(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)公開番号
特開2024-173036
(P2024-173036A)
(43)公開日　令和6年12月12日(2024.12.12)

(51)国際特許分類		FI			テーマコード（参考）
G06T	7/55 (2017.01)	G06T	7/55		5B050
G06V	10/82 (2022.01)	G06V	10/82		5L096
G06T	7/00 (2017.01)	G06T	7/00	350C	
G06V	20/64 (2022.01)	G06V	20/64		
G06T	19/00 (2011.01)	G06T	19/00	A	

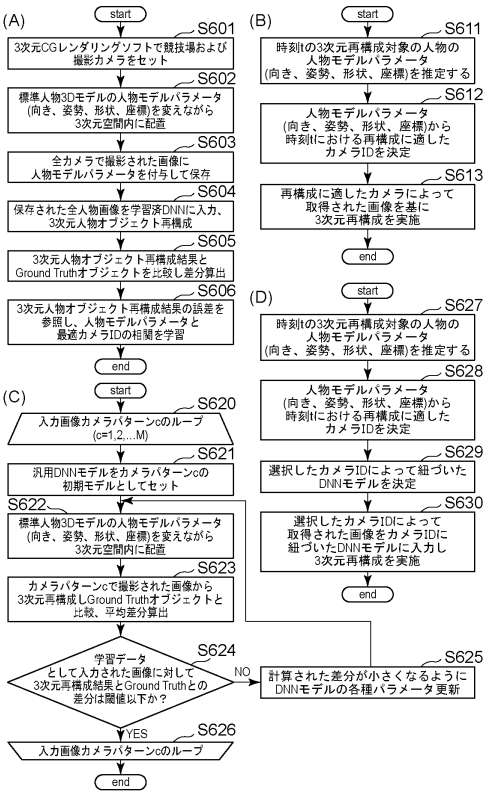
審査請求　未請求　請求項の数　11　ＯＬ　（全20頁）

(21)出願番号　特願2023-91151(P2023-91151)
(22)出願日　令和5年6月1日(2023.6.1)

(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　100223941
弁理士　高橋 佳子
(74)代理人　100159695
弁理士　中辻 七朗
(74)代理人　100172476
弁理士　冨田 一史
(74)代理人　100126974
弁理士　大朋 靖尚
(72)発明者　瀧本 将史
東京都大田区下丸子３丁目３０番２号キ
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】　計算コストを抑えながらも高精度な３次元再構成を実現する。
【解決手段】　情報処理装置は、１以上のカメラによって被写体を撮影し、前記撮影された画像によって前記被写体を３次元再構成する情報処理装置であって、前記被写体の状態を基準に３次元再構成に用いる１以上のカメラを決定する決定手段と、前記決定したカメラで撮影された画像を入力とし、３次元形状を推定する推定器によって３次元形状データを生成する生成手段と、を有する。
【選択図】　図６



【特許請求の範囲】
【請求項１】
１以上のカメラによって被写体を撮影し、前記撮影された画像によって前記被写体を３次元再構成する情報処理装置であって、
前記被写体の状態を基準に３次元再構成に用いる１以上のカメラを決定する決定手段と、
前記決定したカメラで撮影された画像を入力とし、３次元形状を推定する推定器によって３次元形状データを生成する生成手段と、
を有することを特徴とする情報処理装置。
【請求項２】
前記決定手段は、前記被写体の正解データと３次元形状を推定する推定器によって推定した形状の差分と、前記被写体の状態を表す低次パラメータとを基に事前に取得した関数によって決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】
前記被写体の状態を表す低次パラメータは３次元再構成に利用する推定器とは異なる手段によって推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】
前記被写体の状態を表す低次パラメータは、人物の姿勢、形状、方向を表すパラメータであることを特徴とする請求項１から３のいずれか１項に記載の情報処理装置。
【請求項５】
前記決定手段は、前記被写体の状態を表す低次パラメータの他に３次元形状データが生成された後のビデオレンダリングに用いる仮想カメラの状態を表すパラメータを参照して決定することを特徴とする請求項１から４のいずれか１項に記載の情報処理装置。
【請求項６】
前記決定手段は、前記被写体の状態を表す低次パラメータおよび、仮想カメラの状態を表すパラメータに加えて仮想カメラからの視点で仮想カメラから被写体までに遮蔽が発生するかどうかを参照して決定することを特徴とする請求項１から５のいずれか１項に記載の情報処理装置。
【請求項７】
前記決定手段によって選択されたカメラによって３次元再構成を実施する際、各カメラによって撮影される画像の傾向に応じて３次元再構成に利用される推定器が異なるように学習しておくことを特徴とする請求項１から６のいずれか１項に記載の情報処理装置。
【請求項８】
前記決定手段によって選択されたカメラによって３次元再構成を実施する際、時系列に推定結果が安定するために過去に推定した推定時の中間特徴を利用して推定することを特徴とする請求項１から７のいずれか１項に記載の情報処理装置。
【請求項９】
前記決定したカメラで撮影された画像とは２次元画像であることを特徴とする請求項１から８のいずれか１項に記載の情報処理装置。
【請求項１０】
前記決定したカメラで撮影された画像とはデプスマップによって推定された３次元点群であることを特徴とする請求項１から８のいずれか１項に記載の情報処理装置。
【請求項１１】
１以上のカメラによって被写体を撮影し、前記撮影された画像によって前記被写体を３次元再構成する情報処理方法であって、
前記被写体の状態を基準に３次元再構成に用いる１以上のカメラを決定する工程と、
前記決定したカメラで撮影された画像を入力とし、３次元形状を推定する推定器によって３次元形状データを生成する生成工程と、
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、撮影画像に基づく３次元再構成のための情報処理装置、情報処理方法、及びプログラムに関する。
【背景技術】
【０００２】
野球やラグビーなどの試合のテレビ中継時において、瞬時に選手のプレイ内容を３次元再構成することにより通常は撮影することのできない視点から観た映像を視聴者に届ける技術が注目を集めている（以降、仮想視点映像生成技術と呼称する）。
【０００３】
加えて、ディープラーニング技術を用いて３次元再構成を実施する技術がある。これらディープラーニングを利用することにより従来、実現不可能であったレベルの３次元再構成結果を取得することが期待されている。
【０００４】
非特許文献１では、ＳＭＰＬを利用することにより入力となる２次元画像に映った人物領域を検出した後、人物姿勢や形状パラメータをディープラーニングによって推定することで目標人物の３次元再構成を行うことが開示されている。
【０００５】
非特許文献２では、ＮｅＲＦと呼ばれる仮想視点映像生成技術が開示されている。ディープネットワークを利用してシーンを表すこととしてスパースなセットの複数のカメラ入力画像から３次元再構成しさらに仮想カメラから見た２次元画像へとレンダリングすることが開示されている。
【先行技術文献】
【非特許文献】
【０００６】
【非特許文献１】Ｍ．Ｌｏｐｅｒ，Ｎ．Ｍａｈｍｏｏｄ，Ｊ．Ｒｏｍｅｒｏ，Ｇ．Ｐｏｎｓ－Ｍｏｌｌ，ａｎｄ Ｍ．Ｊ．Ｂｌａｃｋ，“ＳＭＰＬ：Ａ ｓｋｉｎｎｅｄ ｍｕｌｔｉ－ｐｅｒｓｏｎ ｌｉｎｅａｒ ｍｏｄｅｌ，”ＡＣＭ ＴＯＧ，ｖｏｌ．３４，ｎｏ．６，ｐ．２４８，２０１５．
【非特許文献２】Ｂｅｎ Ｍｉｌｄｅｎｈａｌｌ，Ｐｒａｔｕｌ Ｐ Ｓｒｉｎｉｖａｓａｎ，Ｍａｔｔｈｅｗ Ｔａｎｃｉｋ，Ｊｏｎａｔｈａｎ Ｔ Ｂａｒｒｏｎ，Ｒａｖｉ Ｒａｍａｍｏｏｒｔｈｉ，ａｎｄ Ｒｅｎ Ｎｇ．Ｎｅｒｆ：Ｒｅｐｒｅｓｅｎｔｉｎｇ ｓｃｅｎｅｓ ａｓ ｎｅｕｒａｌ ｒａｄｉａｎｃｅ ｆｉｅｌｄｓ ｆｏｒ ｖｉｅｗ ｓｙｎｔｈｅｓｉｓ．Ｉｎ Ｅｕｒｏｐｅａｎ ｃｏｎｆｅｒｅｎｃｅ ｏｎ ｃｏｍｐｕｔｅｒ ｖｉｓｉｏｎ，ｐａｇｅｓ ４０５-４２１．Ｓｐｒｉｎｇｅｒ，２０２０．
【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、非特許文献１および２に記載の方法を、例えばＶｉｓｕａｌ ｈｕｌｌ等の従来の幾何的な方式で実施される仮想視点映像生成技術に単に適用する場合、計算コストが莫大になってしまう。
【０００８】
そこで、本発明では、計算コストを抑えながらも高精度な３次元再構成を実現することを目的とする。
【課題を解決するための手段】
【０００９】
本発明の情報処理装置は、１以上のカメラによって被写体を撮影し、前記撮影された画像によって前記被写体を３次元再構成する情報処理装置であって、前記被写体の状態を基準に３次元再構成に用いる１以上のカメラを決定する決定手段と、前記決定したカメラで撮影された画像を入力とし、３次元形状を推定する推定器によって３次元形状データを生

