

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第6912890号
(P6912890)

(45) 発行日 令和3年8月4日 (2021. 8. 4)

(24) 登録日 令和3年7月13日 (2021. 7. 13)

(51) Int. Cl.	F I
HO 4 N 7/18 (2006. 01)	HO 4 N 7/18 D
HO 4 N 5/232 (2006. 01)	HO 4 N 7/18 E
GO 8 B 25/00 (2006. 01)	HO 4 N 5/232 9 9 O
GO 8 B 13/196 (2006. 01)	GO 8 B 25/00 5 1 O M
GO 6 T 7/00 (2017. 01)	GO 8 B 13/196

請求項の数 11 (全 13 頁) 最終頁に続く

(21) 出願番号	特願2017-4616 (P2017-4616)	(73) 特許権者	000001007
(22) 出願日	平成29年1月13日 (2017. 1. 13)		キヤノン株式会社
(65) 公開番号	特開2018-113660 (P2018-113660A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成30年7月19日 (2018. 7. 19)	(74) 代理人	110003281
審査請求日	令和2年1月14日 (2020. 1. 14)		特許業務法人大塚国際特許事務所
		(72) 発明者	矢野 光太郎
			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	河合 智明
			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
		審査官	佐野 潤一

最終頁に続く

(54) 【発明の名称】 情報処理装置、情報処理方法、システム

(57) 【特許請求の範囲】

【請求項 1】

撮像装置による撮像画像から検出したオブジェクトの領域に基づいて、該撮像装置のパン、チルト、ズームのうち少なくとも1つを含む該撮像装置の制御量を推定する推定手段と、

ユーザ操作に応じて指示された前記撮像装置のパン、チルト、ズームのうち少なくとも1つを含む該撮像装置の制御量を取得する取得手段と、

前記取得手段が取得した制御量と前記推定手段が推定した制御量との差分に基づく評価値が小さくなるように、前記推定に用いるパラメータを更新する更新手段と

を備えることを特徴とする情報処理装置。

【請求項 2】

前記更新手段により更新済みのパラメータを用いて前記推定手段が推定した前記撮像装置の制御量に応じて前記撮像装置を制御する制御手段を更に備え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】

前記推定手段は、

前記領域内の画像特徴量を求める第 1 の手段と、

前記画像特徴量と、前記領域の画像座標と、前記領域の尤度と、に基づいて、前記撮像装置のパン、チルト、ズームのうち少なくとも1つを含む該撮像装置の制御量を推定する第 2 の手段と

を備え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】

前記第 2 の手段は、全結合ニューラルネットワークを有し、前記パラメータは該全結合ニューラルネットワークにおけるニューロン間の結合係数であることを特徴とする請求項 3 に記載の情報処理装置。

【請求項 5】

前記制御手段は、更に前記推定手段が推定した前記オブジェクトの領域の注目度に基づいて決定された前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量に応じて前記撮像装置を制御することを特徴とする請求項 2 に記載の情報処理装置。

10

【請求項 6】

前記推定手段は、複数の前記オブジェクトの領域のそれぞれについて、前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量と注目度とを推定し、

前記制御手段は、更に前記複数のオブジェクトの領域のそれぞれに対応する前記注目度に基づいて、前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量を決定することを特徴とする請求項 5 に記載の情報処理装置。

【請求項 7】

前記制御手段は、前記更新手段により更新済みのパラメータを用いて前記推定手段が推定した前記複数のオブジェクトの領域のそれぞれに対応する前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量のうち、対応する前記注目度が最も大きい制御量を、前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量として決定することを特徴とする請求項 6 に記載の情報処理装置。

20

【請求項 8】

前記制御手段は、前記更新手段により更新済みのパラメータを用いて前記推定手段が推定した前記複数のオブジェクトの領域のそれぞれに対応する前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量を、対応する前記注目度を重みとして重み付け平均した結果を、前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量として決定することを特徴とする請求項 6 に記載の情報処理装置。

30

【請求項 9】

撮像装置と、該撮像装置を制御する情報処理装置と、を有するシステムであって、前記情報処理装置は、

前記撮像装置による撮像画像から検出したオブジェクトの領域に基づいて、該撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量を推定する推定手段と、

ユーザ操作に応じて指示された前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量を取得する取得手段と、

前記取得手段が取得した制御量と前記推定手段が推定した制御量との差分に基づく評価値が小さくなるように、前記推定に用いるパラメータを更新する更新手段と

40

を備えることを特徴とするシステム。

【請求項 10】

情報処理装置が行う情報処理方法であって、

前記情報処理装置の推定手段が、撮像装置による撮像画像から検出したオブジェクトの領域に基づいて、該撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量を推定する推定工程と、

前記情報処理装置の取得手段が、ユーザ操作に応じて指示された前記撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量を取得する取得工程と、

