

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7118697号

(P7118697)

(45)発行日 令和4年8月16日(2022.8.16)

(24)登録日 令和4年8月5日(2022.8.5)

(51)国際特許分類

F I

G 0 6 T 7/00 (2017.01)

G 0 6 T 7/00 3 5 0 C

G 0 9 G 5/00 (2006.01)

G 0 9 G 5/00 5 5 0 C

G 0 9 G 5/36 (2006.01)

G 0 9 G 5/00 5 5 0 X

G 0 9 G 5/36 5 2 0 P

G 0 9 G 5/36 5 1 0 M

請求項の数 31 (全23頁) 最終頁に続く

(21)出願番号 特願2018-68120(P2018-68120)

(22)出願日 平成30年3月30日(2018.3.30)

(65)公開番号 特開2019-179390(P2019-179390
A)

(43)公開日 令和1年10月17日(2019.10.17)

審査請求日 令和3年1月27日(2021.1.27)

(73)特許権者 515130201

株式会社 Preferred Netw
orks東京都千代田区大手町1丁目6番1号
大手町ビル

(74)代理人 100091982

弁理士 永井 浩之

(74)代理人 100091487

弁理士 中村 行孝

(74)代理人 100082991

佐藤 泰和

(74)代理人 100105153

弁理士 朝倉 悟

(74)代理人 100118876

弁理士 鈴木 順生

最終頁に続く

(54)【発明の名称】 注視点推定処理装置、注視点推定モデル生成装置、注視点推定処理システム、注視点推定処理方法、プログラム、および注視点推定モデル

(57)【特許請求の範囲】

【請求項1】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定する注視点推定部と、

複数のニューラルネットワークを記憶する記憶部と、を備え、

前記注視点推定部が、前記推定用画像の視聴者の属性に基づき、前記複数のニューラルネットワークから前記注視点に関する情報の推定に用いる前記ニューラルネットワークを選択する、

推定装置。

【請求項2】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定する注視点推定部、を備え、

前記注視点推定部は、前記推定用画像の視聴者の属性を用いて、前記注視点を推定する、推定装置。

【請求項3】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定する注視点推定部と、

前記注視点に関する情報に基づき、前記推定用画像を加工する画像加工部と、を備える、推定装置。

【請求項 4】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定する注視点推定部と、

前記注視点に関する情報に基づき、前記推定用画像を評価する評価部と、を備える、推定装置。

【請求項 5】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定する注視点推定部と、

前記推定用画像と前記注視点に関する情報とに基づき、前記推定用画像の視聴者の注視物を導出する注視物導出部と、を備える、

推定装置。

【請求項 6】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定する注視点推定部、を備え、

前記注視点に関する情報は、前記推定用画像内のピクセルが前記注視点である可能性を示す画像であって、

前記画像は、ヒートマップ、等高線図、及び、3次元グラフのうちの少なくとも一つである、

推定装置。

【請求項 7】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定する注視点推定部、を備え、

前記注視点に関する情報は、前記推定用画像内における領域に関する情報を含む、

推定装置。

【請求項 8】

前記推定用画像からオブティカルフローを算出する算出部

をさらに備え、

前記注視点推定部は、前記オブティカルフローを用いて、前記注視点に関する情報を推定する

請求項 1 ないし 7 のいずれか一項に記載の推定装置。

【請求項 9】

複数のニューラルネットワークを記憶する記憶部

をさらに備え、

前記注視点推定部が、前記推定用画像の視聴者の属性に基づき、前記複数のニューラルネットワークから前記注視点に関する情報の推定に用いる前記ニューラルネットワークを選択する、

請求項 2 ないし 8 のうち 1 に従属しないいずれか一項に記載の推定装置。

【請求項 10】

前記注視点推定部は、前記推定用画像の視聴者の属性を用いて、前記注視点を推定する

請求項 1、3 ないし 9 のうち 2 に従属しないいずれか一項に記載の推定装置。

【請求項 11】

前記注視点に関する情報に基づき、前記推定用画像を加工する画像加工部

をさらに備える請求項 1、2、4 ないし 10 のうち 3 に従属しないいずれか一項に記載

の推定装置。

10

20

30

40

50

【請求項 1 2】

前記注視点に関する情報に基づき、前記推定用画像を評価する評価部
をさらに備える

請求項 1 ないし 3、5 ないし 1 1 のうち 4 に従属しないいずれか一項に記載の推定装置。

【請求項 1 3】

前記推定用画像と前記注視点に関する情報とに基づき、前記推定用画像の視聴者の注視物を導出する注視物導出部

をさらに備える請求項 1 ないし 4、6 ないし 1 2 のうち 5 に従属しないいずれか一項に記載の推定装置。

【請求項 1 4】

前記推定用画像の視聴者に関する入力データに前記注視物に関する情報が含まれている場合、前記注視物に関する返答を生成する返答生成部

をさらに備える請求項 5 または 1 3 に記載の推定装置。

【請求項 1 5】

前記注視点に関する情報は、前記推定用画像内のピクセルが前記注視点である可能性を示す情報を含む、

請求項 1 ないし 5、7 ないし 1 4 のうち 6 に従属しないいずれか一項に記載の推定装置。

【請求項 1 6】

前記注視点に関する情報は、前記推定用画像内のピクセルが前記注視点である可能性を示す画像である、

請求項 1 ないし 5、7 ないし 1 5 のうち 6 に従属しないいずれか一項に記載の推定装置。

【請求項 1 7】

前記画像は、ヒートマップ、等高線図、及び、3次元グラフのうちの少なくとも一つである、

請求項 1 6 に記載の推定装置。

【請求項 1 8】

前記注視点に関する情報は、前記推定用画像内における領域に関する情報を含む、

請求項 1 ないし 1 7 のいずれか一項に記載の推定装置。

【請求項 1 9】

前記推定用画像の視聴者は、前記推定用画像を見る人である、

請求項 1、2、5 のいずれか一項、または、8 ないし 1 8 のうち 1、2、5、9、1 0、1 3、1 4 のいずれかに従属するいずれか一項に記載の推定装置。

【請求項 2 0】

請求項 6、または、8 ないし 1 9 のうち 6、1 6 のいずれかに従属するいずれか一項に記載の推定装置を用いて、前記注視点である可能性を示す画像を生成する方法。

【請求項 2 1】

画像を撮影する撮影装置と、

請求項 1 ないし 1 9 のいずれか一項に記載の推定装置と、

を備え、

前記推定装置が、前記撮影装置によって撮影された画像を、前記推定用画像として受け取り、

前記推定装置が、前記注視点に関する情報を前記撮影装置に送信し、

前記撮影装置が、前記注視点の位置または周辺にある物体を撮影するために自装置の設定を変更する

システム。

【請求項 2 2】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定する注視点推定部、を備え、移動体に搭載された撮影画像により撮影された画像を前記推定用画像として受け取る推定装置と、

10

20

30

40

50

前記推定装置により推定された前記注視点に関する情報に基づき、前記移動体を移動させるように制御する制御装置と、
を備えるシステム。

【請求項 2 3】

移動体に搭載された撮影装置により撮影された画像を前記推定用画像として受け取る請求項 1 ないし 1 9 のいずれか一項に記載の推定装置と、

前記推定装置により推定された前記注視点に関する情報に基づき、前記移動体を移動させるように制御する制御装置と、
を備えるシステム。

【請求項 2 4】

ニューラルネットワークを記憶する少なくとも 1 つのメモリと、
少なくとも 1 つのプロセッサと、を備え、
前記少なくとも 1 つのプロセッサは、

対象画像を前記ニューラルネットワークに入力することによって評価スコアを生成し、
前記対象画像に対するアイトラッキング情報を前記ニューラルネットワークに入力することなく、前記対象画像内のピクセルが注視点である可能性を示す画像を生成し、
生成された前記画像が、ヒートマップ、等高線図、及び、3 次元グラフのうちの少なくとも一つである、
評価装置。

【請求項 2 5】

学習用画像と前記学習用画像に対する注視点に関する情報とに基づく学習を行うことにより、ニューラルネットワークを生成するモデル生成方法であって、

前記ニューラルネットワークは、前記学習用画像に対するアイトラッキング情報が入力されることなく、前記学習用画像が前記ニューラルネットワークに入力されることにより、前記注視点に関する情報を出力し、

前記注視点に関する情報は、前記学習用画像内のピクセルが前記注視点である可能性を示す画像であって、
前記画像は、ヒートマップ、等高線図、及び、3 次元グラフのうちの少なくとも一つである、

モデル生成方法。

【請求項 2 6】

推定用画像に対するアイトラッキング情報をニューラルネットワークに入力することなく、前記ニューラルネットワークを用いて前記推定用画像から前記推定用画像に対する注視点に関する情報を推定するステップ、を備え、

前記注視点に関する情報は、前記推定用画像内のピクセルが前記注視点である可能性を示す画像であって、

前記画像は、ヒートマップ、等高線図、及び、3 次元グラフのうちの少なくとも一つである、

画像の生成方法。

【請求項 2 7】

対象画像をニューラルネットワークに入力することによって評価スコアを生成するステップと、

前記対象画像に対するアイトラッキング情報を前記ニューラルネットワークに入力することなく、前記対象画像内のピクセルが注視点である可能性を示す画像を生成するステップと、

