

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号
特開2020-5202
(P2020-5202A)

(43) 公開日 令和2年1月9日(2020.1.9)

(51) Int.Cl.	F I	テーマコード (参考)
HO4N 13/111 (2018.01)	HO4N 13/111	5B050
GO6T 15/20 (2011.01)	GO6T 15/20 500	5B080
GO6T 19/00 (2011.01)	GO6T 19/00 A	5C061

審査請求 未請求 請求項の数 7 O L (全 13 頁)

(21) 出願番号	特願2018-125276 (P2018-125276)	(71) 出願人	000004352
(22) 出願日	平成30年6月29日 (2018. 6. 29)		日本放送協会
			東京都渋谷区神南2丁目2番1号
		(74) 代理人	100147485
			弁理士 杉村 憲司
		(74) 代理人	230118913
			弁護士 杉村 光嗣
		(74) 代理人	100161148
			弁理士 福尾 誠
		(74) 代理人	100185225
			弁理士 齋藤 恭一
		(72) 発明者	原 一宏
			東京都世田谷区砧一丁目10番11号 日 本放送協会放送技術研究所内

最終頁に続く

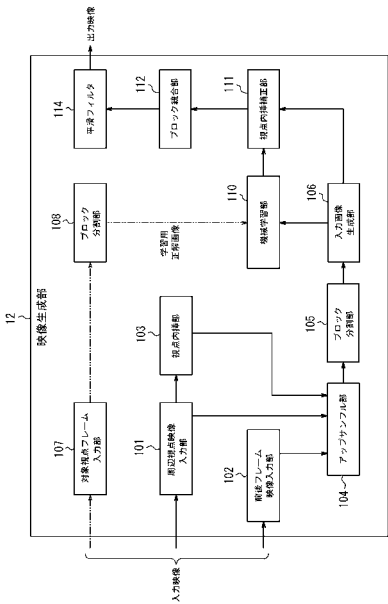
(54) 【発明の名称】映像処理装置

(57) 【要約】

【課題】映像作成時に不足する情報を適切に補完し、映像の品質を改善することのできる、映像処理装置を提供する。

【解決手段】映像処理装置は機械学習機能を持ち、学習モデル及び／又は学習済みパラメータをもとに、作成された映像の不足映像情報を補完する補正処理を行う。作成された映像は、多視点映像から作成された視点内挿映像であり、視点内挿映像と、前記視点内挿映像の周辺視点映像、及び前後フレーム映像を含む入力画像に基づいて、補正処理を行うことを特徴とする。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

機械学習機能を持ち、学習モデル及び／又は学習済みパラメータをもとに、作成された映像の不足映像情報を補完する補正処理を行う映像処理装置。

【請求項 2】

請求項 1 に記載の映像処理装置において、

学習モデル及び／又は学習済みパラメータを持つ補正部を備え、

前記補正部は、作成された映像とその関連映像を含む入力画像に基づいて、機械学習により補正処理を行うことを特徴とする、映像処理装置。

【請求項 3】

請求項 1 又は 2 に記載の映像処理装置において、

作成された映像は、多視点映像から作成された視点内挿映像であることを特徴とする、映像処理装置。

【請求項 4】

請求項 1 乃至 3 のいずれか一項に記載の映像処理装置において、

補正処理のための入力画像は、視点内挿映像と、前記視点内挿映像の周辺視点映像、及び前後フレーム映像をアップサンプリングし、ブロック化した $m \times m$ ピクセルのブロック画像を複数配置した構成であることを特徴とする、映像処理装置。

【請求項 5】

請求項 1 乃至 3 のいずれか一項に記載の映像処理装置において、

補正処理のための入力画像は、視点内挿映像と、前記視点内挿映像の周辺視点映像、及び前後フレーム映像をアップサンプリングし、ブロック化した $m \times m$ ピクセルのブロック画像を $n \times n$ 個配置した画像を、画素の並び替えにより、各ブロック画像の対応する位置のピクセルを集めた $n \times n$ ピクセルの要素画像を $m \times m$ 個配置した構成であることを特徴とする、映像処理装置。