成する生成手段と、を有する。
【発明の効果】
【００１０】
本発明によれば、計算コストを抑えながらも高精度な３次元再構成を実現することができる。
【図面の簡単な説明】
【００１１】
【図１】第１の実施形態に係るシステム構成例を示す図である。
【図２】第１の実施形態に係るカメラ配置の一例を示す図である。
【図３】第１の実施形態に係るＤＮＮの模式図の一例である。
【図４】第１の実施形態に係る仮想カメラ位置の一例を示す図である。
【図５】第１の実施形態に係る人物モデルの一例を示す図である。
【図６】第１の実施形態に係る学習方法および３次元再構成のフローチャートの一例である。
【図７】第５の実施形態に係る３次元再構成のフローチャートの一例である。
【図８】第６の実施形態に係る３次元再構成のフローチャートの一例である。
【図９】第６の実施形態に係るレンダリング結果の一例を示す図である。
【図１０】第７の実施形態に係る人物の推定結果の一例を示す図である。
【発明を実施するための形態】
【００１２】
以下、本発明に係る実施形態を、図面を参照しながら説明する。以下の各実施形態は本発明を限定するものではなく、また、本実施形態で説明されている特徴の組み合わせの全てが本発明の解決手段に必須のものとは限らない。実施形態の構成は、本発明が適用される装置の仕様や各種条件（使用条件、使用環境等）によって適宜修正又は変更され得る。また、以下の実施形態において、同一もしくは同様の構成には同じ参照符号を付し、重複する説明は省略する。
【００１３】
＜第１の実施形態＞
本実施形態では、スポーツ競技場の周囲をＮ台のカメラで撮影し、被写体である競技中の選手の試合の様子を３次元再構成する、仮想視点映像生成技術のカメラ設定にて実施するものとする。３次元再構成は以下で説明する１枚の２次元画像から３次元形状を推定するＤＮＮ（ディープニューラルネットワーク）にて実施するものとする。ここで、ＤＮＮは推定器の一例である。
【００１４】
本実施形態で運用されるカメラ配置の例を図２に示した。図２は競技場２０の周辺にカメラ２１を複数台配置されていることを示す図である。複数のカメラ２１は、競技フィールドを同期して撮影する。このように配置および撮影することによって、競技フィールド内で実施される運動選手によるプレイ内容を、３次元再構成を行うことができる。また、この際の前記多数配置されたカメラ２１を含むシステム構成例を図１（Ａ）に示した。図１（Ａ）に示すように、本実施形態に係るシステムは、情報処理装置１０、キャプチャグループ１１、キャプチャグループ１２…キャプチャグループ１３およびクロック生成器１４を有する。
【００１５】
情報処理装置１０はＣＰＵ１０１、ＲＡＭ１０２、ＲＯＭ１０３、大容量記憶装置１０４、操作部１０５、表示部１０６から成る。
【００１６】
ＣＰＵ１０１は、ＲＡＭ１０２やＲＯＭ１０３に格納されているコンピュータプログラムやデータを用いて各種の処理を実行する。これによりＣＰＵ１０１は、情報処理装置１０全体の動作制御を行うものとして説明する各種の処理を実行もしくは制御する。
【００１７】

ＲＡＭ１０２は、ＲＯＭ１０３や大容量記憶装置１０４からロードされたコンピュータプログラムやデータを格納するためのエリア、上記多数のキャプチャグループから受信したデータを格納するためのエリア、を有する。さらにＲＡＭ１０２は、ＣＰＵ１０１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ＲＡＭ１０２は、各種のエリアを適宜提供することができる。

【００１８】

ＲＯＭ１０３には、情報処理装置１０の設定データ、起動に係るコンピュータプログラムやデータ、基本動作に係るコンピュータプログラムやデータ、などが格納されている。

【００１９】

大容量記憶装置１０４は、ハードディスクドライブ装置などである。大容量記憶装置１０４には、ＯＳ（オペレーティングシステム）や、情報処理装置１０が行うものとして説明する各種の処理をＣＰＵ１０１に実行も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。大容量記憶装置１０４に保存されているデータには、３次元再構成を実施するＤＮＮモデルに係るデータも含まれている。

【００２０】

大容量記憶装置１０４に保存されているコンピュータプログラムやデータは、ＣＰＵ１０１による制御に従って適宜ＲＡＭ１０２にロードされ、ＣＰＵ１０１による処理対象となる。

【００２１】

操作部１０５は、キーボード、マウス、タッチパネルなどの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ＣＰＵ１０１に対して入力することができる。

【００２２】

表示部１０６は、液晶画面やタッチパネル画面などの画面を有し、ＣＰＵ１０１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、表示部１０６は、画像や文字を投影するプロジェクタなどの投影装置であっても良い。

【００２３】

ＣＰＵ１０１、ＲＡＭ１０２、ＲＯＭ１０３、大容量記憶装置１０４、操作部１０５、表示部１０６、は何れもシステムバス１０７に接続されている。なお、情報処理装置１０の構成は図１（Ａ）に示した構成に限らない。

【００２４】

情報処理装置１０は、入出力機器一式を含むＰＣ（パーソナルコンピュータ）、スマートフォン、タブレット端末装置、などのコンピュータ装置である。

【００２５】

次に、上記キャプチャグループ１１、キャプチャグループ１２、キャプチャグループ１３について説明する。本実施形態ではキャプチャグループは図２の２１で示したカメラ台数分用意する。なお、図１（Ａ）（Ｂ）では一例として３つのキャプチャグループのみを示す。ただし、複数のカメラを一つのキャプチャグループに含めても良い。

【００２６】

キャプチャグループ１１は、コントロールボックス１１１、およびカメラ１１２を有する。キャプチャグループ１２は、コントロールボックス１２１、およびカメラ１２２を有する。キャプチャグループ１３は、コントロールボックス１３１、およびカメラ１３２を有する。そして、クロック生成器１４が発行した時刻を上記コントロールボックス１１１、１２１、１３１が受信してカメラ１１２、１２２、１３２で撮影された画像に時刻を与える。付与された時刻を参照することによって情報処理装置１０が各々の画像を受信した後に同期を取りながら３次元再構成を実施することが可能となる。

【００２７】

尚、図１（Ａ）の変形パターンとして図１（Ｂ）に示した構成でも良い。図１（Ａ）との違いはキャプチャグループ１１、１２…１３内にノードＰＣ１１０、１２０…１３０を持つ構成である。特に本実施形態のように入力画像１枚によって３次元再構成を実施する

場合は、カメラ１台につき３次元再構成を各キャプチャグループで実施し、情報処理装置１０側でレンダリングの際の仮想カメラ位置を決める。そして、その際に採用する３次元再構成結果を最適なキャプチャグループを選択して受信する、としても良い。

【００２８】

次に本実施形態で利用する１枚の入力画像から３次元再構成を実行するＤＮＮの構成の一例を図３（Ａ）に示した。図３（Ａ）に示した一例は入力３０がカメラから撮影された１枚の２次元画像であり、出力３２が３次元再構成されたオブジェクト（３次元形状データ）を表している。入力３０は撮影画像そのものでも良いし、入力画像に対してＭａｓｋ－ＲＣＮＮ等の手法によって意味的領域分割を実施することにより人物領域のみを切り出し背景を除去した画像でも良い。

【００２９】

この二者間の変換に用いるＤＮＮ３１は、アーキテクチャとしてダウンサンプリングネットワークとアップサンプリングネットワークが対となったＨｏｕｒｇｌａｓｓ ｎｅｔｗｏｒｋをイメージして表している。Ｈｏｕｒｇｌａｓｓ ｎｅｔｗｏｒｋは、入力画像から重要な構造を低次表現した後に３次元再構成のための高い表現レベルに変換するためのアーキテクチャの一例である。なお、他のアーキテクチャを利用してもよい。該ＤＮＮ３１が前記入力３０の２次元画像を受取り、３次元オブジェクトのボリュームデータを推定し、該ボリュームデータの内／外判定を連続的な確率場として予測することで推定対象のメッシュ表現を取得する。

【００３０】

Ｇｒｏｕｎｄ Ｔｒｕｔｈとして対象形状を与える場合は、数式１に示すように対象オブジェクトの内部を１、外部を０と定義し連続する３Ｄ Ｏｃｃｕｐａｎｃｙ Ｆｉｅｌｄを定めることで該オブジェクトの座標を学習する。この際、出力する対象オブジェクトのサーフェスは上記内／外判定値が０．５となる面とすることで目的を達成する。

【００３１】

【数１】

$$f_s^*(X_w)=\begin{cases}1, & \text{if } X_w \text{ is inside the target object}\\ 0, & \text{otherwise}\end{cases}$$

【００３２】

尚、ここでX_wは世界座標を表す３次元ベクトル、f_s^*はＧｒｏｕｎｄ Ｔｒｕｔｈを基に０／１に変換する関数とする（対して推定時はf_sで表し連続値を推定する関数とする）。

【００３３】

そして、学習における損失L_sは数式２に示すような式により計算する。数式２では基本的に入力とする２次元画像と対応する正解となる３Ｄメッシュデータを対応付け、入力画像から推定された３Ｄボリュームデータの内外判定による上記スコアと正解３Ｄメッシュデータの内外判定によるスコアの２乗誤差の和としている。この正解となる３Ｄメッシュデータは正解データの一例である。