を備え、

生成された前記画像が、ヒートマップ、等高線図、及び、3 次元グラフのうちの少なくとも一つである、

評価方法。

【請求項 2 8】

10

20

30

40

50

複数のニューラルネットワークから、前記対象画像の視聴者の属性に基づき、前記ニューラルネットワークを選択するステップ

をさらに備える

請求項 27 に記載の評価方法。

【請求項 29】

前記対象画像の視聴者の属性が、性別、年齢、出身地、住所、職業、家族構成、年収、趣味、及び、状態のうちの少なくとも一つである、

請求項 28 に記載の評価方法。

【請求項 30】

前記対象画像の視聴者は、前記対象画像を見る人である、

請求項 28 または 29 に記載の評価方法。

【請求項 31】

請求項 25 から請求項 30 のいずれかに記載の方法を少なくとも 1 台のコンピュータに実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、注視点推定処理装置、注視点推定モデル生成装置、注視点推定処理システム、注視点推定処理方法、プログラム、および注視点推定モデルに関する。

【背景技術】

【0002】

複数のカメラを用いてアイトラッキングを行い、アイトラッキングの対象者の注視点を推定する技術の開発が進められている。対象者の注視点を認識することができれば、対象者が注視している物体に関連する情報を提供するというサービスを行うことができる。

【0003】

しかし、対象者の注視点を推定するには、対象者の特徴点の位置情報を収集するための複数の装置が必要となる。例えば、目を撮影するカメラ、頭部の向きを検出するカメラ、対象者の目と基準点との距離を測定する装置などが必要となる。そのため、当該装置が設置された環境でしか注視点を算出することができない。また、対象者が移動している場合、対象者が群衆にまぎれてしまう場合などでは、アイトラッキングによる推定精度が低下してしまう。

【先行技術文献】

【特許文献】

【0004】

【文献】特許 6187155 号公報

【非特許文献】

【0005】

【文献】Phillip Isola、外 3 名、"Image-to-Image Translation with Conditional Adversarial Networks"、[online]、平成 29 年 11 月 22 日、Berkeley AI Research (BAIR) Laboratory University of California、[平成 30 年 3 月 1 日検索]、インターネット (URL : <https://arxiv.org/pdf/1611.07004.pdf>)

【発明の概要】

【発明が解決しようとする課題】

【0006】

アイトラッキングが困難な環境下では、人の注視点を推定し、推定された注視点に基づくサービスを提供することが困難である。

【課題を解決するための手段】

【0007】

本発明の一実施形態である注視点推定処理装置は、注視点推定モデルを記憶する記憶部と、注視点推定部と、を備える。学習用画像と前記学習用画像に対する第 1 注視点に関する

10

20

30

40

50

る情報とに基づく学習により生成された注視点推定モデルを記憶する。前記注視点推定部は、前記注視点推定モデルを用いて、推定用画像から、前記推定用画像に対する第２注視点に関する情報を推定する。

【図面の簡単な説明】

【０００８】

【図１】第１の実施形態に係る注視点推定処理システムの一例を示すブロック図。

【図２】学習用画像および検出注視点画像を説明する図。

【図３】第１の実施形態に係る注視点推定処理システムの処理の概略フローチャート。

【図４】第１の実施形態に係る注視点推定モデル生成装置の処理の概略フローチャート。

【図５】第１の実施形態に係る注視点推定処理装置の処理の概略フローチャート。

10

【図６】第２の実施形態に係る注視点推定処理システムの一例を示すブロック図。

【図７】第２の実施形態に係る注視点推定処理装置の処理の概略フローチャート。

【図８】第３の実施形態に係る注視点推定処理システムの一例を示すブロック図。

【図９】第３の実施形態に係る注視点推定処理装置の処理の概略フローチャート。

【図１０】第４の実施形態に係る注視点推定処理システムの一例を示すブロック図。

【図１１】注視点推定処理装置との会話について説明する図。

【図１２】第４の実施形態に係る注視点推定処理装置の処理の概略フローチャート。

【図１３】本発明の一実施形態におけるハードウェア構成の一例を示すブロック図。

【発明を実施するための形態】

【０００９】

20

以下、図面を参照しながら、本発明の実施形態について説明する。

【００１０】

（第１の実施形態）

図１は、第１の実施形態に係る注視点推定処理システムの一例を示すブロック図である。図１に示された注視点推定処理システム１は、学習用画像表示装置１１と、注視点検出装置１２と、属性データ提供装置１３と、注視点推定モデル生成装置１４と、推定用データ提供装置１５と、注視点推定処理装置１６と、出力先装置１７と、を備える。注視点推定モデル生成装置１４は、学習用画像データ記憶部１４１と、検出注視点画像データ記憶部１４２と、属性データ記憶部１４３と、学習用画像解析部（オプティカルフロー算出部）１４４と、注視点推定モデル生成部（学習部）１４５と、を備える。注視点推定処理装置１６は、注視点推定モデル記憶部１６１と、推定用画像解析部（オプティカルフロー算出部）１６２と、注視点推定部１６３と、を備える。

30

【００１１】

第１の実施形態に係る注視点推定処理システムは、推定用データ提供装置１５から画像が入力されると、当該画像を見た視聴者が示すと推定される当該画像内の注視点を推定し、当該注視点に関する情報を出力するというシステムである。具体的には、まず、注視点推定モデル生成装置１４が、ディープラーニングにより、注視点を推定するための注視点推定モデルを生成する。そして、注視点推定処理装置１６が、注視点推定モデルを用いて、推定用データ提供装置１５から入力された画像から、注視点を推定する。処理の詳細は、各装置および各構成要素とともに説明する。

40

【００１２】

なお、上記では、注視点を推定するために必要と思われる主な装置および構成要素を記載しており、その他の装置および構成要素が含まれていてもよい。また、装置および構成要素が細分化されていてもよいし、一つにまとめられていてもよい。また、各装置は、通信ネットワークを介したデータの送受が可能であり、各装置の入力および出力に係る構成要素（つまり、入力部および出力部）は、省略されている。また、装置および構成要素に係る処理が省略される場合には、当該装置および構成要素も省略されてよい。

【００１３】

例えば、一つの注視点推定モデル生成装置が、生成した注視点推定モデルを複数に複製して、複数の注視点推定処理装置１６に送信し、各注視点推定処理装置１６は、注視点推

50

定モデルをソフトウェアの一部であるプログラムモジュールとして使用することが想定される。そのため、図 1 では、注視点推定モデル生成装置 1 4 と、注視点推定処理装置 1 6 と、が異なる装置として示されている。しかし、注視点推定モデル生成装置 1 4 および注視点推定処理装置 1 6 が、同一の装置であってもよい。また、例えば、注視点推定モデルへの入力情報として、オプティカルフローなどの画像解析結果を用いない場合は、学習用画像解析部 1 4 4 および推定用画像解析部 1 6 2 は省略されてよい。

【 0 0 1 4 】

なお、本説明において、「学習」という用語は、機械学習を意味してもよいし、ディープラーニングを意味してもよい。また、「画像」という用語は、静止画および動画の両方を包括する概念であり、特に問題がなければ、静止画または動画に置き換えられて読まれてもよい。つまり、注視点推定処理システムにより用いられる画像は、動画でも静止画でもよい。また、装置内において、画像はデータとして扱われるが、画像データの表現形式は特に限られるものではない。また、位置およびその位置に対応する情報が示されたデータは画像データとして扱ってよい。ヒートマップなどのグラフも画像に含まれる。なお、画像の表示とともに音が提供されることもある。音のような画像とともに付随する情報により、注視点が変化することもあり得る。そこで、本説明における「画像」には、画像に付随する情報も含まれるものとする。

【 0 0 1 5 】

各装置について説明する。

【 0 0 1 6 】

学習用画像表示装置 1 1 は、視聴者に対し、学習用画像を表示する装置である。学習用画像表示装置 1 1 は、表示した学習用画像を注視点推定モデル生成装置 1 4 に送信する。なお、学習用画像の視聴者を第 1 視聴者と記載する。第 1 視聴者は、複数人を想定するが、1 人でもよい。

【 0 0 1 7 】

注視点検出装置 1 2 は、学習用画像を視聴中の第 1 視聴者に対して、アイトラッキングを行い、学習用画像上の第 1 視聴者の注視点を検出する。例えば、注視点を検出可能なヘッドマウントディスプレイが製品化されている。当該ヘッドマウントディスプレイを学習用画像表示装置 1 1 および注視点検出装置 1 2 として用いてもよい。

【 0 0 1 8 】

注視点の検出後、注視点検出装置 1 2 は、検出結果に基づき、学習用画像上の第 1 視聴者の注視点を示す画像を生成し、当該画像を注視点推定モデル生成装置 1 4 に送信する。当該画像を、検出注視点画像と記載する。検出注視点画像は、学習用画像内の各ピクセルが注視点である可能性（スコア）を示す画像である。スコアは、確率分布などで示すことが可能である。検出注視点画像は、ヒートマップ、等高線図、3 次元グラフといった画像となり得る。