【請求項 6】

請求項 1 乃至 5 のいずれか一項に記載の映像処理装置において、

作成された映像をブロックごとに機械学習により補正処理を行い、補正されたブロック画像をブロック統合し、出力映像を生成することを特徴とする、映像処理装置。

【請求項 7】

請求項 1 乃至 5 のいずれか一項に記載の映像処理装置において、

作成された映像をブロックごとに機械学習により補正処理を行い、補正されたブロック画像をブロック統合した映像を、視点内挿処理で作成された映像のインペイント処理に利用することを特徴とする、映像処理装置。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、映像処理装置に関し、特に、映像の補正処理に機械学習を利用する映像処理装置に関する。

【背景技術】**【0002】**

近年、HMD (Head Mounted Display) による 2 眼の 360 度ビデオをコンテンツとしたサービスが普及している。VR (Virtual Reality) ではユーザーが別の空間に移動したような没入感を得られることが知られる。これらのサービスでは将来のサービス展開として、ユーザーが姿勢を動かしたときの運動視差に対応することが考えられている (例えば、3DoF+: 3 degrees of freedom +)。そしてその先の展望として、ユーザーが姿勢だけではなく位置を移動したときの運動視差に対応するサービス (6DoF+) の実用化が期待されている。また、HMD 以外の表示端末を使用した将来の映像視聴サービスでは、スポーツ中継などでユーザーが見たい視点から映像を楽しむことができる自由視点映像の実用化に向けた開発も進められている。

10

20

30

40

50

【 0 0 0 3 】

これらのサービスの実現に向けて、コンテンツ制作方法の議論が進められている。例えば、自由視点映像サービスに向けた撮影手法では膨大な数の映像を一度に取得する必要があり、複数台のカメラで撮影空間を囲むように配置することが知られている。また、運動視差つき 360 度映像では全天周カメラを所定の位置に複数台配置する取り組みが進められている。

【 0 0 0 4 】

表示側では、撮影された映像をもとにレンダリングを行うことでユーザーが好む任意の視点映像を生成することが考えられている。これらのレンダリング処理では、撮影時に実際のカメラのない視点（仮想ビュー）を生成する視点合成処理、すなわち、カメラ間の視点映像の内挿処理、外挿処理、補間処理等が行われる。この視点合成処理は、インテグラル方式の立体映像等、様々な映像処理で用いられている（特許文献 1）。

10

【 0 0 0 5 】

また、編集などのコンテンツ制作では、3次元モデルシーンやイメージベース（多視点映像及び／又は多視点デプス画像）での編集作業が行われる。伝送と記録では、任意のフォーマットによって符号化が行われデータ量を圧縮した上での記録や、有線や無線ネットワークを介した配信が行われる。

【 先行技術文献 】

【 特許文献 】

【 0 0 0 6 】

20

【 特許文献 1 】 特開 2016 - 158213 号公報

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 7 】

広範囲な運動視差に対応した映像の記録・伝送を行う場合に、情報量が膨大になることから、全ての情報量を記録や伝送することはできず、一部の情報を符号化して記録・伝送する。そのため、復号側では不足した情報を補完することが必要となる。

【 0 0 0 8 】

例えばイメージベース（多視点映像及び／又は多視点デプス画像）のフォーマットでは、視点内挿処理による補間が行われる。ここで、視点内挿処理では奥行き画像（デプスマップ）を利用する方法が一般的に知られており、奥行き画像をもとにそれぞれの参照視点からの視点内挿画像を作成する。しかしながら、視点内挿画像は、参照する視点間の間隔が広い場合や、被写体の形状が複雑でカメラからの距離が近い場合に、多くのオクルージョン領域（陰になって見えない部分）が発生する。これらのオクルージョン領域はインペイント処理によって、周辺の画素などから埋められる。しかし、実際にどの画素値を割り当てるかの判定が難しいことから間違った画素値を埋めてしまうことがあり、それがノイズ成分として発生する。

30

【 0 0 0 9 】

従って、上記のような問題点に鑑みてなされた本発明の目的は、映像作成時に不足する情報を適切に補完し、映像の品質を改善することのできる、映像処理装置を提供することにある。

40

【 課題を解決するための手段 】

【 0 0 1 0 】

上記課題を解決するために本発明に係る映像処理装置は、機械学習機能を持ち、学習モデル及び／又は学習済みパラメータをもとに、作成された映像の不足映像情報を補完する補正処理を行うことを特徴とする。

【 0 0 1 1 】

また、前記映像処理装置は、学習モデル及び／又は学習済みパラメータを持つ補正部を備え、前記補正部は、作成された映像とその関連映像を含む入力画像に基づいて、機械学習により補正処理を行うことが望ましい。

