

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7639931号
(P7639931)

(45)発行日 令和7年3月5日(2025.3.5)

(24)登録日 令和7年2月25日(2025.2.25)

(51)国際特許分類		F I	
G 0 6 T	7/60 (2017.01)	G 0 6 T	7/60 2 0 0 D
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00 3 5 0 B
G 0 6 T	7/55 (2017.01)	G 0 6 T	7/00 6 6 0 B
G 0 6 V	10/82 (2022.01)	G 0 6 T	7/55
		G 0 6 V	10/82

請求項の数 9 (全14頁)

(21)出願番号	特願2023-553831(P2023-553831)	(73)特許権者	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番1号
(86)(22)出願日	令和3年10月13日(2021.10.13)	(74)代理人	110001519 弁理士法人太陽国際特許事務所
(86)国際出願番号	PCT/JP2021/037958	(72)発明者	楊 帆 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
(87)国際公開番号	WO2023/062760	(72)発明者	小田嶋 成幸 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
(87)国際公開日	令和5年4月20日(2023.4.20)	審査官	豊田 好一
審査請求日	令和6年1月11日(2024.1.11)		

最終頁に続く

(54)【発明の名称】 領域検出プログラム、装置、及び方法

(57)【特許請求の範囲】

【請求項1】

人物を各々異なる方向から撮影する複数の撮影装置の各々が撮影した画像を取得し、
取得した画像を、画像に含まれる人物を示す領域を検出するように予め機械学習により生成された機械学習モデルに入力して、前記取得した画像の各々から人物を示す領域を検出し、

前記取得した画像のうち第一の画像から検出された前記人物を示す領域と、前記複数の撮影装置の各々のパラメータとに基づいて、前記取得した画像のうちの第二の画像における前記人物を示す領域を補間する

ことを含む処理をコンピュータに実行させるための領域検出プログラム。

10

【請求項2】

前記第一の画像は、前記取得した画像のうち、前記機械学習モデルにより前記人物を示す領域が検出された画像であり、

前記第二の画像は、前記取得した画像のうち、前記機械学習モデルにより前記人物を示す領域が検出されていない画像である

請求項1に記載の領域検出プログラム。

【請求項3】

前記複数の撮影装置は、同一水平面内に配置されており、

前記第一の画像から検出された前記人物を示す領域の高さと、前記複数の撮影装置の各々のパラメータとに基づいて特定される前記第二の画像における前記人物を示す領域の高

20

さと、前記人物の姿勢に関する統計情報とに基づいて、前記第二の画像における前記人物を示す領域の幅を推定する

請求項 2 に記載の領域検出プログラム。

【請求項 4】

前記第一の画像から検出された前記人物を示す領域の垂直方向の中心線の端点を、前記第一の画像を撮影した撮影装置の各々のパラメータに基づいて、三次元空間における人物の垂直方向の中心線の端点の座標に変換し、変換された三次元空間における座標を、前記第二の画像を撮影した撮影装置のパラメータに基づいて、第二の画像上の座標に変換して、前記第二の画像における前記人物を示す領域の高さを特定する請求項 3 に記載の領域検出プログラム。

10

【請求項 5】

前記人物の姿勢に関する統計情報は、人物が異なる複数の姿勢をとった場合の各々における、前記人物を囲む直方体の高さとの和の平均である請求項 4 に記載の領域検出プログラム。

【請求項 6】

前記三次元空間における人物の垂直方向の中心線の長さを、前記三次元空間における人物の高さとし、前記統計情報が示す平均と、前記三次元空間における人物の高さとの差を、前記三次元空間における人物の幅として推定し、前記三次元空間における人物の高さと幅との比、及び前記第二の画像における前記人物を示す領域の高さに基づいて、前記第二の画像における前記人物を示す領域の幅を推定する請求項 5 に記載の領域検出プログラム。

20

【請求項 7】

前記複数の撮影装置は、同一垂直面内に配置されており、

前記第一の画像から検出された前記人物を示す領域の幅と、前記複数の撮影装置の各々のパラメータとに基づいて特定される前記第二の画像における前記人物を示す領域の幅と、前記人物の姿勢に関する統計情報とに基づいて、前記第二の画像における前記人物を示す領域の高さを推定する

請求項 2 に記載の領域検出プログラム。

【請求項 8】

人物を各々異なる方向から撮影する複数の撮影装置の各々が撮影した画像を取得する取得部と、

30

取得した画像を、画像に含まれる人物を示す領域を検出するように予め機械学習により生成された機械学習モデルに入力して、前記取得した画像の各々から人物を示す領域を検出する検出部と、

前記取得した画像のうち第一の画像から検出された前記人物を示す領域と、前記複数の撮影装置の各々のパラメータとに基づいて、前記取得した画像のうちの第二の画像における前記人物を示す領域を補間する補間部と、

を含む領域検出装置。

【請求項 9】

人物を各々異なる方向から撮影する複数の撮影装置の各々が撮影した画像を取得し、

取得した画像を、画像に含まれる人物を示す領域を検出するように予め機械学習により生成された機械学習モデルに入力して、前記取得した画像の各々から人物を示す領域を検出し、

40

前記取得した画像のうち第一の画像から検出された前記人物を示す領域と、前記複数の撮影装置の各々のパラメータとに基づいて、前記取得した画像のうちの第二の画像における前記人物を示す領域を補間する

ことを含む処理をコンピュータが実行する領域検出方法。

【発明の詳細な説明】

【技術分野】

【0001】

開示の技術は、領域検出プログラム、領域検出装置、及び領域検出方法に関する。

50

【背景技術】

【0002】

従来、人物を撮影した画像から、予め機械学習により生成された認識モデルを用いて、人物の骨格情報等、人物の三次元情報を認識する技術が存在する。このような技術において、認識処理の対象となる人物の領域を囲むバウンディングボックスを検出することが行われている。

【0003】

例えば、複数台のカメラの映像から被写体の3Dモデルを生成する3Dモデル生成装置が提案されている。この装置は、多視点映像から視点毎にシルエット画像を取得し、複数のシルエット画像から視体積交差法によりボクセルサイズが第1サイズの低解像ボクセルモデルを生成する。また、この装置は、低解像ボクセルモデルをその特徴に基づいて分類し、低解像ボクセルモデル毎に、その分類結果に基づいて第1サイズよりも小さい第2サイズを決定する。そして、この装置は、低解像ボクセルモデルの3Dバウンディングボックス毎にボクセルサイズが第2サイズの高解像ボクセルモデルを生成し、高解像ボクセルモデルに基づいて被写体の3DCGモデルを出力する。

10

【0004】

また、複数の視点から撮影した多視点画像から、透視投影を利用して、他の視点における対応する点を推測し、中間視点の画像を生成する技術が存在する。

【先行技術文献】

【特許文献】

20

【0005】

【文献】特開2021-071749号公報

【非特許文献】

【0006】

【文献】Hideo Saito, Makoto Kimura, Satoshi Yaguchi, Naho Inamoto, "View Interpolation of Multiple Cameras Based on Projective Geometry", 2002.