前記情報処理装置の更新手段が、前記取得工程で取得した制御量と前記推定工程で推定した制御量との差分に基づく評価値が小さくなるように、前記推定に用いるパラメータを

50

更新する更新工程と

を備えることを特徴とする情報処理方法。

【請求項 11】

コンピュータを、請求項 1 乃至 8 の何れか 1 項に記載の情報処理装置の各手段として機能させるためのコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、撮像装置の制御技術に関するものである。

【背景技術】

【0002】

従来から、撮像レンズのパン、チルト機構やズーム機構を制御信号によって制御することで撮影方向および撮影倍率を変更可能なカメラが開発されている。このようなカメラは監視用途に有用であり、例えば、カメラで撮影した映像に不審者が写った場合に撮影方向や倍率を変更することで不審者を追尾したりズームアップしたりすることができる。

【0003】

しかしながら、監視者がカメラの映像を見てカメラの制御を行うためには、熟練した操作が必要であり、長時間操作を続けたり、多数のカメラに対して操作したりすることは困難である。このような課題に対応するために、特許文献 1 では、カメラを電動の雲台と電動のズームレンズによって自動的に制御し、人物を検出して追尾する監視装置が提案されている。一方、特許文献 2 では、画像パターンと操作者のカメラ制御との関係をニューラルネットワークで学習し、撮像制御を自動化する監視制御装置が提案されている。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開 2003 - 219225 号公報

【特許文献 2】特開 2004 - 56473 号公報

【非特許文献】

【0005】

【非特許文献 1】Dalal and Triggs. Histograms of Oriented Gradients for Human Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、特許文献 1 では、単に検出した人物をズームアップして追尾するだけであり、対象外の人物であっても追尾制御を行ってしまうため、追尾中により重要なイベントが発生した場合に取りこぼしてしまう恐れがある。

【0007】

また、特許文献 2 では、映像中の人物の有無に関わらず操作者が行う追尾操作と画像パターンの関係を単純に学習するだけなので、人物が映っていない場合にも間違った撮影制御を行ってしまう恐れがある。

【0008】

本発明はこのような問題に鑑みてなされたものであり、操作者の意図を汲んだ撮像制御を精度良く行うための技術を提供する。

【課題を解決するための手段】

【0009】

本発明の一様態は、撮像装置による撮像画像から検出したオブジェクトの領域に基づいて、該撮像装置のパン、チルト、ズームのうち少なくとも 1 つを含む該撮像装置の制御量を推定する推定手段と、

10

20

30

40

50

ユーザ操作に応じて指示された前記撮像装置のパン、チルト、ズームのうち少なくとも1つを含む該撮像装置の制御量を取得する取得手段と、

前記取得手段が取得した制御量と前記推定手段が推定した制御量との差分に基づく評価値が小さくなるように、前記推定に用いるパラメータを更新する更新手段とを備えることを特徴とする。

【発明の効果】

【0010】

本発明の構成により、操作者の意図を汲んだ撮像制御を精度良く行うことができる。

【図面の簡単な説明】

【0011】

10

【図1】システムの構成例を示すブロック図。

【図2】推定パラメータの学習処理のフローチャート。

【図3】制御量推定部140の構成例を示すブロック図。

【図4】深層ニューラルネットワークの構成例を示す図。

【図5】自動制御処理のフローチャート。

【図6】コンピュータ装置のハードウェア構成例を示すブロック図。

【発明を実施するための形態】

【0012】

以下、添付図面を参照し、本発明の実施形態について説明する。なお、以下説明する実施形態は、本発明を具体的に実施した場合の一例を示すもので、特許請求の範囲に記載した構成の具体的な実施例の1つである。

20

【0013】

〔第1の実施形態〕

まず、本実施形態に係るシステムの構成例について、図1のブロック図を用いて説明する。図1に示す如く、本実施形態に係るシステムは、カメラ200と、該カメラ200の動作制御を行う情報処理装置100と、を有する。

【0014】

まず、カメラ200について説明する。カメラ200は、撮像レンズのパン、チルト機構、ズーム機構を有するものであり、カメラ200のパン、チルト、ズームは情報処理装置100から制御することができる。カメラ200は、情報処理装置100からの制御に応じて動画像を撮像する。そしてカメラ200は、撮像した動画像を構成する各フレームの画像（撮像画像）を情報処理装置100に対して出力する。カメラ200は、静止画像を撮像するカメラであっても良い。

30

【0015】

次に、情報処理装置100について説明する。

【0016】

操作部400は、マウスやキーボード、タッチパネル画面などのユーザインターフェースにより構成されており、ユーザが操作することで各種の指示を撮影制御部300に対して入力することができる。

