線形又は非線形訓練アルゴリズムで訓練された二値分類子である。例示的な実施形態では、例えば、Dattat2006に記載されている放射基底関数（RBF）カーネルを用いたSVMが用いられる。これにより、Dattatの結果と直接比較することが可能である。しかし、別の分類子学習法も検討できることを理解されたい。

【0096】

特徴の記述子54、56又は22、24を二つ方法、前期融合と後期融合のうち1つにより組み合わせることができる。

1. 前期融合

30

【0097】

この実施形態では、特徴の記述子 x_a 及び x_c を单一のベクトル x_t に融合することにより、特徴レベルの美的情報とコンテンツの情報を組み合わせる。この実施形態では、異なるソース（例えば、画像の視覚分析からの x_a 、文字標識、手書きの注記、及び自動分類のうちの1つ以上のからの x_c 、）から生じる可能性のある、異なる階級を有する特徴の記述子を組み合わせることができる。このため、線形正規化を記述子の各特徴に別々に適用することができる。

【0098】

特徴の2つのセットに重みを付けるため、重み付けされた集計関数を2つの記述子の線形重み付けされた組み合わせ等の x_a と x_c の混合に適用して画像の品質の特徴を示す单一の記述子 x_t を生成する。分類子110を x_t に供給し、品質コア20を出力する。

40

【0099】

集計関数（融合モジュール112により適用される）は以下の形をとる。

【数7】

$$x_t = [(1 - \alpha)x_a, \alpha x_c] \quad (4)$$

ここで、は0～1の実数の値である。

【0100】

方程式4は、それらの各重みが合計して同じ値、例えば1、になるように記述子 x_a 及び x_c がそれぞれ正規化される制約に基づくことができる。例えば、美的記述子の特徴（

50

重み)は $x_a = (w_{a_1}, w_{a_2}, w_{a_3}, \dots, w_{n_a})$ で表され、コンテンツ記述子の特徴(重み) $x_c = (w_{c_1}, w_{c_2}, w_{c_3}, \dots, w_{n_c})$ はで表される。ここで、 $w_{a_1} + w_{a_2} + w_{a_3} + \dots + w_{n_a} = 1$ 、及び $w_{a_1} + f_{a_2} + w_{a_3} + \dots + w_{n_a} = 1$ 。

【0101】

方程式4において α を0とした場合、美的特徴16だけを用いて分類し、 $\beta = 1$ の場合、画像のコンテンツ特徴14だけを用いて品質を評価することは明らかである。例示的なカテゴライザ18は、 $0 < \alpha < 1$ の動作のモードの少なくとも1つを有するため、両方の特徴が考慮される。ある例示的な実施形態では、0.01、例えば、0.03、であり、ある実施形態では、0.05又は0.1である。別の例示的な実施形態では、0.09、例えば0.08であり、ある特定な実施形態では、0.06又は0.05である。例示的な一実施形態では0.050.05である。10

【0102】

の最適値、すなわちコンテンツ特徴なしで実現できる精度よりも少なくとも高い精度を提供する値は、選択される特徴及び画像データのセットに部分的に依存することができる。従って、例示的な方法では、所与のデータベースに関するの最適値を確認するためのの異なる値を用いてシステム10の結果をテストする。

【0103】

次いで、画像32、34、36毎に組み合わせた記述子 x_t 、及び対応する品質値54は分類子110に入力されて分類子を訓練する。新しい画像12に関して、組み合わされた記述子 x_t を入力して分類子110を訓練し、画像12に関する品質スコア q 出力し、それを品質値20として用いることができる。20

2. 後期融合

【0104】

この実施形態では、特徴レベルでコンテンツと視覚情報を組み合わせる代わりに、融合モジュール112によりコンテンツ従属記述子56を用いて訓練された複数の分類子110の出力を直接組み合わせることができる。例えば、コンテンツカテゴリ毎に N_c 個の異なる分類子110が用いられ、各コンテンツカテゴリ j の画像32、34、36だけを用いて各分類子のパラメータ(随意的に、ハイパー・パラメータのセット s_j を含む)が評価される。従って、訓練中、分類子110に入力された特徴ベクトル58だけが x_a である。30。テスト中、記述子24を全ての訓練された N_c 個の分類子110に入力することにより、新しい画像 $X(i)$ が分類される。分類スコア s_j の対応する数が集められる。最終スコア q は、各分類子スコア及びそれらの関連する重みの関数のスコアである。例えば、 q は各分類子スコアを x_c 記述子22各の特徴の重みで重み付けすることにより得られる。