【 0 0 1 9 】

図 2 は、学習用画像および検出注視点画像を説明する図である。左上側と左下側にある画像は、学習用画像を示す。当該学習用画像は、ある商品のコマーシャルのための画像であることを想定している。左上側の画像には人が、左下側の画像には商品とその人が、表されている。右上側と右下側にある画像は、検出注視点画像を示す。図 2 の例の検出注視点画像では、各ピクセルが注視点である可能性を濃淡で表しており、濃いほうが注視点である可能性が高く、薄いほうが注視点である可能性が低いことを示す。右上側の検出注視点画像が左上側の学習用画像に対応し、右下側の検出注視点画像が左下側の学習用画像に対応している。これにより、人だけが表されている左上の学習用画像では、その人の顔が注視される可能性が高いことが分かる。また、商品と人が表されている左下の学習用画像では、商品が注視される可能性は高いが、人だけが表されている学習用画像における人の顔と比較して、注視される可能性が低いことが示されている。

【 0 0 2 0 】

図 2 のような、学習用画像と、学習用画像に対応し学習用画像上の注視点を示す検出注

10

20

30

40

50

視点画像と、が注視点推定モデル生成装置 14 に送信される。なお、学習用画像と、検出注視点画像と、は、予め対応づけがされており、注視点推定モデル生成装置 14 が学習用画像と検出注視点画像との組み合わせを認識できるものとする。例えば、検出注視点画像の識別子が学習用画像の識別子の一部を含むように、注視点検出装置 12 が、検出注視点画像に識別子を付与すればよい。あるいは、学習用画像の表示時刻と、検出注視点画像の注視点の検出時刻と、に基づき、対応関係が把握されてもよい。

【0021】

なお、本実施形態は、検出注視点画像を注視点検出装置 12 が生成しているが、注視点推定モデル生成装置 14 が生成してもよい。例えば、注視点検出装置 12 は、検出注視点画像ではなく、ピクセルごとのスコアが示されたデータといった注視点に関するデータを注視点推定モデル生成装置 14 に送信し、注視点推定モデル生成装置 14 が、その内部で当該データから検出注視点画像を生成してもよい。

10

【0022】

属性データ提供装置 13 は、第 1 視聴者の属性を示すデータを注視点推定モデル生成装置 14 に提供する。当該データを単に属性データと記載する。例えば、学習用画像表示装置 11 が画像を表示する前に、第 1 視聴者が属性データ提供装置 13 に対し第 1 視聴者の属性を入力する。そして、第 1 視聴者の属性も、学習用画像および検出注視点画像と同様に、これらと対応付けられる。用いられる属性は、特に限られず、適宜に定めてよい。

【0023】

学習用画像における第 1 視聴者の注視点は、第 1 視聴者が興味を示す対象上にあると想定されるが、当該対象は、第 1 視聴者の属性により異なる。例えば、性別、年齢、出身地、住所、職業、家族構成、年収、趣味、状態などが異なれば、興味を示す対象が異なる傾向がある。例えば、スポーツ選手の画像が表示されたときに、第 1 視聴者が興味を示すか否かは、第 1 視聴者の趣味に影響されやすい。また、第 1 視聴者の疲労度が高い場合、第 1 視聴者の注意力が散漫となり、画像を注視しないこともあり得る。ゆえに、第 1 視聴者の属性が、学習用画像および検出注視点画像と対応付けられたほうが好ましい。なお、必ずしも属性が対応付けられていなくともよい。

20

【0024】

また、第 1 視聴者の学習用画像を見る直前の行動、第 1 視聴者の学習用画像を見る場所、第 1 視聴者の学習用画像を見る時間帯なども、注視点に影響すると考えられる。これらの情報も属性として用いてもよい。例えば、第 1 視聴者に対するライフログが記録されている場合、注視点推定モデル生成装置 14 は、ライフログ記録装置からライフログを受信し、受信されたライフログを属性データとして用いてよい。つまり、ライフログ記録装置は、属性データ提供装置 13 に該当する。

30

【0025】

注視点推定モデル生成装置 14 は、学習用画像および検出注視点画像を入力情報として学習を行い、注視点推定モデルを生成する。

【0026】

なお、属性も入力情報に含めてよい。あるいは、注視点推定モデルが、属性ごとに分けて、生成されてもよい。例えば、学習用画像と、男性の第 1 視聴者による検出注視点画像と、を用いて、男性用の注視点推定モデルを生成してもよい。

40

【0027】

注視点推定モデル生成装置 14 が備える、学習用画像データ記憶部 141、検出注視点画像データ記憶部 142、および属性データ記憶部 143 はそれぞれ、学習用画像、検出注視点画像、および属性データを記憶する。

【0028】

学習用画像解析部（オプティカルフロー算出部）144 は、学習用画像を解析することにより、学習用画像に関する情報を生成する。本実施形態では、当該解析により、時系列の学習用画像から、オプティカルフローが算出されることを想定する。なお、オプティカルフロー以外の情報を算出してもよい。

50

【 0 0 2 9 】

オプティカルフローは、動画像などの時系列の画像における（フレーム間における）速度ベクトルの分布状態を示す。つまり、算出されたオプティカルフローは、学習用画像に表された物体の動きを示す。

【 0 0 3 0 】

物体が急に表示された場合、表示された物体が移動した場合などでは、当該物体は注視される傾向にある。つまり、学習用画像に表された物体の動きは、注視点に影響を与える。そのため、オプティカルフローは、注視点推定モデルの生成の際の入力情報として用いられる。なお、本説明において、「物体」には、ヒトなどの生体も含まれる。

【 0 0 3 1 】

オプティカルフローの算出方法としては、様々な方法が知られている。例えば、勾配法、Lucas-Kanade法、マッチング、グラディエント、パーティクルフィルタといったものがある。本実施形態で用いられるオプティカルフローの算出方法は、特に限られるものではなく、公知手法から適宜に定めてよい。

【 0 0 3 2 】

注視点推定モデル生成部 145 は、少なくとも、学習用画像および検出注視点画像に基づく学習を行うことにより、注視点推定モデルを生成する。なお、前述の通り、属性も学習の入力情報に含めてよい。注視点推定モデルは、与えられた画像から別の画像を生成するための、多層のニューラルネットワーク（ディープニューラルネットワーク：DNN）のモデルである。学習用画像、オプティカルフローなどがニューラルネットワークの入力層に入力され、各中間層における演算を経ることにより、模擬画像が出力層から出力される。注視点推定モデル生成部 145 は、模擬画像が検出注視点画像に近づくように、各中間層における演算の重み付け係数（パラメータ）を更新する。模擬画像が検出注視点画像に近づいているかは、特徴ベースマッチング、領域ベースマッチングといった既存のマッチング手法を用いて判断してもよい。こうして、学習済みのニューラルネットワークは、推定用画像から推定注視点画像を生成可能となる。

【 0 0 3 3 】

ニューラルネットワークも様々な種類が提唱されているが、適宜に選んでよい。例えば、敵対生成ネットワーク（GAN）を用いると、与えられた画像から別の画像を精度よく生成できることが知られている。

【 0 0 3 4 】

敵対生成ネットワークを用いた場合、注視点推定モデルは、学習用画像から模擬画像を生成する画像生成モデルと、与えられた画像が模擬画像なのか検出注視点画像なのかを識別する識別モデルと、を含む。学習により、検出注視点画像であると識別モデルが識別してしまうような模擬画像を画像生成モデルが生成できるように、画像生成モデルのパラメータが更新されていく。その一方で、学習により、識別モデルが正しく模擬画像と識別できるように、識別モデルのパラメータも更新されていく。このようにして、画像生成モデルおよび識別モデルのパラメータの更新が繰り返されることにより、生成される模擬画像が、検出注視点画像に近づくこととなる。つまり、学習用画像から検出注視点画像を生成可能となる。

【 0 0 3 5 】

また、敵対生成ネットワークに条件を加えた、条件付き敵対生成ネットワーク（cGAN）を用い、さらに中間層のネットワーク構造を、Encoder-Decoder 構造から U-net 構造に変えることにより、画像生成の精度が高まることが知られている。このようなニューラルネットワークを用いることにより、入力された画像から、注視点を示す画像を生成するモデルを生成することが可能である。

【 0 0 3 6 】

また、注視点推定モデルの出力は、必ずしも画像データとは限られない。例えば、注視点推定モデルのニューラルネットワークが第 1 および第 2 のニューラルネットワークに別れており、第 1 のニューラルネットワークにおいて画像データが出力され、第 2 のニュー

10

20

30

40

50

ラルネットワークにおいて当該画像に対する処理が行われて、注視点推定モデルの出力として、注視点とされるピクセルまたは領域が出力されてもよい。また、第1および第2のニューラルネットワークによる機能を1つのニューラルネットワークで実現してもよい。

【0037】

推定用データ提供装置15は、注視点推定処理装置に、注視点の推定の対象の画像を提供する装置である。当該画像を推定用画像と記載する。また、推定用画像を視聴すると想定された視聴者を、第2視聴者と記載して、学習用画像の第1視聴者と区別する。なお、第2視聴者が、第1視聴者と同じであってもよい。

【0038】

また、推定用データ提供装置15は、推定用画像だけでなく、推定用画像を解析することにより得られたデータを、注視点推定処理装置16に提供してもよい。当該データを推定用データと記載する。例えば、画像認識モデルを用いて、推定用画像に表された物体を認識し、推定用画像とともに当該物体の名称、輪郭などを提供してもよい。また、注視点推定モデルが属性ごとに生成されている場合は、第2視聴者の属性が推定用データに含まれていてもよい。また、以降、推定用画像も推定用データに含まれるものとして記載する。