50

【 0 0 1 2 】

また、前記映像処理装置は、作成された映像が、多視点映像から作成された視点内挿映像であることが望ましい。

【 0 0 1 3 】

また、前記映像処理装置は、補正処理のための入力画像が、視点内挿映像と、前記視点内挿映像の周辺視点映像、及び前後フレーム映像をアップサンプリングし、ブロック化した $m \times m$ (m は整数)ピクセルのブロック画像を複数配置した構成であることが望ましい。

【 0 0 1 4 】

また、前記映像処理装置は、補正処理のための入力画像が、視点内挿映像と、前記視点内挿映像の周辺視点映像、及び前後フレーム映像をアップサンプリングし、ブロック化した $m \times m$ ピクセルのブロック画像を $n \times n$ (n は整数)個配置した画像を、画素の並び替えにより、各ブロック画像の対応する位置のピクセルを集めた $n \times n$ ピクセルの要素画像を $m \times m$ 個配置した構成であることが望ましい。

【 0 0 1 5 】

また、前記映像処理装置は、作成された映像をブロックごとに機械学習により補正処理を行い、補正されたブロック画像をブロック統合し、出力映像を生成することが望ましい。

【 0 0 1 6 】

また、前記映像処理装置は、作成された映像をブロックごとに機械学習により補正処理を行い、補正されたブロック画像をブロック統合した映像を、視点内挿処理で作成された映像のインペイント処理に利用することが望ましい。

【 発明の効果 】

【 0 0 1 7 】

本発明の映像処理装置によれば、機械学習を利用して、映像作成時に不足する情報を適切に補完し、映像の品質を改善することができる。

【 0 0 1 8 】

また、本発明をデプス推定や視点内挿技術に適用すれば、オクルージョン領域(陰になって見えない部分)等におけるノイズの低減をすることができる。さらに、補完処理によってデプス画像や視点内挿画像の画質劣化が低減することにより、より離れたカメラ間隔での視点内挿が可能となり、そのため、符号化対象となる視点数(必要なカメラ台数)を削減することができる。

【 図面の簡単な説明 】

【 0 0 1 9 】

【 図 1 】 本発明の映像処理装置の一例のブロック図である。

【 図 2 】 本発明の第 1 の実施形態における映像生成部のブロック図である。

【 図 3 】 機械学習に用いる画像の一例である。

【 図 4 】 機械学習に用いる入力画像の別の例である。

【 図 5 】 本発明の第 2 の実施形態における映像生成部のブロック図である。

【 発明を実施するための形態 】

【 0 0 2 0 】

以下、本発明の実施の形態について、図面を参照しながら説明する。

【 0 0 2 1 】

(第 1 の実施形態)

図 1 は、本発明の映像処理装置の一例のブロック図である。本発明の映像処理装置は、映像の編集装置であってもよいし、また、映像の受信装置(復号装置)であってもよい。本発明の映像処理装置は、入力された映像から、新たな映像を生成するとともに、機械学習を用いて映像の補正を行う。映像処理装置 10 は、入力部 11、映像生成部 12、及び表示部 13 を備える。

【 0 0 2 2 】

映像処理装置 10 が映像の編集装置である場合には、入力部 11 には、撮影された映像が入力され、必要に応じて編集処理が行われる。また、映像処理装置 10 が映像の受信装置（復号装置）である場合には、入力部 11 には、受信した符号化データを復号した映像が入力される。或いは、入力部 11 内において、映像の符号化データを復号処理してもよい。入力される映像は、立体映像を構成するために複数のカメラで撮影された多視点映像、全天周カメラで撮影された映像等、複数の映像の組み合わせであってもよい。また、多視点映像に対応するデプス画像を含めることができる。

【0023】

映像生成部 12 は、入力部 11 から入力される映像を元に、学習用画像と正解画像情報を作成し、機械学習を行い、映像の補正に最適な学習モデル及び／又は学習パラメータを作成する。また、学習後は、入力部 11 から入力される映像から作成した映像を、機械学習により得られた学習モデル及び／又は学習パラメータ（学習済みパラメータ値）等に基づいて補正し、表示部 13 に出力する。映像生成部 12 の詳細は後述する。

10

【0024】

表示部 13 は、映像生成部 12 で処理された映像を表示する。なお、表示部 13 は、映像処理装置 10 の内部に設けられることは必須ではなく、映像処理装置 20 とは別体の独立した表示装置として構成してもよい。

