(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)公開番号
特開2022-19339
(P2022-19339A)
(43)公開日　令和4年1月27日(2022.1.27)

(51)国際特許分類		FI			テーマコード（参考）
G06T	7/00 (2017.01)	G06T	7/00	350C	5C054
G06T	7/246(2017.01)	G06T	7/246		5L096
G06N	3/08 (2006.01)	G06N	3/08		
H04N	7/18 (2006.01)	H04N	7/18	D	
		H04N	7/18	G	

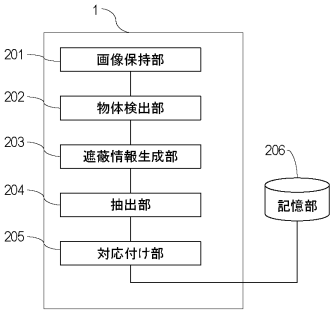
審査請求　未請求　請求項の数　16　ＯＬ　（全21頁）　最終頁に続く

(21)出願番号　特願2020-123119(P2020-123119)
(22)出願日　令和2年7月17日(2020.7.17)

(71)出願人　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100126240
弁理士　阿部 琢磨
(74)代理人　100124442
弁理士　黒岩 創吾
(72)発明者　舘 俊太
東京都大田区下丸子３丁目３０番２号
キヤノン株式会社内
(72)発明者　小川 修平
東京都大田区下丸子３丁目３０番２号
キヤノン株式会社内
(72)発明者　御手洗 裕輔
東京都大田区下丸子３丁目３０番２号
最終頁に続く

(54)【発明の名称】　情報処理装置、情報処理方法及びプログラム

(57)【要約】
【課題】　外見的特徴や姿勢が類似した物体が近接する場合においても安定して追尾を継続できる。
【解決手段】　上記課題を解決する本発明にかかる情報処理装置は、画像から少なくとも１つ以上の物体を検出する情報処理装置であって、遮蔽する物体と遮蔽された物体との遮蔽関係を示す画像特徴を学習した学習済みモデルに基づいて、前記画像から検出された各物体について、前記画像から検出された他の物体との遮蔽関係を推定する推定手段と、前記推定手段によって推定された遮蔽関係に基づいて、前記画像から検出された各物体について、前記画像と異なる時刻に撮像された画像において検出された物体との対応関係を特定する特定手段と、を有する。
【選択図】　図３



【特許請求の範囲】
【請求項１】
画像から少なくとも１つ以上の物体を検出する情報処理装置であって、
遮蔽する物体と遮蔽された物体との遮蔽関係を示す画像特徴を学習した学習済みモデルに基づいて、前記画像から検出された各物体について、前記画像から検出された他の物体との前記遮蔽関係を示す遮蔽情報を推定する推定手段と、
少なくとも前記遮蔽情報に基づいて、前記画像から検出された各物体について、前記画像と異なる時刻に撮像された画像において検出された物体との対応関係を特定する特定手段と、を有することを特徴とする情報処理装置。
【請求項２】
前記推定手段は、入力画像における物体の遮蔽された領域を推定する前記学習済みモデルに基づいて、前記画像から検出された各物体について、前記物体の遮蔽されている部分領域を示す前記遮蔽情報を推定することを特徴とする請求項１に記載の情報処理装置。
【請求項３】
前記特定手段は、前記物体の画像特徴と、前記推定手段によって推定された前記遮蔽情報と、に基づいて、前記画像と異なる時刻に撮像された画像において検出された物体との対応関係を特定することを特徴とする請求項１または２に記載の情報処理装置
【請求項４】
前記推定手段は、前記画像から検出された各物体について、他の物体によって遮蔽されていることを示す尤度である前記遮蔽情報を推定し、
前記尤度を該物体の画像特徴に対応付けて保持する保持手段を更に有することを特徴とする請求項１乃至３のいずれか１項に記載の情報処理装置。
【請求項５】
前記遮蔽情報は、前記画像の各領域について、遮蔽されている前記物体の領域にはより大きい前記尤度を示し、それ以外の領域にはより小さい前記尤度を示す情報であることを特徴とする請求項４に記載の情報処理装置。
【請求項６】
前記特定手段は、前記画像から検出された各物体について、前記画像における位置と、前記画像から検出された画像特徴と、前記画像における前記遮蔽関係と、に基づいて、前記画像と異なる時刻に撮像された画像から検出された物体との対応関係を特定することを特徴とする請求項１乃至５のいずれか１項に記載の情報処理装置。
【請求項７】
前記画像から前記物体毎の領域を取得する取得手段をさらに有し、
前記推定手段は、前記取得手段によって取得された前記物体毎の領域について、遮蔽されている物体の有無を示す前記遮蔽情報を推定することを特徴とする請求項１乃至６のいずれか１項に記載の情報処理装置。
【請求項８】
前記特定手段によって特定された、前記画像から検出された各物体と、前記画像と異なる時刻に撮像された画像において検出された物体と、の対応付けに基づいて、遮蔽されている物体があるか否かを判定する判定手段と、
前記遮蔽されている物体を記憶する記憶手段と、をさらに有することを特徴とする請求項１乃至７のいずれか１項に記載の情報処理装置。
【請求項９】
前記判定手段は、前記画像より前に撮像された画像において検出された物体のうち、前記画像から検出された各物体と対応しない物体を第１の物体として判定し、
前記記憶手段は、前記画像が撮像された時点において、前記第１の物体が遮蔽されたことを記憶することを特徴とする請求項８に記載の情報処理装置。
【請求項１０】
前記判定手段は、前記画像から検出された各物体のうち、前記画像より前に撮像された画像において検出された物体と対応しない物体を第２の物体として判定し、

前記記憶手段は、前記画像から検出された前記第２の物体について、前記記憶手段によって前記画像より前に撮像された画像において遮蔽されていると判定された前記第１の物体との類似度が所定の閾値より大きい場合に、前記画像が撮像された時点において前記第１の物体は遮蔽されていないことを記憶することを特徴とする請求項９に記載の情報処理装置。

【請求項１１】

第１の画像において２つの物体を検出し、前記第１の画像の後で撮像された第２の画像において１つの物体を検出した場合、

前記推定手段は、前記第２の画像において検出された物体について他の物体を遮蔽していることを示す前記遮蔽情報を推定し、

前記特定手段は、前記推定手段によって推定された前記遮蔽情報に基づいて、前記第１の画像において検出された物体のうち他の物体を遮蔽している物体について、前記第２の画像において検出された物体と同一の物体であることを特定することを特徴とする請求項１乃至１０のいずれか１項に記載の情報処理装置。

【請求項１２】

前記第２の画像より後に撮像された第３の画像から２つの物体を検出した場合、

前記推定手段は、前記第３の画像から検出された２つの物体のうち、前記第２の画像から検出された物体の画像特徴と対応付けられた物体について、他の物体を遮蔽していることを示す前記遮蔽情報を推定し、

前記特定手段は、前記第１の画像から検出された物体のうち他の物体によって遮蔽された物体について、前記第３の画像から検出された物体のうち前記第２の画像から検出された物体の画像特徴と対応付けられた物体とは異なる物体を、前記第２の画像において遮蔽された物体と同一の物体であることを特定することを特徴とする請求項１１に記載の情報処理装置。

【請求項１３】

前記学習済みモデルは、ニューラルネットワークであることを特徴とする請求項１乃至１２のいずれか１項に記載の情報処理装置。

【請求項１４】

前記学習済みモデルは、より短い時間間隔で撮像された複数の画像に基づいて、前記物体の遮蔽関係を示す画像特徴を学習させたモデルであることを特徴とする請求項１乃至１３のいずれか１項に記載の情報処理装置。

【請求項１５】

コンピュータを、請求項１乃至１４のいずれか１項に記載の情報処理装置が有する各手段として機能させるためのプログラム。

【請求項１６】

画像から少なくとも１つ以上の物体を検出する情報処理方法であって、

遮蔽する物体と遮蔽された物体との遮蔽関係を示す画像特徴を学習した学習済みモデルに基づいて、前記画像から検出された各物体について、前記画像から検出された他の物体との遮蔽関係を示す遮蔽情報を推定する推定工程と、

少なくとも前記遮蔽情報に基づいて、前記画像から検出された各物体について、前記画像と異なる時刻に撮像された画像において検出された物体との対応関係を特定する特定工程と、を有することを特徴とする情報処理方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、被写体を追尾する技術に関する。