【数8】

$$q = \frac{1}{N_c} \sum_{j=1}^{N_c} w_j s_j \quad (5)$$

ここで各 w_j は、特徴の重み x_c の $w_{c_1}, w_{c_2}, w_{c_3}, \dots, w_{n_c}$ の各1つを表す。別の実施形態では、全ての重み x_c 所与の同じ値である。40

【0105】

次いで、値 q を画像の品質スコア20として用いることができる。実際には、別のセット/特徴の種類により、異なる結果ができる可能性はあったものの、この方法が前期融合法を改良した結果を提供することが分かった。

品質値の使用

【0106】

システム10により出力された品質値20は様々な用途で使用することできる(S13 2にて)。例えば、画像12の品質スコア20を単独又は他のパラメータと組み合わせ使50

用して、サーチエンジンが、ユーザにより入力されるクエリに応答して検索した、そのような画像12のセットをランク付けすることができる情報検索処理において品質スコア20を使用することができる。この実施形態では、セット内の各画像12は品質スコア20を持ち、このパラメータに基づいてスコアが高いほど、その画像は最初にユーザへ提示される画像の中に含まれている可能性が高い。

【0107】

別の実施形態では、品質値20を使用して画像を品質値に基づく配列で（例えば、最も高いスコアの画像を最初に示すよう）ディスプレイ92に表示することができる。

【0108】

またさらに別の実施形態では、全部ではないが少なくとも1つのそれぞれにシステムにより品質値20を割り当てられた画像12のセットがその品質値に基づいて選ぶことができる。例えば、文書に貼り付ける、ウェブコンテンツを作る等のために画像を選択することができる。例えば、最も高いスコアの画像を選択する。

10

【0109】

別の実施形態では、スコア20に基づく標識、例えば、1つ星～5つ星の星評価をユーザ'の画像コレクションに割り当てることができる。次いで、ユーザは、例えば、少なくとも4つ星又は少なくとも5つ星の星評価の画像を検索することができる。

【0110】

別の実施形態では、品質スコア20を使用して新しいカテゴライザの訓練に使用する一連の画像を選択することができる。例えば、少なくとも閾値の品質スコアを有する画像12だけをカテゴライザに入力する。カテゴライザは、分類子114に関して記載した意味的カテゴライザでよい。

20

【0111】

別の実施形態では、訓練されたカテゴライザ18を写真用カメラに直接配置して画像12を撮影する前でさえもリアルタイムの提案をユーザに行うことができる。例えば、カメラのスクリーンの中の画像12を処理して品質スコアを決定する。画像12が撮影されたときその品質が芳しくなさそうな場合（例えば、閾値より低い場合）、カメラはそのスクリーン上に警告を表示する。

【0112】

例示的な実施形態の趣旨を制限することなく、以下の事例は実行する例示的な方法の実施を示している。

30

実施例

【0113】

プロトタイプのシステム10は最初下記文献に記載された美的分類にリスト化された全ての特徴を使用して実行される。Dattat2006(56特徴)、Ke2006(7特徴)、及びECCV(3)(2008年)でのY. LuoとX. Tangによる「Photo and video quality evaluation: Focus ing on the subject.」内のページ386～399。Luoの参考文献に記載された5つの特徴は、他の63の特徴に比べて改善と見なされず、放棄された。

【0114】

40

画像データのセット：訓練画像100及び画像12の両方に関して、写真ネット1のデータセットが使用された。このデータベースのセットは、公的ウェブサイトの写真ネットからランダムに選択した3、581枚の写真のセットである。しかし、何枚かはオリジナルの画像とハイパーリンクされているため、すでに入力の対象ではなく、これらの画像のうちの3、118枚だけがダウンロード可能であった。これら画像を使用してテストを実施した。写真ネット1のデータセット内の画像に1.0から7.0の間（7が最も高い品質スコア）の値でスコアを付け、画像毎に複数の投票を可能にした。画像の被写体は非常に多岐に及ぶ。画像のタイトル及び写真ネットウェブサイトにより行われた画像の分類を調査することにより、7つの主なコンテンツカテゴリを識別した。次いで手作業で各画像に、7つカテゴリのうちの1つを標識した。カテゴリ毎の画像の数を表1に示す。

