

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6544543号
(P6544543)

(45) 発行日 令和1年7月17日(2019.7.17)

(24) 登録日 令和1年6月28日(2019.6.28)

(51) Int.Cl.

G06T 7/00 (2017.01)

F 1

G06T 7/00 Q
G06T 7/00 350C

請求項の数 18 (全 14 頁)

(21) 出願番号 特願2017-563173 (P2017-563173)
 (86) (22) 出願日 平成28年6月3日 (2016.6.3)
 (65) 公表番号 特表2018-516412 (P2018-516412A)
 (43) 公表日 平成30年6月21日 (2018.6.21)
 (86) 國際出願番号 PCT/US2016/035868
 (87) 國際公開番号 WO2016/197026
 (87) 國際公開日 平成28年12月8日 (2016.12.8)
 審査請求日 平成29年12月5日 (2017.12.5)
 (31) 優先権主張番号 14/732,518
 (32) 優先日 平成27年6月5日 (2015.6.5)
 (33) 優先権主張国 米国(US)

(73) 特許権者 000002185
 ソニー株式会社
 東京都港区港南1丁目7番1号
 (74) 代理人 100092093
 弁理士 辻居 幸一
 (74) 代理人 100067013
 弁理士 大塚 文昭
 (74) 代理人 100109070
 弁理士 須田 洋之
 (74) 代理人 100109335
 弁理士 上杉 浩
 (74) 代理人 100120525
 弁理士 近藤 直樹
 (74) 代理人 100151987
 弁理士 谷口 信行

最終頁に続く

(54) 【発明の名称】畳み込みニューラルネットワークに基づいたフルリファレンス画像品質評価方法

(57) 【特許請求の範囲】

【請求項1】

リファレンス画像に対する歪み画像の画像品質を評価するシステムであって、前記システムが、前記歪み画像及び前記リファレンス画像を入力として受け入れて画像品質のメトリックを出力として提供する畳み込みニューラルネットワークを備える、システム。

【請求項2】

前記畳み込みニューラルネットワークが複数の層を含む、請求項1に記載のシステム。

【請求項3】

前記複数の層が、

前記歪み画像及び前記リファレンス画像の各々を構成する画像パッチに正規化関数を適用して、正規化歪み画像及び正規化リファレンス画像を提供するよう構成された入力層と、

N 1 フィルタを用いて前記正規化歪み画像及び前記正規化リファレンス画像の各々を畳み込み、1つのフィルタ処理された正規化歪み画像と1つの対応するフィルタ処理された正規化リファレンス画像とを各々が含む特徴マップのN 1 (1よりも大きい整数)個のペアを提供するよう構成された畳み込み層と、

を含む、請求項2に記載のシステム。

【請求項4】

前記複数の層が更に、

前記歪み画像及び前記リファレンス画像の各々から提供される前記N 1個の特徴マップ

10

20

の N 2 個の線形結合を計算して、フィルタ処理された正規化歪み画像の 1 つの組み合わせと、フィルタ処理された正規化リファレンス画像の対応する 1 つの組み合わせとを各々が含む結合特徴マップの N 2 (1 よりも大きい整数) 個のペアを提供するよう構成された線形結合層と、

前記結合特徴マップの N 2 個のペアのうちの異なるものからの対応するピクセルに各々が基づいた N 2 個の類似性マップを計算するよう構成された類似性計算層と、

前記 N 2 個の類似性マップの各々について平均プーリングを適用して、N 2 個の類似性入力値を提供するよう構成されたプーリング層と、
を備える、請求項 3 に記載のシステム。

【請求項 5】

前記複数の層が更に、

前記 N 2 個の類似性入力値に作用して、M 個 (N 2 よりも大きい整数) の隠れノード値を提供するよう構成された全結合層と、

前記 M 個の隠れノード値を単一の出力ノードにマップして、画像品質のメトリックを提供するよう構成された線形回帰層と、
を含む、請求項 4 に記載のシステム。

【請求項 6】

N 1 = N 2 であり、N 1 = 1 0 である、請求項 5 に記載のシステム。

【請求項 7】

M = 8 0 0 である、請求項 5 に記載のシステム。

10

【請求項 8】

前記畳み込み層によって前記特徴マップの N 1 個のペアが前記複数の層の後続の層に提供される前に、前記特徴マップの N 1 個のペアの各ノードにおいて二乗活性化関数が適用される、請求項 3 に記載のシステム。

【請求項 9】

リファレンス画像に対する歪み画像の画像品質を評価する方法であって、前記方法が、

前記歪み画像及び前記リファレンス画像を処理して画像品質のメトリックとして提供するよう構成された畳み込みニューラルネットワークに対して、前記歪み画像及び前記リファレンス画像を入力するステップを含む、方法。