【背景技術】

【０００２】

画像内の特定の被写体を追尾するための技術としては、輝度や色情報を利用するものやテンプレートマッチングなどが存在する。近年、Ｄｅｅｐ　Ｎｅｕｒａｌ　Ｎｅｔｗｏｒｋ

（以下ＤＮＮと省略）を利用した技術が、高精度な追尾技術として注目を集めている。
【０００３】
非特許文献１は、画像内の特定の被写体を追尾するための方法の１つである。追尾対象が映った画像と、探索範囲となる画像を、重みが同一のＣｏｎｖｏｌｕｔｉｏｎａｌ Ｎｅｕｒａｌ Ｎｅｔｗｏｒｋ（以下ＣＮＮと省略）にそれぞれ入力する。ＣＮＮから得られたそれぞれの特徴量同士の相互相関を算出することによって、探索範囲の画像中で追尾対象が存在する位置を特定するものである。このような追尾手法は追尾対象の位置を正確に同定できる一方、追尾対象に類似した物体が画面の上で重なるような場合に、誤った対象を追尾する失敗が発生し易い。
【０００４】
これを回避するために特許文献１の手法に代表されるように、検出物体の領域の色特徴や奥行き情報からヒストグラムを作成し、その変化等を調べて物体が遮蔽されているか否かを判定する手法がある。
【先行技術文献】
【特許文献】
【０００５】
【特許文献１】米国特許出願第１０１８５８７７（Ｂ２）号広報
【非特許文献】
【０００６】
【非特許文献１】Ｂｅｒｔｉｎｅｔｔｏ ｅｔ ａｌ．，Ｆｕｌｌｙ－Ｃｏｎｖｏｌｕｔｉｏｎａｌ Ｓｉａｍｅｓｅ Ｎｅｔｗｏｒｋｓ ｆｏｒ Ｏｂｊｅｃｔ Ｔｒａｃｋｉｎｇ，ａｒＸｉｖ ２０１６
【発明の概要】
【発明が解決しようとする課題】
【０００７】
しかしながら、特許文献１に示される方法では、同じような姿勢の物体や外見的な特徴の類似した物体が画面上で重なると、色やテクスチャといった特徴量のヒストグラムに差異が出にくいため判定できないという課題がある。例えば、スポーツの集団競技等においては、狭い範囲に存在する複数の人物の服装や姿勢が同一になることも多く、異なる人物を同じ人物と見なして追尾する失敗が起こりうる。本発明は、このような課題に鑑みなされたものであり、外見的特徴や姿勢が類似した物体が近接する場合においても安定して追尾を継続することを目的とする。
【課題を解決するための手段】
【０００８】
上記課題を解決する本発明にかかる情報処理装置は、画像から少なくとも１つ以上の物体を検出する情報処理装置であって、遮蔽する物体と遮蔽された物体との遮蔽関係を示す画像特徴を学習した学習済みモデルに基づいて、前記画像から検出された各物体について、前記画像から検出された他の物体との遮蔽関係を推定する推定手段と、前記推定手段によって推定された遮蔽関係に基づいて、前記画像から検出された各物体について、前記画像と異なる時刻に撮像された画像において検出された物体との対応関係を特定する特定手段と、を有する。
【発明の効果】
【０００９】
本発明によれば、外見的特徴や姿勢が類似した物体が近接する場合においても安定して追尾を継続できる。
【図面の簡単な説明】
【００１０】
【図１】物体検出の一例を説明する模式図
【図２】情報処理装置のハードウェア構成例を示す図
【図３】情報処理装置の機能構成例を示すブロック図

【図４】情報処理装置が実行する処理手順を示すフローチャート
【図５】情報処理装置の処理の結果例を示す図
【図６】情報処理装置の処理の結果例を示す図
【図７】情報処理装置が実行する処理手順を示すフローチャート
【図８】遮蔽に関する情報の派生の例
【図９】情報処理装置の機能構成例を示すブロック図
【図１０】情報処理装置の処理の結果例を示す図
【図１１】情報処理装置が実行する処理手順を示すフローチャート
【図１２】情報処理装置の学習処理の例を示す図
【図１３】情報処理装置の処理の結果例を示す図
【図１４】情報処理装置の処理の結果例を示す図
【図１５】情報処理装置の処理の結果例を示す図
【図１６】情報処理装置の学習処理の例を示す図
【発明を実施するための形態】
【００１１】
＜実施形態１＞
実施形態に係る情報処理装置を、図面を参照しながら説明する。なお、図面間で符号の同じものは同じ動作をするとして重ねての説明を省く。また、この実施の形態に掲載されている構成要素はあくまで例示であり、この発明の範囲をそれらのみに限定する趣旨のものではない。
【００１２】
本実施形態では、動画もしくは連続撮影した静止画フレームから人物を検出し、追尾する機能について説明する。適用範囲は検出・追尾対象の物体のカテゴリを限定しないが、本実施形態１は対象を人物に限定する。本実施形態では、時間的に連続する画像毎に人物を検出し、連続する画像間でそれぞれどの人物がどの人物と同一人物であるかを対応付けることで、人物の追尾を実現する。本実施形態では特に、スポーツイベントなどの撮影を想定し、人物の服装や移動方向等が類似しており、高頻度で近接・交差するとする。このような場合、各画像における人物の位置または服装の色といった外見的な特徴が近い人物同士を対応付けるだけでは、誤った対応付けが発生しやすい。このような失敗をここでは誤マッチングと呼ぶ。
【００１３】
本実施形態では撮影者から見て物体が重なっている時の、遮蔽関係のパターンを学習した学習済みモデルが出力する遮蔽に関する情報に着目する。学習済みモデルによって出力された遮蔽関係を、物体の対応関係の特定に併せて用いることで、手前にいる人物と奥にいる人物同士を対応付ける失敗を抑制し、追尾の精度を向上する。
【００１４】
これを模式的に示した図が図１である。図１（Ａ）の画像２１００，２２１２０，２１４０は同一の絵柄の２枚のトランプカードがテーブル上で交差していく様子を上から写した動画の３フレーム分の静止画を示している（時系列順に左から右に並んでいる）。画像２１００，２１２０，２１４０を観察しだけでは各画像におけるカードがそれぞれどのように移動したかを確定することができない。一方で図１（Ｂ）は（Ａ）よりも高フレームレートで同じ様子を撮影した例である。つまり、より短い時間間隔で撮像された画像群である。図１（Ｂ）の画像を時系列順に観察していけば、どちらのカードが次の画像でどこに移動したかを対応付けることができる。全体としては左側のカード２２０１が右側のカード２２０２の上を通過し、右側に移動したということを比較的容易に推定することができる。画像２２１０や画像２２３０に示すように、物体同士の交差の瞬間に過渡的に生じる見えを観察することで、画像２２２０においてどちらのカードが手前側を通過し、どちらが奥側にあるのかが判定可能となる。この判定に際しては、２．５次元の奥行画像といった特別なセンサーやオプティカルフロー等の生成のコストの高い情報は必ずしも必要でない。物体同士が手前と奥で重なったときに、どのような見えが生じ易いかという、遮蔽関

係と見えの特徴（ａｐｐｅａｒａｎｃｅ　ｆｅａｔｕｒｅ）とのパターン認識の問題として解くことができる。これは図１（Ｃ）および図１（Ｄ）に示す人物の交差のようなシーンでも同様である。本図１（Ｃ）（Ｄ）では人物の服装や姿勢等の見え、移動方向は同一であるとする。このような場合も、物体が交差する前後の見えの状態に着目して観察すれば、図１（Ｄ）の画像２４２０では人物２４０１が手前側に、人物２４０２が奥側にいると判定する。以降の画像において、この遮蔽関係を維持したままであれば、人物２４０１が人物２４０２を一度遮蔽した場合に、手前側の人物２４０１を追尾し、奥側の人物２４０１の遮蔽関係と画像特徴を保持する。そして、遮蔽が解消したときには、人物２４０１の追尾を継続しつつ、奥側にいた人物２４０２を再び検出することが可能である。以上が本実施形態の原理の概要を示す説明である。詳細な処理については後述する。

【００１５】

図２は、本実施形態における、画像認識によって追尾対象を追尾する情報処理装置１のハードウェア構成図である。ＣＰＵ　Ｈ１０１は、ＲＯＭ　Ｈ１０２に格納されている制御プログラムを実行することにより、本装置全体の制御を行う。ＲＡＭ　Ｈ１０３は、各構成要素からの各種データを一時記憶する。また、プログラムを展開し、ＣＰＵ　Ｈ１０１が実行可能な状態にする。記憶部Ｈ１０４は、本実施形態の処理対象となるデータを格納するものであり、追尾対象となるデータを記憶する。記憶部Ｈ１０４の媒体としては、ＨＤＤ，フラッシュメモリ、各種光学メディアなどを用いることができる。入力部Ｈ１０５は、キーボード・タッチパネル、ダイヤル等で構成され、ユーザからの入力を受け付けるものであり、追尾対象を設定する際になどに用いられる。表示部Ｈ１０６は、液晶ディスプレイ等で構成され、被写体や追尾結果をユーザに対して表示する。また、本装置は通信部Ｈ１０７を介して、撮影装置等の他の装置と通信することができる。

【００１６】

図３は、情報処理装置の機能構成例を示すブロック図である。図３ではＣＰＵ　Ｈ１０１において実行される処理を、それぞれ機能ブロックとして示している。情報処理装置１は、画像取得部２０１、物体検出部２０２、遮蔽情報生成部２０３、抽出部２０４、対応付け部２０５を有し、外部の記憶部２０６に接続されている。記憶部２０６は情報処理装置１の内部にあってもよい。それぞれの機能を簡単に説明する。画像取得部２０１は、撮像装置によって特定の物体（本実施形態では人物）を撮像した動画や連続静止画の画像を取得する。物体検出部２０２は、画像取得部２０１によって取得された画像から予め設定された所定の物体を示す画像特徴を検出する。例えば、さまざまな姿勢の人物の画像を用いて人体（頭や動体）を示す画像特徴を予め学習した学習済みモデルに基づいて、画像における人物の領域を検出する。遮蔽情報生成部２０３は、遮蔽する物体と遮蔽された物体との遮蔽関係を示す画像特徴を学習した学習済みモデルに基づいて、画像から検出された各物体について、画像から検出された他の物体との遮蔽関係を示す遮蔽情報を推定する。遮蔽情報とは、注目物体が他の物体によって遮蔽されている可能性を表す尤度（被遮蔽／遮蔽スコア）である。例えば、ある物体について、他の物体によって遮蔽されている可能性が高ければ、被遮蔽／遮蔽スコアを１に近づける。ある物体について、他の物体を遮蔽している可能性が高ければ、被遮蔽／遮蔽スコアを０に近づける。このような遮蔽関係を示す被遮蔽／遮蔽スコアを、学習済みモデルを用いて推定する。抽出部２０４は、ある画像について検出された物体ごとに遮蔽情報を記憶部２０６に記憶する。対応付け部２０５は、複数の画像間で検出された物体の対応付けを行う。すなわち、遮蔽情報に基づいて、ある画像から検出された各物体について、ある画像と異なる時刻に撮像された画像において検出された物体との対応関係を特定する。異なる時間で撮像された画像のそれぞれから検出された物体同士を正しく対応付けることによって物体を追尾できる。また、物体同士の遮蔽関係はある一定の期間において維持されることを仮定することによって、遮蔽関係を使って物体同士を対応付けることができる。記憶部２０６は、各検出物体の被遮蔽スコアを記憶する。各機能部の処理の詳細は図４のフローチャートを用いて説明する。