【発明の概要】

【発明が解決しようとする課題】

【0007】

機械学習モデルを用いて、画像から対象の人物を示す領域としてバウンディングボックスを検出する場合に、バウンディングボックスの未検出や誤検出が生じる場合がある。人物の三次元情報を取得するための多視点画像において、いずれかの画像にこのようなバウンディングボックスの未検出や誤検出が生じると、後段の処理において、精度良く三次元情報を取得することができない。

30

【0008】

また、例えば体操選手等のように、様々な姿勢を取り得る人物を対象とした場合、透視投影を単純に適用しただけでは、未検出又は誤検出のバウンディングボックスを適切に補間することができない場合がある。

【0009】

一つの側面として、開示の技術は、多視点画像における未検出又は誤検出のバウンディングボックスを適切に補間することを目的とする。

40

【課題を解決するための手段】

【0010】

一つの態様として、開示の技術は、人物を各々異なる方向から撮影する複数の撮影装置の各々が撮影した画像を取得する。また、開示の技術は、取得した画像を、画像に含まれる人物の領域を検出するように予め機械学習により生成された機械学習モデルに入力して、前記取得した画像の各々から人物を示す領域を検出する。そして、開示の技術は、前記取得した画像のうち第一の画像から検出された前記人物の領域と、前記複数の撮影装置の各々のパラメータとに基づいて、前記取得した画像のうちの第二の画像における前記人物を示す領域を補間する。

50

【発明の効果】

【0011】

一つの側面として、多視点画像における未検出又は誤検出のバウンディングボックスを適切に補間することができる、という効果を有する。

【図面の簡単な説明】

【0012】

【図1】領域検出装置とカメラとの接続を示す概略図である。

【図2】領域検出装置の機能ブロック図である。

【図3】二次元バウンディングボックスを説明するための図である。

【図4】視点による二次元バウンディングボックスの幅の相違を説明するための図である。 10

【図5】三次元の人物中心線の特定を説明するための図である。

【図6】二次元バウンディングボックスの補間を説明するための図である。

【図7】三次元バウンディングボックスの幅及び高さを説明するための図である。

【図8】三次元バウンディングボックスに関する統計情報を説明するための図である。

【図9】領域検出装置として機能するコンピュータの概略構成を示すブロック図である。

【図10】領域検出処理の一例を示すフローチャートである。

【図11】二次元バウンディングボックスの補間の一例を示す図である。

【図12】二次元バウンディングボックスが検出された多視点画像の適用技術の例を説明するための図である。

【発明を実施するための形態】 20

【0013】

以下、図面を参照して、開示の技術に係る実施形態の一例を説明する。以下の実施形態では、対象の人物が体操選手である場合について説明するが、開示の技術は、他の競技の選手や、一般の歩行者等、様々な人物を対象として適用可能である。

【0014】

図1に示すように、本実施形態に係る領域検出装置10は、各々異なる方向からの視点nにおいて体操選手90を撮影する複数のカメラ30nの各々と接続される。図1の例では、 $n = 1, 2, 3$ であり、視点1から撮影するカメラ301、視点2から撮影するカメラ302、及び視点3から撮影するカメラ303が領域検出装置10と接続されている。なお、領域検出装置10に接続されるカメラ30nの数は図1の例に限定されない。 30

【0015】

カメラ30nは、略同一水平面内の異なる位置に、体操選手90が撮影範囲に収まる角度で設置される。すなわち、カメラ30nは、体操選手90を取り囲むように水平方向に並んで配置されている。なお、ここでは、略同一水平面内とは、カメラ30nの床面からの高さが略同一とみなせる場合であり、カメラ30nの床面からの高さの差が所定値以下になる場合をいう。カメラ30nで撮影された映像は順次領域検出装置10へ入力される。なお、各カメラ30nで撮影される映像に含まれる各フレームには時間情報が対応付けられており、この時間情報に基づいて、各カメラ30nで撮影された映像は同期をとることができる。

【0016】 40

図2に示すように、領域検出装置10は、機能的には、取得部12と、検出部14と、補間部16とを含む。また、領域検出装置10の所定の記憶領域には、検出モデル20が記憶される。

【0017】

取得部12は、カメラ30nから領域検出装置10に入力された映像において、時間情報が対応するフレームが示す画像のセットを多視点画像として取得する。以下では、多視点画像に含まれる画像のうち、カメラ30nで撮影された画像を画像40nとする。

【0018】

検出部14は、取得部12で取得された多視点画像に含まれる画像40nを検出モデル20に入力して、画像40nの各々から人物を示す領域として、二次元バウンディングボ 50

ックス（以下、「2D-BB」とも表記する）を検出する。検出モデル20は、様々な姿勢の体操選手が写った画像に対して、正解の2D-BBが与えられた画像を訓練データとして用いて予め生成されている。なお、検出モデル20は、開示の技術の「機械学習モデル」の一例である。検出部14は、例えば図3に示すように、画像40n内における体操選手90を示す領域の外接矩形を2D-BB42nとして検出する。

【0019】

補間部16は、多視点画像のうち第一の画像から検出された2D-BB42nと、カメラ30nの各々のパラメータとに基づいて、多視点画像のうちの第二の画像における2D-BB42nを補間する。本実施形態では、視点kのカメラ30kで撮影された画像40kからは、2D-BB42kが検出され、視点__missのカメラ30__missで撮影された画像40__missからは、2D-BB42__missが検出されていないものとする。この場合、画像40kは第一の画像の一例であり、画像40__missは第二の画像の一例である。