【0025】

図 2 に、第 1 の実施形態としての映像処理装置 10 の映像生成部 12 のブロック図を示す。ここでは、図 2 の映像生成部 12 が、視点内挿映像を生成すると共に、機械学習を用いて視点内挿映像の補正処理（不足映像情報を補完する処理）を行う例を説明する。

20

【0026】

図 2 において、映像生成部 12 は、周辺視点映像入力部 101、前後フレーム映像入力部 102、視点内挿部 103、アップサンプル部 104、ブロック分割部 105、入力画像生成部 106、対象視点フレーム入力部 107、ブロック分割部 108、機械学習部 110、視点内挿補正部 111、ブロック統合部 112、及び平滑フィルタ 114 を備えている。

【0027】

入力対象となる映像は、例えば、符号化・復号処理をされた後の映像、及び／又は、映像編集された符号化処理を経ていない映像である。入力映像には、学習用の正解画像として視点内挿を行う視点の画像、視点内挿時に参照をされる周辺視点映像などが含まれている。例えば、複数枚の多視点画像とそれに対応する複数枚のデプス画像を入力映像とする。これらの画像は、対象物を 360 度の方向から撮影した多視点映像や、インテグラル立体映像を構成するための複数映像であってもよい。また、それぞれの視点映像は、予め奇数フレームと偶数フレームの映像に分かれていてもよい。

30

【0028】

周辺視点映像入力部 101 は、対象視点の映像（視点内挿映像）に対してその周辺視点の映像を選択し、視点内挿部 103 及びアップサンプル部 104 に出力する。

【0029】

前後フレーム映像入力部 102 は、対象視点フレーム（又は視点内挿が行われるフレーム）の前後のフレームの映像を選択し、アップサンプル部 104 に出力する。なお、対象視点のフレームの前後のフレームからのフレーム間内挿を利用することにより、予測画像の精度が向上する。

40

【0030】

視点内挿部 103 は、周辺視点映像入力部 101 から入力された周辺視点映像（対象視点の周辺の視点の映像）に基づいて、視点の内挿処理を行い、対象視点の仮想ビュー（映像）を作成し、その作成された映像をアップサンプル部 104 に出力する。

【0031】

アップサンプル部 104 は、周辺視点映像入力部 101、前後フレーム映像入力部 102、及び視点内挿部 103 から入力されたそれぞれの映像について、アップサンプリング

50

を行い、映像の解像度を高く（例えば、縦横 2 倍、又は 4 倍に）する。なお、映像をアップサンプリングして処理することにより、予測画像の精度が向上する。

【0032】

ブロック分割部 105 は、アップサンプル処理されたそれぞれの映像をブロックに分割する。ブロックとしては、例えば、 16×16 ピクセル (pixel) を単位ブロックとする。分割対象の映像は 2 倍にアップサンプル処理されているため、このブロックは、元の映像の 8×8 ピクセルのブロック画像に対応する。なお、単位ブロックのサイズは、上記に限定されるものではなく、例えば、 32×32 ピクセルとしてもよい。

【0033】

入力画像生成部 106 は、機械学習部 110 の入力画像を作成する。入力画像は、機械学習の学習用画像（予測の元になる画像）となる。ここでは、ブロック分割された各映像から、対象視点の映像に関連する複数のブロック映像（ブロック画像）をタイル状に並べて、入力画像を構成する。

【0034】

図 3 に、機械学習に用いる画像の一例を示す。図 3 (A) は、入力画像の例である。ここでは 16×16 ピクセル (pixel) のブロック画像が 9 個 (3×3) 配列され、 48×48 ピクセルの入力画像が構成されている。まず、視点補間画像（対象視点の映像に対応する視点内挿画像）を入力画像の中央部に配置する。この視点補間画像は、視点内挿部 103 で作成された対象視点の映像をアップサンプリング（2 倍）してブロック化したものである。また、視点補間画像の周囲には、その関連画像（関連映像）である、隣接画像 1 ~ 4 及び対象視点の映像フレームの前後フレームの映像のブロックを配置する。例えば、隣接画像 1 ~ 4 は、視点補間画像の上下左右に隣接する視点のブロック画像であり、周辺視点映像入力部 101 の出力をアップサンプリングしてブロック化したものから適宜選択して配置する。また、前後フレームの映像としては、例えば、1 フレーム前の画像、1 フレーム後の画像、3 フレーム前の画像、3 フレーム後の画像等であり、前後フレーム映像入力部 102 の出力をアップサンプリングしてブロック化したものから適宜選択して配置する。なお、隣接画像や前後フレーム画像をそれぞれ幾つ選び、どのように配置するか等は、適宜設定することができる。隣接画像としてデプス画像を利用してもよい。