【請求項 10】

30

前記畳み込みニューラルネットワークによって実行される処理が、

前記歪み画像及び前記リファレンス画像の各々を構成する画像パッチに正規化関数を適用して、正規化歪み画像及び正規化リファレンス画像を提供するステップと、

N 1 フィルタを用いて前記正規化歪み画像及び前記正規化リファレンス画像の各々を畳み込み、1 つのフィルタ処理された正規化歪み画像と 1 つの対応するフィルタ処理された正規化リファレンス画像とを各々が含む特徴マップの N 1 (1 よりも大きい整数) 個のペアを提供するステップと、

を含む、請求項 9 に記載の方法。

【請求項 11】

前記歪み画像及び前記リファレンス画像の各々から提供される前記 N 1 個の特徴マップの N 2 個の線形結合を計算して、フィルタ処理された正規化歪み画像の 1 つの組み合わせと、フィルタ処理された正規化リファレンス画像の対応する 1 つの組み合わせとを各々が含む結合特徴マップの N 2 個のペアを提供するステップと、

前記結合特徴マップの N 2 個のペアのうちの異なるものからの対応するピクセルに各々が基づいた N 2 個の類似性マップを計算するステップと、

前記 N 2 個の類似性マップの各々について平均プーリングを適用して、N 2 個の類似性入力値を提供するステップと、

を更に含む、請求項 10 に記載の方法。

【請求項 12】

前記 N 2 個の類似性入力値を全結合層に入力して、M 個 (N 2 よりも大きい整数) の隠

40

50

れノード値を提供するステップと、

回帰層を通じて前記M個の隠れノード値をマップして、画像品質のメトリックを提供するステップと、

を更に含む、請求項1-1に記載の方法。

【請求項1-3】

M=800である、請求項1-2に記載の方法。

【請求項1-4】

リファレンス画像に対する歪み画像の画像品質を評価するために畳み込みニューラルネットワークをトレーニングする方法であって、前記方法が、

前記畳み込みニューラルネットワークにおけるネットワークパラメータの現行のセットとしてネットワークパラメータの初期セットを選択するステップと、10

歪み画像及び対応するリファレンス画像から各々がなる画像の複数のペアの各々について、計算された類似性スコアメトリックS_cを提供するよう前記畳み込みニューラルネットワークを通じて前記画像のペアを処理するステップと、

前記計算された類似性スコアメトリックS_cと前記ペアについての期待類似性スコアメトリックS_pの比較に基づいて、前記ネットワークパラメータの現行のセットのうちの1又はそれよりも多い前記ネットワークパラメータを調整するステップと、

を含み、前記期待類似性スコアメトリックS_pが、人間の知覚によって提供される、方法。20

【請求項1-5】

前記画像の各ペアが、歪み画像パッチI_{dis}及びそのレファレンスパッチI_{ref}を含み、

前記画像のペアを処理するステップが、前記画像のペアを前記畳み込みニューラルネットワークに通して、サポートベクトル回帰(SVR)目的関数を前記複数の画像パッチに適用することによって、結果として得られる歪みスコアメトリックSを前記期待類似性スコアメトリックS_pと比較するステップを含む、請求項1-4に記載の方法。

【請求項1-6】

前記SVR目的関数が

【数1】

$$L(W) = \frac{1}{n} \sum_{i=1}^n L_\epsilon(f(I_{dis}(n), I_{ref}(n); W), y_n)$$

30

式中L_εは次式で定義され、

【数2】

$$L_\epsilon(y, \hat{y}) = \begin{cases} |y - \hat{y}| & |y - \hat{y}| \geq \epsilon \\ 0 & |y - \hat{y}| < \epsilon \end{cases}$$

式中_εは調整可能な小さい正定数であり、nはトレーニングサンプルの数である、請求項1-4に記載の方法。

【請求項1-7】

確率的勾配降下法(SGD)を用いてバックプロパゲーションを実行するステップを更に含む、請求項1-4に記載の方法。40

【請求項1-8】

前記畳み込みニューラルネットワークを通じて前記画像のペアを処理するステップが、結合特徴マップを提供するよう、前記歪み画像及びリファレンス画像の各ペアから生成された特徴マップの線形結合を計算するステップを含み、

前記線形結合を計算するステップが、スペースバイナリマスクを用いて、前記計算に対する特徴マップの入力と前記計算からの前記結合特徴マップの出力との間の結合を制約するステップを含む、請求項1-4に記載の方法。

【発明の詳細な説明】

50

【技術分野】**【0001】**

(関連出願の相互参照)

本出願は、2015年6月5日に提出された、名称「畳み込みニューラルネットワークに基づいたフルリファレンス画像品質評価方法」の米国特許出願第14/732,518号に基づく優先権を主張し、当該出願は、本明細書において本出願に全て記載されるよう引用により本明細書に組み込まれる。