10

【0020】

ここで、上記のように機械学習モデルである検出モデル20を用いる場合、画像40nから2D-BB42nが未検出又は誤検出となる場合がある。これは、画像40n内の体操選手90の一部が障害物で遮蔽されている場合や、対象の体操選手90の姿勢が、検出モデル20の訓練に用いられた画像における体操選手のいずれにも類似しない場合等に生じ得る。多視点画像に含まれる各画像40nから検出された2D-BB42n内の画像と、予め機械学習により生成された骨格情報等の三次元情報を認識する認識モデルとに基づいて、体操選手90の三次元情報を取得することを考える。この場合、2D-BB42nが未検出又は誤検出の画像が多視点画像に含まれると、三次元情報の認識精度が低下してしまう。そのため、未検出又は誤検出のBB42nを補間することが望ましい。

20

【0021】

しかし、体操選手等のように、様々な姿勢を取り得る人物を対象とした場合、透視投影を単純に適用しただけでは、未検出又は誤検出の2D-BB42nを適切に補間することができない場合がある。図4を参照してより具体的に説明する。図4に示すように、水平面及び垂直面からなる直方体であって、三次元空間における体操選手90を完全に囲む最小の直方体を三次元バウンディングボックス（以下、「3D-BB」とも表記する）44として想定する。各カメラ30nの視点nで、3D-BB44を各画像40nに投影したものが2D-BB42nとなる。

30

【0022】

各カメラ30nが体操選手90を取り囲むように水平方向に並んで配置されている場合、各カメラ30nの視点によらず、各画像40nにおける2D-BB42nの高さ（垂直方向の長さ）は共通する。なお、図4において、2D-BB42n及び3D-BB44の高さを、星印を結ぶ線で表している。一方、各画像40nにおける2D-BB42nの幅（水平方向の長さ）は、各カメラ30nの視点によって異なる。図4の例では、画像401から検出された2D-BB421の幅と、画像402から検出された2D-BB422の幅とは、その視点の相違に起因して異なっている。したがって、例えば、画像403において2D-BB423が未検出であったため、従来技術のような透視投影を単純に適用して、検出済みの2D-BB421、422から補間しようとしても、2D-BB423の幅を適切に定めることができない。そのため、画像403における2D-BB423を補間することができない。

40

【0023】

そこで、補間部16は、2D-BBが検出されている画像40kにおける2D-BB42kの高さと、カメラ30kの各々のパラメータとに基づいて、2D-BBが検出されていない画像40__missにおける2D-BB42__missの高さを特定する。そして、補間部16は、体操選手90の姿勢に関する統計情報と、カメラ30__missのパラメータとに基づいて、2D-BBが検出されていない画像40__missにおける2D-BB42__missの幅を推定する。

50

【0024】

図5及び図6を参照して具体的に説明する。補間部16は、画像40kから検出された2D-BB42kを、 $B^k = [x^k, y^k, w^k, h^k]$ とする。 x^k 及び y^k は、2D-BB42kの左上角の点の画像40k内での座標、 w 及び h は、それぞれ2D-BB42kの幅及び高さである。また、補間部16は、画像40__missにおける2D-BB42__missを、 $B^{miss} = [x^{miss}, y^{miss}, w^{miss}, h^{miss}]$ とする。すなわち、 B^{miss} を特定することが、2D-BB42__missを補間することに相当する。なお、図5及び図6の例では、 $k = 1, 2$ であり、__miss = 3である。

【0025】

補間部16は、2D-BB42kの垂直方向の中心線の上部端点の座標 $[x^k + w^k / 2, y^k]$ 、及下部端点の座標 $[x^k + w^k / 2, y^k + h^k]$ を特定する。そして、補間部16は、三次元座標を、カメラ30kで撮影された画像平面上の座標へ変換するためのカメラ30kのパラメータを用いて、上部端点及び下部端点の座標を三次元座標に変換する。例えば、補間部16は、OpenCV (https://docs.opencv.org/3.4/d0/dbd/group__triangulation.html)で定義された`cv::sfm::trianglePoints`を利用して座標を変換してよい。この場合、補間部16は、上部端点及び下部端点の各々に対応する三次元空間での点 P_{3d_top} 及び P_{3d_bot} の三次元座標を、カメラ30kのパラメータ行列 $ProjMat_{cam_k}$ を用いて、下記(1)式及び(2)式に示すように算出する。なお、三次元空間において、 P_{3d_top} と P_{3d_bot} とを結ぶ線を、人物中心線という。

【0026】

$P_{3d_top} = [X_{top}, Y_{top}, Z_{top}]$
 $= cv::sfm::trianglePoints(\{[x^1 + w^1 / 2, y^1], [x^2 + w^2 / 2, y^2], \dots, [x^k + w^k / 2, y^k]\}, \{ProjMat_{cam_1}, ProjMat_{cam_2}, \dots, ProjMat_{cam_k}\})$ (1)

$P_{3d_bot} = [X_{bot}, Y_{bot}, Z_{bot}]$
 $= cv::sfm::trianglePoints(\{[x^1 + w^1 / 2, y^1 + h^1], [x^2 + w^2 / 2, y^2 + h^2], \dots, [x^k + w^k / 2, y^k + h^k]\}, \{ProjMat_{cam_1}, ProjMat_{cam_2}, \dots, ProjMat_{cam_k}\})$ (2)

【0027】

また、補間部16は、図6に示すように、 P_{3d_top} 及び P_{3d_bot} の三次元座標を、カメラ30__missのパラメータに基づいて、画像40__miss上の座標に変換する。これにより、補間部16は、2D-BB42__missの垂直方向の中心線の上部端点及び下部端点の座標を算出する。例えば、補間部16は、カメラ30__missのパラメータ行列 $ProjMat_{cam_miss}$ 、及び三次元座標と画像40のサイズとのスケール比を表すパラメータ s を用いて、下記(3)式及び(4)式に示すように座標変換を行う。

【0028】

$[s \times (x^{miss} + w^{miss} / 2), s \times y^{miss}, s]$
 $= ProjMat_{cam_miss} \times [X_{top}, Y_{top}, Z_{top}, 1]$ (3)

$[s \times (x^{miss} + w^{miss} / 2), s \times (y^{miss} + h^{miss}), s]$
 $= ProjMat_{cam_miss} \times [X_{bot}, Y_{bot}, Z_{bot}, 1]$ (4)