【0039】

注視点推定処理装置16は、注視点推定モデルを用いて、推定用画像から、第2視聴者の注視点に関する情報を推定する。

【0040】

注視点推定モデル記憶部161は、注視点推定モデルを属性ごとに分類して記憶する。これにより、注視点推定部163が、属性に応じた注視点推定モデルを抽出することができる。

【0041】

推定用画像解析部（オプティカルフロー算出部）162は、推定用画像を解析することにより、推定用画像に関する情報を生成する。本実施形態では、学習用画像解析部144と同様に、当該解析により、時系列の推定用画像から、推定用画像に対するオプティカルフローが算出されることを想定する。

【0042】

注視点推定部163は、注視点推定モデル記憶部161から注視点推定モデルを取得する。推定用データ提供装置15から第2視聴者の属性を取得した場合は、複数の注視点推定モデルのうちから、当該属性に応じた注視点推定モデルを抽出する。そして、注視点推定モデルを用いて、推定用画像から、推定用画像内の各ピクセルが注視点である可能性を示す情報を生成する。当該情報は、各ピクセルが注視点であるか否かを示す情報であればよく、画像を示すデータでもよいし、当該画像に基づき加工が行われたデータでもよい。

【0043】

なお、ここでは、画像が生成されるとして説明を行う。当該画像を、推定注視点画像と記載する。注視点推定モデルの入力層に推定用画像が入力されると、推定モデルの各中間層において、学習済みのパラメータに基づく演算が行われ、注視点推定モデルの出力層から推定用注視点画像が出力される。推定用画像とオプティカルフローから、推定注視点画像を生成してもよい。

【0044】

推定注視点画像は、推定用画像内の各ピクセルが注視点である可能性（スコア）を示す画像である。推定注視点画像は、ヒートマップ、等高線図、3次元グラフなどのような、検出注視点画像と同じ表示形態の画像となる。

【0045】

生成された推定注視点画像を推定された注視点とみなしてもよいが、注視点推定部163は、推定注視点画像に基づき、推定注視点画像の一部を注視点として認定してもよい。注視点は、1つのピクセルとしてもよいし、複数のピクセルを含む領域としてもよい。例えば、推定注視点画像のスコアが最も高いピクセルを注視点としてもよい。あるいは、所定の単位円または単位矩形を用いて、単位円または単位矩形内のスコアの合計が最も高い

10

20

30

40

50

領域を算出し、当該領域の中心のピクセルを注視点としてもよいし、当該領域を注視点としてもよい。あるいは、含まれる全てのピクセルのスコアが閾値以上である領域を注視点としてもよい。例えば、推定注視点画像がヒートマップで表されている場合に、最も濃い領域を注視点としてもよい。

【0046】

推定注視点画像、注視点といった注視点に関する情報は、指定された出力先装置17に出力される。出力先装置17は、特に限られるものではない。これにより、出力先装置17において、推定された注視点に関する情報に基づき、様々な処理を実行することができる。また、注視点は、画像で表示されてもよいし、位置座標などで表示されてもよい。

【0047】

例えば、出力先装置17が、モニタなどの画像表示装置であって、注視点推定処理装置16による処理結果がモニタなどに表示されてもよい。あるいは、外部装置または外部装置を制御する制御装置に推定結果が送信されて、推定結果に基づき、当該外部装置が制御されてもよい。

【0048】

例えば、推定用データ提供装置15および出力先装置17が同一の撮影装置（例えばデジタルカメラ）であるとする。当該撮影装置が注視点推定処理装置16に推定用画像を提供し、注視点が当該撮影装置に返される場合、当該撮影装置が、注視点を含む物体または注視点に含まれる物体（注視点が領域の場合）を精度よく撮影するために、設定を変更することが可能になる。例えば、フォーカスを合わせる、ズームアップするといった制御が可能になる。

【0049】

また、出力先装置17が、ドローン、レスキューロボットといった移動体の制御装置である場合、移動体に搭載されたカメラからの画像を推定用画像として用いることが考えられる。そして、推定注視点画像を当該制御装置に送信することにより、制御装置が推定注視点画像に基づき、移動体の移動を制御することが可能となる。例えば、当該移動体を注視点のほうに移動させるといった制御を行うこともできる。これにより、人による操縦なしで、当該移動体を人が興味を持つほうへ移動させることもできる。

【0050】

次に、本実施形態の処理の流れについて説明する。図3は、第1の実施形態に係る注視点推定処理システム1の処理の概略フローチャートである。本フローチャートでは、属性データを用いる場合を記載する。

【0051】

属性データ提供装置13が、学習用画像を視聴する第1視聴者の属性を取得する（S101）。学習用画像表示装置11が学習用画像を第1視聴者に表示する（S102）。注視点検出装置12が学習用画像上の第1視聴者の注視点を測定し、検出注視点画像を生成する（S103）。そして、第1視聴者の属性、学習用画像、および検出注視点画像の各データが各装置から送信され、注視点推定モデル生成装置14がこれらのデータを受信する（S104）。

【0052】

注視点推定モデル生成装置14は、これらの受信データから注視点推定モデルを生成する（S105）。生成された注視点推定モデルは、注視点推定処理装置16に送られる。その後、推定用画像、第2視聴者の属性といった推定用データが、推定用データ提供装置15から送られ、注視点推定処理装置16が当該推定用データを受信する（S106）。注視点推定処理装置16は、第2視聴者の属性に応じた注視点推定モデルを用いて、推定用画像から注視点を推定する（S107）。そして、注視点に関する情報が出力先装置17に送信されて、出力先装置17が注視点に関する情報を用いて処理を実行する（S108）。このようにして、推定された注視点に基づくサービスが提供される。

【0053】

次に、注視点推定モデル生成装置14の処理の流れについて説明する。図4は、第1の

10

20

30

40

50

実施形態に係る注視点推定モデル生成装置 1 4 の処理の概略フローチャートである。本フローチャートでは、敵対生成ネットワークを用いる場合を記載する。

【 0 0 5 4 】

学習用画像解析部 1 4 4 が、時系列の推定用画像からオプティカルフローを算出する（S 2 0 1）。一方、注視点推定モデル生成部 1 4 5 は、属性データと対応する注視点推定モデルを選択する（S 2 0 2）。注視点推定モデル生成部 1 4 5 が、選択された注視点推定モデルを用いて、学習用画像および対応するオプティカルフローに基づき、模擬画像を生成する（S 2 0 3）。注視点推定モデル生成部 1 4 5 が、模擬画像を検出注視点画像として認識可能か判定する（S 2 0 4）。注視点推定モデル生成部 1 4 5 が判定結果に基づき、注視点推定モデルのパラメータを更新する（S 2 0 5）。これらの処理が何度も行われることにより、注視点推定モデルのパラメータが改善されていき、注視点推定モデルの精度が向上する。

10

【 0 0 5 5 】

次に、注視点推定処理装置 1 6 の処理の流れについて説明する。図 5 は、第 1 の実施形態に係る注視点推定処理装置 1 6 の処理の概略フローチャートである。

【 0 0 5 6 】

推定用画像解析部 1 6 2 が、時系列の推定用画像からオプティカルフローを算出する（S 3 0 1）。一方、注視点推定部 1 6 3 は、注視点推定モデル記憶部 1 6 1 に記憶されている複数の注視点推定モデルから、属性に応じた注視点推定モデルを選択する（S 3 0 2）。注視点推定部 1 6 3 が、選択された注視点推定モデルを用いて、推定用画像および対応するオプティカルフローに基づき、推定注視点画像を生成する（S 3 0 3）。さらに、注視点推定部 1 6 3 が推定注視点画像に基づき、注視点を認定する（S 3 0 4）。こうして、注視点に関する情報が生成される。

20

【 0 0 5 7 】

なお、本説明におけるフローチャートは一例であり、上記の例に限られるものではない。実施形態の求められる仕様、変更などに応じて、手順の並び替え、追加、および省略が行われてもよい。例えば、オプティカルフローを用いない場合は、オプティカルフローの算出は省略される。また、上記では、説明の便宜上、順番に処理が行われたが、オプティカルフローの算出と、注視点推定モデルの選択と、は並行に処理されてもよい。以降のフローチャートについても同様である。

30

【 0 0 5 8 】

以上のように、本実施形態によれば、画像に基づき、注視点を推定することができる。アイトラッキングを行う場合は、視聴者を撮影するためのカメラなどの装置が必要であり、視聴者と基準点との相対的位置関係を認識し続けなければならないといった欠点があったが、本実施形態は、これらの欠点を有しない。

【 0 0 5 9 】

また、第 1 視聴者の属性に応じた注視点推定モデルを生成することもでき、第 2 視聴者の属性に応じた注視点推定モデルを用いて注視点推定を行うことにより、注視点の推定の精度が向上する。

【 0 0 6 0 】

40

また、オプティカルフローを考慮した注視点推定モデルを生成することもでき、画像内の物体の動きに対応した注視点推定が行われるため、注視点の推定の精度が向上する。

【 0 0 6 1 】

（第 2 実施形態）

図 6 は、第 2 の実施形態に係る注視点推定処理システムの一例を示すブロック図である。図 6 に示された注視点推定処理システム 1 では、注視点推定処理装置 1 6 が、画像加工部 1 6 4 をさらに備える。なお、第 1 の実施形態と同様な点は、説明を省略する。