【００３４】

【数２】

$$L_s(c)=\frac{1}{P}\sum_{p=1}^{P}\left|f_s\left(g\left(I(x_{ci})\right),h_c(X_{wi})\right)-f_s^*(X_{wi})\right|^2$$

【００３５】

数式２では内外判定で用いる座標に係る点の数はＰとし、これは推定対象の空間を占めるボクセル全てまたはそこからサンプリングされた数で良い。ｉは観測点のＩＤを表す。

x_{ci}はカメラＩＤをｃとするカメラによる入力３０の２次元画像上の座標を表し、かつ世界座標上の３次元点X_{wi}を入力画像面へ射影した点を表す。ｇ（Ｉ（x_{ci}））は、カメラＩＤをｃとするカメラにより撮影された入力３０の２次元画像から算出された特徴量を表している。本実施形態においてカメラＩＤのｃは、カメラごとに割り振られる一意の番号（１以上カメラ台数以下の数）である。

【００３６】

また、h_c（X_{wi}）は世界座標上の３次元点X_{wi}を入力３０の２次元画像を撮影したカメラＩＤがｃのカメラ自体で定義されるカメラ座標系に変換する変換式を表す。よって算出された２乗誤差は撮影カメラｃによって定義される値として計算される。

【００３７】

以上、数式２で計算される損失を最小化するように学習することによって２次元の画像入力から３次元のオブジェクトを推定するＤＮＮが学習される。

【００３８】

該学習されたＤＮＮモデルは、図２に示されたカメラ一つについて対応する一つのモデルを保持する。全てのカメラで同一のＤＮＮモデルが共有されていたとしても、全く同一の瞬間に撮影された画像を入力した場合、当然のことながら３次元再構成の結果が全く同じになる可能性は低い。推定対象が人物である場合は特に形状、姿勢および向きが随時変わるため都度変動する推定対象の状態に合わせて最適な入力を決定することが必要となる。以降では学習済ＤＮＮによる３次元再構成の精度が良くなるように最適な入力画像を随時選択する方法について説明する。

【００３９】

まず、簡単のため推定対象を人物としたとき、該人物が１名のみの場合を例として説明する。このとき、図２で説明した競技場内に３次元再構成対象の人物が位置しているものとする。該人物を図４（Ａ）に模式的に４４で表した位置に存在するものとする。競技場内の人物の座標を４５に示したｘ軸と４６に示したｙ軸のｘ、ｙ座標で定義する。そして３次元再構成の後にビデオレンダリングする際の仮想カメラ（ＶＣ）の移動パスを４７と決定する。そして、仮想カメラ（ＶＣ）位置は前記パス上を時系列に４１、４２、４３の順に移動しながら対象人物４４の中心方向を向きながら撮影した画角の映像を出力するものとする。

【００４０】

図４（Ａ）では図示の簡易さを優先するため人物の位置を４４のみで表しているが該人物は自由に競技場内を動いているものとする。

【００４１】

今、当然のことながら仮想カメラ（ＶＣ）４１、４２、４３の位置にカメラは実在しないため各々の仮想カメラ（ＶＣ）位置から見えるはずの映像を生成するために最適なカメラを競技場周辺に設置されているカメラから選択する。

【００４２】

リアルタイムに３次元再構成を行うために最適なカメラを選択するためには選択するための判別式を事前に学習しておく必要がある。学習に関するステップを図６（Ａ）のフローチャートを用いて説明する。

【００４３】

競技場では再構成対象の空間は限定されている。図４（Ａ）の例であれば競技場内部のみである。そして撮影に用いるカメラも競技場周囲に設置されたカメラ台数のみに限定される。設置されたカメラ台数をＭ台（Ｍは２以上の整数）とする。

【００４４】

そこで図６（Ａ）のＳ６０１にて実環境と同様の競技場と撮影カメラを３次元ＣＧレンダリングソフトで再現する。この時、各カメラの内部パラメータ、外部パラメータ、歪み係数も実環境のパラメータとできる限り一致させ、実環境で撮影される画像が該仮想環境においても再現されるようにしておく。次にＳ６０２において３次元人物モデルを１つＣＧで再現された競技場内部に多様なパラメータを変動させて配置する。ここでいう多様な

パラメータとは人物の向きρ、姿勢θ、形状β、座標（ｘ、ｙ）のことである。人物の向きρとは人物の正面向きで定義される角度のことであり、正面向きとは３次元人物モデルのＣａｎｏｎｉｃａｌ　Ｔ－ｐｏｓｅと呼称されることの多い標準ポーズに変換した際の人物正面方向の向きで定義する。例えば、図５（Ａ）の状態の人物モデルの正面向きはこの人物モデルを図５（Ｂ）のＣａｎｏｎｉｃａｌ　Ｔ－ｐｏｓｅに逆変換した際の正面方向５０の向きで定義することができる。この際の正面向きと図４（Ａ）の４５で定義したｘ軸向きの間の角度を前記人物の向きとし値域は０≦ρ＜２πで定義される。次に姿勢θ、形状βで定義されるパラメータは人物の姿勢を定義するパラメータおよび人物形状を定義するパラメータであれば何でも良い。最適な実施形態としては人物形状としてベースとなるテンプレートメッシュを保持しておく。該テンプレートメッシュを上記βによって多様な人物の体形に合うように変形させ、さらに多様な姿勢に合うようにθによって変形させる。

【００４５】

図５（Ａ）の入力に対して、簡易なテンプレートメッシュをβとθによって変形させた結果の例を図５（Ｃ）に示した。テンプレートメッシュの変形では概ね形状を近似できれば良い。ここで定義したβは人物形状の各部位を太くしたり細くしたり、長くしたり短くしたりするパラメータであるため、１０次元程度のパラメータを想定している。また姿勢を定義するθについては想定する３次元人物モデルの関節数と自由度に依存するが１００次元未満程度を想定している。

【００４６】

以上、３次元人物モデルをテンプレートメッシュの変形によって表現するパラメータとしてρ、θ、βの説明を行った。該パラメータは前記テンプレートメッシュを多様に変形させた結果、３次元再構成に最適なカメラを決定するためのものであるため、最終的に求める３次元再構成結果となる３次元オブジェクトの詳細レベルと比較すると低次表現であり簡易なものになっている。ここで、ρ、θ、βは低次パラメータの一例である。ρ、θ、β以外で重要なパラメータは座標（ｘ、ｙ）である。これは既に説明した図４（Ａ）のｘ軸４５およびｙ軸４６で定義される２次元座標で表される（ｘ、ｙ）の２次元のパラメータことである。

【００４７】

よって、競技場に設置されたカメラによって撮影される人物は上記説明した人物の向きρ、姿勢θ、形状β、座標（ｘ、ｙ）で全て定義することが可能となる。

【００４８】

次にＳ６０３にて上記説明した人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）を適切なステップで変化させながら配置した３次元人物モデルを全カメラで撮影した際の２次元画像をレンダリングする。そして、各々の画像に撮影カメラのＩＤ（１≦ｃ≦Ｍ）および上記３次元人物モデルパラメータを付与して保存する。

【００４９】

Ｓ６０４では前記保存された全２次元画像を図３（Ａ）に例示した学習済ＤＮＮに入力することによって３次元再構成を実施する。これによって、推定された３次元人物モデル（３次元形状データ）と入力した２次元画像に付与されている３次元人物モデルパラメータから再現可能なＧｒｏｕｎｄ　Ｔｒｕｔｈとなる３次元人物モデルの比較が可能となる。

【００５０】

Ｓ６０５では再構成結果とＧｒｏｕｎｄ　Ｔｒｕｔｈのオブジェクトを比較し差分を算出する。全てのカメラＩＤと適切なステップで変動させた３次元人物モデルパラメータの組み合わせで算出される前記差分を算出する。ここで言う差分は数式２で定義したスコアＬｓ（ｃ）とする。なお、ここでカメラＩＤは単純にカメラを識別するＩＤであれば何でも良いが、以降の計算を簡易にするためには例えば競技場周囲を時計回りまたは反時計回りに設置順に連番に付与して用いることが好ましい。

【００５１】

推定された３次元人物モデルとＧｒｏｕｎｄ　Ｔｒｕｔｈとなる３次元人物モデルとの

差分が小さい３次元形状を再構成することができた入力画像を撮影したカメラが該人物の撮影条件に合致したカメラである。よって、撮影対象の人物に関する３次元人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）を入力した際にＧｒｏｕｎｄ　Ｔｒｕｔｈとの差分スコアＬｓを最小にするカメラＩＤ　ｃを求める判別関数Ｑを求めることで当該人物の撮影条件に合致したカメラを決定できる。このときの式を数式３のように定義することができる。

【００５２】

【数３】

$$Q(\rho,\theta,\beta,x,y) = arg\min_{c}(L_s(c))$$

【００５３】

さて、スコアＬｓ（ｃ）は既に数式２で定義済であり、この値は３次元人物モデルの内外判定で決定する。既に説明した通り該３次元人物モデルは人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）で決定するため、判別関数Ｑを求めることは可能である。入力となる人物モデルパラメータを１次元のカメラＩＤに次元削減を実施する。既に数式３の右辺は一意に算出可能であるため一般的な教師有次元削減を実施し適切な尺度空間を求めることで入力に対して適切なカメラＩＤであるｃを求める判別式を得ることができる。例えばフィッシャー判別分析（ＦＤＡ）によって、Ｑ（ρ、θ、β、ｘ、ｙ）が求まる。