【背景技術】**【0002】**

デジタル撮像デバイス及び通信技術の発展並びに普及に伴って、デジタル画像は、情報の表現及び伝達において益々重要になっている。デジタル画像は、その寿命の間に様々な段階で劣化する可能性があり、このような品質の劣化は、その後の段階で活用できなくなる恐れがある。従って、数多くの画像及びビデオ処理システムにおいて画像品質を維持及び監視することが重要であり、このような画像品質評価(IQA)の第1の目的は、観る人によって知覚されるような視覚的品質を予測することである。画像品質測度は、伝送速度のようなパラメータの関数として、また画像向上方法の最適パラメータを選択するため、知覚された歪みの依存性を評価するのに用いることができる。IQAを実行するために、主観テストを実験室設定にて実施することができるが、このようなテストは、コストと時間がかかり、リアルタイムの自動システムでは用いることはできない。従って、画像品質を自動的且つ効率的に測定する客観的IQAメトリックを構築し実現することが重要である。

10

【0003】

フルリファレンスIQA(FR-IQA)モデルは、歪み画像と対応する初期のリファレンス画像の両方からの情報を利用して視覚的品質を推定する。平均二乗誤差(Mean Squared Error(MSE))又はピーク信号対雑音比(Peak Signal-to-Noise Ratio(PSNR))などの従来のFR-IQAメトリックは、空間領域において歪み画像とリファレンス画像との間のピクセル毎の差違を直接測定している。これらのタイプのメトリックは、信号忠実度を測定するが、多くの場合、特に雑音が加法性ではない場合には、人間の知覚との相関性が低くなる。

20

【0004】

人間の知覚とより良好に一致する知覚による視覚的品質メトリック(PVQM)の構築に向けて、2つのタイプの手法、すなわち、ボトムアップとトップダウン手法がとられてきた。ボトムアップ手法は、コントラスト感度、輝度順応、種々のマスク効果などを含む関連の精神物理学的特性及び生理学的特性をシミュレートすることによって、人間の視覚システム(HVS)の視覚路における様々な処理段階をモデル化する試みである。しかしながら、これらの特性の知識が限定的なこと、並びに最終的な知覚への複合的な影響を考慮すると、HVSは複雑過ぎて、このようにして正確にモデル化することはできない。

30

【0005】

より最近の研究への取り組みは、自然画像の統計的特性などの様々な情報源からの知識並びに画像歪みがHVSによって処理される方法に関するデータを組み込むことによって入出力関係をモデル化するトップダウン機構に向けられている。最新のほとんどのFR-IQAモデルは、このカテゴリーに分類され、更に、構造類似性(Structural Similarity(SSIM))インデックス及びその変形形態(マルチスケールSSIM(Multi-Scale SSIM(MS-SSIM))及び情報重み付けSSIM(Information Weighted SSIM(IW-SSIM))を含む)、特徴類似性(Feature Similarity(FSIM))インデックス及び勾配度類似性偏差(Gradient Magnitude Similarity Deviation(GMSD))などは、ある程度の成功を収めており、平均強度、強度の標準偏差、位相一致及び勾配度などの低レベル視覚的特徴が有効な品質指標であることを示唆している。しかしながら、これら低レベルのキーは、異なる

40

50

歪みカテゴリー全体にわたり一様に良好には機能しない場合がある。その結果、対応するFR測度の性能は、異なるタイプの歪みにわたり大きく異なる可能性がある。

【0006】

従って、異なるタイプの歪みにわたって人間の知覚とより良好に一致し、客観性があり、測定可能なデータによって決定され且つ効率的である改善された画像品質評価方法に対する要求がある。一部では、畳み込みニューラルネットワーク（ConvNet）を利用した学習ベースの手法を適用することに向けて取り組まれているが、これらは、リファレンス画像が品質推定に利用可能ではない状況、すなわちノンリファレンスIQA（No-Reference IQA（NR-IQA））に限られている。従って、リファレンス画像及び歪み画像の対応するペアが分析に利用可能である、FR-IQAに対するこのような方法の適用を研究し構築する必要性が依然としてある。10

【発明の概要】

【0007】

本実施形態は、一般に、リファレンス画像に対する歪み画像の画像品質を評価するためのシステム及び方法を提供することに関する。1つの実施形態において、本システムは、歪み画像及びリファレンス画像を入力として受け入れて画像品質のメトリックを出力として提供する畳み込みニューラルネットワークを備える。別の実施形態において、本方法は、歪み画像及びリファレンス画像を処理して画像品質のメトリックとして提供するよう構成された畳み込みニューラルネットワークに対して、歪み画像及びリファレンス画像を入力するステップを含む。20

【0008】

更に別の実施形態において、リファレンス画像に対する歪み画像の画像品質を評価するために畳み込みニューラルネットワークをトレーニングする方法が提供される。本方法は、畳み込みニューラルネットワークにおけるネットワークパラメータの現行のセットとしてネットワークパラメータの初期セットを選択するステップと、歪み画像及び対応するリファレンス画像から各々がなる画像の複数のペアの各々について、計算された類似性スコアメトリック S_c を提供するよう畳み込みニューラルネットワークを通じて画像のペアを処理するステップと、歪みスコアメトリック S とペアについての期待類似性スコアメトリック S_p の比較に基づいて、ネットワークパラメータの現行のセットのうちの1又はそれよりも多いネットワークパラメータを調整するステップと、を含み、期待類似性スコアメトリック S_p は、人間の知覚によって提供される。30