【0029】

補間部16は、上記の座標変換に基づいて、2D-BB42__missの垂直方向の中心線の上部端点及び下部端点の座標を特定するための y^{miss} 、 h^{miss} 、及び $x^{miss} + w^{miss} / 2$ を下記(5)式~(7)式に示すように算出する。

【0030】

$y^{miss} = s \times y^{miss} / s$ (5)

$h^{miss} = s \times (y^{miss} + h^{miss}) / s - y^{miss}$ (6)

10

20

30

40

50

$$x^{miss} + w^{miss} / 2 = s \times (x^{miss} + w^{miss} / 2) / s \quad (7)$$

【0031】

補間部16は、特定した上部端点と下部端点とを結ぶ線を2D-BB42__missの垂直方向の中心線として特定し、その中心線の長さを2D-BB42__missの高さとして特定する。

【0032】

また、補間部16は、特定した2D-BB42__missの高さと、体操選手90の姿勢に関する統計情報とに基づいて、2D-BB42__missの幅を推定する。統計情報は、例えば、体操選手が異なる複数の姿勢をとった場合の各々における、体操選手を囲む3D-BB44の高さ及び幅の和の平均としてよい。以下では、図7に示すように、3D-BB44の高さをHeight__3D、幅のうち大きい方をWidth__max__3D、小さい方をWidth__min__3Dとする。

10

【0033】

統計情報についてより具体的に説明する。まず、図8に示すように、モーションキャプチャ、人手によるアノテーション、公開されているデータ等から取得される、実物大の体操選手の三次元モデルを複数の姿勢(ポーズ)について用意する。図8の例では、Mポーズ分の三次元モデルが用意されている。そして、各三次元モデルについて、3D-BB44を特定し、それぞれHeight__3D、Width__max__3D、及びWidth__min__3Dを算出し、下記(8)式に示す平均Mean__3Dを統計情報として算出しておく。

20

【0034】

$$\text{Mean_3D} = (1/M) \sum_M ((\text{Width_max_3D} + \text{Width_min_3D}) / 2 + \text{Height_3D}) \quad (8)$$

【0035】

補間部16は、上記統計情報Mean__3Dと、既に算出しているZ_{top}及びZ_{bot}とを用いて、下記(9)式に示すように、対象の体操選手90の3D-BB44の幅Width__3Dを算出する。

$$\text{Width_3D} = \text{Mean_3D} - ||Z_{top} - Z_{bot}|| \quad (9)$$

【0036】

3D-BBを画像上に透視投影する場合、投影された2D-BBの高さと幅との比は、投影前の3D-BBの高さと幅との比と同様である。そこで、補間部16は、算出したWidth__3Dと、既に算出しているZ_{top}、Z_{bot}、及びh^{miss}とを用いて、下記(10)式に示すように、2D-BB42__missの幅w^{miss}を算出する。さらに、補間部16は、下記(11)式に示すように、算出したw^{miss}及び(7)式により、x^{miss}を算出する。

30

$$w^{miss} = (\text{Width_3D} / ||Z_{top} - Z_{bot}||) \times h^{miss} \quad (10)$$

$$x^{miss} = x^{miss} + w^{miss} / 2 - w^{miss} / 2 \quad (11)$$

【0037】

補間部16は、(5)式、(6)式、(10)式、及び(11)式により算出される[x^{miss}, y^{miss}, w^{miss}, h^{miss}]で特定される2D-BB42__missを、画像40__missにおいて補間する。そして、補間部16は、補間した2D-BB42__missと、2DBB42kとをまとめて、2D-BB検出済みの多視点画像として出力する。

40

【0038】

領域検出装置10は、例えば図9に示すコンピュータ50で実現されてよい。コンピュータ50は、CPU(Central Processing Unit)51と、一時記憶領域としてのメモリ52と、不揮発性の記憶部53とを備える。また、コンピュータ50は、カメラ30n、入力装置、表示装置等の外部装置が接続される入出力I/F(Interface)54と、記憶媒体59に対するデータの読み込み及び書き込みを制御するR/W(Read/Write)部55とを備える。また、コンピュータ50は、イン

50

ターネット等のネットワークに接続される通信 I / F 5 6 を備える。CPU 5 1、メモリ 5 2、記憶部 5 3、入出力 I / F 5 4、R / W 部 5 5、及び通信 I / F 5 6 は、バス 5 7 を介して互いに接続される。

【0039】

記憶部 5 3 は、HDD (Hard Disk Drive)、SSD (Solid State Drive)、フラッシュメモリ等によって実現されてよい。記憶媒体としての記憶部 5 3 には、コンピュータ 5 0 を、領域検出装置 1 0 として機能させるための領域検出プログラム 6 0 が記憶される。領域検出プログラム 6 0 は、取得プロセス 6 2 と、検出プロセス 6 4 と、補間プロセス 6 6 とを有する。また、記憶部 5 3 は、検出モデル 2 0 を構成する情報が記憶される情報記憶領域 7 0 を有する。

10

【0040】

CPU 5 1 は、領域検出プログラム 6 0 を記憶部 5 3 から読み出してメモリ 5 2 に展開し、領域検出プログラム 6 0 が有するプロセスを順次実行する。CPU 5 1 は、取得プロセス 6 2 を実行することで、図 2 に示す取得部 1 2 として動作する。また、CPU 5 1 は、検出プロセス 6 4 を実行することで、図 2 に示す検出部 1 4 として動作する。また、CPU 5 1 は、補間プロセス 6 6 を実行することで、図 2 に示す補間部 1 6 として動作する。また、CPU 5 1 は、情報記憶領域 7 0 から情報を読み出して、検出モデル 2 0 をメモリ 5 2 に展開する。これにより、領域検出プログラム 6 0 を実行したコンピュータ 5 0 が、領域検出装置 1 0 として機能することになる。なお、プログラムを実行する CPU 5 1 はハードウェアである。

20

【0041】

なお、領域検出プログラム 6 0 により実現される機能は、例えば半導体集積回路、より詳しくは GPU (Graphics Processing Unit) や ASIC (Application Specific Integrated Circuit) 等で実現することも可能である。

【0042】

次に、本実施形態に係る領域検出装置 1 0 の作用について説明する。領域検出装置 1 0 にカメラ 3 0 n から多視点画像が入力されると、領域検出装置 1 0 において、図 1 0 に示す領域検出処理が実行される。なお、領域検出処理は、開示の技術の領域検出方法の一例である。