【００５４】

次に、実運用時のフローを説明する。３次元再構成対象の人物に関する人物モデルパラメータが分かると自動的に３次元再構成に用いるカメラが決定する関数Ｑが求まったため、実運用時は図６（Ｂ）のフローチャートに従って推定を実行する。

【００５５】

Ｓ６１１では推定時刻ｔにおける３次元再構成対象の人物４４の向きρ、姿勢θ、形状β，座標（ｘ，ｙ）を求める。このとき２次元画像の入力に対してρ、θ、βを求める際は図３（Ｂ）に例示したネットワークアーキテクチャによる。図３（Ｂ）と図３（Ａ）は似ているが、異なっている。既に説明した図３（Ａ）は２次元画像入力（３０）から詳細な解像レベルの３次元人物モデル（３２）を出力する砂時計型のＤＮＮ（３１）であった。対して図３（Ｂ）は３３に示すように入力は同じ２次元画像であるが３４に示したエンコーダを経て出力される情報は図３（Ａ）に比して低次情報となっておりそのまま向きρ（３５）、形状β（３６）、姿勢θ（３７）を出力する。ここで簡易な出力とするＤＮＮを用いる意味は、本ステップは３次元再構成に最適なカメラを決定するための情報を取得する目的で行うためであり、汎化性能を高く維持しながら高速に処理を実施するためである。よって、図３（Ｂ）の出力は簡易なものとなり、前記テンプレートメッシュをρ、β、θによって変形したものとなるため、図３（Ａ）の高次出力結果となる３２と比して３８に示したようなレベルのものに留まる。さらに人物座標（ｘ、ｙ）は一般的な人物検出器によって求める。

【００５６】

以上、ρ、θ、β、ｘ、ｙを各々求める方法を説明したが、該推定処理はカメラ内に対象が写っていれば精度は大きく変わらないため、例えばランダムに選択した数個のカメラから算出された上記パラメータの中央値等によって各パラメータを決定すると良い。

【００５７】

尚、以上では図３（Ｂ）に示したように人物の姿勢や形状をＤＮＮによって推定する方法として説明したが、同様の推定ができる方法であれば他の方法によっても良い。例えば２次元画像入力による姿勢推定をスケルトン推定によって実施し、最適な撮影条件と対応付ける等によって目的を達成しても良い。

【００５８】

次にＳ６１２にて上記人物モデルパラメータρ、θ、β、ｘ、ｙを既に取得済の判別関

数Ｑに入力しカメラＩＤを決定する。
【００５９】
そして最後にＳ６１３にて前記決定したカメラによって撮影された画像を３次元再構成用ＤＮＮに入力することで３次元人物モデルを推定する。
【００６０】
以上によって３次元人物モデルの形状の推定精度が良い結果が得ることができる。最終的なレンダリング結果を得るためには図４（Ａ）の仮想カメラパスが４１、４２、４３と移動した際に各カメラ方向の角度が近接するカメラから作成したテクスチャを対象オブジェクトに貼り付けることによって最終的なレンダリング結果を取得する。
【００６１】
以上によって、ＤＮＮによって３次元再構成する場合に精度良く再構成するための方法が示された。
【００６２】
尚、上記テクスチャを決定する方法以外にも各推定オブジェクトの表面上の各点について色情報推定を行うＤＮＮによって目的を達成しても良い。
【００６３】
また、本実施形態は入力を２次元画像のみとしたが、例えば図２に記載のカメラ構成によって実施可能なＶｉｓｕａｌ　Ｈｕｌｌの推定を行い、ここで推定された３次元点群を入力としても同様の結果を得ることができる。その他、デプスマップを取得して入力し該デプス情報と出力する３次元形状を紐づけた学習によって達成しても良い。
【００６４】
＜第２の実施形態＞
第２の実施形態は、第１の実施形態の変形形態である。第１の実施形態では３次元再構成に用いるカメラＩＤの判定を人物モデルパラメータのみの推定結果から決定する方法を説明した。本実施形態では人物モデルパラメータだけでなく、ビデオレンダリングの際の仮想カメラに関わるパラメータも参照して決定する方法を採用する実施形態とする。
【００６５】
第２の実施形態に係る情報処理装置が適用されるシステム構成は第１の実施形態と同様であり図１で表すことができ、また具体的な競技場とカメラの設定も図２に示した例と同じものである。
【００６６】
以降は、第１の実施形態と異なる点のみに焦点を当て説明する。第１の実施形態では数式１によって対象オブジェクトの内外判定を定義し、連続する３Ｄ　Ｏｃｃｕｐａｎｃｙ　Ｆｉｅｌｄを定めることで該オブジェクトの座標を学習する方法を説明した。この際、数式２によってＧｒｏｕｎｄ　Ｔｒｕｔｈの３次元形状と推定した３次元形状の比較によって推定時の誤差を評価するスコアを定義することで該誤差が最小となるカメラが形状推定に最適なカメラとして選択する式を数式３により示した。本実施形態では仮想カメラでレンダリングした場合の差異を計算するため、３次元形状同士の比較によるスコアを扱わず、３次元再構成後の２次元ビデオレンダリング時の形状推定誤差を考慮する。
【００６７】
２次元ビデオレンダリングにおいては、仮想カメラ（ＶＣ）視点で２次元画像に変換された後の誤差のみを考慮する。よって、図４（Ａ）の４５で示したｘ軸および４６で示したｙ軸とこれら軸に直交するｚ軸によって定義される仮想カメラ（ＶＣ）の３次元座標（x_v，y_v，z_v）および仮想カメラ（ＶＣ）の光軸中心が向く角度（ρ_v，ϕ_v）によって随時決定される撮影画像のパラメータによって３次元オブジェクト情報を仮想カメラ（ＶＣ）によってレンダリングする２次元平面射影する関数をπ_{vc}（ρ_v，ϕ_v，x_v，y_v，z_v）と定義すると、この際に２次元平面状の誤差は数式４によって定義することができる。
【００６８】
【数４】

$$L_s{}'\quad(c)=\frac{1}{P}\sum_{p=1}^{P}\left|\pi_{VC(\rho_v,\varphi_v,x_v,y_v,z_v)}\left(f_s\left(g\left(I(x_{ci})\right),h_c(X_{wi})\right)\right)-\pi_{VC(\rho_v,\varphi_v,x_v,y_v,z_v)}\left(f_s^{*}(X_{wi})\right)\right|^2$$

【００６９】

この場合は、数式３で定義した３次元再構成対象自体のパラメータのみならず、仮想カメラ（ＶＣ）の位置姿勢を定義するパラメータによって、３次元再構成に用いるカメラＩＤを求める。その場合は数式５に示す判別関数Ｑを求めることにより取得される。数式５で追加した仮想カメラ（ＶＣ）に関わるパラメータ数は対象人物に関わるパラメータと比較すると低次であるため、第１の実施形態で説明した方法と同様の方法によって、一般的な教師有次元削減を実施し適切な尺度空間を求める。これにより適切なカメラＩＤを求める判別関数Ｑを得ることができる。

【００７０】

【数５】

$$Q(\rho,\theta,\beta,x,y,\rho_v,\varphi_v,x_v,y_v,z_v)=arg\min_{c}(L_s{}'\quad(c))$$

【００７１】

推定時には図４に示したカメラパス４７上で仮想カメラ（ＶＣ）位置が４１，４２，４３と変化していくことで仮想カメラパラメータ（ρ_v，φ_v，x_v，y_v，z_v）は自動的に決定する。よって第１の実施形態で説明済の対象人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）を求めることによって数式５の右辺の変数は全て決定する。これにより推定時は判別関数Ｑによって適切なカメラＩＤを求めることができる。

【００７２】

＜第３の実施形態＞

第１および２の実施形態では図２に示した競技場周囲に設置されたカメラＭ台（Ｍは２以上の整数）で各々で撮影された画像を入力とした際に３次元再構成するＤＮＮモデルは全て共通のものとした。すなわち、どのカメラで撮影された場合であっても同じ汎用ＤＮＮモデルによって人物を再構成するものとして説明した。しかし、場合によっては各々のカメラで撮影される領域が狭い場合などにおいて、各々のカメラ内で撮影される人物の撮影角度等に偏りが発生する場合が有る。

【００７３】

例えば競技場の天井から真下に向かって撮影するカメラが存在した場合、このカメラは人物を真横から観測した際の画像を撮影する可能性は低い。この場合に入力画像が対象人物をあらゆる方向から撮影された場合に全般的に３次元再構成の推定精度の良い一つのＤＮＮモデルを利用するより、入力画像として入力される画像の傾向に従って学習したＤＮＮモデルを利用する方が３次元再構成の精度が良くなる。

【００７４】

これらは図２に示したカメラＭ台についても撮影される人物の傾向の偏りが発生する可能性が有る。そのため、第３の実施形態ではカメラＭ台についてカメラごとに利用するＤＮＮモデルを別個に学習し用意する方法を説明し、また実運用時はそれらをどのようにして利用するかを説明する。以下、図６（Ｃ）のフローチャートに従って説明する。