【0009】

本明細書で開示された特定の実施形態の性質及び利点の更なる理解は、本明細書の残りの部分及び添付図面を参照することによって実現することができる。

【図面の簡単な説明】

【0010】

【図1】1つの実施形態によるリファレンス画像に対する歪み画像の画像品質を評価するシステムの利用を概略的に示す図である。

【図2】1つの実施形態によるリファレンス画像に対する歪み画像の画像品質を評価するシステムのアーキテクチャを示す図である。40

【図3】リファレンス画像に対する歪み画像の画像品質を評価する階層システムの1つの実施態様を示す図である。

【図4】1つの実施形態によるリファレンス画像に対する歪み画像の画像品質を評価する方法のステップを示すフローチャートである。

【図5】1つの実施形態によるリファレンス画像に対する歪み画像の画像品質を評価するための畳み込みニューラルネットワークをトレーニングする方法を示す図である。

【図6】1つの実施形態によるリファレンス画像に対する歪み画像の画像品質を評価するための畳み込みニューラルネットワークをトレーニングする方法のステップを示すフローチャートである。

【発明を実施するための形態】

10

20

30

40

50

【0011】

図1は、1つの実施形態によるリファレンス画像104に対する歪み画像102の画像品質を評価するシステム100の利用を概略的に示している。システム100は、トレーニングFR-IQAモデルを具現化する相互接続されたモジュール又は層のネットワークを含み、これは図2を参照して以下で更に説明される。歪み画像102及びリファレンス画像104の各々からの画像データがシステム100に提供され、該システムは、観る人によって知覚される可能性が高い歪み画像102の品質を示す出力メトリック106を生成する。

【0012】

図2は、1つの実施形態によるリファレンス画像214に対する歪み画像212の画像品質を評価するのに用いることができるシステム200のアーキテクチャの概略図を示す。システム200は、複数の層201～207を含む。システム200を2段階システムとみなすことが有用とすることができ、ここでは、第1段階が層201～205からなり、第2段階である層206及び207にデータを全体的に提供する。この第2段階は、「標準」ニューラルネットワークと考えることができ、第1段階と第2段階の組み合わせにより、システム200が畠み込みニューラルネットワークとなる。

10

【0013】

歪み画像212及びリファレンス画像214からの画像データが入力層201に提供され、該入力層201は、画像データの両方のセットを正規化して、正規化リファレンス画像及び正規化リファレンス画像を提供する機能を果たす。正規化歪み画像及び正規化リファレンス画像からの画像データは、畠み込み層202に提供され、該畠み込み層202は、N1個の複数のフィルタを用いて正規化歪み画像及び正規化リファレンス画像の各々を畠み込む機能を果たし、また、二乗活性化関数(activation function)を各画像の各ピクセルに適用して、特徴マップのN1個のペアをもたらす。特徴マップの各ペアは、1つのフィルタ処理された正規化歪み画像と1つの対応するフィルタ処理された正規化リファレンス画像とを含む。

20

【0014】

特徴マップのN1個のペアからの画像データは、線形結合層203に提供され、該線形結合層203は、歪み画像212に対応するN1個の特徴マップのN2個の線形結合と、リファレンス画像214に対応する対応N1個の特徴マップのN2個の線形結合とを計算して、結合特徴マップのN2個のペアを提供する。結合特徴マップの各ペアは、フィルタ処理された正規化歪み画像の1つの組み合わせと、フィルタ処理された正規化リファレンス画像の対応する1つの組み合わせとを含む。

30

【0015】

類似性計算層204は、線形結合層203から受け取った結合特徴マップのN2個のペアからのデータに作用して、N2個の類似性マップを計算する。各類似性マップは、結合特徴マップの1つのペアからのピクセルの対応するパッチによるデータに基づいて計算され、各類似性マップは、結合特徴マップのN2個のペアのうちの異なるものに対応する。

【0016】

類似性マップからのデータは、ブーリング層205に提供され、該ブーリング層は、N2個の類似性マップの各々について平均ブーリングを適用して、N2個の類似性入力値をもたらす。N2個の類似性入力値は、全結合層206に提供され、該全結合層206は、N2個の類似性入力値に作用して、M個の隠れノード値(ここでMは、N2よりも大きな整数)を提供する。M個の隠れノード値は、線形回帰層207によって单一の出力ノードにマップされる。出力ノードでの値は、システム200のトレーニングに基づいて、観る人によって知覚される可能性が高い歪み画像202の品質を示す画像品質のメトリックである。