【 0 0 6 2 】

画像加工部 1 6 4 は、推定用画像内の注視点に基づき、推定用画像を加工する。そして、加工された推定用画像は、出力先装置 1 7 に出力される。つまり、本実施形態の注視点

50

推定処理装置 16 は、画像加工装置でもあり、出力先装置 17 は画像表示装置を想定する。また、推定用画像は、出力先装置 17 にて表示される予定の画像である。

【0063】

推定用画像の加工は、予め定められた規則に基づき、行われればよい。当該規則は、推定用データ提供装置 15 から提供されてもよいし、予め注視点推定処理装置 16 内に設定されていてもよい。

【0064】

加工規則は適宜に定めてよい。例えば、推定用データ提供装置 15 から、推定用画像内の物体の輪郭を示す情報を受け取り、推定用画像内の注視点に基づき、推定用画像内の物体のいずれかを、精細にレンダリングするといった加工を行ってもよい。

【0065】

例えば、ロボットなどの移動体からの送信画像を見ながら、人が遠隔通信にて当該移動体を操作する場合に、送信画像を推定用画像として用いて注視点を検出した上で、当該送信画像を注視点に基づき加工し、操縦士に表示することが想定される。当該想定においては、注視点から離れている物体は、操縦士が認識していない可能性があり、移動体が当該物体にぶつかる恐れがある。ゆえに、画像加工部 164 が、注視点から離れている物体を強調するように、送信画像を加工するといったことが考えられる。このような処理を行うことにより、運転、操作などの補助を行うことができる。

【0066】

また、推定用画像の注視点の周辺の画像を、推定用画像に対応する属性に応じた広告に変更してもよい。当該広告は、推定用データに含まれていればよい。また、推定用データに複数の広告が含まれている場合に、画像加工部 164 が第 2 視聴者の属性に応じた広告を抽出してもよい。これにより、推定用画像を提供する業者、広告業者などのニーズに応えることができる。

【0067】

また、第 2 視聴者の注視点に応じて、これから表示する画像の一部または全体を変更してもよい。例えば、推定用画像に物体 A と物体 B とが表示されている場合に、画像加工部 164 が、注視点に基づき、第 2 視聴者が物体 A のほうを注視すると認識したとする。その場合において、次に加工する推定用画像として、物体 A に係る推定用画像と、物体 B に係る推定用画像と、を受け取ったときは、画像加工部 164 は、物体 A に係る推定用画像を用いる。あるいは、その場合において、画像加工部 164 は、次に加工する推定用画像から物体 B を消去する加工を行う。こうして、推定用画像を第 2 視聴者用に特化することにより、第 2 視聴者の画像に対する満足度、集中度などが向上する。

【0068】

図 7 は、第 2 の実施形態に係る注視点推定処理装置 16 の処理の概略フローチャートである。本実施形態の注視点推定処理装置 16 の処理の流れは、S301 から S304 までは、図 5 に示したフローチャートと同じである。そして、注視点推定部 163 が注視点を認定した後 (S304) に、画像加工部 164 が推定注視点画像または注視点に基づき、推定用画像を加工する (S401)。これにより、出力先装置 17 が加工された推定用画像を用いて処理を実行することができる。

【0069】

以上のように、本実施形態によれば、注視点推定処理装置 16 が注視点に基づき推定用画像の加工まで行う。これにより、推定用画像の第 2 視聴者が注視する物体などに応じて、推定用画像を変更することができ、第 2 視聴者または提供者のニーズに応えた画像を提供することができる。

【0070】

(第 3 実施形態)

図 8 は、第 3 の実施形態に係る注視点推定処理システムの一例を示すブロック図である。図 8 に示された注視点推定処理システム 1 では、注視点推定処理装置 16 が、画像評価部 165 をさらに備える。なお、第 1 の実施形態と同様な点は、説明を省略する。

10

20

30

40

50

【 0 0 7 1 】

画像評価部 1 6 5 は、推定用画像内の注視点に基づき、推定用画像を評価する。そして、画像評価部 1 6 5 による評価が、出力先装置 1 7 に出力される。つまり、本実施形態の注視点推定処理装置 1 6 は、画像評価装置でもある。

【 0 0 7 2 】

推定用画像の評価は、所定の評価規則に基づき、行われればよい。評価規則は、推定用データ提供装置 1 5 から提供されてもよいし、予め注視点推定処理装置 1 6 に設定登録されていてもよい。

【 0 0 7 3 】

評価規則内容は適宜に定めてよい。例えば、推定用データ提供装置 1 5 から、注視点として好ましい領域が指定されて、当該領域と、推定された注視点との差分により、評価されてもよい。また、評価は、スコアで表されてもよいし、優、良、可、不可といった所定の分類項目に分類されてもよい。

10

【 0 0 7 4 】

図 9 は、第 3 の実施形態に係る注視点推定処理装置 1 6 の処理の概略フローチャートである。本実施形態の注視点推定処理装置 1 6 の処理の流れは、S 3 0 1 から S 3 0 4 までは、図 5 に示したフローチャートと同じである。そして、そして、注視点推定部 1 6 3 が注視点を認定した後 (S 3 0 4) に、画像評価部 1 6 5 が推定注視点画像または注視点に基づき、推定用画像を評価する (S 5 0 1)。これにより、出力先装置 1 7 が推定用画像の評価を用いて処理を実行することができる。

20

【 0 0 7 5 】

以上のように、本実施形態によれば、注視点推定処理装置 1 6 が注視点に基づき推定用画像の評価まで行う。これにより、推定用画像の提供者は、推定用画像を第 2 視聴者に見せる前に、推定用画像の客観的な評価を得ることができ、推定用画像の改善を行うことができる。

【 0 0 7 6 】

(第 4 実施形態)

図 1 0 は、第 4 の実施形態に係る注視点推定処理システムの一例を示すブロック図である。図 1 0 に示された注視点推定処理システム 1 では、音データ提供装置 1 8 をさらに備える。また、注視点推定処理装置 1 6 が、注視物導出部 1 6 6 と、返答生成部 1 6 7 と、をさらに備える。なお、第 1 の実施形態と同様な点は、説明を省略する。

30

【 0 0 7 7 】

本実施形態では、画像と音声の両方を利用するマルチモーダル機能を有するシステムを想定し、推定された注視点に関する情報が、装置と人との会話に応用される。ここでは、第 2 視聴者と会話する装置が、推定された注視点に関する情報を利用して、第 2 視聴者が示唆する物体を認識することを想定する。

【 0 0 7 8 】

人の指示は、不明確になることが多い。例えば「あれを取ってください」というような指示があるが、「あれ」が意味する物体が何であるかを他者が理解することは困難である。しかし、このような指示代名詞を用いた指示を出す場合、指示者は「あれ」が意味する物体を注視している可能性が高い。ゆえに、本実施形態では、注視点に関する情報を、特定が困難な語句に対する補助情報として用いることにより、ユーザが伝えようとする物体を特定する。

40

【 0 0 7 9 】

音データ提供装置 1 8 は、マイクなどによって拾われた音データ化して、注視点推定処理装置 1 6 に送信する装置である。当該データは、例えば、第 2 視聴者の音声をテキスト化したものでもよい。あるいは、音データ提供装置 1 8 は、マイクに拾われた音が何の音であることを判定し、判定された結果が注視点推定処理装置 1 6 に送信されてもよい。音のデータ化は、汎用の音認識ソフトにより行うことが可能である。

【 0 0 8 0 】

50

なお、本実施形態では、音データに含まれている、「この」、「あの」といった指示語と、当該指示語に対応する語句と、を必要とする。ゆえに、音データ提供装置 18 は、音データを全て送信してもよいし、必要とされる語句に関するデータだけを送信してもよい。なお、当該指示語に対応する語句は、単に、指示語の後に続く語句としてもよい。例えば、「あのレストランについて教えて」という音データが得られた場合、音データ提供装置 18 は、少なくとも、「あの」および「レストラン」を、注視点推定処理装置 16 に送信する。なお、音データが全て送信される場合は、注視点推定処理装置 16 が、予め登録されたリストに含まれる指示語と、マッチングを行うことにより、指示語と、指示語に対応する語句を抽出すればよい。

【0081】

10

注視点推定処理装置 16 の注視物導出部 166 は、推定用画像内の注視点に基づき、第 2 視聴者が注視すると推定される物体を導出する。当該物体を注視物と記載する。注視点を含む物体を注視物としてもよい。注視点が領域である場合は、当該領域内の物体のいずれかにしてもよい。例えば、当該領域内の物体のうち、最も大きく表示されている物体としてもよい。

【0082】

注視物導出部 166 は、注視物の名称を導出する。注視物の名称は、推定用データから導出すればよい。あるいは、注視物導出部 166 が、予め生成済みの画像認識モデルを用いて、当該物体の名称を推定してもよい。

【0083】

20

また、注視物導出部 166 は、注視物の位置を示す語句を導出する。位置を示す語句は、ピクセルの座標値などではなく、日常会話においてよく用いられる語句にする方が好ましい。例えば、注視点のピクセルが、推定用画像の中心座標を基準とした四象限の第 1 象限にある場合、ユーザ認識推定部は、第 1 象限に予め割り当てられていた語句「右上」を導出する。割り当て語句は、予め記憶された、位置および割り当て語句の対応関係を示すリストから導出されればよい。