30

【0043】

ステップ S 1 0 で、取得部 1 2 が、領域検出装置 1 0 に入力された多視点画像を取得する。次に、ステップ S 1 2 で、検出部 1 4 が、取得された多視点画像に含まれる各画像 4 0 n を検出モデル 2 0 に入力して、画像 4 0 n の各々から 2 D - BB 4 2 n を検出する。次に、ステップ S 1 4 で、検出部 1 4 が、多視点画像に含まれる画像 4 0 n のうち、2 D - BB 4 2 n が検出されていない画像 4 0 n が存在する否かを判定する。2 D - BB 4 2 n が検出されていない画像 4 0 n が存在する場合には、ステップ S 1 6 へ移行し、存在しない場合には、ステップ S 2 4 へ移行する。

【0044】

ステップ S 1 6 では、補間部 1 6 が、検出された 2 D - BB 4 2 k の垂直方向の中心線の上端点の座標 $[x^k + w^k / 2, y^k]$ 、及下部端点の座標 $[x^k + w^k / 2, y^k + h^k]$ を特定する。次に、ステップ S 1 8 で、補間部 1 6 が、カメラ 3 0 k (図 1 0 では「OKカメラ」と表記) のパラメータ行列を用いて、上部端点及び下部端点の座標を三次元座標に変換した点 $P3d_{top}$ 及び $P3d_{bot}$ を特定する。そして、補間部 1 6 が、 $P3d_{top}$ と $P3d_{bot}$ とを結ぶ線を人物中心線として特定する。

40

【0045】

次に、ステップ S 2 0 で、補間部 1 6 が、 $P3d_{top}$ 及び $P3d_{bot}$ の三次元座標を、カメラ 3 0 _miss (図 1 0 では「missカメラ」と表記) のパラメータ行列に基づいて、画像 4 0 _miss (図 1 0 では「miss画像」と表記) に投影する。これにより、2 D - BB 4 2 _miss の垂直方向の中心線が特定され、その中心線の長さが 2

50

D - B B 4 2 __ m i s s の高さとして特定される。

【 0 0 4 6 】

次に、ステップ S 2 2 で、補間部 1 6 が、特定した 2 D - B B 4 2 __ m i s s の高さ
、体操選手 9 0 の姿勢に関する統計情報とに基づいて、2 D - B B 4 2 __ m i s s の幅を
推定する。これにより、上記ステップ S 2 0 で特定された 2 D - B B 4 2 __ m i s s の垂
直方向の中心線及び高さ、本ステップで推定された 2 D - B B 4 2 __ m i s s の幅とで
特定される 2 D - B B 4 2 __ m i s s が画像 4 0 __ m i s s において補間される。

【 0 0 4 7 】

次に、ステップ S 2 4 で、補間部 1 6 が、各画像 4 0 n から 2 D - B B 4 2 n が検出さ
れた多視点画像を出力する。検出された 2 D - B B 4 2 n には、上記ステップ S 2 2 で補
間された 2 D - B B が含まれる。次に、ステップ S 2 6 で、取得部 1 2 が、次の多視点画
像が領域検出装置 1 0 に入力されたか否かを判定する。次の多視点画像が領域検出装置 1
0 に入力された場合には、ステップ S 1 0 に戻り、入力されていない場合には、領域検出
処理は終了する。

【 0 0 4 8 】

図 1 1 に、4 つの画像を含む多視点画像において、検出されなかった 2 D - B B を補間
した一例を示す。なお、図 1 1 において、c a m __ i d : n (n = 1 , 2 , 3 , 4) は、
i d が n のカメラで撮影された画像を表す。また、f r a m e : 8 5 2 は、各画像に対応
付けられたフレーム番号であり、上記実施形態の時間情報に相当する。図 1 1 左図に示す
ように、フレーム番号 8 5 2 の多視点画像に含まれる c a m __ i d : 3 において欠落して
いた 2 D - B B が、本実施形態を適用することにより、図 1 1 右図に示すように補間され
る。

【 0 0 4 9 】

以上説明したように、本実施形態に係る領域検出装置は、人物を各々異なる方向から撮
影する複数のカメラの各々が撮影した画像のセットである多視点画像を取得する。また、
領域検出装置は、取得した多視点画像に含まれる画像の各々を、画像に含まれる人物の領
域を示すバウンディングボックスを検出するように予め機械学習により生成された検出モ
デルに入力して、画像の各々からバウンディングボックスを検出する。そして、領域検出
装置は、取得した画像のうち第一の画像から検出されたバウンディングボックスと、複数
のカメラの各々のパラメータとに基づいて、取得した画像のうちの第二の画像におけるバ
ウンディングボックスを補間する。

【 0 0 5 0 】

より具体的には、領域検出装置は、検出された 2 D - B B をカメラの内部及び外部パラ
メータを使用して二次元から三次元へ投影し、三次元空間における垂直方向の人物中心線
を特定する。そして、領域検出装置は、様々な姿勢の体操選手の三次元モデルから事前
に算出しておいた、3 D - B B の高さ及び幅の平均を示す統計情報と、人物中心線の長さで
ある 3 D - B B の高さとに基づいて、3 D - B B の幅を推定する。さらに、領域検出装置
は、人物中心線、3 D - B B の高さ及び幅から特定される 3 D - B B を、カメラの内部及
び外部パラメータを使用して三次元から二次元へ投影することにより、画像内の 2 D - B
B を補間する。これにより、多視点画像における未検出のバウンディングボックスを適切
に補間することができる。

【 0 0 5 1 】

本実施形態に係る領域検出装置から出力される 2 D - B B 検出済みの多視点画像は、例
えば図 1 2 に示すように、体操選手の学習型骨格認識等に利用される。図 1 2 の例では、
2 D - B B が既知である多視点画像と、その多視点画像が表す体操選手の各関節の三次元
座標（以下、「3 D 関節座標」という）の正解とのペアを訓練データとして、骨格認識モ
デルを事前に機械学習により生成しておく。骨格認識モデルは、例えばニューラルネット
ワーク等である。そして、認識対象の体操選手の多視点画像として、本実施形態に係る領
域検出装置から出力された 2 D - B B 検出済みの多視点画像が取得されると、機械学習済
みの骨格認識モデルに多視点画像が入力され、3 D 関節座標が出力される。そして、骨格