40

【0017】

ニューラルネットワーク200の複数のハイパーパラメータは、特定の画像評価アプリケーションに好適な固有の異なるアーキテクチャ又は実施態様を決定するよう選択するこ

50

とができる。1つのハイパーパラメータは、オリジナルの歪み画像及びリファレンス画像におけるピクセルのパッチサイズである。他のハイパーパラメータは、畳み込み層で使用されるフィルタのサイズ（ピクセルの数に関する）、畳み込み層で使用されるフィルタの数N1、畳み込み層で使用される活性化関数、線形結合層にて計算された線形結合の数N2、全結合層における隠れノードの数M、及び類似性計算層において類似性を計算するのに使用される式を含む。更に、一部の実施形態において、2以上の畳み込み層、2以上の線形結合層、及び2以上の全結合層が存在してもよく、使用されるそれぞれの数は、追加のハイパーパラメータである。

【0018】

画像品質評価を実施するのに好適であることが分かっている畳み込みニューラルネットワークの1つの固有の実施態様が図3に例示され、 $32 \times 32 - 7 \times 7 \times 10 - 1 \times 1 \times 10 - 800 - 1$ 構造を示している。各層の詳細について以下で説明する。

【0019】

入力層は、2つの 32×32 グレースケール画像パッチ、すなわち、歪み画像パッチ I_{dis} 及びそのリファレンスパッチ I_{ref} からなる。両パッチは、線形変換関数 $f(x) = (x - 128) / 128 * 1.6$ により正規化される。

【0020】

畳み込み層において、10個の 7×7 線形フィルタ

$$W_1^k (k=1, \dots, 10)$$

が歪み及び画像パッチに適用されて、歪み画像及びリファレンス画像それぞれに対してフィルタ処理された特徴マップ

【数1】

$$D_1^k$$

及び

$$R_1^k, k=1, \dots, 10$$

をもたらし、ここで、

【数2】

$$D_1^k = W_1^{k \oplus I_{dis}}$$

及び、

【数3】

$$R_1^k = W_1^{k \oplus I_{ref}}$$

である。他の実施形態において、対象の低レベル特徴の最小サイズに応じて、 7×7 よりも小さい又は大きいフィルタを用いてもよい。同様に、一部の実施形態において、関連する歪みの複雑さに応じて、10個よりも少ない又は多いフィルタを用いてもよい。この層では、二乗活性化関数が、各画像パッチの各ピクセルに適用される。

【0021】

線形結合層において、畳み込み層によって生成されるフィルタ処理された特徴マップの線形結合が計算される。

【数4】

$$D_2^k = \sum_{i=1}^{10} W_2^k(i)(D_1^i)^2 \quad (1)$$

$$R_2^k = \sum_{i=1}^{10} W_2^k(i)(R_1^i)^2 \quad (2)$$

【0022】

類似性計算層において、対応する歪みパッチ及びリファレンスパッチからの結合特徴マップ間のポイント毎（すなわち、ピクセル毎）の方式で類似性が計算される。

【数5】

$$S^k = \frac{2D_2^k R_2^k + C}{(D_2^k)^2 + (R_2^k)^2 + C} \quad (3)$$

ここで、 S^k は第 k の類似性マップ、 C は小さい正定数である。

【0023】

プーリング層において、平均プーリングが各類似性マップに適用される。プーリング層の出力を

s_{ave}^k
とすると、

【数6】

$$s_{ave}^k = \text{mean}(S^k)$$

となる。プーリング層は、10次元プーリングされた類似性を入力として全結合層に提供する。10を超えるフィルタが畳み込み層で使用されて10を超えるフィルタ処理された特徴マップが生成されるなどの場合には、これに応じて、プーリングされた類似性入力に対しより多くの次元性が存在することになる。

【0024】

全結合層において、類似性入力が800個の隠れノードに密に結合される。隠れノードに対する活性化関数として、正規化線形関数 (Rectified Linear Unit (ReLU)) ($f(x) = \max(x, 0)$) が使用される。全結合層における重み付けを

【数7】

$$W_3(i, j) \quad (i = 1, \dots, 10; j = 1, \dots, 800)$$

とすると、全結合層の出力は、

【数8】

$$h(j) = \max \left(\sum_{i=1, \dots, 10} W_3(i, j) s_{ave}^i + b_j, 0 \right)$$

となり、ここで b_j はバイアス項である。一部の実施形態において、活性化関数は他の非線形動作を含むことができる点に留意されたい。

【0025】

線形回帰層において、800個の隠れノードが单一の出力ノードにマップされる。線形回帰層の重み付けが、

【数9】

$$W_4(i) \quad (i = 1, \dots, 800)$$

で記述される場合、ネットワークの出力は、

【数10】

$$y = \sum_{i=1, \dots, 800} W_4(i) h(i)$$

+ b

であり、ここで b はバイアス項である。

【0026】

図4は、1つの実施形態によるリファレンス画像に対する歪み画像の画像品質を評価するための方法400のステップを示すフローチャートである。ステップ402において、歪み画像及びリファレンス画像（例えば、歪み及びリファレンスビデオストリームからの対応するフレームとすることができる）がトレーニング畳み込みニューラルネットワークへの入力として受けられる。ステップ404において、歪み画像及びリファレンス画像は、トレーニング畳み込みニューラルネットワークを通じて処理される。