【0084】

返答生成部 167 は、指示語と対応する語句と、注視物の名称と、が一致しているかを判定する。一致している場合は、指示語と対応する語句が、注視物を指しているとして、注視物に関する返答を、音データ、推定用データに含まれる注視物の情報などを用いて、生成する。

30

【0085】

図 11 は、注視点推定処理装置 16 との会話について説明する図である。図 11 の例では、第 2 視聴者が車両を運転中の運転手であることを想定している。図 11 (A) は、運転手が見ている画像が示されている。当該画像は、車両に搭載されたカメラから取得される。そして、当該画像が推定用画像として、注視点推定処理装置 16 に送信されているとする。図 11 (B) は、注視点推定部 163 が生成した推定注視点画像を示す図である。図 11 (B) の左側に、黒の濃度が高い領域が存在する。当該領域を注視点とすると、注視物導出部 166 は、例えば、「注視物」が「レストラン」であり、「注視物の位置」が「左側」であると導出して、返答生成部 167 に出力する。

40

【0086】

返答生成部 167 に、上記の注視点に関する情報が入力されているときに、「あのレストランについて教えて」という音データが返答生成部 167 に入力されたとする。返答生成部 167 は、当該音データに注視物の名称「レストラン」が含まれていることを認識し、返信を生成する。例えば、音データに含まれる「あのレストラン」と、注視点に関する情報に含まれる「左側」と、を用いて、「あのレストランは左側のレストランですか」という質問文を生成する。

【0087】

あるいは、推定用データ提供装置 15 から提供された、推定用画像内の物体の情報を用いて、回答文を生成してもよい。例えば、推定用データに、前述のレストランの分類区分

50

(例えば、和食、イタリアン、ファストフードなど)、企業名、営業時間などが含まれていることが考えられる。これらの情報を用いて、「左側のレストランは、和食レストランです」という回答を生成することもできる。

【0088】

図11(A)には、右側にもレストランが示されている。注視点に関する情報がなければ、第2視聴者が言う「あのレストラン」がどちらのレストランを示唆しているかを認識することはできない。そのため、どのレストランかを確認するための質問文を生成することになるが、「あのレストランは左側のレストランですか」といった同意だけを求める質問文は、第2視聴者に与える不満は少ないと想定される。例えば、第2視聴者に対して「どちらのレストランですか」と質問すると、第2視聴者がレストランを説明することになるため、第2視聴者の不満が増加する。また、「左側のレストランですか、右側のレストランですか」と質問することもあるが、第2視聴者が右側のレストランを認識していないと、第2視聴者が反応できない恐れもある。また、根拠なく「あのレストランは左側のレストランですか」と聞いた場合には、注視していないほうを尋ねる可能性が本実施形態よりも高い。つまり、本実施形態よりも、第2視聴者の不満が増加する可能性が高い。

10

【0089】

したがって、音データに含まれる名称に係る物体が、第2視聴者が見ている画像内に複数ある場合にも、同意だけを求める質問文、または、回答文を生成できる本実施形態は、第2視聴者に与える不満を抑え、円滑な会話を可能にする。

【0090】

20

返答生成部167は生成した返答データを出力先装置17に送信する。本実施形態では、出力先装置17が音声生成装置であることを想定する。出力先装置17は返答データを音声に変換して、スピーカ等で出力する。これにより、第2視聴者との会話が成立する。音声生成装置は、テキストから音声を合成する汎用ソフトを用いて実現することが可能である。

【0091】

図12は、第4の実施形態に係る注視点推定処理装置16の処理の概略フローチャートである。注視点の推定に係るS301からS304までの処理は、これまでの実施形態と同じであり、ここではS304の後の処理を説明する。

【0092】

30

注視物導出部166が推定された注視点に基づき、注視物の名称を導出する(S601)。注視物導出部166が、注視物の位置を示す語句を導出する(S602)。返答生成部167が、注視物導出部166から注視物の名称および位置を、音データ提供装置18から音データを取得する(S603)。指示語と対応する語句が、物体の名称と一致している場合(S604のYES)に、返答生成部167が物体に係る返答を生成する(S605)。こうして、返答が生成されて、音声生成装置に送られることにより、第2視聴者との会話が成立する。

【0093】

なお、上記では、注視物の導出と、返答の生成と、を注視点推定処理装置16が行うとしたが、注視点推定処理装置16が注視物の導出までを行い、導出された注視物を出力先装置17に送信し、返答の生成は、出力先装置17が行ってもよい。

40

【0094】

なお、上記では、装置と人との会話が、音声により行われることを想定した。しかし、耳が不自由な人、言語が異なる人に対しては、音声ではなく、手話、筆談(テキスト)などにより会話を行う場合もあり得る。そのような場合においては、音データ提供装置18の代わりに、カメラにより撮影されたジェスチャーを解読するジェスチャー解読装置、またはカメラにより撮影された文字を解読する光学文字認識装置(OCR)を用いる。そして、音データの代わりに、これらの装置が解読した内容を示すデータが、注視点推定処理装置16に送信されればよい。これにより、注視点推定処理装置16が返答を生成する処理は上記と同様に行うことができる。つまり、返答生成部167への入力データは、音デ

50

ータに限られるものではない。なお、返答の送信先は、画像表示装置でもよいし、音声生成装置でもよい。返答内容を異なる言語に置き換えることは、返答の送信先で行われればよい。

【0095】

以上のように、本実施形態によれば、ロボットなどの装置に音声で指示を出すような場合において、注視点に関する情報が補助情報として用いられる。これにより、指示代名詞などが含まれ特定が困難な語句に対しても、ユーザが伝えようとする物体を特定して返答することができる。

【0096】

なお、上記の各実施形態は、組み合わせられてもよい。例えば、第2の実施形態の画像加工部164と、第3の実施形態の画像評価部165と、の両方を備える注視点推定処理装置に係る実施形態も存在する。画像評価部165による評価に基づき、画像加工部164が加工内容を変化させてもよい。

10

【0097】

なお、上記の実施形態では、注視点推定処理装置16は、注視点に関する情報を各ピクセルに対するスコアで表現していたが、注視点に関する情報を各ピクセルの座標で表現してもよい。例えば、注視点が存在する領域をバウンディングボックスで表現する場合、注視点推定部163が、矩形に対応する4つの座標に関する情報を出力するように構成してもよい。この場合、注視点推定モデル生成装置14では、注視点に係るバウンディングボックスの座標に関する情報を出力するように注視点推定モデルを学習する。また、注視点に関する情報として、矩形領域とその領域に対応するスコアの情報とを出力するようにしてもよい。

20

【0098】

なお、上記の実施形態の少なくとも一部は、プロセッサ、メモリなどを実装しているIC(Integrated Circuit:集積回路)などの専用の電子回路(すなわちハードウェア)により実現されてもよい。また、上記の実施形態の少なくとも一部は、ソフトウェア(プログラム)を実行することにより、実現されてもよい。例えば、汎用のコンピュータ装置を基本ハードウェアとして用い、コンピュータ装置に搭載された中央処理装置(CPU:Central Processing Unit)、画像処理装置(GPU:Graphics Processing Unit)などのプロセッサ(処理回路)にプログラムを実行させることにより、上記の実施形態の処理を実現することが可能である。言い換えると、当該プログラムの実行により、プロセッサ(処理回路)が、注視点推定モデル生成装置14、注視点推定処理装置16などの各処理を実行できるように構成される。

30

【0099】

例えば、コンピュータが読み取り可能な記憶媒体に記憶された専用のソフトウェアをコンピュータが読み出すことにより、コンピュータを上記の実施形態の装置とすることができる。記憶媒体の種類は特に限定されるものではない。また、通信ネットワークを介してダウンロードされた専用のソフトウェアをコンピュータがインストールすることにより、コンピュータを上記の実施形態の装置とすることができる。こうして、ソフトウェアによる情報処理が、ハードウェア資源を用いて、具体的に実装される。

40

【0100】

図13は、本発明の一実施形態におけるハードウェア構成の一例を示すブロック図である。注視点推定モデル生成装置14および注視点推定処理装置16は、プロセッサ21と、主記憶装置22と、補助記憶装置23と、ネットワークインタフェース24と、デバイスインタフェース25と、を備え、これらがバス26を介して接続されたコンピュータ装置2として実現できる。

【0101】

なお、図13のコンピュータ装置2は、各構成要素を一つ備えているが、同じ構成要素を複数備えていてもよい。また、図13では、1台のコンピュータ装置2が示されている

50

が、ソフトウェアが複数のコンピュータ装置にインストールされて、当該複数のコンピュータ装置それぞれがソフトウェアの異なる一部の処理を実行してもよい。

【0102】

プロセッサ21は、コンピュータの制御装置および演算装置を含む電子回路（処理回路）である。プロセッサ21は、コンピュータ装置2の内部構成の各装置などから入力されたデータやプログラムに基づいて演算処理を行い、演算結果や制御信号を各装置などに出力する。具体的には、プロセッサ21は、コンピュータ装置2のOS（オペレーティングシステム）や、アプリケーションなどを実行することにより、コンピュータ装置2を構成する各構成要素を制御する。プロセッサ21は、上記の処理を行うことができれば特に限られるものではない。注視点推定モデル生成装置14および注視点推定処理装置16内の画像解析部（144および162）、注視点推定モデル生成部145、注視点推定部163などは、プロセッサ21により実現される。また、プロセッサ21は、注視点推定モデルを読み込むと、推定用画像から推定注視点画像などを出力するように機能する。