10

20

30

40

50

認識モデルから出力された 3D 関節座標を一次骨格認識結果とし、各関節間の長さや位置関係等の制約に基づいて、各関節位置の位置探索を行った 3D 関節座標がフィッティング結果として出力される。

【 0 0 5 2 】

また、上記実施形態では、複数のカメラが略同一水平面に配置されている場合について説明したが、これに限定されない。例えば、選手が行う競技によっては、略同一垂直面内に配置された複数のカメラで撮影される多視点画像の方が、骨格認識等で精度の高い認識結果を得られる場合もある。この場合、第一の画像から検出された 2D - BB の幅と、第一の画像を撮影したカメラのパラメータとに基づいて 3D - BB の幅を特定し、3D - BB の幅と、体操選手の三次元モデルに関する統計情報とに基づいて 3D - BB の高さを推定するようにすればよい。

10

【 0 0 5 3 】

また、上記実施形態では、第一の画像を 2D - BB が検出された画像とし、第二の画像を 2D - BB が検出されなかった画像とする場合、すなわち、2D - BB が欠落している画像において 2D - BB を補間する場合について説明したが、これに限定されない。例えば、多視点画像に含まれる各画像をそれぞれ第一の画像とし、他の画像を第二の画像とし、第一の画像において検出された 2D - BB を、上記実施形態と同様に補間される 2D - BB に基づいて補正するようにしてもよい。また、検出モデルがバウンディングボックスの検出と共に、その検出の信頼度を出力する場合、検出の信頼度が所定値以下の場合に、上記実施形態における 2D - BB が検出されなかった場合と同様に扱うようにしてもよい。

20

【 0 0 5 4 】

また、上記実施形態では、領域検出プログラムが記憶部に予め記憶（インストール）されている態様を説明したが、これに限定されない。開示の技術に係るプログラムは、CD - ROM、DVD - ROM、USBメモリ等の記憶媒体に記憶された形態で提供することも可能である。