【0035】

次に、対象視点フレーム入力部 107 は、入力映像から、対象視点のフレームの映像を選択し、ブロック分割部 108 に出力する。なお、対象視点のフレームの映像は編集された符号化前の映像に限定されるものではなく、符号化され更に復号された映像から、対象視点の映像を選択してもよい。対象視点フレーム入力部 107 から入力された映像からは、機械学習における学習用正解画像が作成される。

【0036】

ブロック分割部 108 は、対象視点フレーム入力部 107 から入力された対象視点フレームの映像を、ブロックに分割する。ブロックとしては、例えば、 8×8 ピクセル (pixel) を単位ブロックとする。対象視点フレーム入力部 107 からの映像はアップサンプル処理されていないため、このブロックは、入力画像の 16×16 ピクセルのブロック画像に対応する。なお、単位ブロックは、上記のサイズに限定されるものではなく、学習用画像のサイズと対応させればよい。ブロック分割部 108 で分割され出力された映像が、9 ブロックの入力画像の中央にある 16×16 ピクセルの視点補間画像（視点内挿画像）に対応する正解画像となっている。図 3 (B) に、正解画像の一例（イメージ）を示す。

【0037】

機械学習部 110 は、入力画像生成部 106 からの入力画像を学習用画像（訓練用画像）とし、ブロック分割部 108 から入力された正解画像との比較を行うことで、機械学習を進め、最適な学習モデルと学習パラメータを算出する。なお、ここでの機械学習は広い概念でとらえて良い。例えば、畳み込みニューラルネットワークを用いて構築される機械学習であって良く、特徴量の抽出とモデリングを自動的に行うディープラーニング等を含むことができる。また、畳み込みニューラルネットワーク以外に、例えば、SVM (Supp

10

20

30

40

50

ort Vector Machine) などを用いて構築してもよい。

【0038】

学習により、最適な学習モデルと学習パラメータを習得した機械学習部110は、図3(A)に示す入力画像が入力されると、正解画像に近似した補間画像を出力することができる。図3(C)に、出力画像(補間画像)の一例(イメージ)を示す。出力画像(補間画像)は、正解画像と同じく、8×8ピクセル(pixel)の単位ブロックである。

【0039】

なお、学習用画像に対象視点を内挿して作成した映像を用い、正解画像に対象視点における実際の映像を用いることにより、機械学習部は、内挿映像で不足する映像情報(例えば、オクルージョン領域の映像情報)を生成する学習機能を得ることができる。また、学習用画像に符号化及び復号された映像を用い、正解画像に符号化前の編集映像を用いることにより、機械学習部は、符号化で失われた映像情報(画質改善のための映像情報)を生成する学習機能を得ることができる。

10

【0040】

学習を終了すると、機械学習部110は、最適な学習モデル及び/又は学習パラメータを視点内挿補正部111に出力する。機械学習部110が学習を終了し、最適な学習モデルと学習パラメータを視点内挿補正部111に出力した後は、対象視点フレーム入力部107及びブロック分割部108からの入力信号(学習用正解画像)は停止してよい。或いは、学習モデルと学習パラメータを出力後も、機械学習部110は継続的に学習を行い、より精度の高い学習モデルと学習パラメータを算出して、視点内挿補正部111の学習モデルと学習パラメータを更新してもよい。

20

【0041】

視点内挿補正部111は、機械学習部110からの最適の学習モデル及び/又は学習パラメータに基づいて、入力画像生成部106から出力される入力画像を受けて機械学習により視点内挿画像の補正をブロック毎に行う。視点内挿補正部111の出力は、例えば、図3(C)に示される8×8ピクセルの補間画像(補正画像)である。補正された画像(映像)は、作成された映像で不足する映像情報を補完することができる。