【００７５】

Ｓ６２０にてカメラＭ台分の入力に対して各々異なるＤＮＮモデルとなるように以降でＭ回分のループが実行される。本ループ内においてＳ６２１では特定のカメラｃ（１≦ｃ≦Ｍ）に関するＤＮＮモデルの学習を開始する。特定のカメラＩＤがｃであるカメラの初期ＤＮＮモデルとしてあらゆる入力画像に対して汎化性能の高いパラメータをセットする。次にＳ６２２において第１の実施形態のＳ６０２で実施した方法と同様に３次元人物モ

デルを１つＣＧで再現された競技場内部において人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）を多様に変動させて配置する。

【００７６】

次にＳ６２３にて前記多様な３次元人物モデルをカメラｃで撮影された画像としてレンダリングし、本レンダリング画像から現在のＤＮＮモデルによって３次元再構成を実施し人物形状を推定する。そして、前記配置した３次元人物モデルをＧｒｏｕｎｄ　Ｔｒｕｔｈとして推定した人物形状と比較、これら差分を算出する。人物モデルパラメータを多様に変更した各種結果について平均差分を算出する。次にＳ６２４で前記平均差分が事前に設定した閾値以下になっているかどうか比較する。閾値以下でない場合、Ｓ６２５によって前記差分が小さくなるようにＤＮＮモデルのパラメータセットを微小に更新しＳ６２２に戻る。前記Ｓ６２４にて平均差分が閾値以下である場合、カメラｃに関する処理はＳ６２６にて終了し、ｃ＝ｃ＋１としてセットしてＳ６２１に戻る。これら処理がｃ＝Ｍになるまで繰り返す。

【００７７】

以上によってＭ台分のカメラに対応するＤＮＮモデルが取得される。

【００７８】

次に、推定時のフローを図６（Ｄ）のフローチャートによって説明する。Ｓ６２７にて時刻ｔの瞬間の３次元再構成対象の人物の人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）を既に第１の実施形態で説明した方法によって取得する。次にＳ６２８にて既に説明した判別関数Ｑによって再構成に適したカメラＩＤを決定する。尚、ここで用いる判別関数Ｑを求める方法は、既に第１の実施形態で説明した数式、方法と同様で導出可能である。つまり、既に説明した数式２または数式４で定義されるＧｒｏｕｎｄ　Ｔｒｕｔｈと推定結果の誤差についての式はＤＮＮモデルの違いがカメラＩＤのｃで表され同じ意味となるため、同様の方法にて実施される。Ｓ６２９ではＳ６２８で決定したカメラＩＤに紐づいたＤＮＮが決定しＳ６３０にて前記カメラで撮影された画像を前記選択したＤＮＮに入力することによって最終的に時刻ｔにおける対象人物の３次元再構成を実施する。

【００７９】

＜第４の実施形態＞

第１、第２、第３の実施形態では、簡単のため入力となる撮影画像が単一のカメラで撮影された画像として図３（Ａ）に示した構成を想定した実施形態としていた。しかし、３次元再構成を目的とする場合、複数の画像を入力とした方が精度向上の期待ができる。特に図２に示したように現在想定しているカメラ設定は競技場の周囲に多数のカメラが設置されており、図１に示したシステムによって多数のカメラは同期して撮影されている。よってこれら同期して撮影した画像から複数枚を選択してＤＮＮに入力することは精度良い３次元再構成を実施する上で良い結果をもたらす。

【００８０】

例えば画像３枚を入力とした場合のＤＮＮに対する入力と出力の関係を図３（Ｃ）に示した。この図では同期されて３次元再構成対象の人物が異なるカメラで撮影されており、各入力画像を３０１，３０２，３０３で表している。今、これら３カメラの関係は固定とする。該３つのカメラから撮影された画像を既に第１の実施形態で説明したのと同等のＤＮＮ３０４に入力することで、３次元再構成結果３０５を出力している。

【００８１】

さて、上記ではＤＮＮに入力される３つのカメラは固定と述べた。つまり３カメラの関係性をＤＮＮ３０４は学習済である。よって異なる３つのカメラを選んでＤＮＮ３０４に入力しても適切な３次元再構成は難しい。３次元再構成対象となる人物は姿勢を変えながら移動するため常に同じ３カメラのセットで全てのカメラで全身撮影され続けることは難しい。そこで対象人物の人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）を推定し、このパラメータ（特にｘ、ｙ）において良好な条件で撮影している３カメラで学習されたＤＮＮモデルを利用した３次元再構成を実施することが必要となる。

【００８２】

入力カメラ数が３となっても、カメラセットに対して一意に定義されるカメラセットＩＤを与えることで入力カメラ数が１の場合の実施形態と同様のフローで３次元再構成を実施可能である。例えばＭ台のカメラから３台のカメラを選ぶ全ての組み合わせはＭ（Ｍ－１）（Ｍ－２）／３×２×１通りである。そのため、全ての組み合わせについてＩＤを付与する場合は、各組合せについて第３の実施形態に示したフローに則って各ＤＮＮを学習し、各学習における損失は数式２または数式４に示したスコアによって算出する。この場合パラメータｃは上記カメラセットＩＤとして計算する。これら計算によって算出された損失により最適なカメラセットＩＤを判別する判別関数は数式３または数式５によって定義した式によって算出される。

【００８３】

尚、第４の実施形態では同期されて撮影した画像複数枚を入力とする方法として具体的なＤＮＮの構成として３カメラで撮影された事例を基に説明したが、ここで入力する画像枚数は３に限らず何枚になっても同様の方法によって目的を達成できる。

【００８４】

他、単一のカメラで対象を撮影する際には時系列に連続するフレームを図３（Ｃ）に例示したアーキテクチャのＤＮＮに複数枚入力する方法としても良い。

【００８５】

＜第５の実施形態＞

第１～第４の実施形態では簡易に説明する実施形態とするため、図４（Ａ）に示したように３次元再構成対象とする人物が１名のみの例を挙げた。本実施形態では複数の人物を３次元再構成対象とする場合として説明する。

【００８６】

尚、本実施形態ではシステム構成例は図１同様であるものとし、競技場やカメラの設定は図２に示したものと同様の設定とする。

【００８７】

複数人数が競技場に登場する場合の具体的な例を図４（Ｂ）に示した。ビデオレンダリングする際の仮想カメラ（ＶＣ）の移動パスを４０７と決定し、その上を仮想カメラ（ＶＣ）が時系列に４０１、４０２、４０３の順に移動するものとする。その際に撮影対象となる人物４０４、４０５、４０６は姿勢や座標を変える。第１の実施形態と異なるのは、仮想カメラ位置の変動によってビデオレンダリングの対象人数が増減することと、複数人物が存在することによって競技場内に設置されたカメラにおける遮蔽が発生することである。

【００８８】

よって、第１の実施形態に加えてレンダリング対象人物を特定し、該当人物についてのみ３次元再構成を実施して対応する必要があることと、遮蔽が発生した場合の３次元再構成では各々選択するカメラ優先度を変えることによって目的を達成する。

【００８９】

以降、図７に示したフローチャートによって本実施形態に係る処理フローを説明する。

【００９０】

まず、Ｓ７０１にて３次元再構成対象となる人物全ての座標（ｘ、ｙ）を推定する。本ステップはＭ台のカメラで撮影された画像どれを利用しても良いし、全体を撮影しているカメラ１台によって決定してもどちらでも良い。

【００９１】

次にＳ７０２にて仮想カメラ（ＶＣ）のパス４０７とカメラ方向によって決定する第２の実施形態で説明した仮想カメラパラメータ（ρ_v，Φ_v，x_v，y_v，z_v）を参照し、該パラメータによって決定する仮想カメラの画角内に描画される人物を特定する。例えば仮想カメラ位置が４０１に有る場合は、人物４０４のみを描画対象とすれば良く、仮想カメラ位置が４０２に有る場合は人物４０４，４０５，４０６を描画対象とする。また仮想カメラ位置が４０３に移動すれば描画対象を４０４および４０５のみとする、等決定する。

【００９２】
次にＳ７０３にて特定した全対象人物について既に説明済の人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）を推定する。次にＳ７０４にて第１の実施形態および第２の実施形態で説明した判別関数Ｑによって各々の対象人物を再構成するのに用いるカメラＩＤを決定する。しかし、判別関数Ｑは既に説明した通り競技場内に１人の人物モデルが存在した場合に再構成に適したカメラＩＤを特定する関数として学習したものである。よって現在の実施形態では競技場内に複数人の再構成対象人物が存在するため、上記判別関数Ｑで決定したカメラＩＤで撮影した画像に複数の人物が撮影され、片方の人物によって再構成対象の人物が遮蔽される場合が発生しうる。そこでＳ７０５にて前記決定したカメラＩＤのカメラで撮影された画像内の人物のうち再構成対象の人物が遮蔽されているか否かを判定する。

【００９３】
遮蔽されているか否かの判定はＭａｓｋ－ＲＣＮＮ等で人物領域検出し人物領域に重なりが有るかどうかを判定する等によって実施する。もし遮蔽が有ると判定された場合、該カメラによる３次元再構成結果は良好なものになりにくい。よってＳ７０６にて判別関数によるカメラＩＤの候補を次点のＩＤに修正し再度Ｓ７０５に戻る。遮蔽が無いと判定されれば、Ｓ７０７によって３次元再構成に利用するカメラを決定し、本画像を３次元再構成用のＤＮＮに入力することによって３次元再構成を実施する。