【００１７】

図４は本実施形態の処理の流れを示したフローチャートである。以下の説明では、各工程

（ステップ）について先頭にＳを付けて表記することで、工程（ステップ）の表記を省略する。ただし、情報処理装置はこのフローチャートで説明するすべての工程を必ずしも行わなくても良い。図４のフローチャートに示した処理は、コンピュータである図２のＣＰＵ　Ｈ１０１により記憶部Ｈ１０４に格納されているコンピュータプログラムに従って実行される。

【００１８】

Ｓ３０１では、情報処理装置１が、各動画フレームについて繰り返すループ処理を開始する。Ｓ３０２では、画像取得部２０１が人物を撮像した動画や連続静止画の画像フレームを順次取得する。以降の処理はＳ３０１～Ｓ３１１まで各画像について順次処理がなされる。なお、画像取得部２０１は、情報処理装置に接続された撮像装置によって撮像された画像を取得してもよいし、記憶部Ｈ１０４に記憶された画像を取得してもよい。図５（Ａ）中の動画フレーム３１００，３１１０，３１２０，３１３０，３１４０が取得した画像フレームの例である。

【００１９】

次にＳ３０３では、物体検出部２０が、所定の物体（ここでは人物）の画像特徴に基づいて、前記取得された画像から少なくとも１つ以上の所定の物体を検出する。画像内から物体を検出する公知技術としては、Ｌｉｕによる手法等が挙げられる（Ｌｉｕ，ＳＳＤ：Ｓｉｎｇｌｅ　Ｓｈｏｔ　Ｍｕｌｔｉｂｏｘ　Ｄｅｔｅｃｔｏｒ．Ｉｎ：ＥＣＣＶ２０１６）。画像内から候補物体を検出した結果を図５（Ａ）に示す。図５（Ａ）中の矩形枠３１０１，３１０２，３１０３，３１１１，３１１２，３１１３，３１２１，３１２２，３１３１，３１３２，３１４１，３１４２が検出された物体領域を示すＢｏｕｎｄｉｎｇ　Ｂｏｘ（以下ＢＢ）である。

【００２０】

Ｓ３０４では、遮蔽マップ生成部２０３が、各画像について、領域毎に遮蔽されているか否か（遮蔽関係）についての遮蔽情報を示した遮蔽マップを生成する。遮蔽マップ生成部２０３が、各画像について、遮蔽されている物体のうちの見えている領域（被遮蔽物体領域）を推定する。ここでは各人物が他の人物と重なっているか、重なっている場合に奥側にいるか、手前側にいるかを判定し、その結果を遮蔽状態のスコア（尤度）として領域ごとに出力する。これは意味的領域分割の認識タスクの一種であり、Ｃｈｅｎらの手法等の公知の手法を使って実現することができる。（Ｃｈｅｎ，ＤｅｅｐＬａｂ：　Ｓｅｍａｎｔｉｃ　Ｉｍａｇｅ　Ｓｅｇｍｅｎｔａｔｉｏｎ　ｗｉｔｈ　Ｄｅｅｐ　Ｃｏｎｖｏｌｕｔｉｏｎａｌ　Ｎｅｔｓ，　Ａｔｒｏｕｓ　Ｃｏｎｖｏｌｕｔｉｏｎ，　ａｎｄ　Ｆｕｌｌｙ　Ｃｏｎｎｅｃｔｅｄ　ＣＲＦｓ，２０１６）。

【００２１】

図６（Ａ）に遮蔽マップの生成処理を説明する模式図と結果の一例を示す。ニューラルネットワーク４０２は入力された画像から、入力画像の各画素について遮蔽状態を判定するニューラルネットワークである。ＲＧＢ画像４０１が入力されると、ニューラルネットワーク４０２は画像中に人物がいるか否か、さらにその人物が遮蔽されているか否かを推定した結果を遮蔽マップ４０４として出力する。同マップは遮蔽されていない人物および人物以外の領域と推定された場合は０、遮蔽されている人物の領域には１、の被遮蔽スコアが出力される。遮蔽マップ４０４中の黒い領域ほど高い被遮蔽スコアであることを示す。すなわち、黒い領域は遮蔽された人物の領域であると推定されたことを示している。ニューラルネットワーク４０２は入力画像に対してこのような出力ができるように事前に学習を行っている（学習については後述する）。なお、図に示した遮蔽マップ４０４は推定結果として理想的な出力状態の一例を示したものである。

【００２２】

なお、ＲＧＢ画像４０１のほかに、専用センサー等を使って２．５次元奥行画像４０５を別途取得するような派生的な形態も考えられる。前記奥行画像４０５を３チャンネルのＲＧＢ画像４０１と連結した４チャンネルの情報をＲＧＢ画像の代わりに画像入力として学習・認識する。これにより遮蔽領域の情報をより高精度にすることも可能である。

【００２３】
次に、Ｓ３０５では、情報処理装置１が、Ｓ３０３で検出された各物体について、Ｓ３０６からＳ３０７のループ処理を実行する。Ｓ３０５～Ｓ３０８では、抽出部２０４が、生成された遮蔽マップから検出物体ごとに遮蔽関係を示す情報を抽出し、記憶部２０６に記憶する。Ｓ３０６では、抽出部２０４が、遮蔽マップから検出物体毎に遮蔽関係を示す情報を抽出する。具体的には、図６（Ａ）の人物検出枠４０７の中の遮蔽マップ４０４の被遮蔽スコアを平均する。この被遮蔽スコアが１に近いほどその物体は遮蔽されている可能性が高く、被遮蔽スコアが０に近いほどその物体は遮蔽されていない可能性が高いことを示す。なお、検出枠の位置のずれや、遮蔽マップ４０４にノイズが含まれることを想定して、図６（Ｂ）に示すように枠の中央付近を重視した重み付き平均で取得する。図中の演算４１２は各画像の部分領域毎ごとの要素積（アダマール積）を意味する。マップ４１３は中央にピークがあり、画像ブロックの総和が１となる２次元ガウス関数である（縦横サイズを人物検出枠に合わせて変形してある）。取得結果の例を図６（Ａ）に記号ｏｃｃを付して被遮蔽スコア値４０９と４１０として示す。左側の検出枠は奥側にいる人物のため被遮蔽スコアが高く、右側の検出枠は手前側のため被遮蔽スコアが低いと判定されている。以上のような処理を先ほどの図５の入力画像に対して処理した結果例を図５（Ｂ）（遮蔽マップ）および図５（Ｃ）（各枠の被遮蔽スコア推定結果）として示す。交差開始～終了の間、奥側に位置する人物３１０１に対応する被遮蔽スコアは人物３１０２のそれよりも相対的に高いことを示している。
【００２４】
Ｓ３０７では、記憶部２０６が、抽出部２０４によって取得された各検出物体の被遮蔽スコアを記憶する。同時に、各検出物体の位置・サイズの情報、および色やテクスチャのヒストグラムといった物体の見えに関する特徴量、も記憶する。ここではこれら複数種類の数量を一括して検出物体の特徴量と呼ぶ。なお、見えに関する特徴量としてはこの他にニューラルネットワークの中間層情報等を利用してもよい。（例えば”Ｈａｒｉｈａｒａｎ，ｅｔ．ａｌ，Ｈｙｐｅｒｃｏｌｕｍｎｓ ｆｏｒ Ｏｂｊｅｃｔ Ｓｅｇｍｅｎｔａｔｉｏｎ ａｎｄ Ｆｉｎｅ－ｇｒａｉｎｅｄ Ｌｏｃａｌｉｚａｔｉｏｎ，ｉｎ ＣＶＰＲ２０１５”）。
【００２５】
Ｓ３０８では、情報処理装置１が、各画像について繰り返すループと、各画像において検出された人物について繰り返すループを終了する。このループは画像毎に、その画像から検出された人物すべてについて遮蔽情報を取得したときに終了する。次に、Ｓ３０９～Ｓ３１０では、対応付け部２０５が、前後の画像間の物体の対応付けを行う（ただし一つ目の動画フレームの場合は過去のフレームがないためこれを行わない）。まず、Ｓ３０９で、対応付け部２０５が記憶部２０６に記憶された過去の物体の特徴量である、被遮蔽スコア、位置サイズおよび見えの特徴量を取得する。次にＳ３１０で、対応付け部２０５が過去の動画フレーム中に検出された物体と、現在処理しているフレーム中に検出された物体の対応付けを行う。
【００２６】
Ｓ３１０における対応付け部２０５の詳細な処理フローを図７に示す。Ｓ５０１で、対応付け部２０５は、まず現フレームの検出物体と一つ前のフレームで検出された物体の間で全組み合わせのペアを作る。前後のフレームでそれぞれｎ人とｍ人の人物が検出されていれば、全部でｎ×ｍ個のペアが生成される。次に、Ｓ５０２で、対応付け部２０５は全ての物体ペアについて類似度を算出する。類似度としては検出物体同士の特徴量の差分に基づいた指標を用いることができる。一例として過去の検出物体ｃ１と現在の検出物体ｃ２の類似度を下式のように算出する。
（数式１）
$$L(c_1, c_2) = -W_1 \| BB_1 - BB_2 \| - W_2 \| f_1 - f_2 \| - W_3 \| occ_1 - occ_2 \|$$
ここで、ＢＢとは各物体の（中心座標値ｘ、中心座標値ｙ、幅、高さ）の４変数をまとめ