10

20

30

40

50

【0027】

図5は、1つの実施形態によるリファレンス画像に対する歪み画像の画像品質を評価するための畳み込みニューラルネットワークモデルをトレーニングする方法を示す。モデルについてネットワークパラメータの初期セットが選択され、観る人によって決定された、期待又は予測される類似性スコアメトリックを有する対応する歪み画像(502)及びリファレンス画像(504)のペアがネットワークへの入力として提供される。ネットワークによって提供される類似性スコアメトリックS'は、コストモジュール506において、予測されたスコアメトリックSと比較され、この比較に応じて、モデル500の1又はそれよりも多いパラメータが調整されてモデルを更新する。次いで、歪み画像及びリファレンス画像の第2のペアが、調整されたパラメータを用いてネットワーク500を通じて処理されて、生成された第2のメトリックとこの第2のペアで期待されるスコアメトリックとの間の比較が行われ、これに応じて更に調整を行うことができる。このプロセスは、最終のモデルパラメータが設定されて、モデルネットワークがトレーニングされたとみなされるまで、トレーニング画像ペアの利用可能なセットに対して必要に応じて繰り返される。10

【0028】

図6は、1つの実施形態によるリファレンス画像に対する歪み画像の画像品質を評価するための畳み込みニューラルネットワークをトレーニングする方法600のステップを示すフローチャートである。ステップ602において、歪み画像及びリファレンス画像のペア(利用可能な全T個のペアのうちのi番目のペア)が、パラメータの以前に選択されたセットによって特徴付けられるネットワークへの入力として受けられる。ステップ604において、歪み画像及びリファレンス画像がネットワークを通じて処理される。ステップ606において、畳み込みニューラルネットワークによって計算された出力類似性スコアScと、観る人によって知覚される、画像の当該ペアに対して予測された類似性スコアメトリックSpとの間の比較に応じて、ネットワークパラメータが目的関数を用いて調整される。ステップ608において、未だ処理されていないトレーニング画像の他のペアが利用可能であるかどうか判定される。利用可能である場合、このような1つのペアがステップ602及び604を通じて処理されて、ステップ606において、ネットワークパラメータが更に調整され、ステップ608において、全てのトレーニング画像ペアの処理が完了したかどうかの更なる判定が行われる。ステップ608において、全てのトレーニング画像の処理が完了したと判定されると、ニューラルネットワークがトレーニングされたとみなすことができる。場合によっては、トレーニング画像は、トレーニングを改善するために2回以上ネットワークに提供されてもよい。トレーニング画像が処理される順序はランダムであってもよい。2030

【0029】

1つの実施形態において、例えば、方法600のステップ606でモデルパラメータがどのように調整されるかを正確に判定する際に、ネットワークをトレーニングするのに使用される目的関数は、標準の-SVR(サポートベクトル回帰)で使用されるのと同じである。

$$(I_{ref}(n), I_{dis}(n))$$

が入力パッチのペアを表し、y_nが計算されたw類似性スコアメトリック(グラウンドトリース品質スコアとも呼ばれる)を表し、40

$$f(I_{ref}(n), I_{dis}(n); W)$$

がネットワークパラメータWとの入力ペアについての予測類似性スコアメトリックを表すとすると、目的関数は次式で記述することができる。

【数11】

$$L(W) = \frac{1}{n} \sum_{i=1}^n L_\epsilon(f(I_{dis}(n), I_{ref}(n); W), y_n)$$

この場合、 L は次式で定義される。

【数12】

$$L_\epsilon(y, \hat{y}) = \begin{cases} |y - \hat{y}| & |y - \hat{y}| \geq \epsilon \\ 0 & |y - \hat{y}| < \epsilon \end{cases}$$

ここで ϵ は、調整可能な小さい正定数であり、 n はトレーニングサンプルの数である。ネットワークは、確率的勾配降下法 (SGD) を用いてバックプロパゲーションを実施することによってトレーニングすることができる。

【0030】

正則化法を用いて、ニューラルネットワークの過学習を回避することができる。1つの実施形態において、正則化法は、目的関数の線形回帰層において重み付けの L_2 ノルムを付加するステップを含む。これは、例えば SVM (サポートベクタマシン) で使用されてきた正則化に広く使用されている方法である。詳細には、目的関数は、次式のように修正することができる。

【数13】

$$L(W) = \frac{1}{n} \sum_{i=1}^n L_\epsilon(f(I_{dis}(n), I_{ref}(n); W), y_n) + \lambda_1 \|W\|_{L_2}$$