【0042】

ブロック統合部112は、視点内挿補正部111でブロック毎に生成された補正画像(ブロック画像)を統合(ブロック統合)し、一枚の補正された視点内挿映像(対象視点の映像)を生成する。なお、補正された視点内挿映像は、機械学習の結果(学習モデル結果)により、オクルージョン領域が予測され補完されている。

30

【0043】

平滑フィルタ114は、ブロック統合部112から入力された映像に対して、フィルタ処理を行う。ブロック境界に生じるノイズを平滑フィルタ処理により除去することができる。この平滑フィルタ114の出力が、映像生成部12の出力映像となる。

【0044】

このように、学習モデル結果(機械学習による予測)を使用することにより、視点内挿映像の精度を向上させることができる。また、補正された画像は符号化・復号処理で失った画像情報も補完するから、出力映像の画質を向上させることができる。

40

【0045】

(入力画像の改良)

図4に、入力画像の別の例を示す。入力画像生成部106は、まず、アップサンプリングされてブロック分割された各映像から、対象視点の映像に関連する複数のブロック映像(ブロック画像)をタイル状に並べて、図4の左側の画像を構成する。この左側の画像は、図3(A)と同様であり、ここでは16×16ピクセル(pixel)のブロック画像が9個(3×3)配列され、48×48ピクセルの画像が構成される。まず、視点補間画像(視点内挿部103で作成された対象視点の映像をアップサンプリングしてブロック化したもの)を中央部に配置し、その周囲には、隣接画像1~4及び対象視点の映像フレームの前後フレームのブロック画像を配置する。例えば、隣接画像1~4は、視点補間画像の上

50

下左右に隣接する視点のブロック画像であり、前後フレームの映像としては、例えば、1フレーム前の画像、1フレーム後の画像、3フレーム前の画像、3フレーム後の画像等である。これらは、周辺視点映像入力部101及び前後フレーム映像入力部102の出力をアップサンプリングしてブロック化したものから適宜選択して配置する。なお、隣接画像や前後フレーム画像をそれぞれ幾つ選び、どのように配置するか等は、適宜設定することができる。隣接画像としてデプス画像を利用してもよい。

【0046】

次に、画素（ピクセル）の並び替えを行う。各 16×16 ピクセル（ $m \times m$ ピクセル）の 3×3 個（ $n \times n$ 個）のブロック画像から同じ位置（例えば、左上）にある1ピクセル（pixel）を抽出して、 3×3 ピクセルのピクセル集合を作成し（以下、これを「要素画像」と呼ぶ。）、これを 48×48 ピクセルの対応する位置（左上の位置）に配置する。また、9個のブロック画像それぞれの次のピクセル（例えば、上の行の左から2番目の1ピクセル）を抽出し、 3×3 ピクセルのピクセル集合（要素画像）として、前の要素画像の右隣の位置に配置する。同様の並び替えを繰り返し、各ブロック画像から同じ位置にある1ピクセルを抽出して要素画像を作成し、対応する位置に順に配置する。こうして 3×3 ピクセル（ $n \times n$ ピクセル）の要素画像が 16×16 個（ $m \times m$ 個）配置され、 48×48 ピクセルの新たな入力画像（図4の右側）を作成することができる。

【0047】

この新たな入力画像を、機械学習部110の学習用画像（訓練用画像）とする。なお、学習用正解画像は、ブロック分割部108から入力されるものであり、同じとする。この場合、機械学習モデルにおける、畳み込み処理の中で、それぞれの畳み込み層の中に全ての映像の成分が含まれることとなるため、補正の精度が向上する。学習終了後は、この新たな入力画像を、視点内挿補正部111の入力画像とする。

【0048】

（第2の実施形態）

図5に、第2の実施形態としての映像生成部12のブロック図を示す。ここでは、映像生成部12が、視点内挿映像を生成すると共に、インペイント処理で割り当てる画素値を機械学習結果によって得る映像生成部12の例を説明する。第2の実施形態では、学習方法は第1の実施形態と同じだが、インペイント部において、視点内挿時のオクリュージョン領域のみを、機械学習で得られた視点内挿補正結果の画素値で置き換えられることが異なる。

【0049】

図5において、映像生成部12は、周辺視点映像入力部101、前後フレーム映像入力部102、視点内挿部103、アップサンプル部104、ブロック分割部105、入力画像生成部106、対象視点フレーム入力部107、ブロック分割部108、機械学習部110、視点内挿補正部111、ブロック統合部112、インペイント部113、及び平滑フィルタ114を備えている。図2と共通の構成は、説明を簡略化する。