【 符号の説明 】

【 0 0 5 5 】

1 0 領域検出装置

1 2 取得部

1 4 検出部

1 6 補間部

2 0 検出モデル

3 0 1、3 0 2、3 0 3 カメラ

4 0 1、4 0 2、4 0 3 画像

4 2 1、4 2 2 2D - BB

5 0 コンピュータ

5 1 CPU

5 2 メモリ

5 3 記憶部

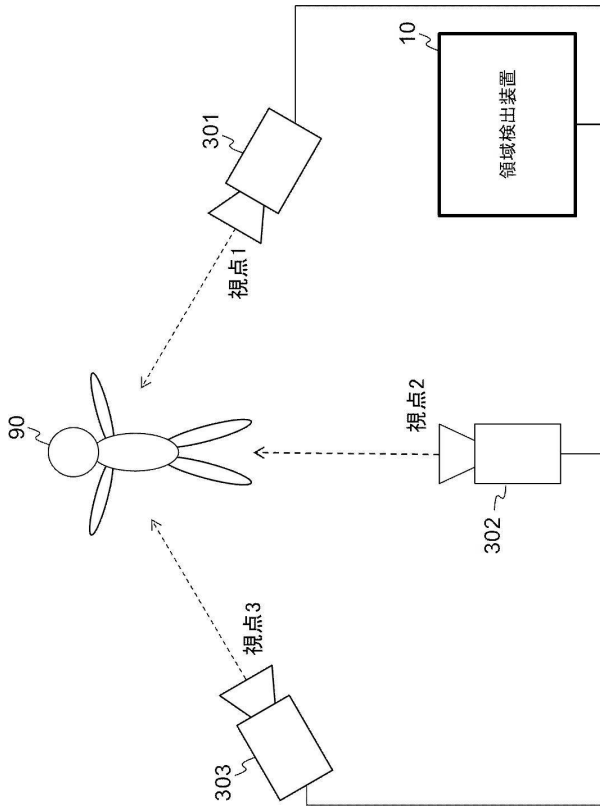
5 9 記憶媒体

6 0 領域検出プログラム

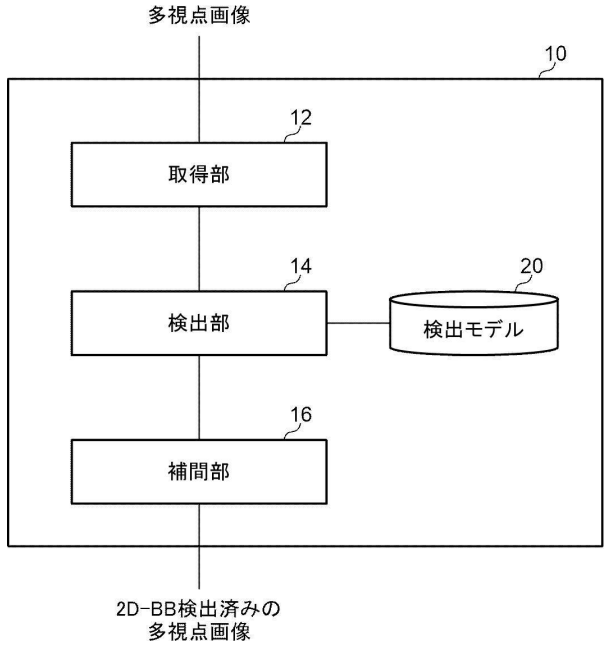
30

40

【図面】
【図 1】



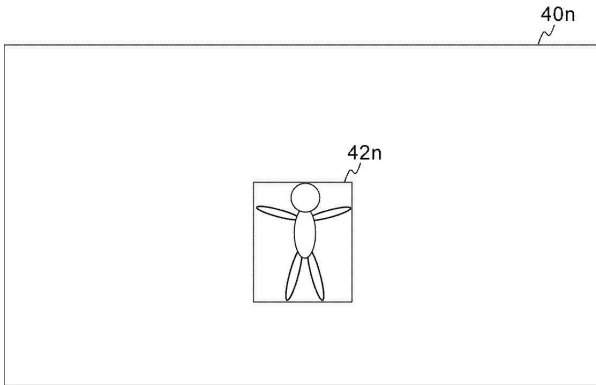
【図 2】



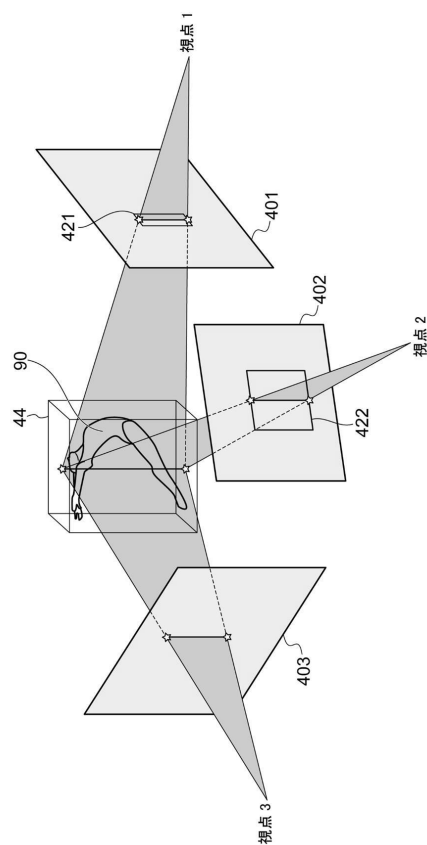
10

20

【図 3】



【図 4】

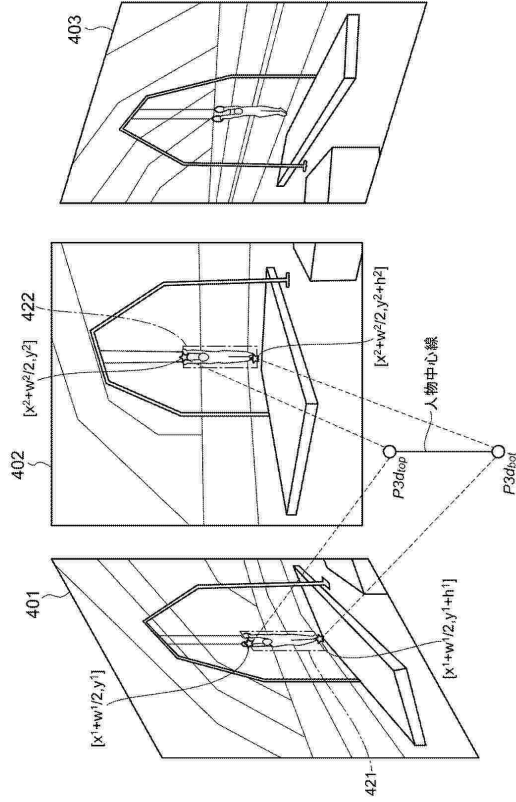


30

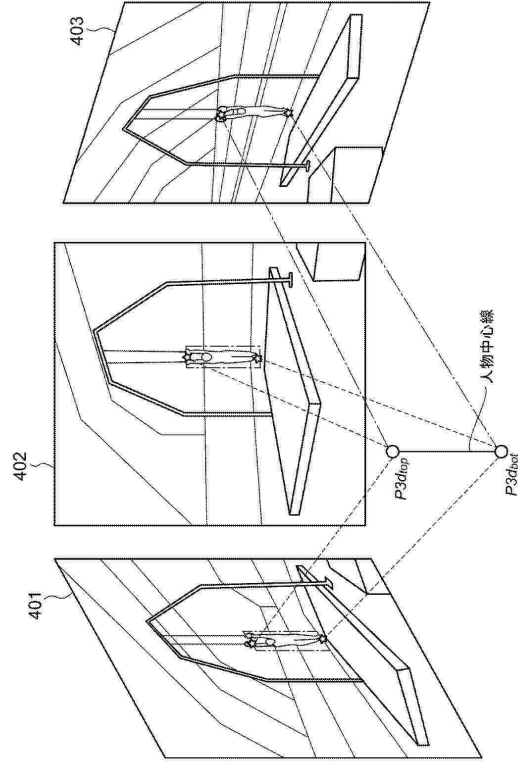
40

50