【0017】

40

撮影制御部300は、操作部400からの操作指示、若しくは制御量推定部140が後述する推定処理によって推定した「パン、チルト、ズーム等の制御量」に従って、カメラ200のパン、チルト、ズーム等を制御するための制御信号を生成する。以下では、カメラ200の「パン、チルト、ズーム等の制御量」を単に制御量と呼称する場合がある。そして撮影制御部300は、該生成した制御信号をカメラ200に対して出力する。カメラ200は、この制御信号に従って、撮像レンズのパン、チルト、ズームを制御する。

【0018】

操作情報取得部120は、撮影制御部300が生成した制御信号から、該制御信号が示す制御量を取得する。画像取得部110は、カメラ200から出力された撮像画像を取得する。

50

【 0 0 1 9 】

人検出部 1 3 0 は、画像取得部 1 1 0 が取得した撮像画像から人物が写っている領域（人物領域）を検出する。表示部 5 0 0 は、C R T や液晶画面などにより構成されており、画像取得部 1 1 0 が取得した撮像画像を表示する。

【 0 0 2 0 】

制御量推定部 1 4 0 は、画像取得部 1 1 0 が取得した撮像画像、人検出部 1 3 0 による該撮像画像からの検出結果、記憶部 1 6 0 に格納されている推定パラメータ、を用いて、該撮像画像中の人物領域ごとに、制御量と、該人物領域内の人物に対する注目の度合いを示す値（注目度）と、を推定する。

【 0 0 2 1 】

学習部 1 5 0 は、画像取得部 1 1 0 が取得した撮像画像中の人物領域ごとに、注目度を取得する。学習部 1 5 0 は、画像取得部 1 1 0 が取得した撮像画像中の人物領域の位置や、その人物領域の人物へのユーザの操作（ズームアップ）等に基づいて、ユーザがその人物領域をどの程度注目しているのかを推定することにより、その注目度を取得する。そして学習部 1 5 0 は、該取得した注目度、操作情報取得部 1 2 0 が取得した制御量、制御量推定部 1 4 0 が推定した制御量及び注目度、を用いて、記憶部 1 6 0 に格納されている推定パラメータを更新（学習）する。

【 0 0 2 2 】

情報処理装置 1 0 0 が行う、推定パラメータの学習処理について、同処理のフローチャートを示す図 2 を用いて説明する。

【 0 0 2 3 】

ステップ S 1 0 0 では、画像取得部 1 1 0 は、カメラ 2 0 0 から出力された撮像画像を取得する。本実施形態では、撮像画像は、各画素の R（赤）、G（緑）、B（青）の各色成分の輝度値が 8 ビットで表されるカラー画像データであるものとする。しかし、撮像画像はカラー画像データに限らず、モノクロ画像データであっても良いし、各画素の色成分の種類やビット数もまた特定の種類、ビット数に限らない。

【 0 0 2 4 】

ステップ S 1 1 0 では、人検出部 1 3 0 は、ステップ S 1 0 0 で取得した撮像画像から人物領域を検出する。画像から人を検出する方法としては、例えば非特許文献 1 に記載の方法がある。非特許文献 1 に記載の方法では、画像から勾配方向ヒストグラム特徴（Histograms of Oriented Gradients）を抽出し、抽出した特徴量をサポートベクターマシンで学習したモデルを用いて人か否かを識別するようにしている。なお、撮像画像から人物領域を検出するための方法は、非特許文献 1 に開示されている方法に限らない。例えば、抽出する特徴量は勾配方向ヒストグラム特徴に限らず、Haar-like 特徴、LBPH 特徴（Local Binary Pattern Histogram）等を用いてもよいし、それらを組み合わせてもよい。また、人を識別するモデルはサポートベクターマシンに限らず、アダブースト識別器、ランダム分類木（Randomized Tree）等を用いてもよい。なお、人検出部 1 3 0 は、撮像画像中に複数人の人が写っている場合には、それぞれの人を検出することになる。

【 0 0 2 5 】

そして人検出部 1 3 0 は、人物領域を検出すると、該人物領域の四隅の画像座標と、該人物領域に対する尤度と、を出力する。人物領域に対する尤度とは、該人物領域から抽出した特徴量と人を識別するモデルとを照合した結果であり、モデルとの一致度を表す。

【 0 0 2 6 】

ステップ S 1 2 0 では、制御量推定部 1 4 0 は、ステップ S 1 0 0 で取得した撮像画像中の人物領域について、制御量と、該人物領域内の人物に対する注目の度合いを示す値（注目度）と、を推定する。ステップ S 1 2 0 における処理は、ステップ S 1 0 0 で取得した撮像画像中のそれぞれの人物領域について行われる。そして、ステップ S 1 0 0 で取得した撮像画像中のそれぞれの人物領域についてステップ S 1 2 0 の処理が完了すると、処理はステップ S 1 7 0 に進む。ここで、制御量推定部 1 4 0 の構成例について、図 3 のブロック図を用いて説明する。