たベクトルであり、ｆは各物体の特徴を示したものである。｜｜ｘ｜｜はｘのＬＰノルムである。ｏｃｃは各物体の被遮蔽スコアである。Ｗ１，Ｗ２，Ｗ３はそれぞれ経験的あるいは機械学習的に調整して設定される０以上のバランス係数である。ここで各特徴量のばらつきを事前に統計的に求めておいて各特徴量を正規化する等してもよい。物同士が交差する場合であっても、他の物体を遮蔽する側の人物を追尾することによって、被遮蔽側の人物が再び画像で確認されたときに、直前で他の物体を遮蔽する側の人物と対応付けようとすると数式１の３つめの項の値が小さくなり、類似度が低く算出される。つまり、この処理によって、遮蔽関係が異なる人物同士はマッチングされる可能性が低くなり、追尾の誤マッチングが抑制できる。

【００２７】

次に、Ｓ５０３において、対応付け部２０５が、過去の物体と現在の物体との類似度に基づいて物体間の対応関係を特定するための対応付け（マッチング）を行う。マッチングの方法にはいくつか存在する。例えば、類似度が高い候補同士から優先的にマッチングする方法や、ハンガリアンアルゴリズムを用いる方法等がある。ここでは前者を用いる。

【００２８】

Ｓ５０３では、対応付け部２０５が、現フレームの全物体について対応付けが終了していなければＳ５０６で類似度最大のペアから同一人物として対応付けていく。対応付けの終わったペアの物体は対応付けの候補から省いていく。上記の処理の際に、その時点で残っているペアの中の最大の類似度の大きさが所定の閾値を下回った場合は、もはや類似した物体ペアが残っていないことを意味する。その場合はそれ以上無理に対応付けることなく（Ｓ５０５）、対応付けを終了する。

【００２９】

以上の処理Ｓ３０１～Ｓ３１１を動画フレームごとに行う。その結果、図５（Ｄ）に結果例を示すように、動画中から人物を検出し、それぞれの物体がどこに移動したかの一連の追尾結果が得られる（フレーム間の同一の人物に記号Ａ，Ｂ，Ｃで付して追尾の結果を示している）。

【００３０】

＜変形例＞

本実施形態では物体ペア同士のマッチングの類似度として差分に基づき、被遮蔽スコアや見えといった各指標の距離を重み付け和した。ここで例えばＫＬダイバージェンスを使うことも考えられる。またメトリック学習を行ってより精度の高い距離指標を求めることも考えられる。また単一の類似度を一度だけ用いるのでなく、まず見えの特徴で類似度を判定し、条件を満たしたものは次に遮蔽状態のスコアの類似度に基づいて判定する、等のルールベースによる方法や段階的な判定方法も考えられる。またさらにニューラルネットやサポートベクトルマシンといった公知の識別器の手法を用い、説明変数を特徴量、目的変数を同一物体か否かの結果、として学習・識別し、この値によってマッチングを判定することも可能である。以上のようにフレーム間の物体間の対応付けは特定の形態に限定されない。

【００３１】

またさらに別の派生形態として、被遮蔽スコアの推定値を安定させるために、下式のように過去のスコアを移動平均した値を用いる工夫も考えられる。

（数式２）

$$occ^{EMA}(t) = (1-\alpha) \times occ^{EMA}(t-1) + \alpha \times occ(t)$$

上式は指数移動平均値と呼ばれる値であり、$occ^{EMA}(t)$は時刻ｔの被遮蔽スコアの指数移動平均値、$occ(t)$は時刻ｔの被遮蔽スコア、αは$0 < \alpha \leqq 1$の係数である。過去の複数フレームで追尾ができている物体については上式で指数移動平均値を算出しておき、類似度を比較する際には元の被遮蔽スコアではなく、指数移動平均被遮蔽スコアを用いる。これにより、交差時に複数のフレームにまたがって徐々に重畳状態が起こるような場合に、複数フレームの被遮蔽スコアの平均値に基づいてマッチングできるので、より物体間の対応付けが安定する。

【００３２】
またさらに別の派生形態として、マッチングの際に前後フレーム間の類似度だけでなく、ｎステップ前の過去の複数のフレームの特徴量・位置を用いてマッチングを行うような形態も考えられる。この方法を用いることで、一度物体が遮蔽されて追尾できないフレームが発生しても、その後のフレームで遮蔽が解消されれば再び追尾が可能になる。この形態では例えば、ｎフレームまでさかのぼって物体の特徴量の平均値を求め、これに基づいて現フレームから検出された物体との類似度の算出を行う。もしくは、過去のｎフレームの物体と現フレームの物体間でそれぞれ類似度を求め、得られたｎ個の類似度の平均値が最も高い物体に対応付ける。また、過去だけでなく、ｎステップの未来のフレームの結果も使って双方向で判定を行うことも考えられる。この形態は未来のフレームを処理するまで結果が判明しないため処理のリアルタイム性には劣るが、過去のみを見る方法よりも高精度である。
【００３３】
またさらに別の派生形態として、検出の失敗に対応するための形態が考えられる。物体検出・追尾においては物体の姿勢が特殊な形状に変化した、等の理由で物体検出の段階で一時的に失敗するようなことも起こり得る。このような未検出が起こると、フレーム間の対応付けの際に、前のフレームに存在した物体が、現フレームでは対応なしと判定される。すると追尾はそこで途切れることになる。このような失敗を防ぐために、以下のような工夫もありえる。すなわち、マッチングで未対応の人物が発生したら、その情報をリストに記憶しておき、次のフレームのマッチングのときに対応付けの候補に加える（一定時間が経過してもまだ未対応であれば物体自体がもう存在しないと判断し、リストから除去する。ここではこれをタイムアウト処理と呼ぶ）。
【００３４】
このように動画フレームをまたがる物体の対応付けについては種々のやり方が考えられ、特定の形態に限定されない。
【００３５】
＜遮蔽情報の形態のバリエーションおよび学習方法＞
本実施形態では、遮蔽マップとして、遮蔽されている物体のうちの見えている領域（被遮蔽物体領域）を推定した。この形態についても様々な派生形態が考えられる。一例を図８に示す。ここでは図８（Ｂ）に示すように、画像８０１のように奥側の物体の見えている領域を推定する以外でもよい。例えば、画像８０２のように奥側の物体の全領域を推定する。また、画像８０３のように、手前側の遮蔽物体の領域を推定する（図の領域４４０のように他物体と重なっていない物体も手前側領域として含めて推定している。ただし別の形態としてこのような単独の物体は手前側の領域に含めないことも考えられる）。また、画像８０１～８０３のように前景領域を推定するのではなく、画像８０４のように物体の中心や重心の位置を推定することも考えられる。画像８０４の場合においては被遮蔽物体の中心付近の領域に大きな正の値を、遮蔽物体の中心付近に小さな負の値を推定するようにする（ここでいう物体の中心領域は図示するようにガウス関数状の領域を推定させるような形態が考えられる）。
【００３６】
ここで遮蔽状態の情報の学習方法について図６（Ｄ）を用いて説明する。前述のＣｈｅｎらの手法等で示されるニューラルネット４０２は、入力画像であるＲＧＢ画像４０１に対して遮蔽物体の被遮蔽スコアマップ４０３を出力する。４０３の結果例を４３０に示す。ＣＨＥＮらの手法等は特定カテゴリ物体の前景領域を推定する手法であるが、ここでは遮蔽情報の教師値４３１を与えて、教師値４３１と同じようなマップが推定によって得られるようニューラルネット４０２の学習を行う。具体的には出力結果のマップ４０３と教師値４３１を比較し、交差エントロピーや二乗誤差などの公知の方法で損失値算出４３２を行う。損失値が漸減するように誤差逆伝搬法等でニューラルネット４０２の重みパラメーターを調整する（この処理についてはＣｈｅｎらの手法と同一のため詳細は略す）。入力画像と教師値は十分な量を与える必要がある。重なった物体の領域の教師値を作成するの

はコストがかかるため、ＣＧを用いることや、物体画像を切り出して重畳する画像合成の方法を用いて学習データを作成するようなことも考えられる。以上が学習方法になる。

【００３７】

またさらに、本実施形態では上記で求めた物体の枠の中で取得して被遮蔽スコアと呼ぶ指標を求めた。遮蔽情報の取得の形態の様々な例を図８（Ｃ）に示す。図８（Ｃ１）は本実施形態の形態である。この他に、（Ｃ２）奥側の被遮蔽スコアと手前側の被遮蔽スコアの差分値を物体枠内で取得する、（Ｃ３）物体の中心のスコアを１点だけ参照する、等様々に考えられる。また、枠内で取得する際に、物体の枠内で取得する際に、他の物体枠と重なっている領域についてはどちらの物体の領域か判然としないために取得から省くような方法も考えられる。

【００３８】

またさらに、上述の＜遮蔽状態の推定＞と＜各物体の被遮蔽スコアの取得＞を同時に行う方法も考えられる。例として、Ｌｉｕの手法等で使われている公知な方法であるアンカーと呼ばれる手法があげられる。この手法では物体の候補枠の集合が求められるので、これを利用して各候補枠が遮蔽物体か被遮蔽物体かの被遮蔽スコアを推定し対応付けることが考えられる（この形態の詳細については実施形態３で示すのでここでは説明を略す）。

【００３９】

またさらに、上で示したような複数の形態の遮蔽情報をそれぞれ取得し、これを遮蔽に関する多次元の特徴として後段の物体の対応付けに用いてもよい。もしくは前記の遮蔽に関する多次元の特徴から機械学習によって物体の遮蔽されている面積の割合を推定して用いてもよい。この場合は前記の遮蔽に関する多次元の特徴を説明変数とし、物体が遮蔽されている面積の割合を目標変数とし、ロジスティック回帰等の公知技術で回帰推定を行う等すればよい。