50

【表1】

カテゴリ	カテゴリ内の 画像の数	$\delta = 0$ に関する「良」 の標識の画像の パーセンテージ	$\delta = 0$ に関する「不良」 の標識の画像の パーセンテージ
動物	436	70.41%	29.59%
風景	832	51.80%	48.20%
人々	735	44.76%	55.24%
都市	505	40.40%	59.60%
花	210	47.62%	52.38%
対象	241	44.81%	55.19%
その他	118	31.36%	68.64%

【0115】

分類子110として、標準のサポートベクタマシン分類のツールキットを使用した。(libSVMツールキット、C.-C. ChangとC.-J. Linによる、LIBSVM：サポートベクタマシンに関するライブラリ、2001年、ソフトウェアはhttp://www.csie.ntu.edu.tw/~cjlin/libsvm.で入手可能) 10

【0116】

訓練画像及びテスト画像を提供するために、全部で3118枚の画像を五つのフォルダに分け、各フォルダは続いてセットされた5つフォルダのクロス検証で画像12をテストするために使用する。 20

【0117】

画質注記プロトコルを作成する。2つの美的品質カテゴリ(良、不良)を規定することにより画質に関する二項分類を確立する。データベース内の各画像*i*に割り当てる平均美的スコア $q_m(i)$ (すなわち、1~7の値)。2つの閾値 $\tau_1 = 5 + / 2$ 及び $\tau_2 = 5 - / 2$ をセットする。各画像が、 $q_{av}(i) \geq \tau_1$ の場合「良」($= 1$)の標識で注記を付けられ、 $q_{av}(i) \leq \tau_2$ の場合、「不良」($= -1$)で注記を付けられる。 30

は数字である。これらの標識は訓練画像に関して用いられ、正確に評価するためにも用いられる。

【0118】

Datta2006の同じ分類枠組みも使用し、それには下記の形のRBFカーネル関数を有するSVM含まれる。

【数9】

$$\exp(-\gamma * |u - v|^2)$$

40

【0119】

以下の実験では、放射基底関数のパラメータについて、 $\gamma = 3.7$ 、コスト $c = 1.0$ 、 u と v をカーネル関数に入力した。

【0120】

$\gamma = 0.15$ 又は $\gamma = 0.35$ を用いてコンテンツ特徴及び美的特徴を前期融合において組み合わせることによりカテゴリライザを訓練した(「前期融合」と標識される)。訓練及びテスト中に用いたコンテンツ記述子は、表1に示した7つカテゴリのうちの1つだけを画像に手書きで注記することにより生成された7つ特徴ベクトルである。

【0121】

システムの精度は、テスト画像ごとに、二項分類でシステムにより出力されたその画像

50

に関する品質値を比較し、エラー、すなわち 2 つの値の間の差を決定することにより決定される。次いでフォルダ内の全てのテスト画像に関するエラーを平均化して 0 ~ 1 の値にしてパーセンテージで表すことができる。いくつかの実施形態では、5 つのフォルダに関するパーセンテージを平均化することができる。100 % の精度は全ての画像が二項品質分類と一致する品質スコアを有していることを意味する。しかし、実際にはユーザにより適用されるスコアにはばらつきがあるため、これは期待できない。

【 0 1 2 2 】

図 4 には、 $\gamma = 0.35$ を用いた例示的なシステムの結果が示されている。結果を「基本線」システム（コンテンツ特徴が使用され、融合の必要がないときを除いた「前期融合」と同じ）及び、「最適化基本線」システム（SVM のハイパーパラメータが最適化されたときを除いた「基本」と同じ）と比較する。図 4 の結果では基本線システム及び最適化基本線システムに対して、精度が改善されていることが示されている。10

【 0 1 2 3 】

図 5 には、変動する γ の前期融合の結果に対する影響が示されている。この例では比較的単純なコンテンツ特徴を用いており改善は 0.05 ~ 0.55 の範囲に渡って示されているが、 γ の最適値は約 1.5 である。

【 0 1 2 4 】

図 6 では、基本の結果と様々な数のガボール特徴（2、3、及び 6 個の方向）を加えることにより得られた結果とを比較している。ガボール特徴を追加することにより、基本の結果が改善されている。方向性の数は著しく精度に影響しないようである。20