【 0 0 2 7 】

領域抽出部 1 4 1 は、ステップ S 1 0 0 で取得した撮像画像から、人検出部 1 3 0 が検出した人物領域（人検出部 1 3 0 が検出した四隅の画像座標で規定される領域）内の画像を抽出し、該抽出した画像を規定サイズに正規化した正規化画像を生成する。

【 0 0 2 8 】

特徴抽出部 1 4 2 及び推定部 1 4 3 は、図 4 に示す深層ニューラルネットワークで構成されている。図 4 に示す深層ニューラルネットワークでは、縦 H 画素 × 横 W 画素を有する入力画像（正規化画像）を入力として 5 層構成の畳込み型ニューラルネットワークの演算を行い、その演算結果を第 6 層及び第 7 層の全結合ニューラルネットワークに入力して出力を得る。f 1 ~ f 5 はそれぞれ、第 1 層（C o n v 1）~ 第 5 層（C o n v 5）の畳込み演算のフィルタサイズを表し、d 1 ~ d 7 はそれぞれ、第 1 層 ~ 第 7 層（第 6 層及び第 7 層はそれぞれ F c 6 , F c 7）の出力チャンネル数を表す。

10

【 0 0 2 9 】

第 1 層 ~ 第 5 層の畳込み型ニューラルネットワークは特徴抽出部 1 4 2 に含まれており、特徴抽出部 1 4 2 は、第 1 層 ~ 第 5 層の畳込み型ニューラルネットワークによって入力画像から画像特徴量を抽出する。そして特徴抽出部 1 4 2 は、該入力画像から抽出した画像特徴量を出力する。

【 0 0 3 0 】

第 6 層及び第 7 層の全結合ニューラルネットワークは推定部 1 4 3 に含まれている。推定部 1 4 3 は、第 6 層及び第 7 層の全結合ニューラルネットワークによって、特徴抽出部 1 4 2 から出力された人物領域の画像特徴量、人検出部 1 3 0 から出力された四隅の画像座標、尤度から該人物領域に対応する制御量及び注目度を求める。

20

【 0 0 3 1 】

図 3 に戻って、統合部 1 4 4 は、情報処理装置 1 0 0 が推定パラメータの学習処理を行っている際には動作せず、推定部 1 4 3 からの出力（人物領域に対応する制御量及び注目度）をそのまま学習部 1 5 0 に対して出力する。情報処理装置 1 0 0 が推定パラメータの学習処理を行っていないときの統合部 1 4 4 の動作については後述する。

【 0 0 3 2 】

以上説明した図 3 の構成を用いてステップ S 1 2 0 の処理を撮像画像中のそれぞれの人物領域について行うことで、該人物領域に対応する制御量及び注目度を推定することができる。

30

【 0 0 3 3 】

一方、ステップ S 1 0 0 で取得した撮像画像は、ステップ S 1 3 0 において表示部 5 0 0 に表示される。ここでユーザが操作部 4 0 0 を操作して、カメラ 2 0 0 のパン、チルト、ズームなどを操作する指示（操作指示）を入力すると、ステップ S 1 4 0 において撮影制御部 3 0 0 は、操作部 4 0 0 からの操作指示を取得する。

【 0 0 3 4 】

ステップ S 1 5 0 では、撮影制御部 3 0 0 は、ステップ S 1 4 0 で取得した操作指示に従って、カメラ 2 0 0 のパン、チルト、ズーム等を制御するための制御信号を生成し、該生成した制御信号をカメラ 2 0 0 に対して出力する。これによりカメラ 2 0 0 は、撮影制御部 3 0 0 から出力された制御信号に従って、パン、チルト、ズーム等を変更する。

40

【 0 0 3 5 】

ステップ S 1 6 0 では、操作情報取得部 1 2 0 は、ステップ S 1 5 0 において撮影制御部 3 0 0 が生成した制御信号から、該制御信号が示す制御量を取得する。

【 0 0 3 6 】

ステップ S 1 7 0 では、学習部 1 5 0 は、ステップ S 1 2 0 において制御量推定部 1 4 0 がそれぞれの人物領域について推定した制御量及び注目度と、ステップ S 1 6 0 において操作情報取得部 1 2 0 が取得した制御量と、を取得する。更に学習部 1 5 0 は、人検出部 1 3 0 による検出結果と操作情報取得部 1 2 0 が取得した制御量とから、撮像画像においてユーザがどの人物に注目したのかを判定して、該撮像画像内のそれぞれの人物領域に

50

ついて注目度を取得する。ユーザが操作部 400 を操作して撮像画像の中央に近づけたりズームアップした人物（人物領域）の注目度を「1」、その他の人物（人物領域）の注目度を「0」とする。また、何も操作を行わなかった場合は検出した全ての人物の注目度は「0」となる。