10

【0103】

主記憶装置22は、プロセッサ21が実行する命令および各種データなどを記憶する記憶装置であり、主記憶装置22に記憶された情報がプロセッサ21により直接読み出される。補助記憶装置23は、主記憶装置22以外の記憶装置である。なお、これらの記憶装置は、電子情報を格納可能な任意の電子部品を意味するものとし、メモリでもストレージでもよい。また、メモリには、揮発性メモリと、不揮発性メモリがあるが、いずれでもよい。注視点推定モデル生成装置14および注視点推定処理装置16内の各記憶部（141、142、143、161）は、主記憶装置22または補助記憶装置23により実現されてもよい。

20

【0104】

ネットワークインタフェース24は、無線または有線により、通信ネットワーク3に接続するためのインタフェースである。ネットワークインタフェース24は、既存の通信規格に適合したものをを用いればよい。ネットワークインタフェース24より、通信ネットワーク3を介して、コンピュータ装置2と外部装置4Aとを接続することができる。

【0105】

デバイスインタフェース25は、外部装置4Bと直接接続するUSBなどのインタフェースである。つまり、コンピュータ装置2と外部装置4との接続は、ネットワークを介してでもよいし、直接でもよい。

30

【0106】

なお、外部装置4（4Aおよび4B）は、注視点推定モデル生成装置14および注視点推定処理装置16以外の各装置でもよいし、外部記憶媒体でもよいし、ストレージ装置でもよい。

【0107】

なお、外部装置4は入力装置でもよい。入力装置は、キーボード、マウス、タッチパネルなどのデバイスを備え、これらのデバイスにより入力された情報をコンピュータ装置2に与える。入力装置からの信号はプロセッサ21に出力される。

【0108】

上記に、本発明の一実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

40

【符号の説明】

【0109】

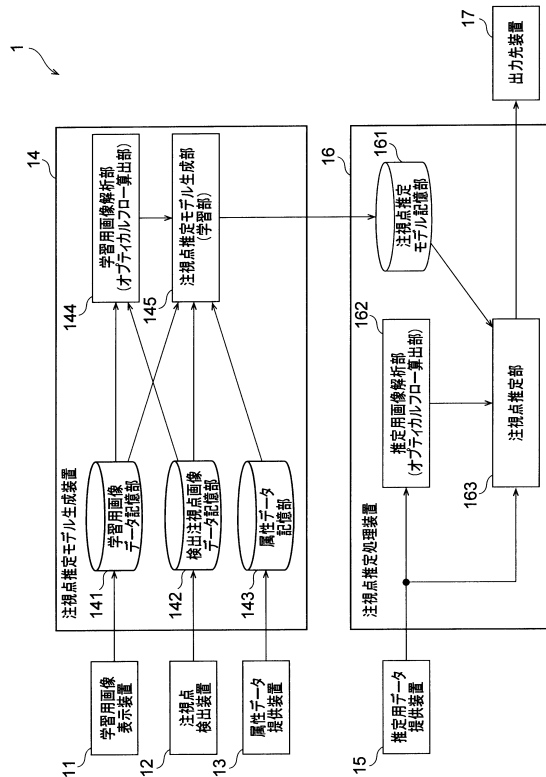
1：注視点推定処理システム、11：学習用画像表示装置、12：注視点検出装置、13：属性データ提供装置、14：注視点推定モデル生成装置、141：学習用画像データ記憶部、142：検出注視点画像データ記憶部、143：属性データ記憶部、144：学習

50

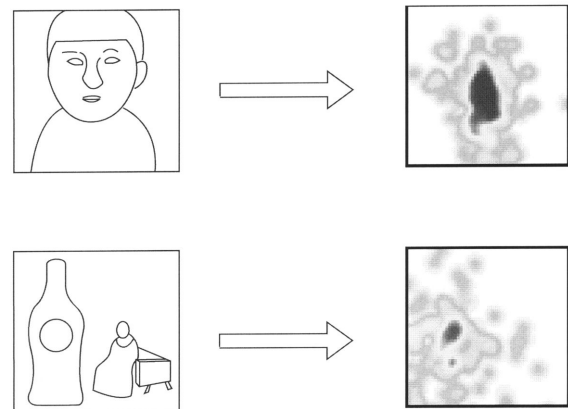
用画像解析部（オプティカルフロー算出部）、145：注視点推定モデル生成部（学習部）、15：推定用データ提供装置、16：注視点推定処理装置、161：注視点推定モデル記憶部、162：推定用画像解析部（オプティカルフロー算出部）、163：注視点推定部、164：画像加工部、165：画像評価部、166：注視物導出部、167：返答生成部、17：出力先装置、18：音データ提供装置、2：コンピュータ装置、21：プロセッサ、22：主記憶装置、23：補助記憶装置、24：ネットワークインタフェース、25：デバイスインタフェース、26：バス、3：通信ネットワーク、4（4A、4B）：外部装置