【 0 1 2 5 】

ガボール特徴と例示的なコンテンツ特徴を組み合わせることにより精度がさらに改善されることが期待できる。

【図1】

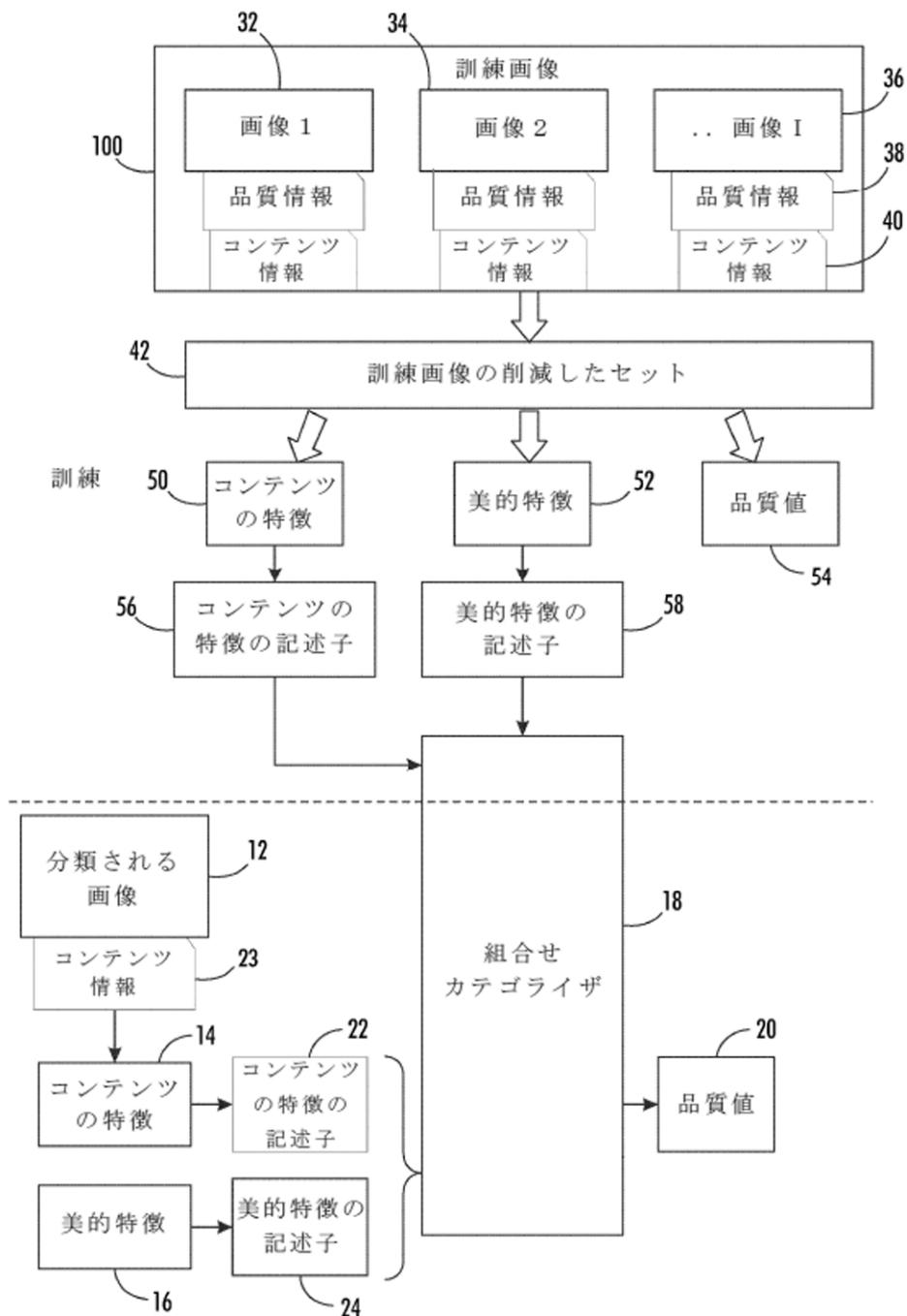