【００４０】

＜実施形態２＞

本実施形態では実施形態１と同様に人物の検出と追尾を行う。ハードウェア構成は実施形態１の図２と同様である。本実施形態における機能構成例を示すブロック図は図９（Ａ）になる。実施形態１の構成に新たに遮蔽状態判定部３０１が追加されている。実施形態１では追尾中に人物の枠は人物同士の重なりによって、人物の検出ができないことがある。例えば図１０（Ａ）中の動画フレーム４１２０に示すように、人物間で重なった面積が大きいときには、奥側の人物が検出できないことは多い。このような時に遮蔽状態判定部３０１が、人物は存在しているが被遮蔽状態にある、と判定する。

【００４１】

実施形態１で説明したような物体検出部の一時的な検出の失敗による未検出と異なり、人物の集団が同じ方向に同じ速度で移動しているような場合、長時間未検出の状態が続く。さらに被遮蔽状態から脱した画面上の位置が、被遮蔽状態が開始した位置から離れることがある。このため被遮蔽状態であると積極的に判定し、推定した前記状態に応じた処理を行うことで追尾の成功率を高めることが望ましい。

【００４２】

本実施形態も全体の処理フローは実施形態１の図４と同じであるが、Ｓ３１０の処理の詳細が下記のように異なる。ここでは、実施形態１と異なるＳ３１０の処理についてのみ説明する。図１１を用いて遮蔽状態判定部３０１が行うＳ３１０処理の詳細なフローについて説明する。まずこれまでと同じようにＳ６０１で現フレームと前フレームで物体の対応付けを行う。Ｓ６０２で対応付けられなかった前フレームの物体がある場合、被遮蔽状態に入った可能性がある。そこでＳ６０３で当該物体のそれまでの被遮蔽スコアの高さが閾値以上かを調べる。これは動画フレームのフレームレートが十分に高ければ、遮蔽により未検出になる前後で被遮蔽スコアが高くなることが多いためである。さらにＳ６０４で当該物体の周辺領域で現フレームの物体の検出数の数が減っていないかを調べ、上記の二つの結果が真であれば当該物体は被遮蔽状態に入ったと推定し被遮蔽状態のリストに記憶する（Ｓ６０５）。被遮蔽状態のリストに記憶された物体については前回検出されたときの

特徴量と位置も合わせて記憶する。これによって、遮蔽が解消されて再び検出されたときに追尾できる可能性が向上する。

【００４３】

Ｓ６０６～Ｓ６１０は被遮蔽状態の物体が再出現したかどうかを判定する処理である。Ｓ６０３で対応付けられなかった現フレームの物体がある場合、被遮蔽状態を脱して再度検出できるようになった可能性がある。そこでＳ６０７で当該物体の被遮蔽スコアの高さが閾値以上かを調べる。さらにＳ６０８で当該物体の周辺領域で現フレームの物体の検出数の数が増えていいないかを調べる。両方の結果が真で、且つ被遮蔽状態のリストに記憶されている物体のいずれかと当該物体が所定閾値以上に類似度が高い場合（Ｓ６０８）、当該物体は被遮蔽状態から脱して再度出現したと推定する。そのとき、対応付けた物体を被遮蔽状態のリストから除去する（Ｓ６０９）被遮蔽状態のリストから除去された物体については、現在の入力画像から検出された特徴量と位置を取得する。

【００４４】

ここで、対応付けの処理の工夫として、例えば、フレーム間の物体のマッチングの際に、被遮蔽状態にある人物とのマッチングは距離による類似度のペナルティを減ずる。再出現を待つタイムアウトの時間を長く取る。遮蔽状態の物体との対応付けの閾値は、通常の物体間のマッチングよりも閾値を低く設定する、等が考えられる。

【００４５】

またさらに、ここでは二人の人物の重なりを想定して説明を行ったが、３人以上の人物の間で重なりが生じることもある。この場合は、遮蔽状態に入ったと判定されれば被遮蔽状態のリストに加えておき、再出現したら前フレームとの対応付けを行い、被遮蔽状態のリストから都度除去する。これにより３人以上についてもある程度の対応が可能である。

【００４６】

＜実施形態３＞

本実施形態では、ユーザが指定した単一の物体を追尾する形態について説明する。ここでは追尾対象は人体等の特定カテゴリに限らず、ユーザが指定した不特定の物体を追尾する形態を扱う。例えば、犬などの動物や、車などの乗り物であってもよい。

【００４７】

機能ブロックの図は図９（Ｂ）になる。これまでの構成に新たに追尾物体指定部３０２が追加されている。ここで追尾物体指定部３０２と物体検出部２０２の機能は非特許文献１の方法を用いることで容易に実現することができる。追尾物体指定部３０２はユーザが動画フレーム中で追尾対象物体の枠位置を指定する機能部である。これにより追尾すべき物体の特徴が初期化される。物体検出部２０２は各動画中で最も対象物体と一致度の高い画像領域を同定する。同定した結果例を図１２（Ａ）に示す。図１２（Ａ）の動画フレーム５１１０上の枠５１１１がユーザによって指示された追尾物体の枠である。動画フレーム５１２０ではこの物体が画面中で右側に移動しており、物体検出部２０２によって枠５１２１として検出されている。非特許文献１の方法は物体の追尾手法として優れるが、類似物体間で容易に誤スイッチが生じる。そこで本実施形態ではこれまでの実施形態と同様に、追尾物体に対して遮蔽状態に関する情報を推定し、誤スイッチが生じていないかを判定する。

【００４８】

このために遮蔽情報生成部２０３として図１３（Ｂ）に示すようなニューラルネット６３００を用いる。これは検出された追尾物体の画像６３０１（ここでは処理の簡単のために正方形の画像に縦横比率を正規化している）を入力すると、画像パターンを見て、遮蔽されている（Ｙｅｓ）かされていない（Ｎｏ）かの分類結果６３０２を出力する分類器である。遮蔽の有無の定義としては、物体の面積が何％以上遮蔽されているか否かとして定義する。この２クラスの値を教師値として与えてニューラルネット６３００を学習させる。この技術は通常の画像分類タスクと同様の広く公知な方法のため詳細を略す。また、教師値（目標変数）を遮蔽の有無の２値ではなく遮蔽面積の割合として与えて回帰学習を行えば、推定結果６３０３のように遮蔽の割合を推定することができる。この回帰学習には学

習時に与える損失値として二乗誤差等を用いる。
【００４９】
遮蔽情報生成部２０３で追尾物体候補の遮蔽度を推定した結果が図１４（Ａ）（Ｂ）である。図１４（Ａ）に示す物体の検出結果に対して、物体検出部２０２が図１４（Ｂ）に符号ｏｃｃを付して示したのが被遮蔽面積の推定値である。同図では被遮蔽スコアの変動幅は所定値（例えば０．３等の値）より小さく、追尾に失敗していないと判定できる（ここで、被遮蔽スコアだけでなく実施形態１で用いたような位置や見えの特徴量の類似度も併用して追尾の成功・失敗を判定してもよい）。
【００５０】
一方で図１４（Ｃ）では、動画フレーム７２２０から７２３０にかけて物体７２０１が物体７２０２の向こう側を通過しており、その結果、物体検出部２０２が動画フレーム７２３０における物体の位置を枠７２３１として誤って推定している。この場合の遮蔽スコアは図１４（Ｄ）に示すように０．４から０．０へと大きく変動しているため、交差によって誤追尾が発生したと判定することができる。誤追尾が発生したことが分かれば、そこで検出を止めたり、後段で修正する等の工夫を行うことができる。
【００５１】
以上が本実施形態の説明となる。
【００５２】
なお、遮蔽情報生成部２０３の学習は図１２（Ａ）５１１０～５１５０に示すように、不特定の物体について遮蔽状態が判定できるように様々な物体の遮蔽状態を推定できるように学習しておくことが望ましい。
【００５３】
なお他の派生の形態としては、図１３（Ｂ）では、物体枠で切られた画像６３０１を入力画像として示している。しかし、被遮蔽状態にあるか否かの判定には当該物体だけでなくその周辺を観察することが重要なため、入力画像としてはより広い範囲を入力することも考えられる（その場合、推定時にも同様の範囲を切り取って入力する）。
【００５４】
なお他の派生の形態としては、図１３（Ｃ）に示すように、上述のＬｉｕの手法のようなアンカーと言われる候補枠を使って物体の検出と遮蔽度の推定を同時に行う形態も考えられる。アンカー枠は図１３（Ｄ）に示すような複数のサイズ・縦横比率の候補枠の集合である（ここでは３種類のアンカー枠を図示している）。アンカー枠は図１３（Ｃ）の結果画像６４５０に示すように、画像中の各ブロック領域に配置されている。ニューラルネット６４００は画像が入力されたら、各ブロック領域の各アンカーに当該物体があるか否かの被遮蔽スコアマップ６４３０を生成する。被遮蔽スコアマップ６４３０はアンカー枠の種類の３個に対応した３枚のマップである。推定結果の例を図１３（Ｃ）６４５０に示す（以上の手法は広く公知のため詳細は上述のＬｉｕの方法を参照されたい）。
【００５５】
ここで本実施形態の派生の形態として、物体が存在するか否かの推定と同時に、物体の被遮蔽スコアマップ６４４０を生成する。これは各アンカー枠に、もしそこに物体がある場合、その被遮蔽割合がいくつになるかを推定したマップである。同マップもアンカーの種類の数に対応した３枚からなる（学習時には画像の各ブロックにおいて、各アンカー枠に被遮蔽スコアの教師値を与えてニューラルネット６４００を学習すればよい）。結果例を図１３（Ｃ）６４６０に示す。二つの推定マップを最終的に統合した例を統合結果例６４７０として図示する。
【００５６】
上記の説明は物体検出の例になるが、非特許文献１の方法もアンカー候補枠ベースの手法であるため、物体を追尾しながら同時にその被遮蔽スコアを推定する派生形態を構成することが可能である。
【００５７】
＜実施形態４＞