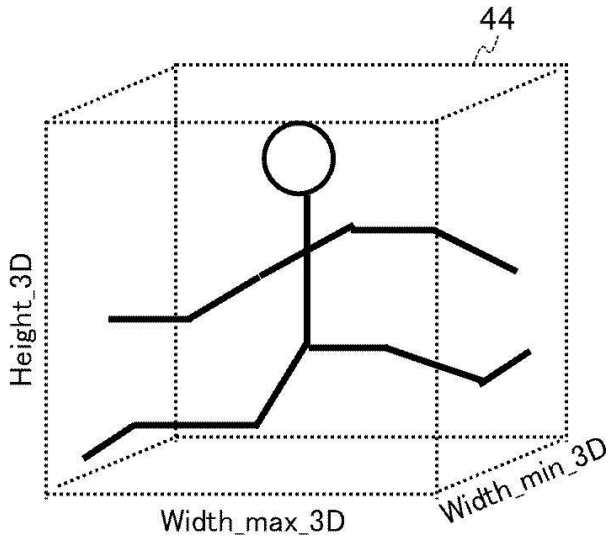
【図 5】



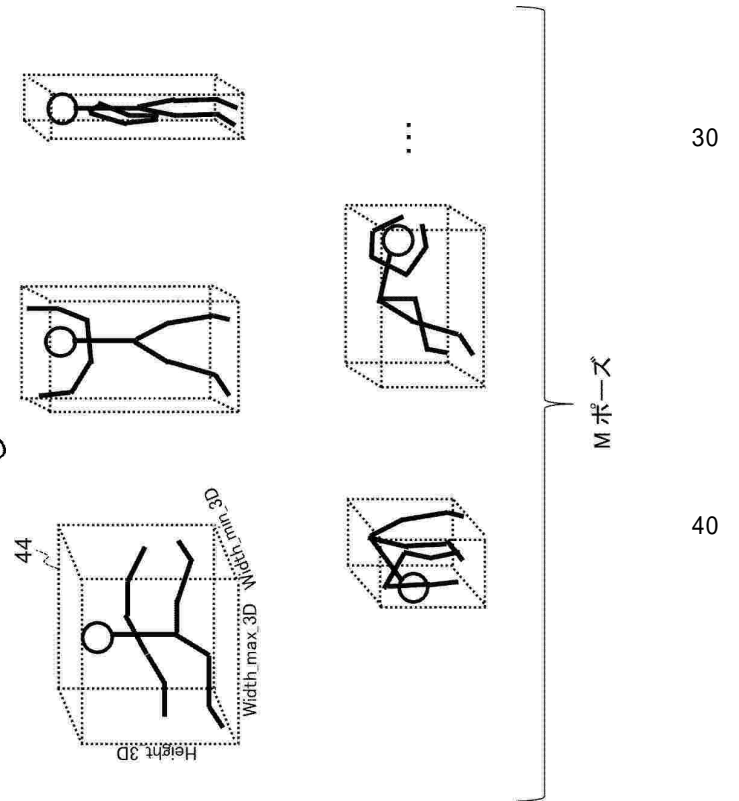
【図 6】



【図 7】



【図 8】



10

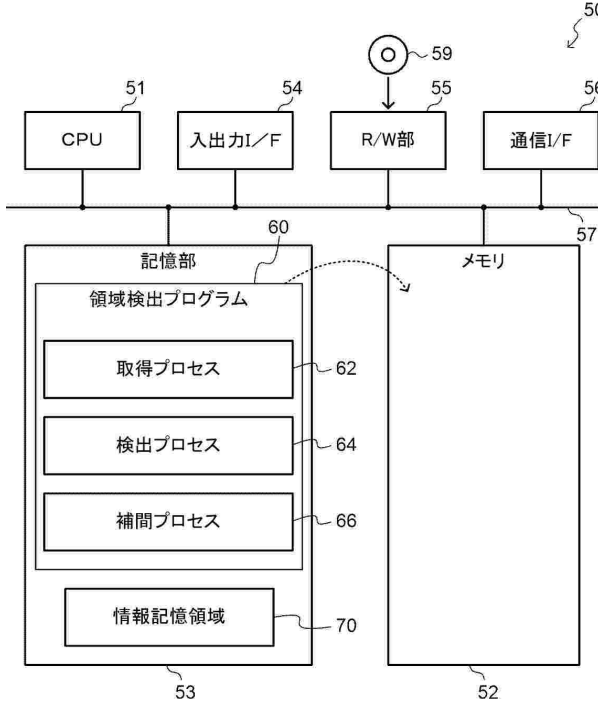
20

30

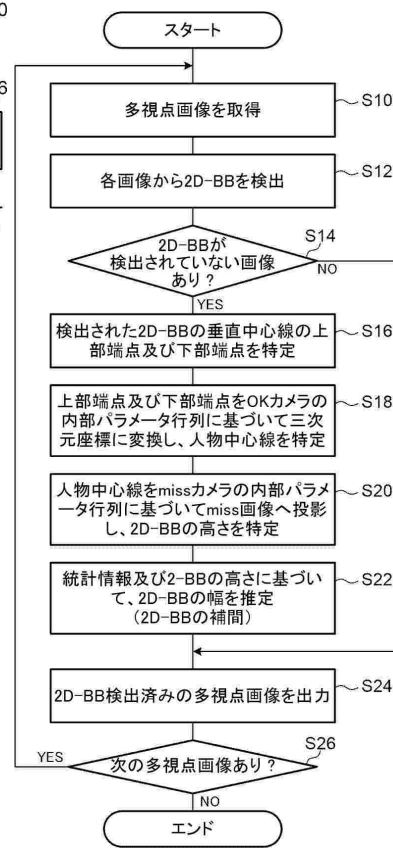
40

50

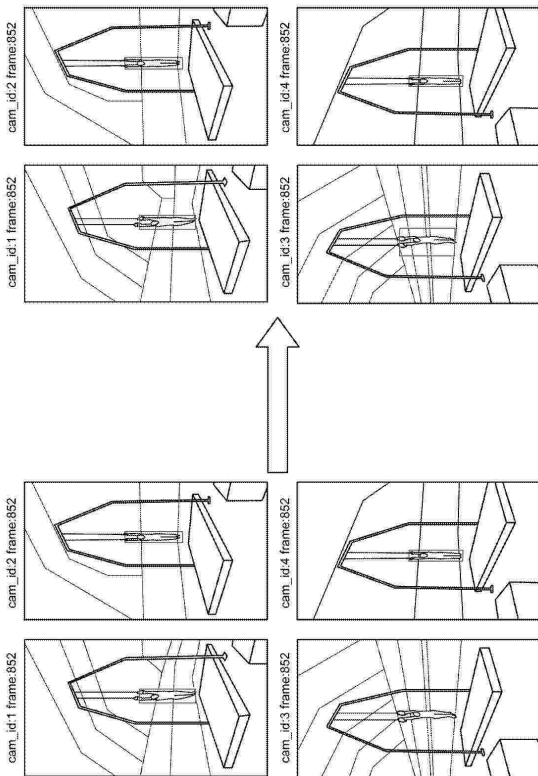
【図9】



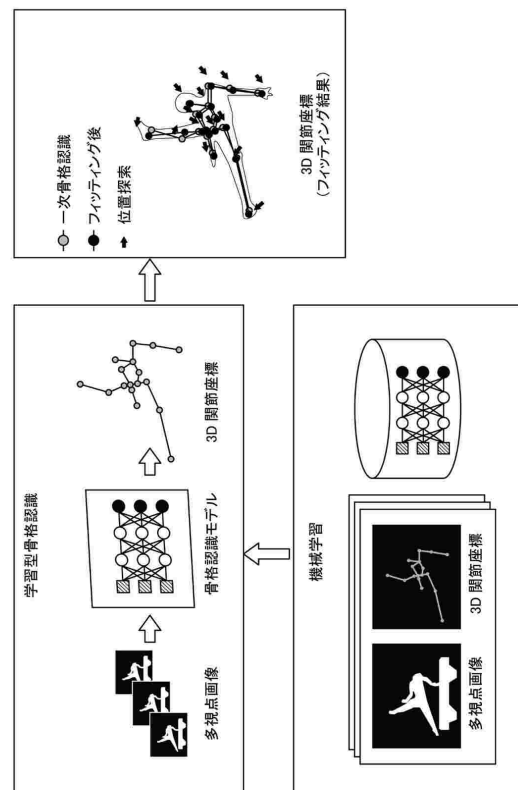
【図10】



【図11】



【図12】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開2009-025874(JP,A)
特開2021-152724(JP,A)
特開2009-143722(JP,A)
特開2002-290962(JP,A)

- (58)調査した分野 (Int.Cl., DB名)
G06T 7/60
G06T 7/00
G06T 7/55
G06V 10/82