【図面】

【図1】



【図2】



10

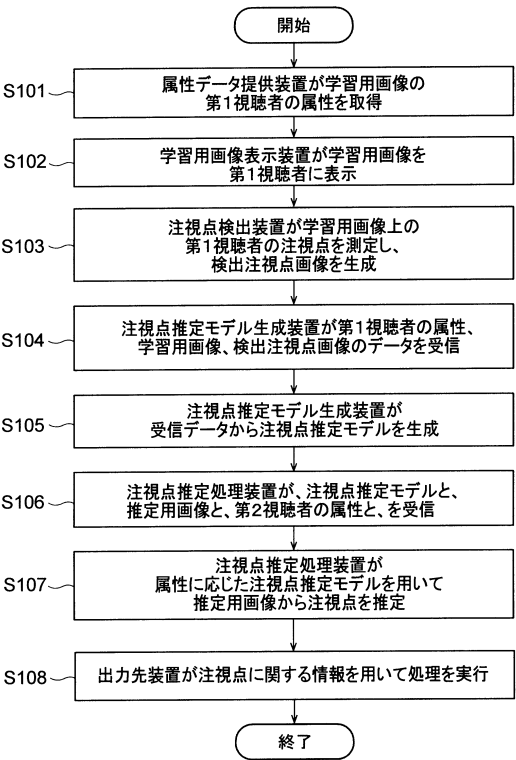
20

30

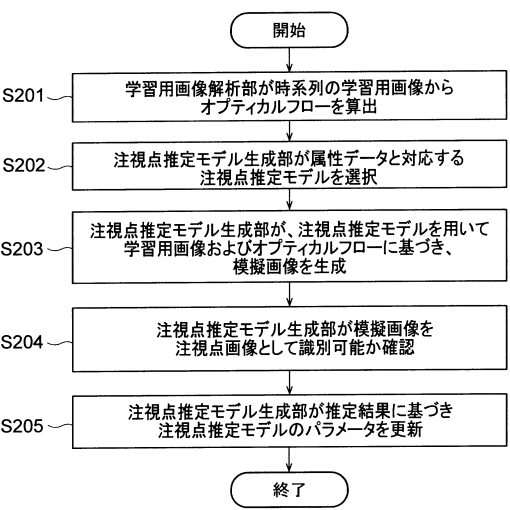
40

50

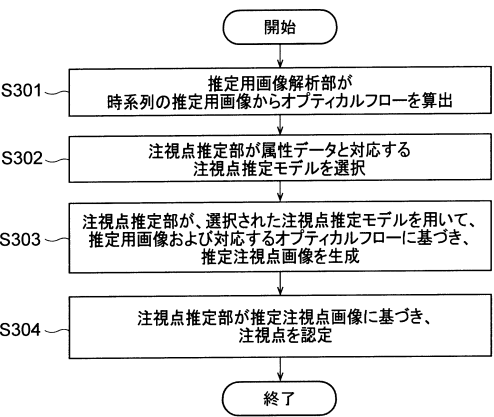
【図 3】



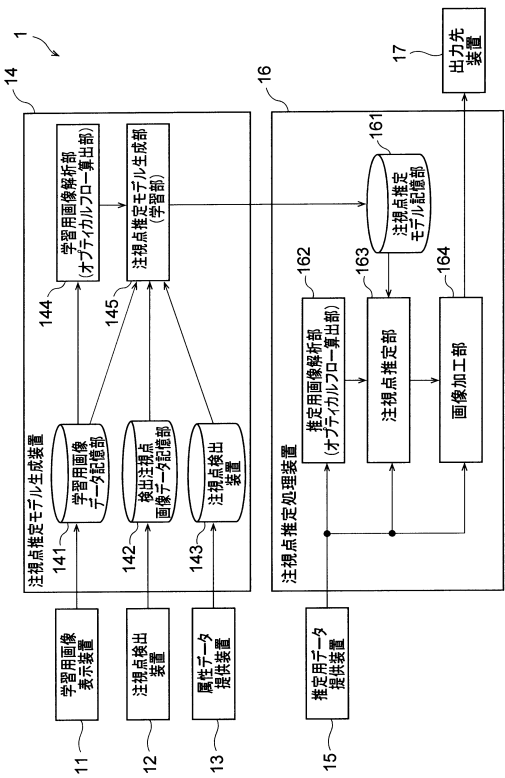
【図 4】



【図 5】



【図 6】



10

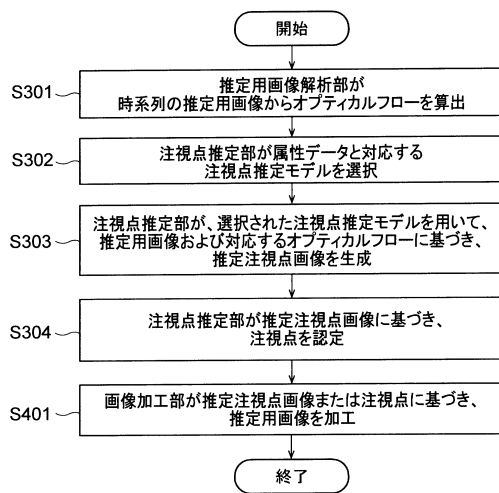
20

30

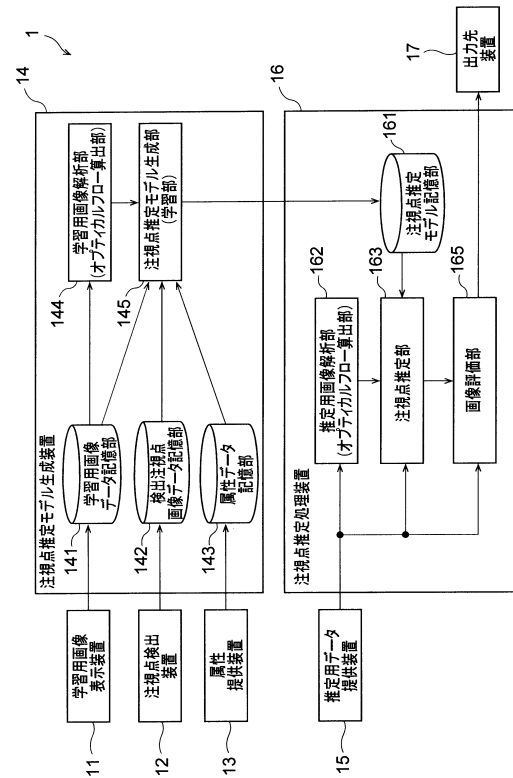
40

50

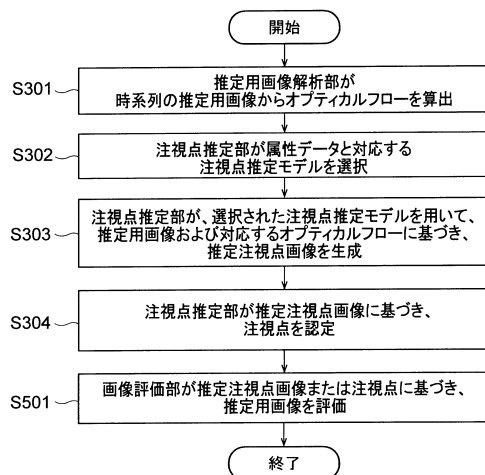
【図 7】



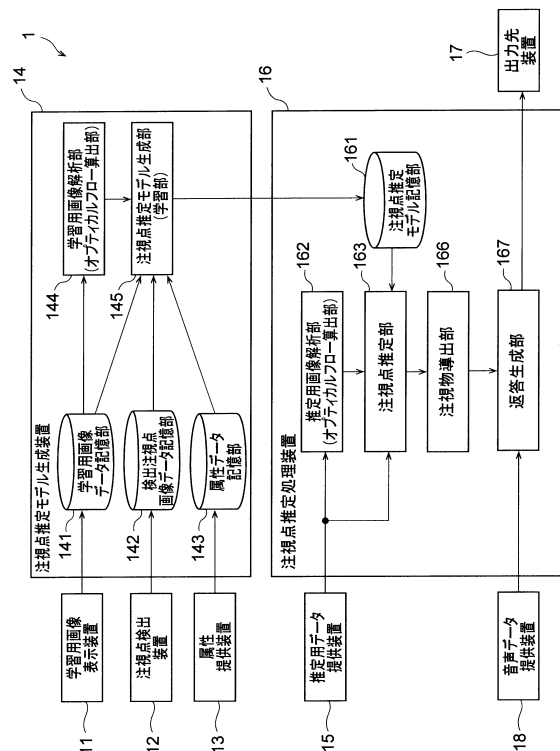
【図 8】



【図 9】



【図 10】



10

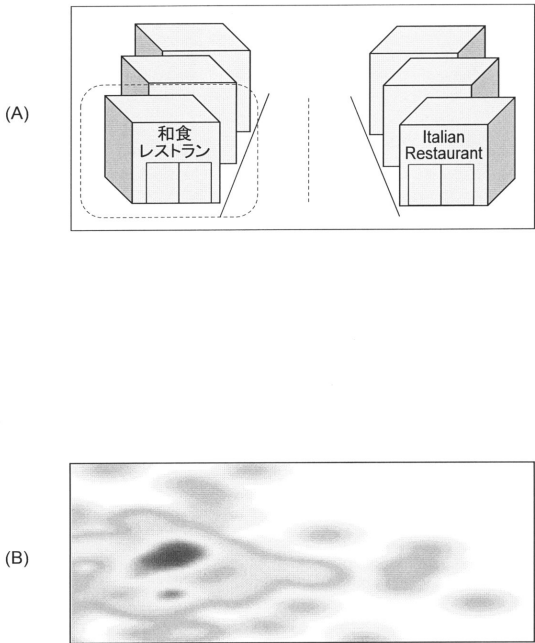
20

30

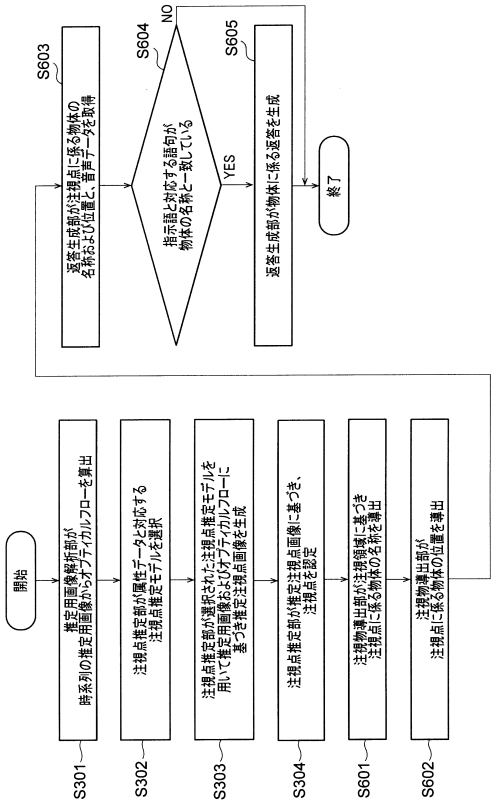
40

50

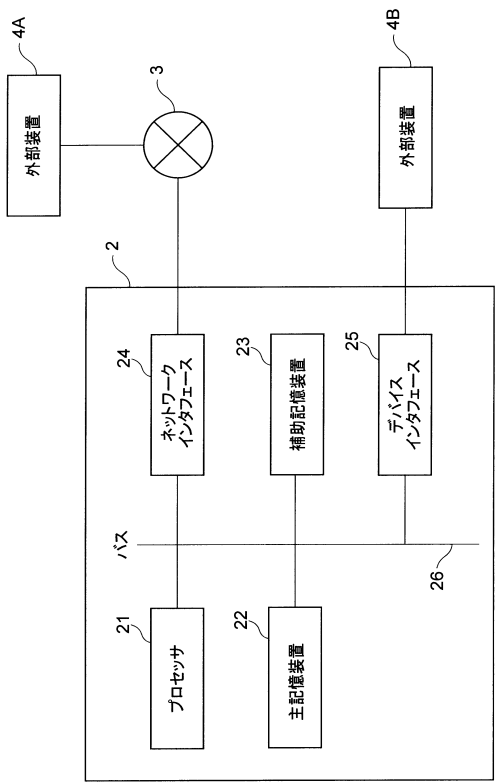
【図 1 1】



【図 1 2】



【図 1 3】



フロントページの続き

(51)国際特許分類

F I
G 0 9 G 5/00 5 5 0 H

(74)代理人 100206243

弁理士 片桐 貴士

(72)発明者 福 田 昌 昭

東京都千代田区大手町1丁目6番1号 大手町ビル2階 株式会社Preferred Networks内

審査官 新井 則和

(56)参考文献

特開2018-022360(JP, A)

国際公開第2017/155663(WO, A1)

米国特許第09025880(US, B2)

TILKE Judd et al., Learning to predict where humans look, 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009年09月29日, pp. 2106-2113, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5459462>

渡辺 隆, シグモイド型パルスジェネレータを取り入れたPulse-Coupled Neural Networkによる注視点探索法, 計測自動制御学会論文集 第38巻 第8号, 日本, 計測自動制御学会, 2002年08月31日, pp. 726-732

(58)調査した分野 (Int.Cl., DB名)

G 0 6 T 7 / 0 0 - 7 / 9 0

G 0 9 G 5 / 0 0

G 0 9 G 5 / 3 6