本実施形態では、ユーザが指定した単一の物体を追尾する形態について説明する。機能ブロックの図は実施形態３と同じで図９（Ｂ）である。これまでの実施形態では類似度を比較する際に、直前と直後のフレームで特徴量を比較することや、前後のｎフレームを用いて比較すること等、ルールベースでフレーム間の物体の対応付けを行った。本実施形態では、この部分を機械学習に置き換えることでより精度の高い対応付けを行う。

【００５８】

リカレントニューラルネットは時系列データを処理して識別・分類等を行うことができる技術であり、Ｂｙｅｏｎらの方法などで公知なＬｏｎｇ ｓｈｏｒｔ ｔｅｒｍ ｍｅｍｏｒｙネットワーク（以下ＬＳＴＭ）が代表的手法である。（Ｂｙｅｏｎ ｅｔ ａｌ．，Ｓｃｅｎｅ ｌａｂｅｌｉｎｇ ｗｉｔｈ ＬＳＴＭ ｒｅｃｕｒｒｅｎｔ ｎｅｕｒａｌ ｎｅｔｗｏｒｋｓ， ＣＶＰＲ ２０１５）。当該手法で物体の特徴の経時的な変化を判別して物体間の対応付けを行うことができる。本実施形態の構成と結果例の模式図を図１５に示す。ここでは１つの物体９１０２が追尾対象として指定され、Ｂｅｒｔｉｎｅｔｔｏら等の手法で追尾されている（ｔ＝２の動画フレームで誤スイッチが起こっている）。図１５（Ｃ）のＬＳＴＭユニット９５０１～９５０４は、各時刻で追尾している物体の特徴９４０１～９４０４を受け取って、追尾が成功しているか、失敗しているかを判定して出力９７０１～９７０４として出力する。ここでは図示上ＬＳＴＭユニットを複数書いているが、ここでは複数のユニットが存在するのではなく同一のユニットの各時刻の状態を示している。各時刻のＬＳＴＭユニットは次の時刻のＬＳＴＭユニットにリカレント入力９８０２を送る。ＬＳＴＭユニットはその時点の物体の特徴とそれまでの過去の情報を含むリカレント入力９８０２を元に内部状態を必要に応じて変更する。これにより物体のパターンが経時的にどのように変化しているかを踏まえた上で現時点の追尾が成功しているか否かを判断することができる。

【００５９】

ＬＳＴＭユニットへの入力の特徴量は実施形態３で説明したニューラルネットの特徴量などを用いることができる。例えば図１３（Ｂ）の物体の被遮蔽スコアを判定するニューラルネット６３１０の最終層６３２０への入力値を用いる。ここでは前記層の出力値（１値のスカラー）でなく入力値（多次元ベクトル）を用いている。これは遮蔽状態を判断するのに用いたのと同じ多次元特徴を用いることで、遮蔽に関する多種の情報をＬＳＴＭに取り込むためである。これにより様々な遮蔽のパターンを判定できることが期待できる。

【００６０】

ＬＳＴＭユニットの学習時には、教師値として各瞬間の追尾が成功しているか失敗しているかを与え、ＬＳＴＭの各重みパラメーターを調整する。また別の形態として図１６（Ｄ）に示すように、追尾の成功・失敗ではなく、教師値として遮蔽状態にあるか否かを与えて学習すれば、被遮蔽状態にあるか否かを判定させることも可能である。

【００６１】

また別の形態として、実施形態３で説明した派生の形態と同様に、追尾物体をアンカー枠ベースで検出し、９４０１として図１３（Ｃ）の物体の位置および被遮蔽スコアを同時に判定するニューラルネット６４００の特徴量６４２０を使ってもよい。この形態であれば、物体の追尾や検出と当該物体の被遮蔽スコアを同時・高速に判定することができる。

【００６２】

＜実施形態５＞

本実施形態では、ユーザが指定した単一の物体を追尾する形態について説明する。基本機能構成は実施形態１と同様である。本実施形態では物体の遮蔽情報として、相対的な物体間の遠近情報を用いる。

【００６３】

図１６（Ａ１）にその例を示す。ここでは学習画像としてＲＧＢ画像８０１を用意する。さらにレーザーレンジファインダー装置やステレオ計測等によりＲＧＢ画像８０１に対応した距離画像８３３が得られている。距離画像８３３はカメラからの距離の絶対値をグレースケールで表したものであり、白い色ほど近い距離の物体を意味する。本実施形態では

ＲＧＢ画像８０１を入力画像とし、距離画像を教師値８３１として、ニューラルネット４０２の重みを学習する。ただし絶対値としての距離画像８３１と全く同じ出力結果８３０を得ることはパターン認識としては比較的難しい問題であり、本実施形態に用いる遮蔽情報としてはそこまで高精度であることを必要としない。そこで本実施形態では近傍の物体間の相対的な遠近関係を推定するような学習を行う。

【００６４】

例えば同図の出力結果８３０に示すように、人物８０１１と８０１２の距離の推定結果８３０１と８３０２は絶対値としては正しくない。人物８０１１と離れた人物８０１３に対応する推定結果８３０１と８３０３も正しくない遠近関係になっている。しかし近傍の二人の人物８０１１と８０１２の、遠近の順序関係だけに限定すれば、正しい結果である。このように＜局所の物体間＞の＜遠近順序の関係＞は正しく推定できるように学習し、これらを物体の遮蔽情報として集計して用いる。

【００６５】

以上は、学習時の損失値計算に以下の工夫を施すことで実現される。図１６（Ａ２）に図１６（Ａ１）の教師値８３１上の記号＊の付近の領域を拡大した教師値領域８３１ａを示す。対応する出力結果の領域８３０ａも示す。ここで領域８３１ａ上の各画素ｉと画素ｊに注目し、その遠近関係が正しいか否かで当該画素ペアの損失を求める。ここでは領域８３０ａ上の画素ｉと画素ｊの遠近関係は教師値と一致するので損失は発生しない。対してもし領域８３０ｂのような推定結果であった場合は、遠近関係が正しくないので損失を計上する。このような判断を、所定距離内にある全画素ペアで行う。最終的に遠近関係を誤ったペア数を全ペア数で割った値を損失値の総計とする。このようにして学習したニューラルネット８０２が学習終了し、推定した距離の出力結果８３４を図１６（Ｂ）に示す。

【００６６】

次に、相対的な距離の出力結果８３４を集計して物体の被遮蔽尤度を求める。ここでは別途検出しておいた人物検出枠８３５１と８３５２を用いて検出枠ごとに集計する。各枠内でそれぞれの距離の値を平均し、$dave_1$と$dave_2$とする。次にこの距離の値を隣接した物体枠間で比較して正規化して被遮蔽尤度のスコア値ｏｃｃへと変換する。例えば下式で変換する。

（数式３）

$$occ_i = Sigmoid(Log(dave_i / dave_j)) = 1/(1 + dave_j / dave_i),$$

$$occ_j = 1/(1 + dave_i / dave_j),$$

ただし

$$Sigmoid(x) = 1/(1 + exp(-x)).$$

ここでｉとｊは重なり部分のある二つの隣接した検出物体枠である。３つ以上の物体が重なっている場合は、それぞれ上記の式で被遮蔽スコアocc_iを求め、そのうちの最大値をその物体の被遮蔽スコアとしてもよい。

【００６７】

以上が相対的な距離推定を行い、被遮蔽スコアを集計するまでの処理内容となる。被遮蔽スコアを用いた追尾処理は実施形態１と同様になるためここでは割愛する。

【００６８】

なお派生的な学習の工夫として下記のようなものが考えられる。（１）距離の教師値の差分が所定閾値Θ以上のペアのみに限定して損失を集計する。これにより距離画像の観測時のノイズに対しロバストに学習できる。（２）（１）を行い、且つマージン領域を設定する。例えばペアの遠近関係が正しいか正しくないかのみならず、遠近関係が正しく、且つ所定閾値Θ以上値が相対的に離れていない場合に損失を発生させる。（３）距離の教師値の差分が閾値Θ未満の画素ペアに対する出力値が、閾値Θ以上に大きなケースも誤りとして損失を与える。これによりノイズ的な出力を抑制する。

【００６９】

以上、さまざまな形態があり得るが、相対的・局所的に距離を学習できるような形態であ

ればいずれでもよく、一つの形態に限定されない。本発明は、以下の処理を実行することによっても実現される。即ち、上述した実施形態の機能を実現するソフトウェア（プログラム）を、データ通信用のネットワーク又は各種記憶媒体を介してシステム或いは装置に供給する。そして、そのシステム或いは装置のコンピュータ（またはＣＰＵやＭＰＵ等）がプログラムを読み出して実行する処理である。また、そのプログラムをコンピュータが読み取り可能な記録媒体に記録して提供してもよい