【0037】

この様に学習部 150 は 1 フレーム分の撮像画像について「制御量推定部 140 が推定した人物領域ごとの制御量及び注目度、操作情報取得部 120 が取得した制御量、学習部 150 が撮像画像から取得した人物領域ごとの注目度」を学習データとして取得する。

【0038】

そして、学習部 150 が学習データを規定フレーム数分収集できた場合には、処理はステップ S180 に進む。一方、学習データを規定フレーム数分収集できていない場合には、次のフレームについてステップ S100 以降の処理を繰り返す。

【0039】

なお、ステップ S180 に進むための条件は特定の条件に限らない。例えば、制御量推定部 140 が推定したデータ量が規定量以上になった場合に、ステップ S180 に進むようにしても良い。

【0040】

ステップ S180 では、学習部 150 は、記憶部 160 に格納されている推定パラメータ、すなわち、上記の第 6 層及び第 7 層の全結合ニューラルネットワークにおけるニューロン間の結合係数を、学習データを用いて更新（学習）する。

【0041】

ここで、学習データを用いた推定パラメータの更新処理について説明する。制御量推定部 140 が規定フレーム数の撮像画像から収集した制御量及び注目度をそれぞれ、 $C = \{C_1, C_2, \dots, C_n\}$ 、 $a = \{a_1, a_2, \dots, a_n\}$ とする。 n は 2 以上の整数である。 n が大きいほど精度の高い学習が可能であるが、その分だけ学習に時間がかかる。ここで、 C_i 、 a_i ($1 \leq i \leq n$) はそれぞれ、同フレームにおける撮像画像において同じ人物領域に対して制御量推定部 140 が推定した制御量、注目度である。なお、 $C_i = (P_i, T_i, Z_i)$ であり、 P_i はパンの制御量、 T_i はチルトの制御量、 Z_i はズームの制御量を表す。また、 C_i を求めた撮像画像について操作情報取得部 120 が取得した制御量を $C^{\wedge}i$ とする。 $C^{\wedge}i = (P^{\wedge}i, T^{\wedge}i, Z^{\wedge}i)$ であり、 $P^{\wedge}i$ はパンの制御量、 $T^{\wedge}i$ はチルトの制御量、 $Z^{\wedge}i$ はズームの制御量を表す。また、 a_i を求めた人物領域について学習部 150 が取得した注目度を $a^{\wedge}i$ とする。

【0042】

本実施形態では、平均損失の勾配から推定パラメータを求める確率的勾配降下法を用いる。本実施形態では、平均損失として制御量及び注目度の差異（差分）を評価する。損失関数（評価値）は以下に示す（式 1）で求める。

【0043】

$$L = \{w_1 \times (P_i - P^{\wedge}i)^2 + w_2 \times (T_i - T^{\wedge}i)^2 + w_3 \times (Z_i - Z^{\wedge}i)^2 + w_4 \times (a_i - a^{\wedge}i)^2\} \quad (\text{式 1})$$

w_1 、 w_2 、 w_3 、 w_4 は規定の重み係数である。また、 i は全ての i ($= 1 \sim n$) についての総和を表す。学習に用いるデータは全てを用いてもよいし、ランダムに所定数分選択してもかまわない。

【0044】

学習部 150 は、上記の第 6 層及び第 7 層における結合係数（推定パラメータ）を微量だけ変化させて得た学習データから（式 1）に基づく勾配をそれぞれ求めて、平均損失が小さくなるように推定パラメータを学習する。学習した推定パラメータは、記憶部 160 に格納済みの推定パラメータに上書き保存され、これにより、記憶部 160 に格納されている推定パラメータが更新される。

【0045】

推定パラメータの学習の終了条件については様々な条件が考えられる。すなわち、損失

関数の値の変化量が規定値未満となった場合や、学習回数が規定値に達した場合に、学習を終了させても良い。また、ユーザが操作部 400 を操作して学習の終了指示を入力した場合に、学習を終了させても良い。

【0046】

次に、上記の学習が完了した後、情報処理装置 100 が推定パラメータを用いてカメラ 200 のパン、チルト、ズームなどを制御する自動制御処理について、同処理のフローチャートを示す図 5 を用いて説明する。

【0047】

ここで、ステップ S200 ~ S220 の各ステップにおける処理はそれぞれ、次の点を除き、上記のステップ S100 ~ S120 と同様である。ステップ S220 で動作する上記の第 6 層及び第 7 層の全結合ニューラルネットワークの結合係数は、上記の学習によって更新された（更新済みの）推定パラメータである。そして、ステップ S200 で取得した撮像画像中のそれぞれの人物領域についてステップ S220 の処理が完了すると、処理はステップ S230 に進む。