ここで、 λ_1 は小さい正定数である。

【0031】

一部の実施形態において、線形結合層における入力ノード及び出力ノードの結合は制約されるので、ほとんどの出力ノードは、入力ノードの小さなセット (例えば、 $N_1 = 10$ の場合、全ての 10 個の入力ノードではなく) とのみ結合される。この制約は、スパースバイナリマスク W_{mask} を付加することによって実装することができ、ここで $W_{mask}(k, i) = 0$ の場合、

$$W_{mask}(i)$$

は強制的にゼロでなければならない。結合パターンは、ランダムに選ぶことができる。1 つの特定の実施形態において有用であることが分かっているマスクは、以下で定義される。

【数14】

$$W_{mask} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

【0032】

本明細書で記載される実施形態は、様々な利点をもたらす。詳細には、本実施形態は、リファレンス画像及び歪み画像の対応するペアが利用可能である用途において、リアルタイムの自動システムで容易に実装され且つ異なるタイプの歪みにわたり人間の知覚とより良好に一致する結果をもたらすシステム及び方法を用いて画像品質を評価することを可能にする。本発明は、入力としてリファレンス画像及び歪み画像を受け入れて、出力として有意義な画像品質メトリックを提供する、自動画像品質評価の「エンドツーエンド」解決手段をもたらす。

【0033】

本発明の詳細な説明では、その特定の実施形態について説明してきたが、これらの特定の実施形態は、単に例証に過ぎず、限定ではない。

【0034】

10

20

30

40

50

特定の実施形態のルーチンを実行するために、C、C++、Java(登録商標)、アセンブリ言語等を含むあらゆる好適なプログラム言語を使用できる。手続き形又はオブジェクト指向等の別のプログラミング手法を使用することができる。ルーチンは、単一の処理デバイス又は複数のプロセッサで実行することができる。しかしながら、ステップ、動作、又は計算は、特定の順番で提示されるが、この順番は、別の特定の実施形態では変更することができる。特定の実施形態において、本明細書では順次的に示される複数のステップは、同時に実行してもよい。

【0035】

特定の実施形態は、命令実行システム、装置、システム、又はデバイスによって、又はこれに関連して使用するコンピュータ可読記憶媒体で実施することができる。特定の実施形態は、ソフトウェア又はハードウェアもしくはこれらの組み合わせにおける制御ロジックの形態で実施することができる。制御ロジックは、1又はそれよりも多いプロセッサで実行される場合、特定の実施形態で説明されたものを実行するように作動可能である。10

【0036】

特定の実施形態は、プログラム汎用デジタルコンピュータ、特定用途向け集積回路、プログラム可能倫理デバイス、フィールドプログラマブルゲートアレイ、光学的、化学的、生物学的、量子力学的又はナノ加工システム、構成要素、及び機構を使用して実行できる。一般に、特定の実施形態の機能は、従来から公知の任意の手段で実現することができる。分散型、ネットワークシステム、構成要素及び/又は回路を使用することができる。データの通信又は伝送は、有線、無線、又は他の何れかの手段とすることができます。20

【0037】

また、図面に示される要素の1又はそれよりも多い要素は、特定の用途に有用であるよう、より分離された又は統合された様式で、或いは場合によっては取り除かれるか又は動作不能にされた状態で実施できることは理解されるであろう。また、マシン可読媒体に格納されてコンピュータが上述の方法の何れかを実行可能にすることができるプログラム又はコードを実行することは、本発明の精神及び範囲に含まれる。

【0038】

「プロセッサ」は、データ、信号、又は他の情報を処理する何らかの好適なハードウェア及び/又はソフトウェアシステム、機構又は構成要素を含む。プロセッサは、汎用中央処理装置、マルチ処理装置、特定機能を実現する専用回路、又は他のシステムを含むことができる。処理は、地理的位置に限定されるか又は時間的限界を有する必要はない。例えば、プロセッサは、「リアルタイム」、「オフライン」、「バッチモード」等での機能を実行することができる。処理の一部は、別の(又は同じ)処理システムによって別の時間及び場所で実行することができる。処理システムの実施例は、サーバ、クライアント、エンドユーザデバイス、ルータ、スイッチ、ネットワークストレージなどを含むことができる。コンピュータは、メモリと通信するあらゆるプロセッサとすることができます。メモリは、プロセッサにより実行する命令を格納するのに好適な、ランダムアクセスメモリ(RAM)、読み出し専用メモリ(ROM)、磁気又は光ディスク、又は他の有形媒体などの何れかの好適なプロセッサ可読記憶媒体とすることができます。30

【0039】

本明細書及び請求項を通して使用する場合、単数形態は、前後関係から明らかに別の意味を示さない限り、複数形態も含む。本明細書及び請求項を通して使用する場合、前後関係から明らかに別の意味を示さない限り、「における」の意味は「における」と「上で」の意味を含む。

【0040】

従って、本明細書において特定の実施形態を説明してきたが、許容範囲の変形例、種々の変更、及び代用例は前述の開示内容の対象であり、場合によっては、特定の実施形態の一部の特徴は、本発明の範囲及び精神を逸脱することなく、他の特徴を対応して使用することなく利用されることは理解される。従って、特定の状況又は内容を実質的な範囲及び精神に適合させるように、多くの変形を行うことができる。4050