【符号の説明】

【００７０】

１　情報処理装置

２０１　画像取得部

２０２　物体検出部

２０３　遮蔽情報生成部

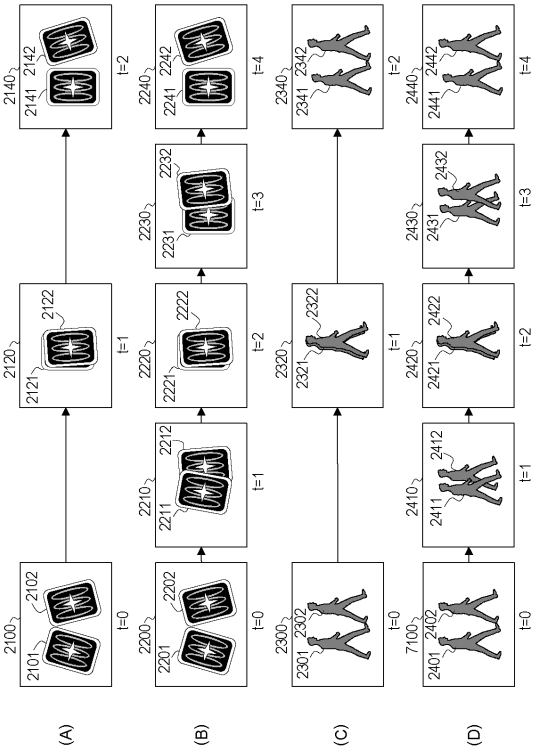
２０４　特徴量取得部

２０５　対応付け部

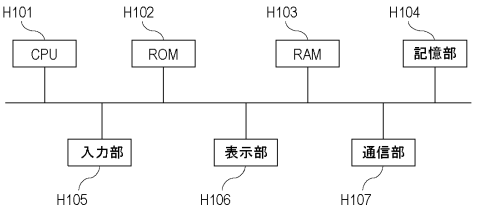
２０６　記憶部

【図面】

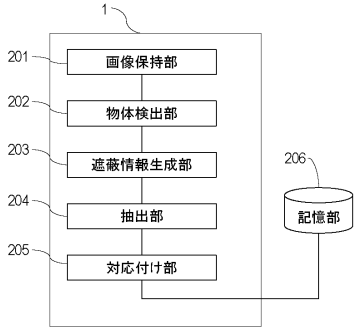
【図１】



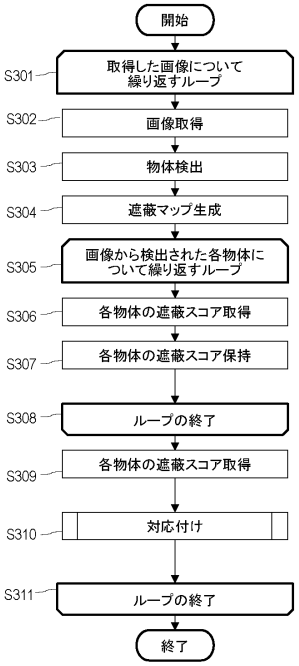
【図２】



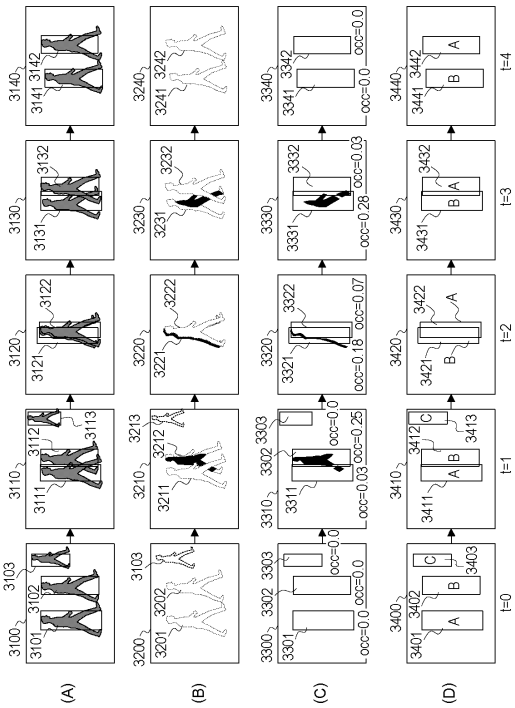
【図３】


1
201
画像保持部
202
物体検出部
203
遮蔽情報生成部
204
抽出部
205
対応付け部
206
記憶部

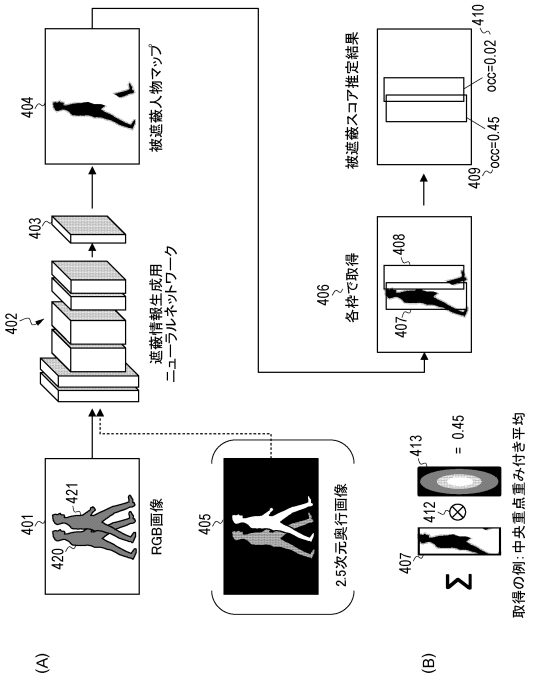
【図４】


開始
S301
取得した画像について繰り返すループ
S302
画像取得
S303
物体検出
S304
遮蔽マップ生成
S305
画像から検出された各物体について繰り返すループ
S306
各物体の遮蔽スコア取得
S307
各物体の遮蔽スコア保持
S308
ループの終了
S309
各物体の遮蔽スコア取得
S310
対応付け
S311
ループの終了
終了

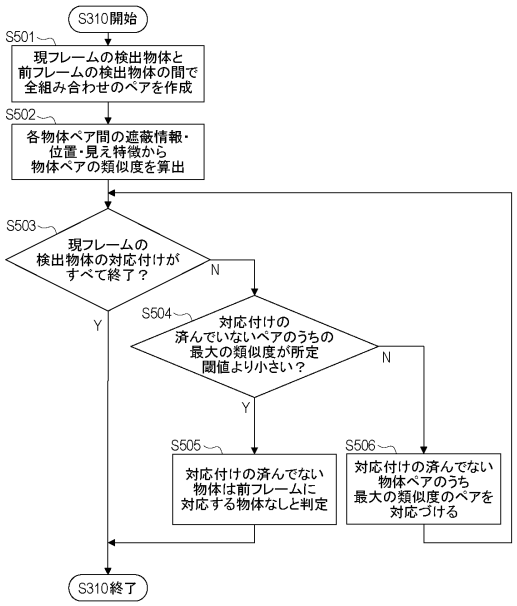
【図５】


(A)
(B)
(C)
(D)
t=0
t=1
t=2
t=3
t=4
occ=0.0
occ=0.03
occ=0.18
occ=0.28
occ=0.45

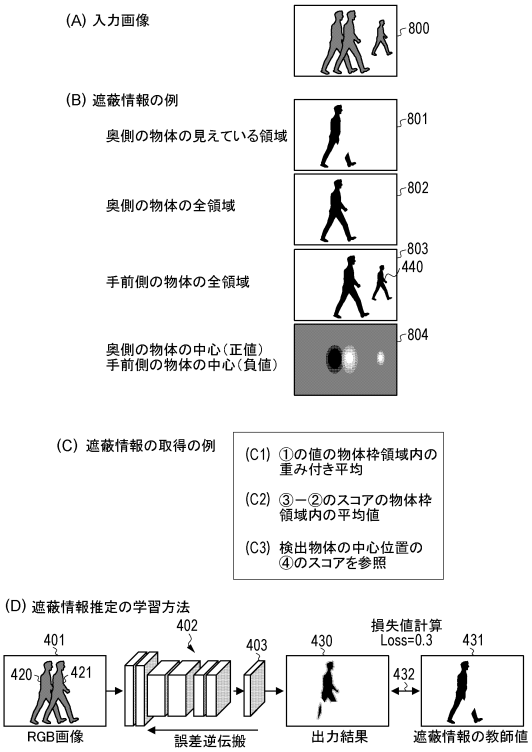
【図６】


(A)
(B)
RGB画像
2.5次元奥行画像
遮蔽情報生成用ニューラルネットワーク
被遮蔽人物マップ
各枠で取得
被遮蔽スコア推定結果
occ=0.02
occ=0.45
= 0.45
取得の例：中央重点重み付き平均

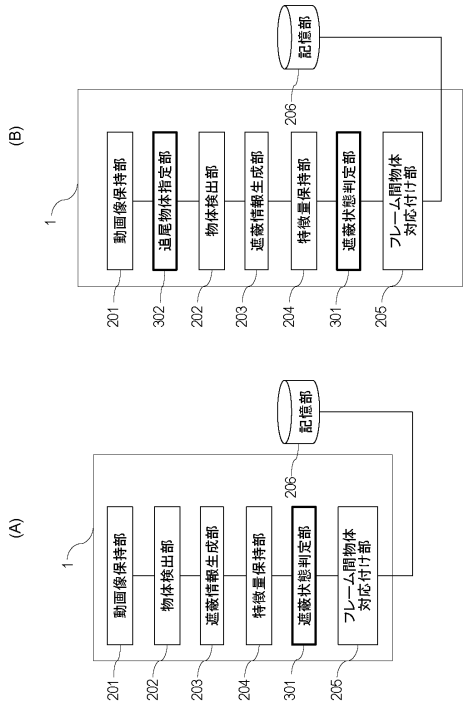
【図７】


S310開始
S501
現フレームの検出物体と前フレームの検出物体の間で全組み合わせのペアを作成
S502
各物体ペア間の遮蔽情報・位置・見え特徴から物体ペアの類似度を算出
S503
現フレームの検出物体の対応付けがすべて終了？
N
Y
S504
対応付けの済んでいないペアのうちの最大の類似度が所定閾値より小さい？
N
Y
S505
対応付けの済んでない物体は前フレームに対応する物体なしと判定
S506
対応付けの済んでない物体ペアのうち最大の類似度のペアを対応づける
S310終了