【0048】

ステップ S230 では、制御量推定部 140 の統合部 144 は、推定部 143 が人物領域毎に出力した制御量を統合することで、カメラ 200 の制御量を決定する。統合する方法には様々な統合方法がある。例えば統合部 144 は、推定部 143 から出力された人物領域ごとの制御量のうち、対応する注目度が最も高い制御量を統合結果として出力する。また統合部 144 は、複数の人物領域から推定した制御量を、対応する注目度を重みとして重み付け平均した結果を統合結果として出力する。

【0049】

ステップ S240 では、撮影制御部 300 は、ステップ S230 で統合結果として統合部 144 から出力された制御量を表す制御信号を生成し、該生成した制御信号をカメラ 200 に対して出力する。これによりカメラ 200 は、撮影制御部 300 から出力された制御信号に従って動作する。

【0050】

図 5 のフローチャートに従った処理は、1 フレーム分の撮像画像についての処理であるため、実際には、カメラ 200 から入力される撮像画像毎に図 5 のフローチャートに従った処理が行われる。なお、図 5 のフローチャートに従った処理の終了条件については特定の終了条件に限らない。例えば、ユーザが操作部 400 を操作して図 5 のフローチャートに従った処理の終了指示を入力した場合に、図 5 のフローチャートに従った処理を終了させるようにしても良い。なお、上述の説明では、制御量及び注目度の両方を学習（更新）するようにしているが、その一方のみを学習（更新）する態様であっても構わない。

【0051】

[第 2 の実施形態]

図 1 に示した情報処理装置 100 を構成する各機能部はハードウェアで構成しても良いし、一部をソフトウェア（コンピュータプログラム）で構成しても良い。後者の場合、撮影制御部 300、操作情報取得部 120、画像取得部 110、人検出部 130、制御量推定部 140、学習部 150 をソフトウェアで構成しても良い。このような場合、該ソフトウェアを実行可能なプロセッサを有するコンピュータ装置であれば、情報処理装置 100 に適用可能である。

【0052】

情報処理装置 100 に適用可能なコンピュータ装置のハードウェア構成例について、図 6 のブロック図を用いて説明する。なお、情報処理装置 100 に適用可能なコンピュータ装置のハードウェア構成例は、図 6 に示した構成に限らない。また、情報処理装置 100 は、1 台のコンピュータ装置で構成しても良いし、複数台のコンピュータ装置で構成しても良い。

【0053】

CPU 601 は、RAM 602 や ROM 603 に格納されているコンピュータプログラ

10

20

30

40

50

ムやデータを用いて処理を実行する。これによりCPU601は、コンピュータ装置全体の動作制御を行うと共に、情報処理装置100が行うものとして上述した各処理を実行若しくは制御する。

【0054】

RAM602は、ROM603や外部記憶装置606からロードされたコンピュータプログラムやデータ、I/F（インターフェース）607を介して外部（例えばカメラ200）から受信したデータを格納するためのエリアを有する。更にRAM602は、CPU601が各種の処理を実行する際に用いるワークエリアを有する。このようにRAM602は、各種のエリアを適宜提供することができる。ROM603には、書換不要のコンピュータプログラムや設定データなどが格納されている。

10

【0055】

操作部604は、上記の操作部400に適用可能なユーザインターフェースであり、ユーザが操作することで各種の指示をCPU601に対して入力することができる。表示部605は、上記の表示部500に適用可能な表示装置であり、CPU601による処理結果を画像や文字などでもって表示することができる。なお、操作部604と表示部605とを一体化させてタッチパネル画面を構成しても良い。

【0056】

外部記憶装置606は、ハードディスクドライブ装置に代表される大容量情報記憶装置である。上記の記憶部160は、RAM602や外部記憶装置606によって実装することができる。外部記憶装置606には、OS（オペレーティングシステム）や、情報処理装置100が行うものとして上述した各処理をCPU601に実行若しくは制御させるためのコンピュータプログラムやデータが保存されている。外部記憶装置606に保存されているコンピュータプログラムには、上記のソフトウェアが含まれている。また、外部記憶装置606に保存されているデータには、上記の説明において既知の情報として説明したデータが含まれている。外部記憶装置606に保存されているコンピュータプログラムやデータは、CPU601による制御に従って適宜RAM602にロードされ、CPU601による処理の対象となる。

20

【0057】

I/F607は、情報処理装置100を外部の機器と接続するためのインターフェースとして機能するものであり、例えば、上記のカメラ200を情報処理装置100に接続するためのインターフェースとして機能する。CPU601、RAM602、ROM603、操作部604、表示部605、外部記憶装置606、I/F607は何れもバス608に接続されている。

30

【0058】

このように、上記の実施形態では、検出結果から推定した制御量とユーザ操作との差異が小さくなるように学習を行うので、ユーザの意図を汲んだ撮影制御の学習が可能となる。さらに、損失を制御量の差異で評価すると同時に注目度も評価しており、意図しない人物の追尾や人物が映っていない場合の間違った撮影制御を回避することができる。