【００９４】
＜第６の実施形態＞
第１～第５の実施形態では３次元再構成対象の人物を３次元再構成する際の推定方法について、入力とする最適なカメラＩＤを随時変えながら最適な結果を得る方法やカメラＩＤに応じて学習済ＤＮＮモデルを変え、最適な結果を得る方法を説明した。

【００９５】
一方で各実施形態では基本的に利用するＤＮＮのアーキテクチャは固定で実施することを前提とした方式であった。

【００９６】
本実施形態ではレンダリング時に時系列に変動する仮想カメラの画角とその画角内に入る人物の状態に応じて各３次元再構成のＤＮＮのアーキテクチャも変えて対応する方法を説明する。また、ＤＮＮを利用しない、Ｖｉｓｕａｌ　Ｈｕｌｌ等に代表される幾何的な手法を動的に切り替え随時最適な手法によって３次元再構成を実施する方法を説明する。

【００９７】
この場合の処理フローを図８のフローチャートに沿って説明する。初めにＳ８０１にて３次元再構成を実施する時刻ｔにおける再構成対象人物全ての座標（ｘ、ｙ）を推定する。次にＳ８０２にて３次元再構成結果を最終的にビデオレンダリングして出力する際の仮想カメラのパスやカメラの動きから時刻ｔにおける仮想カメラパラメータ（ρ_v，ϕ_v，x_v，y_v，z_v）を取得し、このときに第５の実施形態で説明した方法によって仮想カメラ内に描画される人物を特定する。そしてＳ８０３にて前記特定した全ての３次元再構成対象の人物の人物モデルパラメータを推定する。そしてＳ８０４にて上記３次元再構成対象人物の人物モデルパラメータと仮想カメラパラメータによって以降で各対象人物を３次元再構成する際の方式や利用するカメラＩＤを決定する。尚、各種方式決定には第１～５の実施形態では参照されなかった情報が勘案される。ここで仮想カメラによりレンダリングされる複数人物の様子の例を図９に示し、図９を用いて以下説明する。

【００９８】
図９では仮想カメラパラメータ（ρ_v，ϕ_v，x_v，y_v，z_v）により決定した仮想カメラ画角内に描画される人物が２名となった場合のレンダリング結果を図示した例である。ここで描画される２名は９２および９３の人物となる。既に説明した通り、各々の人物を３次元再構成する場合、各人物の３次元形状を推定する際に最適なカメラは通常異なる。各々の人物を３次元再構成する際に最適なカメラは既に説明済の判別関数Ｑで選択可能である。

【００９９】
しかし、判別関数Ｑは描画対象人物の３次元再構成結果を良好にするための基準として取得された関数であった。しかし、仮想カメラで描画する際に仮想カメラから遠方に存在する９２の人物の再構成結果は９１の再構成結果と同レベルで推定する必要は無い。９１の人物でビデオレンダリング時に１ピクセル以下のレベルの誤差精度を達成していたとしたら、９２の人物はそのスケールの２分の１以下で描画されるため、描画スケールに合わせた３次元再構成を実施すれば充分である。
【０１００】
よって、この場合のカメラ画角における人物描画スケールは仮想カメラパラメータの内の（x_v，y_v，z_v）と人物モデルパラメータの内の（ｘ、ｙ）によって決まる。例えば、３次元再構成に利用する図３（Ａ）のＤＮＮ３１や（Ｃ）のＤＮＮ３０４に入力する画像数が異なるＤＮＮモデルを混在させて事前に保持しておき運用する場合を考える。この場合、仮想カメラに近い人物９１を３次元再構成する場合は入力画像数の多いＤＮＮモデルを用いて形状推定しメッシュ情報に変換する。またこの場合、遠方の人物９２を３次元再構成する場合は入力画像数の少ないＤＮＮモデルや簡易なネットワークによる推定、または輪郭情報のみで計算可能なＶｉｓｕａｌ　Ｈｕｌｌによる形状推定しメッシュ情報に変換する。これら採用する方式は例えば各種手法の計算コストと精度に対してレンダリング時の対象スケールを対応付けて事前に自動で決定することのできる決定木やルックアップテーブル等で定義しておくと良い。
【０１０１】
最後にＳ８０５にてＳ８０４で決定した手法およびカメラによって撮影された１以上の画像によって最終的な３次元再構成を描画対象人数分実施し、時刻ｔにおけるシーンのレンダリングに必要な３次元再構成の処理を終える。
【０１０２】
＜第７の実施形態＞
第１～第６の実施形態では多視点から同期され撮影された画像から３次元再構成に適したカメラを随時決定して再構成する方法を説明した。これによって、３次元再構成した後に取得されるビデオレンダリングの結果は毎フレーム高品質な結果を取得することが可能となる。一方、毎フレームの３次元再構成を異なる方法で実施した結果を時系列に並べて閲覧することによって再構成時の結果に各種異なる劣化が発生した場合、画像上のチラつきのような現象としてビデオを閲覧している人にとって好ましくない映像となる可能性がある。
【０１０３】
図１０に簡単な具体例を挙げて説明する。図１０（Ａ）に示したように３次元再構成対象の人物１０００が存在し、該人物を仮想カメラ１００６から見たシーンを時系列方向に一定の長さを持つシーンとして３次元再構成する場合を考える。今、簡単のため仮想カメラ位置は１００６に限定するが、３次元再構成対象の人物１０００は随時姿勢等を変えるため人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）は時系列で遷移する。既に説明した方法に則って人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）に応じて、または人物モデルパラメータ（ρ、θ、β、ｘ、ｙ）と仮想カメラ１００６に関わる仮想カメラパラメータ（ρ_v，Φ_v，x_v，y_v，z_v）の組み合わせによって既に説明済の判別関数Ｑによって３次元再構成に適した最適カメラが選択される。このとき、人物１０００の挙動に応じて選択されたカメラがカメラ１００４になったりカメラ１００５になったり随時切替わる。この際、カメラ１００４で撮影された実画像例は１００１に例示した画角で撮影された画像となりカメラ１００５で撮影された実画像例は１００３に例示した画角で撮影された画像となる。このとき各々のカメラ撮影画像入力から仮想カメラ１００６による撮影画像１００２を学習済のＤＮＮモデルによって推定するとき、推定結果が完全に真値と一致することはあり得ないため、各々の推定結果が真の正解値と僅かに異なる結果を出力する。
【０１０４】
推定結果の異なる様子の例を図１０（Ｂ）に示した。通常、仮想カメラ１００６の位置

に実カメラは設置していないため真の正解画像を取得することはできないが、仮に真の正解画像を取得できた場合の画像例を１００００１に示している。正解画像は正解データの一例である。これに対し、カメラ１００４で撮影された画像を入力として３次元再構成を実施し仮想カメラ１００６の位置から見た画像をレンダリングした結果の例を１００００２に示した。１００００４の斜線領域は真の正解画像１００００１と比較して推定結果が異なる領域である。この領域が真の正解画像と形状が異なっている原因は、カメラ１００４から実際には観測されない領域に対する推定結果のオブジェクトサーフェスに該当するためである。

【０１０５】

また、同様にカメラ１００５で撮影された画像を入力として３次元再構成を実施し仮想カメラ１００６の位置から見た画像をレンダリングした結果の例を１００００３に示した。１００００５の斜線領域は真の正解画像１００００１と比較して推定結果が異なる領域である。この領域が真の正解画像と形状が異なっている原因は、カメラ１００５から実際には観測されない領域に対する推定結果のオブジェクトサーフェスに該当するためである。

【０１０６】

実用上は仮想カメラ１００６の位置が固定される場合に３次元再構成に利用する画像を常にカメラ１００４で撮影された画像としていれば、再構成後の誤差は１００００４で例示した領域のみに発生するため、鑑賞上気になる結果となりにくい。同様に常にカメラ１００５撮影された画像を用いていても常に再構成後の誤差は一貫して１００００５で例示したような領域のみに発生するためこの誤差が鑑賞上問題になる可能性は低い。

【０１０７】

しかし、対象人物１０００の姿勢等パラメータが時系列で変化するのに対して前記判別関数Ｑによる選択カメラが随時１００４と１００５を交互に切り替えるように選択すると誤差の傾向に一貫性が無くなる。時系列で切替わるフレームごとに１００００４の領域が膨張したフレーム、１００００５の領域が膨張したフレームが交互に発生しこれが繰り返されることになり、これは再構成形状のチラつきとして視聴者に不快な印象を与え得る。

【０１０８】

そのため、この現象を抑制するために最適なＤＮＮアーキテクチャ例を図３（Ｄ）に示した。

【０１０９】

図３（Ｄ）は第１～第６の実施形態を全て含む設定において前述した弊害対策のための運用を想定したアーキテクチャとなっている。３次元再構成の時系列処理で時系列方向にｔ＝Ｔ－１、Ｔ、Ｔ＋１と遷移した際の各時刻ｔにおけるＤＮＮの間で情報が伝播することで上述した３次元再構成間の不安定な劣化揺らぎを抑制する様子を表している。ｔ＝Ｔ－１の際に最適なカメラＩＤが１つ以上選択され、これら画像から３次元再構成に必要な情報表現に変換するエンコーダ３１１が有る。そして、３次元再構成を要約した中間特徴表現３１３を経て、デコーダ３１２により高次特徴表現に変換され最終的に３次元再構成結果３１４を得る。中間特徴表現３１３は前記エンコーダおよびデコーダを通じて学習された任意の表現で良いが、例えば明示的に人物モデルの形状を表現するρ、θ、β等と一致するように誘導した学習によって取得される表現であっても良い。