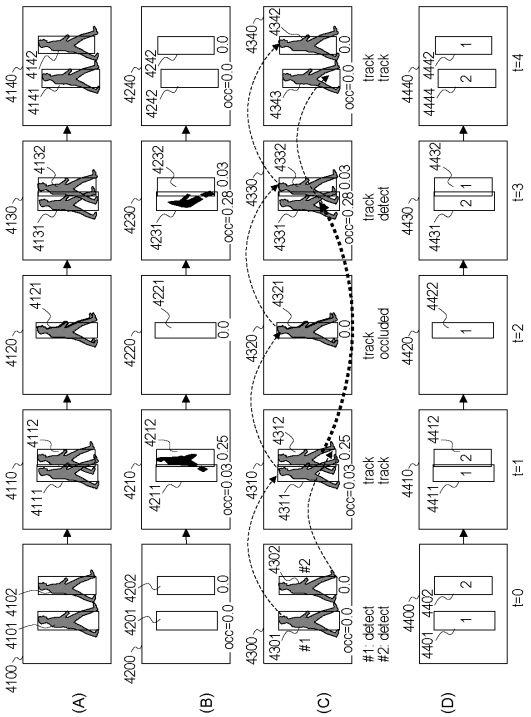
【図８】


(A) 入力画像
800
(B) 遮蔽情報の例
奥側の物体の見えている領域
801
奥側の物体の全領域
802
手前側の物体の全領域
803
440
奥側の物体の中心（正値）
手前側の物体の中心（負値）
804
(C) 遮蔽情報の取得の例
(C1) ①の値の物体枠領域内の重み付き平均
(C2) ③－②のスコアの物体枠領域内の平均値
(C3) 検出物体の中心位置の④のスコアを参照
(D) 遮蔽情報推定の学習方法
401
420
421
RGB画像
402
403
誤差逆伝搬
430
出力結果
損失値計算
Loss=0.3
431
432
遮蔽情報の教師値

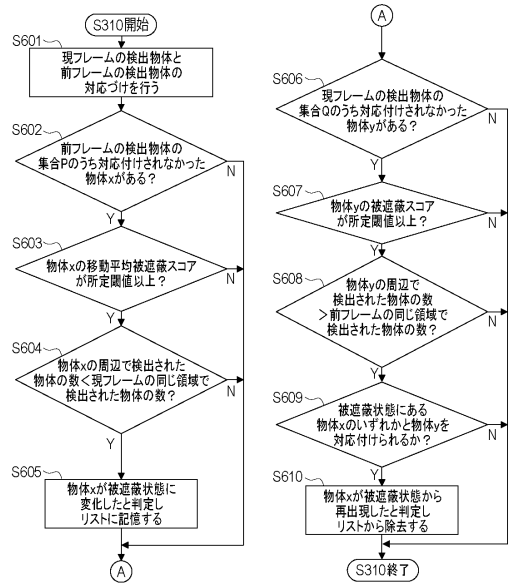
【図９】


(B)
記憶部
206
1
動画像保持部
追尾物体指定部
物体検出部
遮蔽情報生成部
特徴量保持部
遮蔽状態判定部
フレーム間物体対応付け部
201
302
202
203
204
301
205
(A)
記憶部
206
1
動画像保持部
物体検出部
遮蔽情報生成部
特徴量保持部
遮蔽状態判定部
フレーム間物体対応付け部
201
202
203
204
301
205

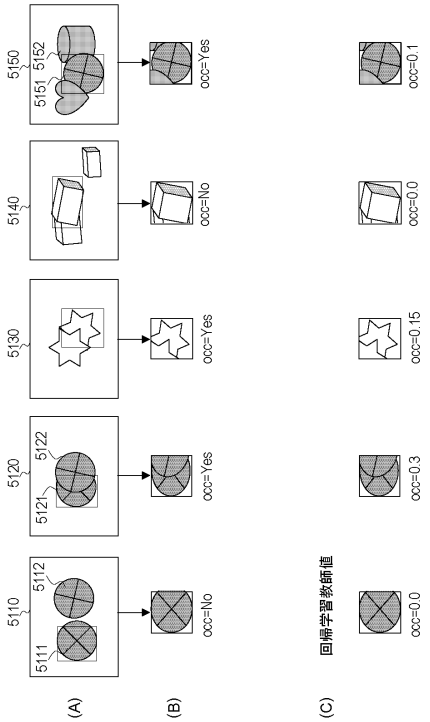
【図１０】


4100
4101
4102
4110
4111
4112
4120
4121
4130
4131
4132
4140
4141
4142
4200
4201
4202
4210
4211
4212
4220
4221
4230
4231
4232
4240
4242
4300
4301
4302
#1
#2
4310
4311
4312
4320
4321
4330
4331
4332
4340
4343
4400
4401
4402
4410
4411
4412
4420
4422
4430
4431
4432
4440
4442
4444
occ=0.0
occ=0.03
occ=0.28
track
detect
occluded
#1: detect
#2: detect
t=0
t=1
t=2
t=3
t=4
(A)
(B)
(C)
(D)

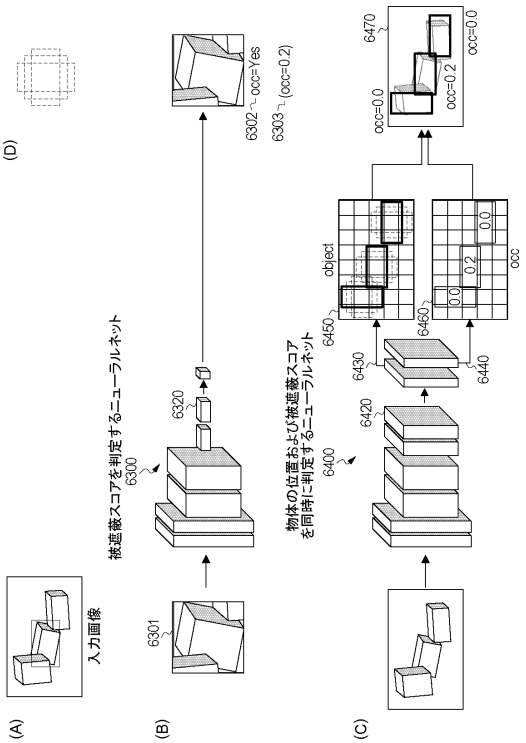
【図１１】



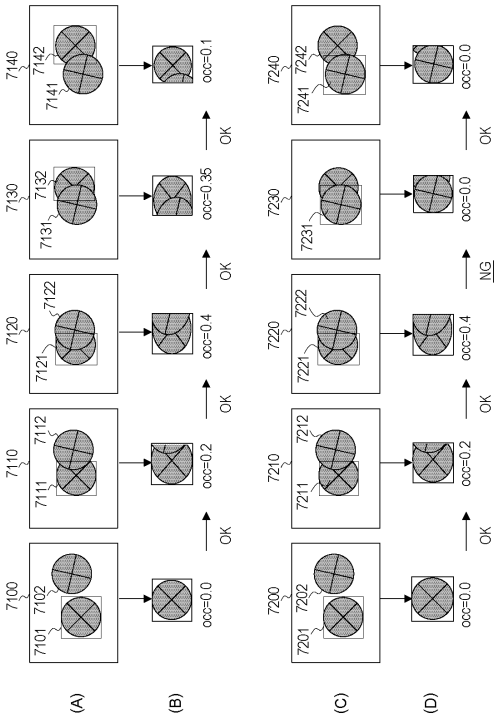
【図１２】



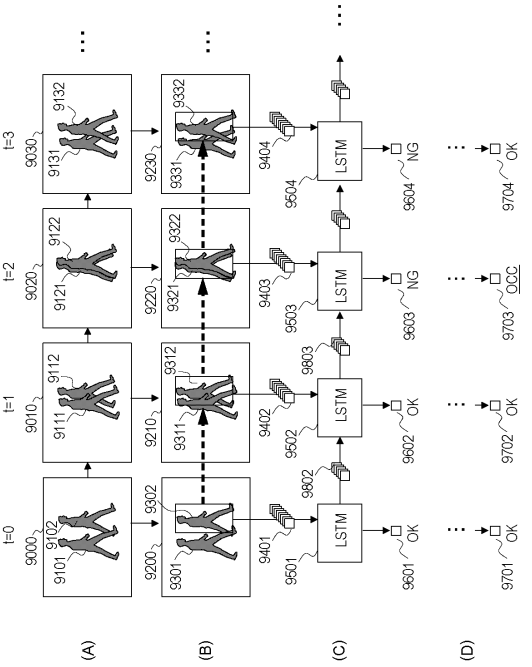
【図１３】



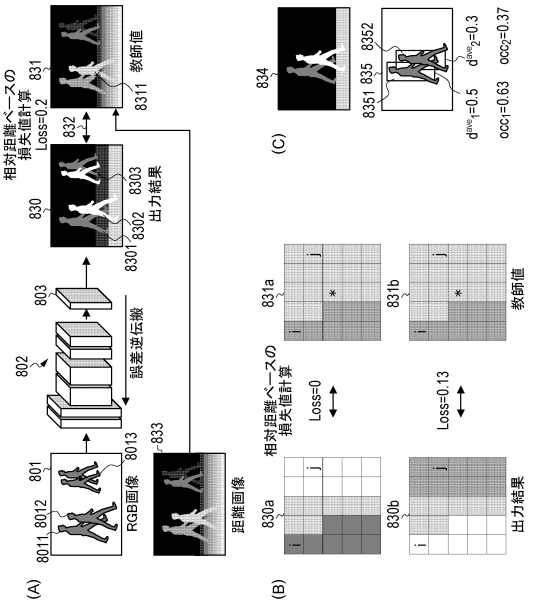
【図１４】



【図１５】


t=0
t=1
t=2
t=3
LSTM
OK
NG
OCC
(A)
(B)
(C)
(D)

【図１６】


相対距離ベースの
損失値計算
Loss=0.2
教師値
出力結果
誤差逆伝搬
RGB画像
距離画像
Loss=0
Loss=0.13
(A)
(B)
(C)

フロントページの続き

(51)国際特許分類	FI		テーマコード（参考）
	H04N	7/18	K

キヤノン株式会社内

Fターム（参考） 5C054 CA04 CC02 FC12 FC13 HA19
5L096 AA02 AA06 AA09 BA18 CA04 DA02 FA15 FA35 FA59 FA69
FA77 GA40 HA05 HA11 JA11