【0050】

映像生成部12の入力対象となる映像は、符号化・復号処理をされた後の映像、及び/又は、映像編集された符号化処理を経ていない映像である。例えば、複数枚の多視点画像とそれに対応する複数枚のデプス画像を入力映像とする。

【0051】

周辺視点映像入力部101は、対象視点の映像（視点内挿映像）に対してその周辺視点の映像を選択し、視点内挿部103及びアップサンプル部104に出力する。

【0052】

前後フレーム映像入力部102は、対象視点フレーム（又は視点内挿が行われるフレーム）の前後のフレームの映像を選択し、アップサンプル部104に出力する。

【0053】

視点内挿部103は、周辺視点映像入力部101から入力された周辺視点映像（対象視点の周辺の視点の映像）に基づいて、視点の内挿処理を行い、対象視点の仮想ビュー（映

10

20

30

40

50

像)を作成する。この映像は、第2の実施形態においては映像生成部12の出力映像の基礎となる映像であり、その作成された映像をアップサンプル部104とインペイント部113に出力する。

【0054】

アップサンプル部104は、周辺視点映像入力部101、前後フレーム映像入力部102、及び視点内挿部103から入力されたそれぞれの映像について、アップサンプリングを行い、映像の解像度を高くする。なお、映像をアップサンプリングして処理することにより、予測画像の精度が向上する。

【0055】

ブロック分割部105は、アップサンプル処理されたそれぞれの映像をブロックに分割する。ブロックとしては、例えば、 16×16 ピクセル(pixel)を単位ブロックとする。なお、単位ブロックのサイズは、例えば、 32×32 ピクセルとしてもよい。

【0056】

入力画像生成部106は、機械学習部110の入力画像を作成する。入力画像は、ブロック分割された各映像から、対象視点の映像に関連する複数のブロック映像をタイル状に並べた、図3(A)の入力画像として良い。また、ブロック映像からなる画像の画素を並び変えた、図4(右側)の入力画像としてもよい。

【0057】

次に、対象視点フレーム入力部107は、入力映像から、対象視点のフレームの映像を選択し、ブロック分割部108に出力する。

【0058】

ブロック分割部108は、対象視点フレーム入力部107から入力された対象視点フレームの映像を、ブロックに分割する。ブロックとしては、例えば、 8×8 ピクセル(pixel)を単位ブロックとする。ブロック分割部108で分割され出力された映像が、 16×16 ピクセルの視点補間画像(視点内挿画像)に対応する正解画像となっている。

【0059】

機械学習部110は、入力画像生成部106からの入力画像を学習用画像(訓練用画像)とし、ブロック分割部108から入力された正解画像との比較を行うことで、機械学習を進め、最適な学習モデルと学習パラメータを算出する。

【0060】

学習を終了すると、機械学習部110は、最適な学習モデル及び/又は学習パラメータを視点内挿補正部111に出力する。機械学習部110が学習を終了し、最適な学習モデルと学習パラメータを視点内挿補正部111に出力した後は、対象視点フレーム入力部107及びブロック分割部108からの入力信号(学習用正解画像)は停止してもよい。

【0061】

視点内挿補正部111は、機械学習部110からの最適の学習モデル及び/又は学習パラメータに基づいて、入力画像生成部106から出力される入力画像を受けて機械学習により視点内挿画像の補正をブロック毎に行う。

【0062】

ブロック統合部112は、視点内挿補正部111でブロック毎に生成された補正画像(ブロック画像)を統合し、一枚の補正された視点内挿映像(対象視点の映像)を生成する。なお、補正された視点内挿映像は、機械学習の結果(学習モデル結果)により、オクルージョン領域が予測され補完されている。

【0063】

インペイント部113には、視点内挿部103からの対象視点の内挿映像と、ブロック統合部112からの補正された視点内挿映像とが入力される。視点内挿部103で生成された内挿映像は、オクルージョン領域等を生じる可能性があり、インペイント部113は、ブロック統合部112からの補正された映像により、このオクルージョン領域を埋める補完処理を行う。すなわち、視点内挿部で発生したオクルージョン領域を埋めるためのインペイント領域の画素値に、学習モデル結果を使用することができる。補完された映像は