【０１１０】

ｔ＝Ｔにおいてもｔ＝Ｔ－１の際と同様に１つ以上のカメラで撮影された画像がエンコーダ３１５に入力される。そして中間特徴表現３１７に変換されるが、ここにｔ＝Ｔ－１の時点での特徴表現３１３が入力される。そしてｔ＝Ｔ－１からの特徴表現とｔ＝Ｔにおける入力画像が参照された結果３１６のデコーダを経て最終的に３１８に示された３次元再構成結果を出力する。

【０１１１】

同様にｔ＝Ｔ＋１においてもｔ＝Ｔの時点の中間特徴表現３１７がＤＮＮの中間で入力され、これら中間特徴が時系列で伝播され過去フレームにおける推定結果の参照が繰り返される。これにより、第１～第５の実施形態で想定としていた、各フレームで独立した３

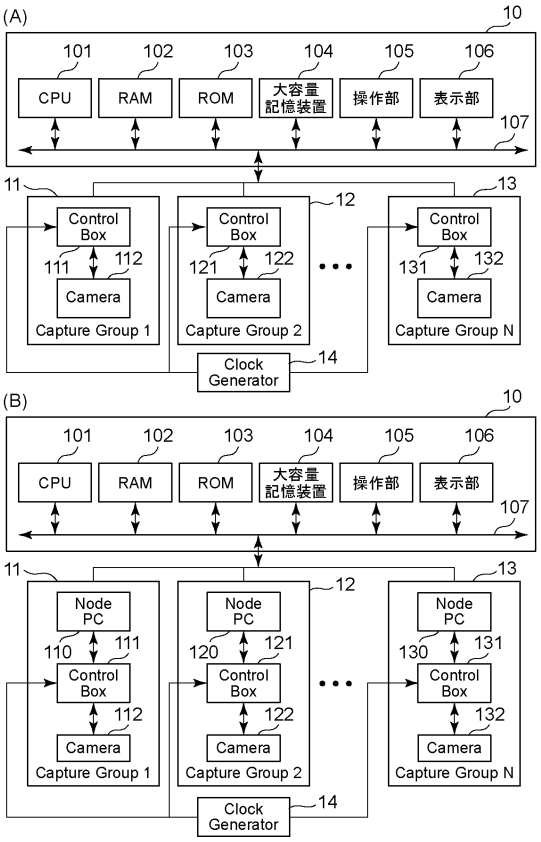
次元再構成を実施していたときと比較し、時系列的に過去の情報が参照されるようになるため一貫性の無い揺らぎ等による画像のチラつき等の弊害を抑制する効果が得られる。

【０１１２】

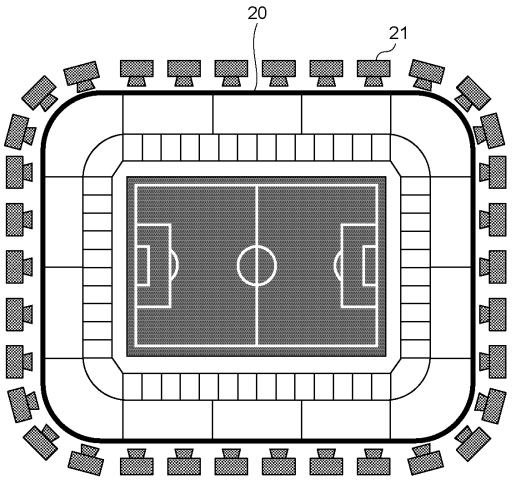
尚、図３（Ｄ）は一般的なＲＮＮ（Ｒｅｃｕｒｒｅｎｔ　Ｎｅｕｒａｌ　Ｎｅｔｗｏｅｋ）のように同一のネットワーク構造の一部が循環して次の入力に入ってくるもののみを表しているがほかの方法でもよい。例えば時刻ｔ＝Ｔ－１、Ｔ，Ｔ＋１と時系列方向に変遷するにつれて第１～第６の実施形態で説明したように、入力カメラＩＤが毎時刻で変わる場合やＤＮＮモデルが変わる場合においても、中間特徴表現を合わせることで本アーキテクチャは運用可能である。また、入力画像数も毎時刻変わっても良い。例えばｔ＝Ｔ－１における入力画像数は３でありｔ＝Ｔにおける入力画像数が１となっても、各入力に合わせて学習されたＤＮＮモデルによって３次元再構成を行い、各中間特徴表現を時系列に伝播させて用いれば同等の効果を発揮することができる。また時刻ｔにおいて仮想カメラから遠方に存在する人物をＶｉｓｕａｌ　Ｈｕｌｌ等によって三次元再構成をした場合でも、３次元再構成結果を低次の中間特徴表現（特にρ、θ、β）に変換するＤＮＮを予め学習しておくことで達成可能となる。

【図面】

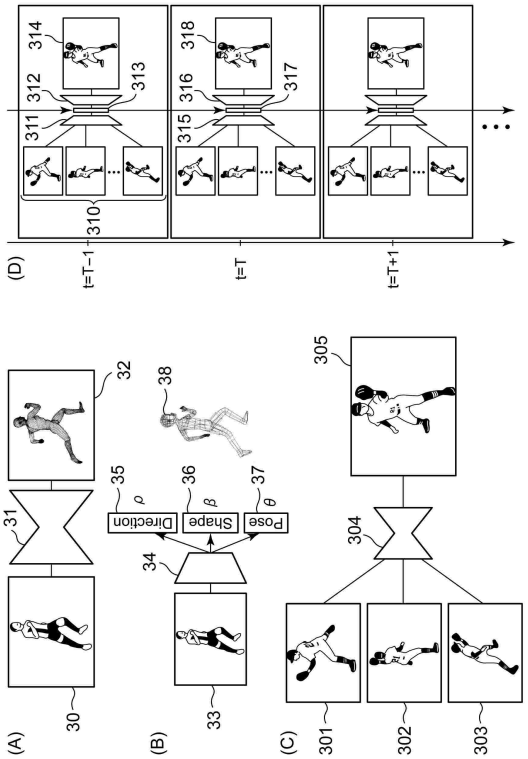
【図１】



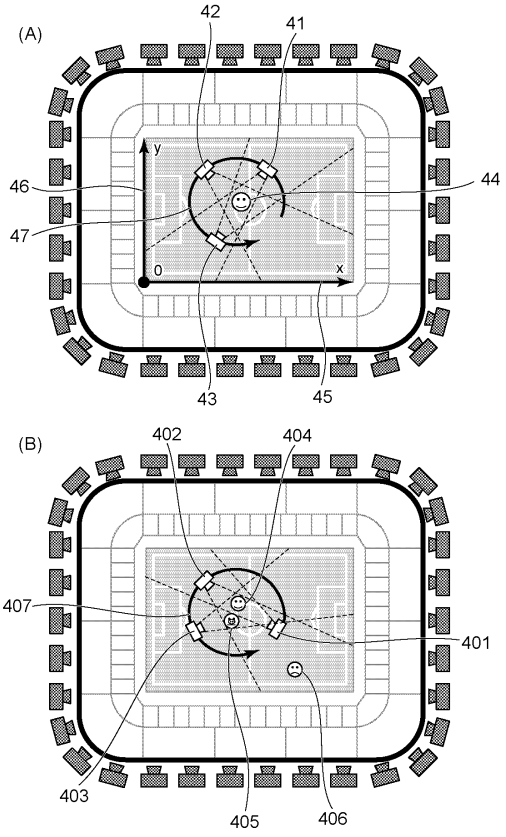
【図２】



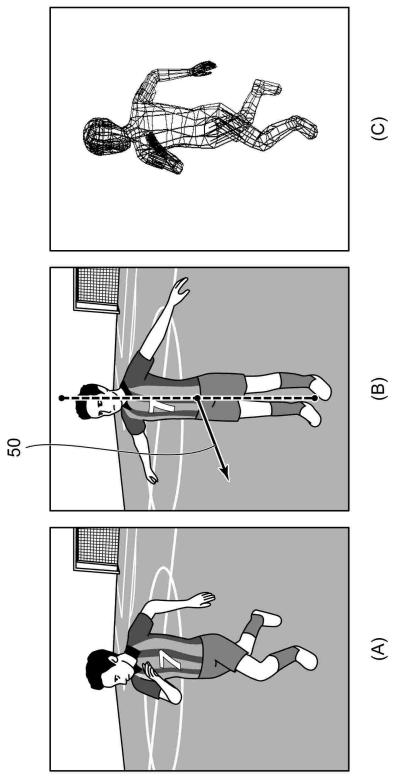
【図３】


(D)
t=T−1
t=T
t=T+1
310
311
312
313
314
315
316
317
318
(A)
(B)
(C)
30
31
32
33
34
35
36
37
38
Direction
Shape
Pose
301
302
303
304
305