【符号の説明】

【0041】

104：リファレンス

102：歪み

100：FR-IQAモデル（トレーニング）

【図1】

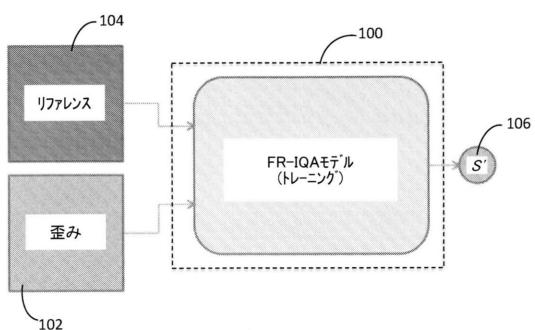


Figure 1

【図2】

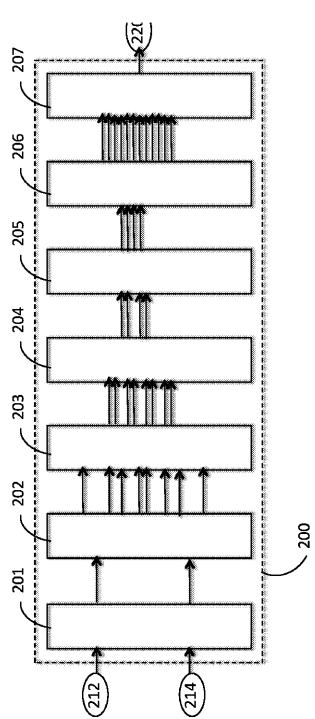


Figure 2

【図3】

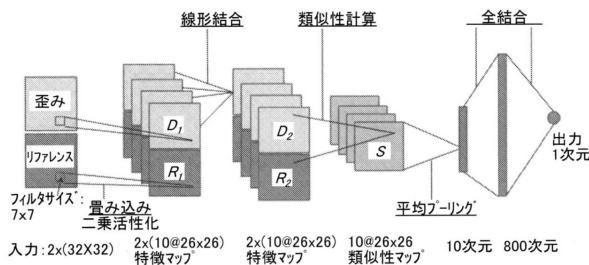


Figure 3

【図5】

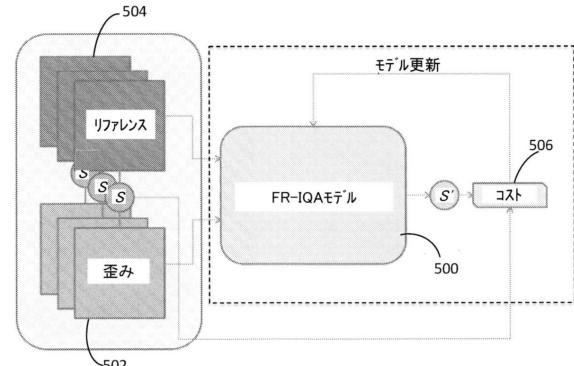


Figure 5

【図4】

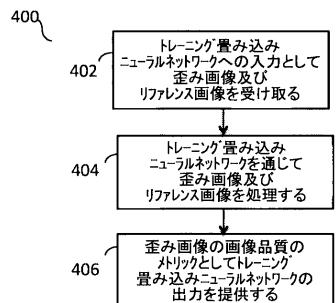


Figure 4

【図6】

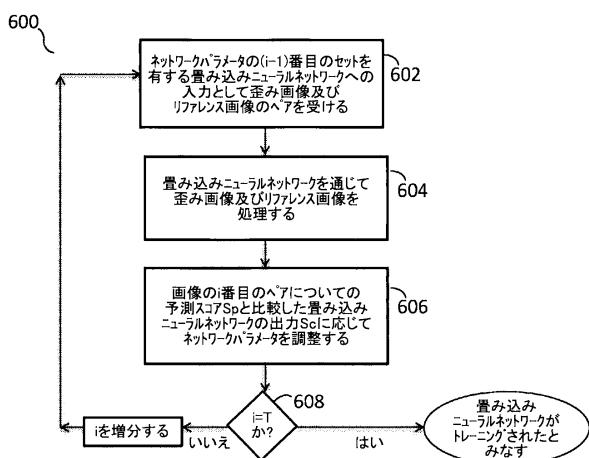


Figure 6

フロントページの続き

(72)発明者 シュー シュン

アメリカ合衆国 カリフォルニア州 95014 クバチーノ パリッシュ ブレイス 1016
4

(72)発明者 イエ ペン

アメリカ合衆国 カリフォルニア州 94404 フォスター シティ キャボット レーン 8
63

審査官 松田 直也

(56)参考文献 特表2006-505853(JP,A)

特表2015-503778(JP,A)

Kang et al., Convolutional Neural Networks for No-Reference Image Quality Assessment,
2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014年 6
月28日, p.1733~1740, U R L, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6909620>

(58)調査した分野(Int.Cl., DB名)

G06T 7/00