図1

【図2】

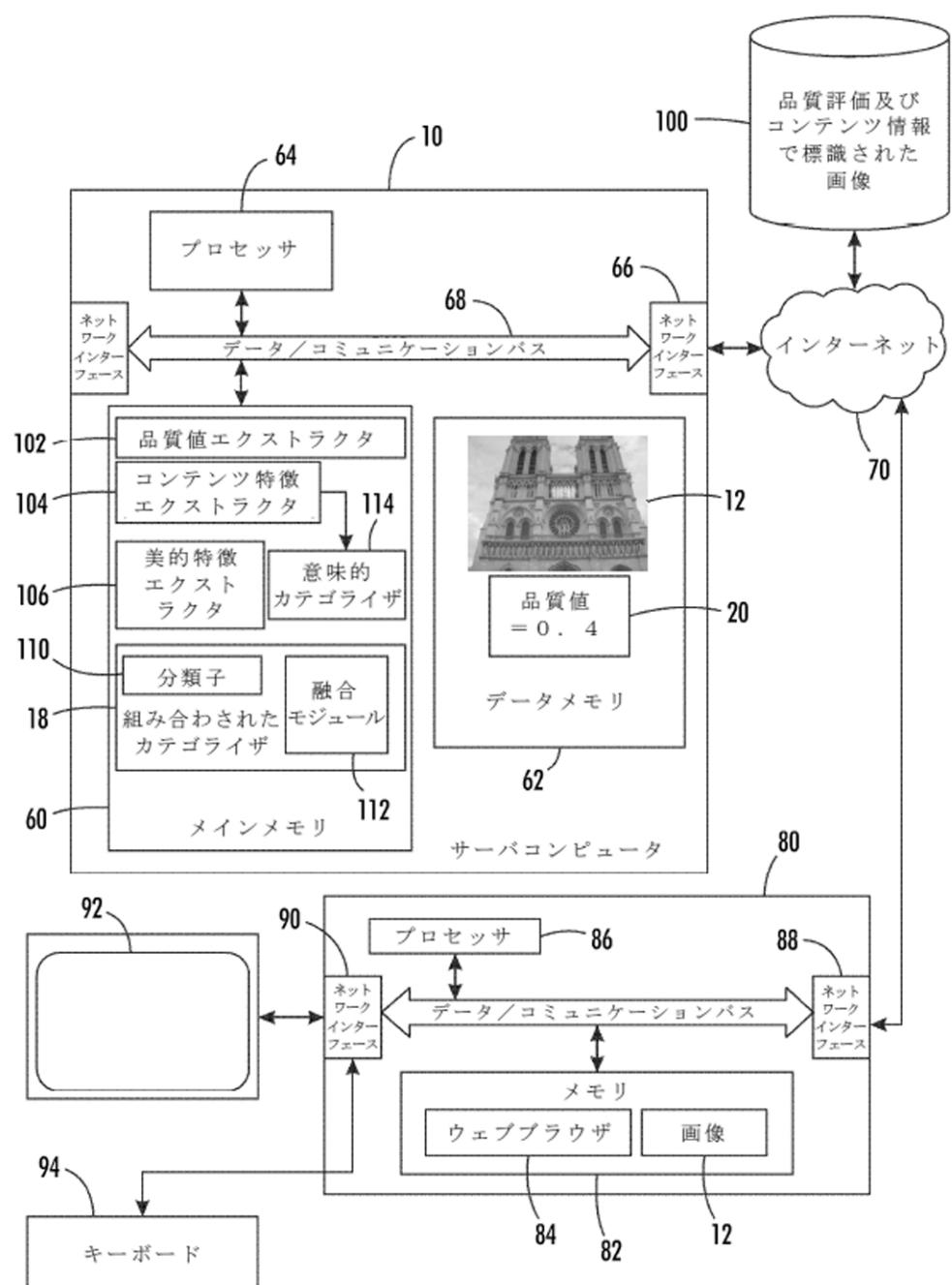


図2

【図3】

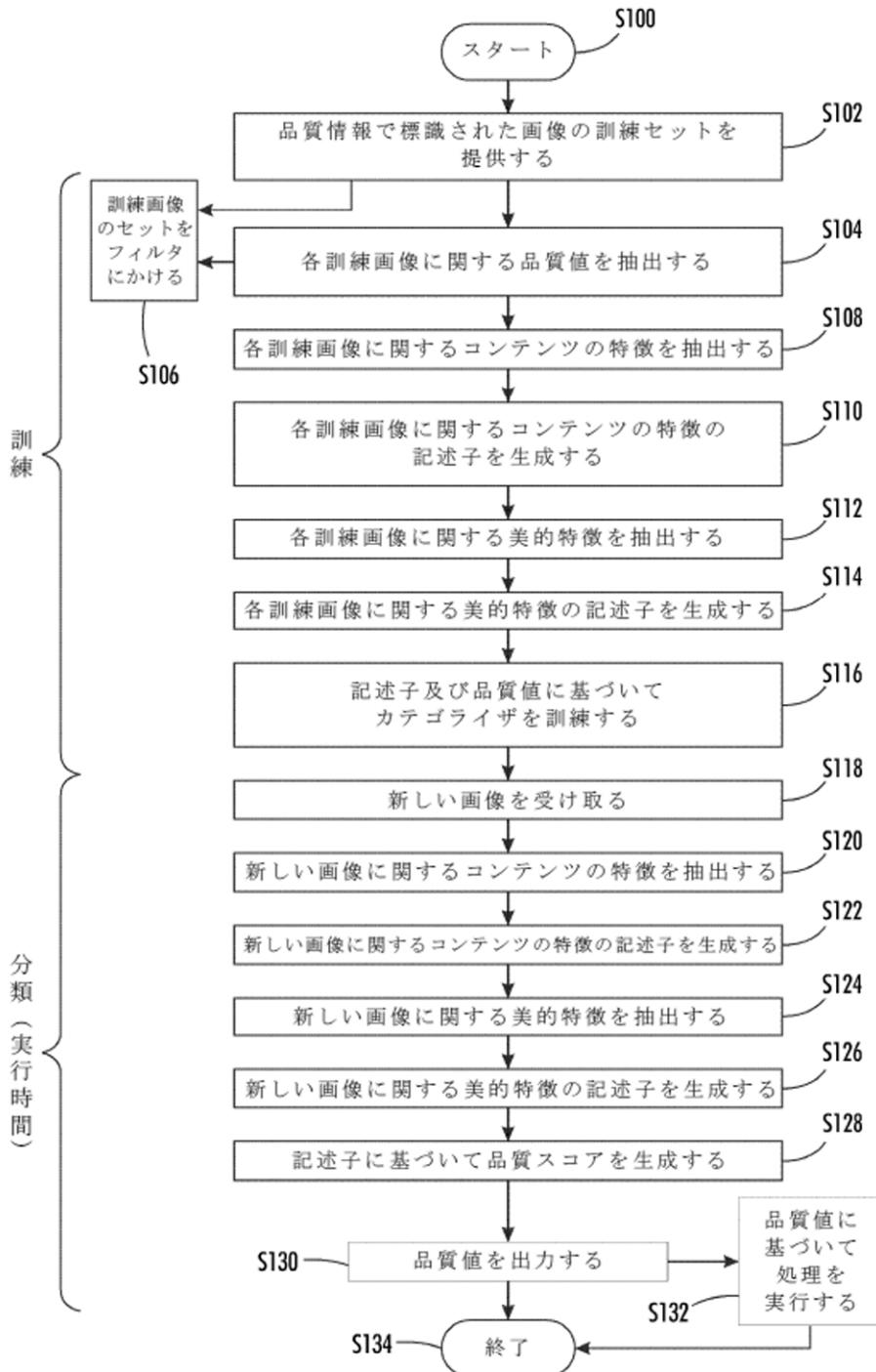


図3