【0059】

なお、上記の実施形態では、制御量推定部140はニューラルネットワークを含むものとしたが、人検出部130も同様にニューラルネットワークを含むようにしても良い。このとき、上記の特徴抽出部142を人検出部130と共有することが可能である。また、制御量推定部140をサポートベクター回帰等の他の機械学習による推定部で構成することも可能である。

40

【0060】

また、上記の実施形態では、制御量推定部140は人検出部130の結果と画像から制御量を推定するようにしたが、人検出部130の結果のみを用いても制御量を推定することは可能である。

【0061】

また、上記の実施形態では、制御量推定部140は静止画における人検出部130の結

50

果と画像から制御量を推定するようにしたが、時系列画像の複数フレームの人検出部 1 3 0 の結果を結合した時空間画像から制御量を推定するようにしてもよい。これにより、ユーザが人のどのような動きに注目して操作したかを学習することができる。

【 0 0 6 2 】

また、上記の実施形態では、制御量推定部 1 4 0 は人検出部 1 3 0 が出力する人物領域の四隅の画像座標と尤度とを用いて制御量を推定するようにしたが、この推定に用いる情報は、画像中の人の位置を表す情報であればよい。例えば、人の存在確率を表す尤度を二次元の座標位置に対応させた尤度マップのようなものでもよい。

【 0 0 6 3 】

また、上記の実施形態では、学習部 1 5 0 は、制御量推定部 1 4 0 が推定する画像中の複数の推定結果を別々に学習データとして取得するようにしたが、制御量推定部 1 4 0 の統合部 1 4 4 で一つの推定結果として統合した後に学習データとするようにしてもよい。あるいは、複数の推定結果をRNN (Recurrent Neural Network) やLSTM (Long short-term memory) 等の再帰型のニューラルネットワークを用いて統合して推定するようにしてもよい。この場合、学習部 1 5 0 でその出力を学習データとして取得する。

【 0 0 6 4 】

また、上記の実施形態では、検出対象として人物 (人物領域) を例にとり説明したが、検出対象は人物に限らず、人物以外のオブジェクトを検出対象としても良い。また、図 1 ではカメラの台数を 1 としているが、これに限らず、複数台のカメラを制御対象としても良い。また、上記の実施形態では、制御量は、カメラ 2 0 0 のパン、チルト、ズームの 3 つを含むものとしたが、これに限らず、パン、チルト、ズームうち少なくとも 1 つを含むようにしても良い。なお、上記の様々な変形例の一部若しくは全部を適宜組み合わせても構わない。

【 0 0 6 5 】

(その他の実施例)

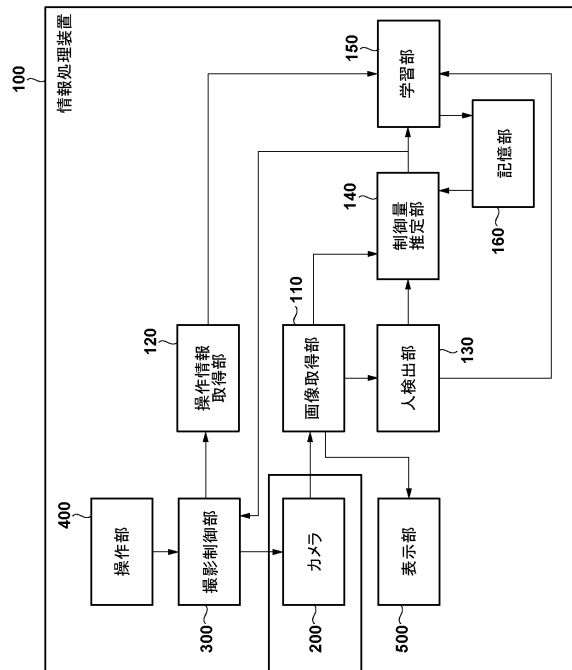
本発明は、上述の実施形態の 1 以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける 1 つ以上のプロセッサがプログラムを読み出し実行する処理でも実現可能である。また、1 以上の機能を実現する回路 (例えば、ASIC) によっても実現可能である。