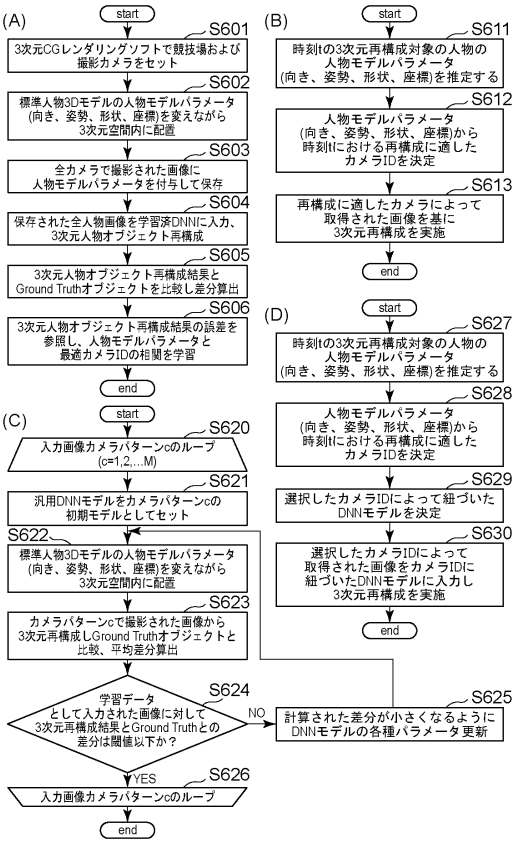
【図４】


(A)
(B)
41
42
43
44
45
46
47
x
y
0
401
402
403
404
405
406
407

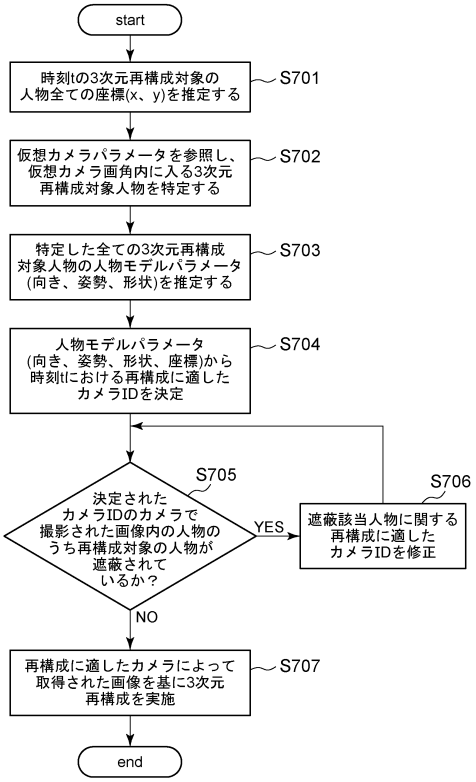
【図５】


(A)
(B)
(C)
50

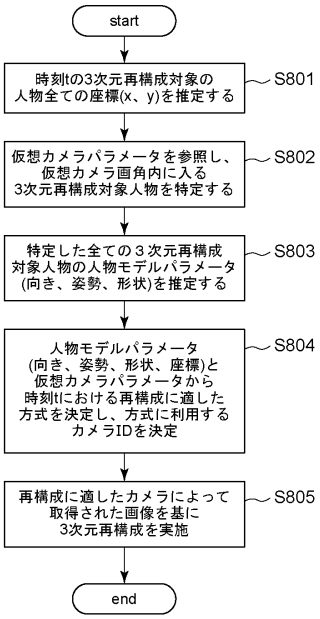
【図６】


(A)
start
S601
3次元CGレンダリングソフトで競技場および撮影カメラをセット
S602
標準人物3Dモデルの人物モデルパラメータ(向き、姿勢、形状、座標)を変えながら3次元空間内に配置
S603
全カメラで撮影された画像に人物モデルパラメータを付与して保存
S604
保存された全人物画像を学習済DNNに入力、3次元人物オブジェクト再構成
S605
3次元人物オブジェクト再構成結果とGround Truthオブジェクトを比較し差分算出
S606
3次元人物オブジェクト再構成結果の誤差を参照し、人物モデルパラメータと最適カメラIDの相関を学習
end
(B)
start
S611
時刻tの3次元再構成対象の人物の人物モデルパラメータ(向き、姿勢、形状、座標)を推定する
S612
人物モデルパラメータ(向き、姿勢、形状、座標)から時刻tにおける再構成に適したカメラIDを決定
S613
再構成に適したカメラによって取得された画像を基に3次元再構成を実施
end
(C)
start
S620
入力画像カメラパターンcのループ(c=1,2,…M)
S621
汎用DNNモデルをカメラパターンcの初期モデルとしてセット
S622
標準人物3Dモデルの人物モデルパラメータ(向き、姿勢、形状、座標)を変えながら3次元空間内に配置
S623
カメラパターンcで撮影された画像から3次元再構成しGround Truthオブジェクトと比較、平均差分算出
S624
学習データとして入力された画像に対して3次元再構成結果とGround Truthとの差分は閾値以下か？
NO
S625
計算された差分が小さくなるようにDNNモデルの各種パラメータ更新
YES
S626
入力画像カメラパターンcのループ
end
(D)
start
S627
時刻tの3次元再構成対象の人物の人物モデルパラメータ(向き、姿勢、形状、座標)を推定する
S628
人物モデルパラメータ(向き、姿勢、形状、座標)から時刻tにおける再構成に適したカメラIDを決定
S629
選択したカメラIDによって紐づいたDNNモデルを決定
S630
選択したカメラIDによって取得された画像をカメラIDに紐づいたDNNモデルに入力し3次元再構成を実施
end

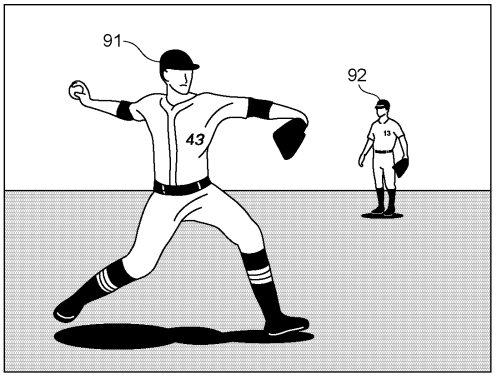
【図７】


start
時刻tの3次元再構成対象の人物全ての座標(x、y)を推定する
S701
仮想カメラパラメータを参照し、仮想カメラ画角内に入る3次元再構成対象人物を特定する
S702
特定した全ての3次元再構成対象人物の人物モデルパラメータ(向き、姿勢、形状)を推定する
S703
人物モデルパラメータ(向き、姿勢、形状、座標)から時刻tにおける再構成に適したカメラIDを決定
S704
S705
決定されたカメラIDのカメラで撮影された画像内の人物のうち再構成対象の人物が遮蔽されているか？
YES
S706
遮蔽該当人物に関する再構成に適したカメラIDを修正
NO
再構成に適したカメラによって取得された画像を基に3次元再構成を実施
S707
end

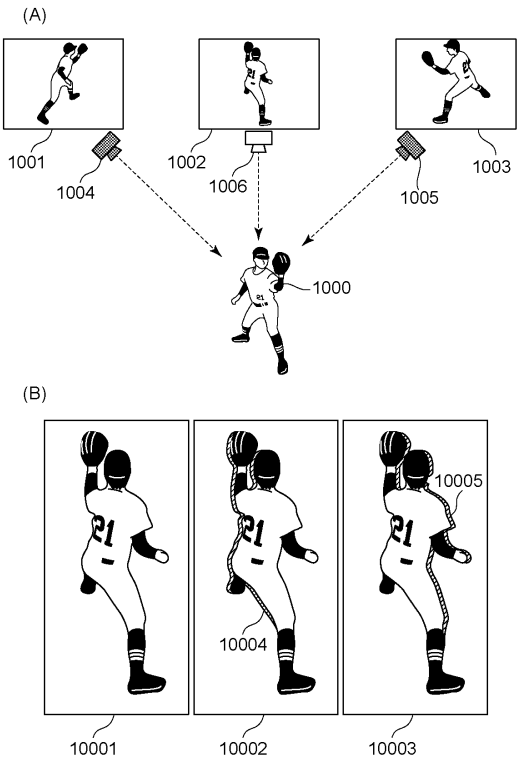
【図８】


start
時刻tの3次元再構成対象の人物全ての座標(x、y)を推定する
S801
仮想カメラパラメータを参照し、仮想カメラ画角内に入る3次元再構成対象人物を特定する
S802
特定した全ての３次元再構成対象人物の人物モデルパラメータ(向き、姿勢、形状)を推定する
S803
人物モデルパラメータ(向き、姿勢、形状、座標)と仮想カメラパラメータから時刻tにおける再構成に適した方式を決定し、方式に利用するカメラIDを決定
S804
再構成に適したカメラによって取得された画像を基に3次元再構成を実施
S805
end

【図９】


91
92
43

【図１０】


(A)
1001
1002
1003
1004
1006
1005
1000
(B)
10005
10004
10001
10002
10003

フロントページの続き

ヤノン株式会社内

Ｆターム（参考） 5B050 AA08 BA09 BA12 DA07 EA07 EA27 EA28 FA02
5L096 CA05 FA34 FA67 FA69 GA06 HA11 KA01 KA04