【図4】

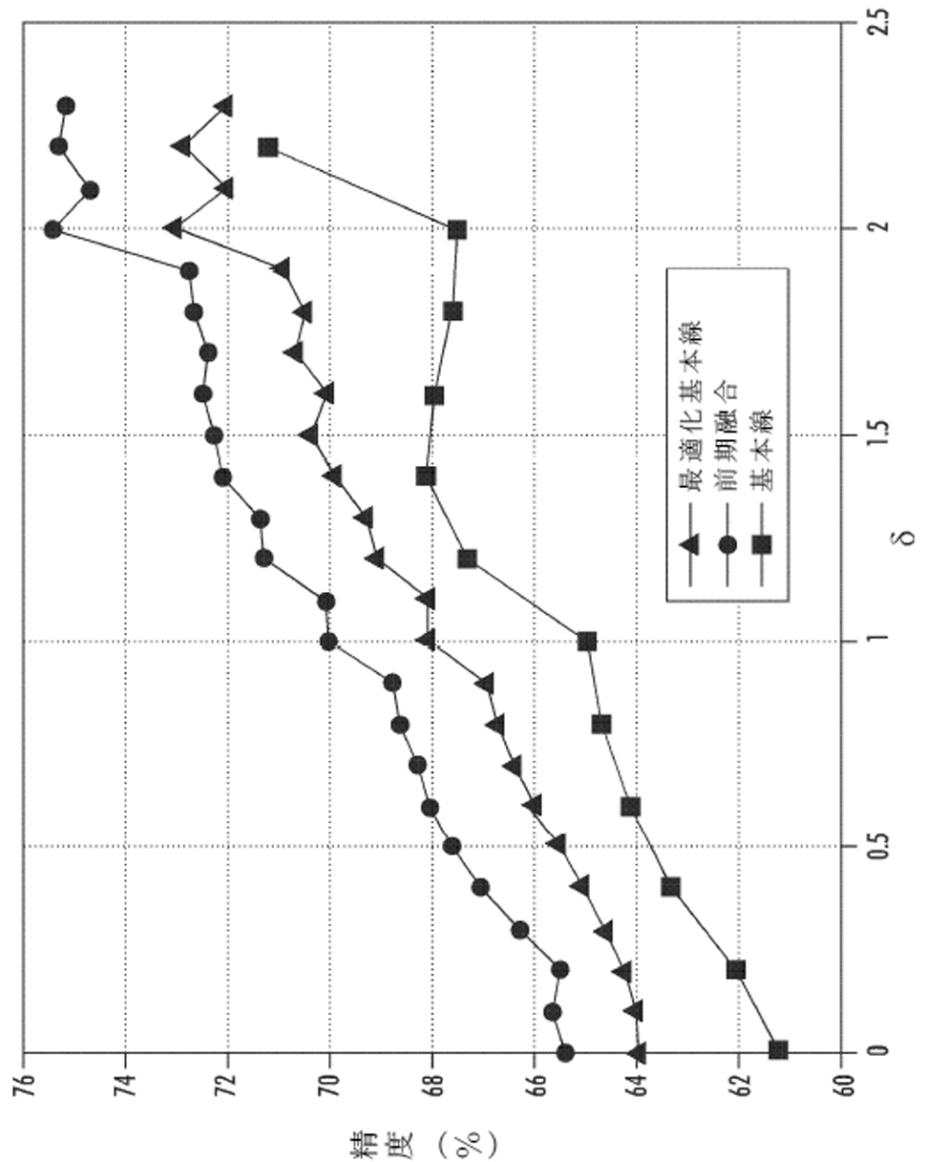


図4

【図5】

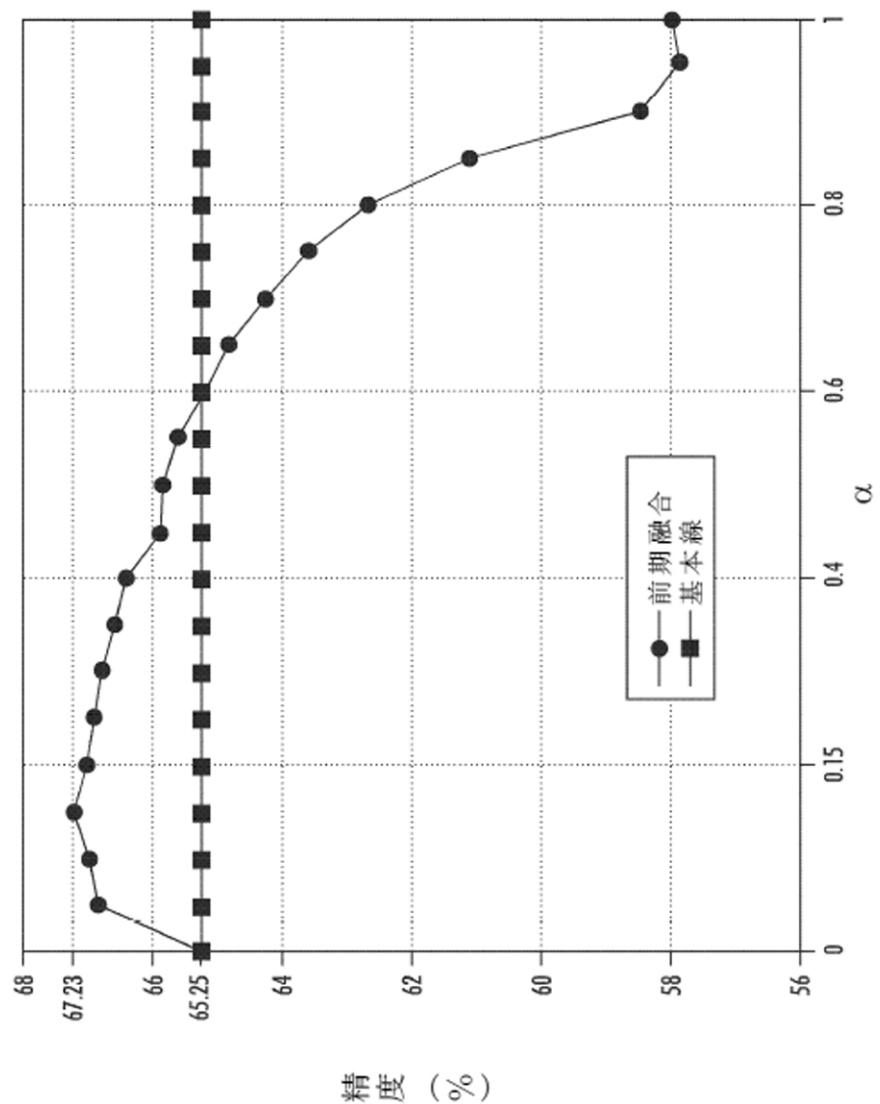


図5

【図 6】