10

20

30

40

50

、平滑フィルタ 1 1 4 に出力される。

【 0 0 6 4 】

平滑フィルタ 1 1 4 は、インペイント部 1 1 3 からの入力された映像に対して、フィルタ処理を行う。入力された映像は、ブロック統合処理やインペイント処理により、ブロックや領域の境界にノイズ（つなぎ目）を生じやすいが、平滑フィルタ処理によりこのノイズを除去することができる。この平滑フィルタ 1 1 4 の出力が、映像生成部 1 2 の出力映像となる。

【 0 0 6 5 】

このように、視点内挿で得られた映像のオクルージョン領域を埋めるためのインペイント領域の画素値に、学習モデル結果（機械学習による予測）を使用することにより、視点内挿映像の精度を向上させることができる。

10

【 0 0 6 6 】

上記の実施の形態では、映像処理装置 1 0 の構成と動作について説明したが、本発明はこれに限らず、機械学習を用いた映像の処理方法として構成されてもよい。

【 0 0 6 7 】

なお、上述した映像処理装置 1 0 として機能させるためにコンピュータを好適に用いることができ、そのようなコンピュータは、映像処理装置 1 0 の各機能を実現する処理内容を記述したプログラムを該コンピュータの記憶部に格納しておき、該コンピュータの CPU によってこのプログラムを読み出して実行させることで実現することができる。なお、このプログラムは、コンピュータ読取り可能な記録媒体に記録可能である。

20

【 0 0 6 8 】

上述の実施形態は代表的な例として説明したが、本発明の趣旨及び範囲内で、多くの変更及び置換ができることは当業者に明らかである。したがって、本発明は、上述の実施形態によって制限するものと解するべきではなく、特許請求の範囲から逸脱することなく、種々の変形や変更が可能である。例えば、実施形態に記載の複数の構成ブロックを 1 つに組み合わせたり、あるいは 1 つの構成ブロックを分割したりすることが可能である。

【 符号の説明 】

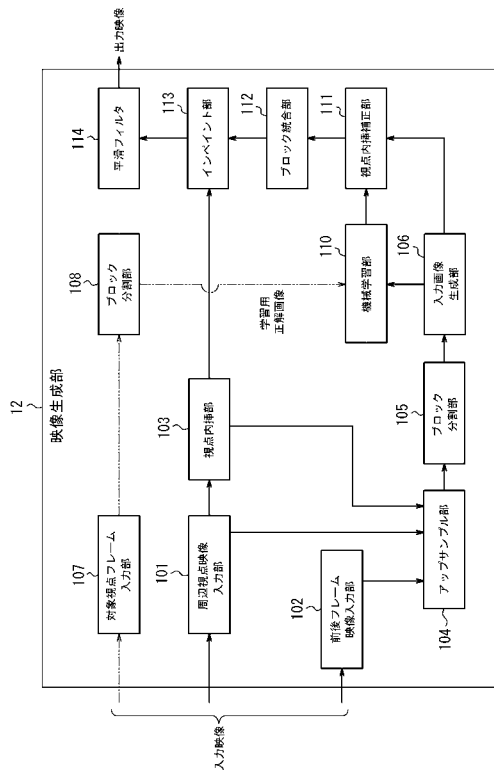
【 0 0 6 9 】

- 1 0 映像処理装置
- 1 1 入力部
- 1 2 映像生成部
- 1 3 表示部
- 1 0 1 周辺視点映像入力部
- 1 0 2 前後フレーム映像入力部
- 1 0 3 視点内挿部
- 1 0 4 アップサンプル部
- 1 0 5 ブロック分割部
- 1 0 6 入力画像生成部
- 1 0 7 対象視点フレーム入力部
- 1 0 8 ブロック分割部
- 1 1 0 機械学習部
- 1 1 1 視点内挿補正部
- 1 1 2 ブロック統合部
- 1 1 3 インペイント部
- 1 1 4 平滑フィルタ

30

40

【図 5】



フロントページの続き

- (72)発明者 井口 義則
東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
- (72)発明者 河北 真宏
東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
- (72)発明者 三科 智之
東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
- (72)発明者 菊池 宏
東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
- (72)発明者 妹尾 孝憲
東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
- (72)発明者 藤井 俊彰
愛知県名古屋市千種区不老町1番 国立大学法人名古屋大学内
- Fターム(参考) 5B050 BA04 BA11 DA07 EA19 EA27 FA02 FA05
5B080 BA02 FA08
5C061 AA29 AB12