【符号の説明】

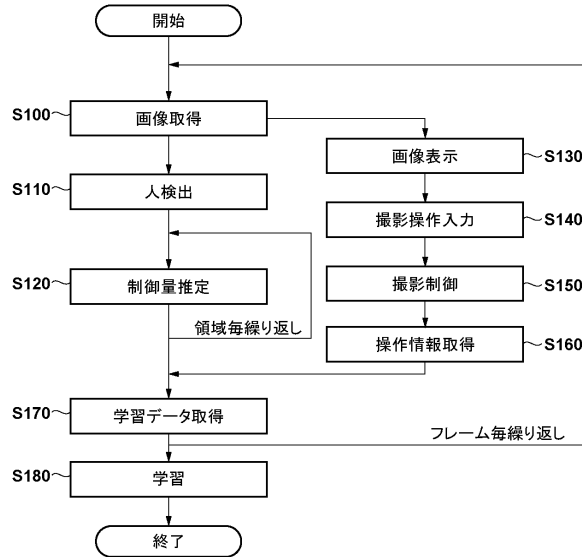
【 0 0 6 6 】

4 0 0 : 操作部 3 0 0 : 撮影制御部 1 2 0 : 操作情報取得部 1 1 0 : 画像取得部
1 3 0 : 人検出部 1 4 0 : 制御量推定部 1 5 0 : 学習部 1 6 0 : 記憶部

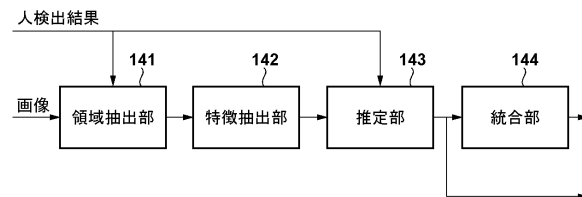
【図 1】



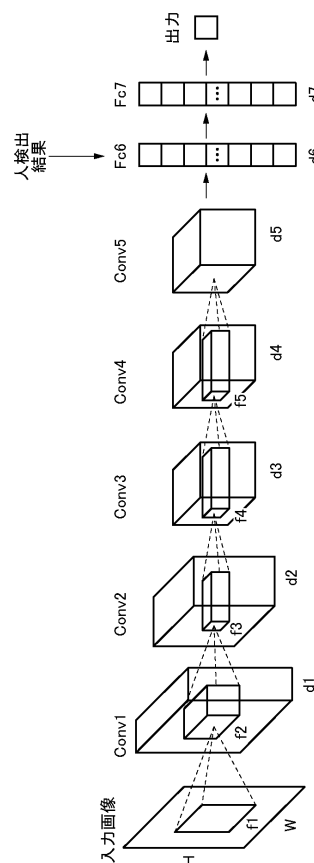
【図 2】



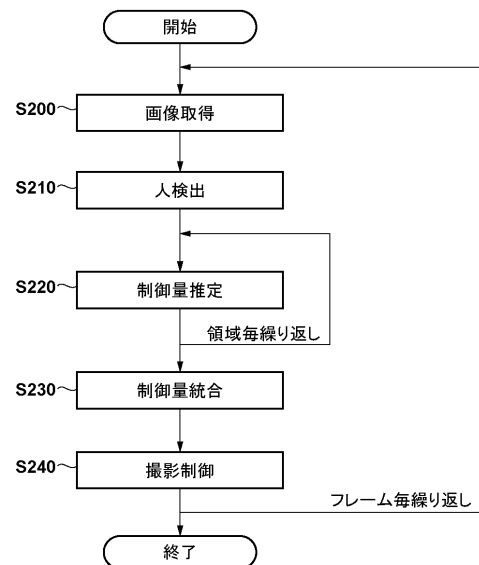
【図 3】



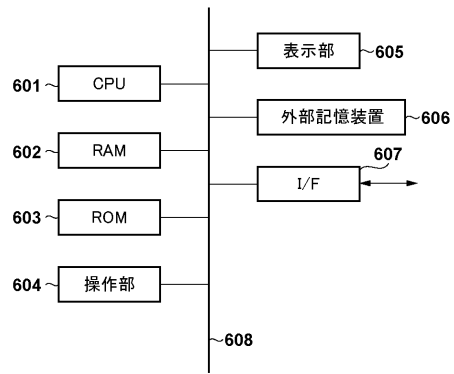
【図 4】



【図 5】



【図 6】



フロントページの続き

(51)Int.Cl. F I
G 0 6 T 7/00 3 5 0 B

(56)参考文献 特開平 0 7 - 0 3 8 7 9 8 (J P , A)
特開 2 0 1 0 - 1 8 3 3 8 4 (J P , A)
特開 2 0 0 1 - 1 2 8 0 3 2 (J P , A)
特開 2 0 1 1 - 1 6 0 0 4 4 (J P , A)
特開 2 0 0 9 - 0 9 4 5 8 5 (J P , A)
特開平 1 1 - 2 5 2 4 5 0 (J P , A)
特開 2 0 1 5 - 1 9 1 3 3 4 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

H 0 4 N 7 / 1 8
H 0 4 N 5 / 2 2 2 - 5 / 2 5 7
G 0 6 T 1 / 0 0 - 7 / 0 0
G 0 8 B 1 3 / 0 0 - 3 1 / 0 0