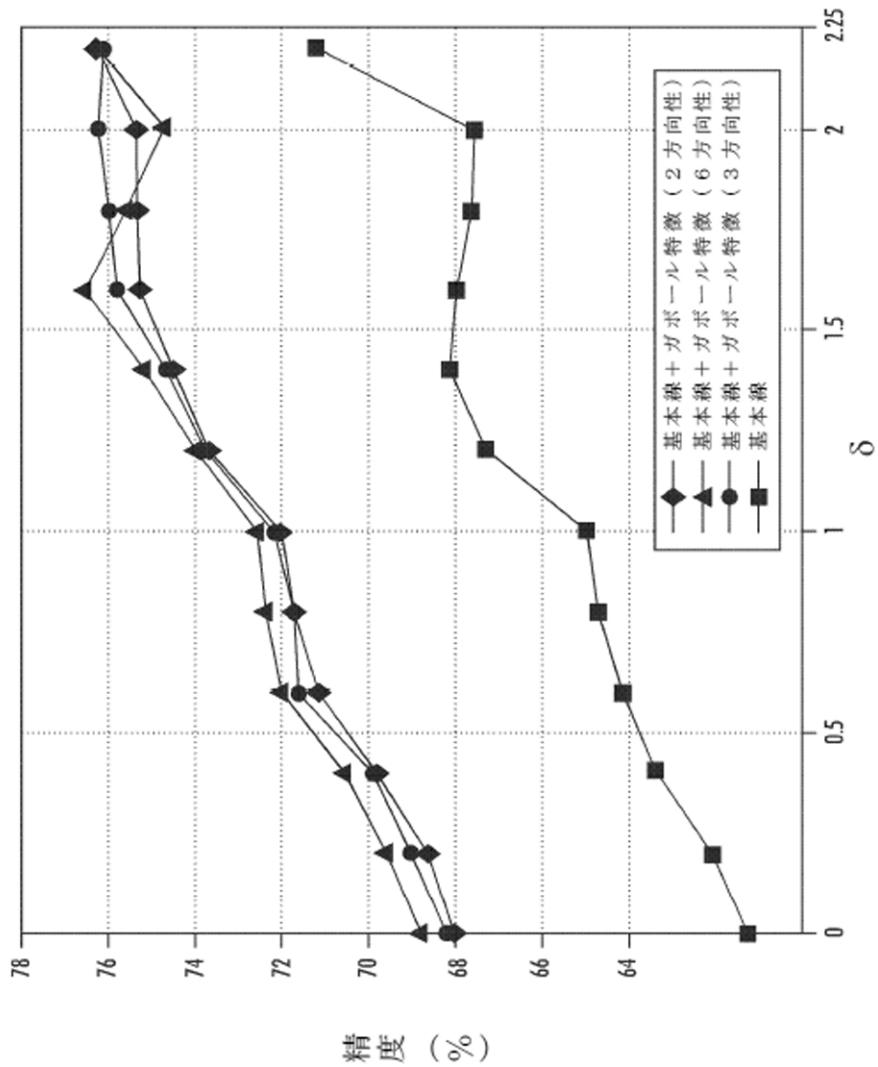


図 6

フロントページの続き

(72)発明者 ロドリグ・ンクッチ

フランス国 ヌイイ・シュル・マルヌ 93330 アレ・デ・スフレンヌ 5

審査官 佐藤 実

(56)参考文献 特開2003-87442 (JP, A)

米国特許出願公開第2011/0075917 (US, A1)

(58)調査した分野(Int.Cl., DB名)

G 06 T 7 / 00