

## 研究論文

## SIGMA Retriever：イメージベースドモデリング・レンダリングのための検索機構の設計と実装

山崎 賢人<sup>1,a)</sup> 关 斯琨<sup>1</sup> 松木 輝<sup>1</sup> 木村 朝子<sup>1</sup> 柴田 史久<sup>1</sup>

受付日 2021年7月15日, 採録日 2021年11月26日

**概要：**本論文では、多視点幾何を用いたコンテンツ生成手法であるイメージベースドモデリング・レンダリング (IBMR) のための検索機構 “SIGMA Retriever” の設計について述べる。将来、いたるところにカメラが設置され、常時、実世界を撮影するようになることが予想される。撮影された画像をオープンデータとして活用するためには、利用目的に合った画像を抽出する手法が必要である。こうした背景から、我々は、一般的な IBMR を概観し、これらの処理の流れを抽象化したうえで、検索手法を体系的にまとめた “SIGMA Retriever” を設計し、SIGMA フレームワークに実装した。“SIGMA Retriever” を用いて鎌倉駅旧駅舎時計台を撮影した画像を蓄積したデータベースから IBMR の例である 3D モデルやパノラマ画像の生成に必要な画像を抽出できることを確認した。

**キーワード：**多視点幾何, 視錐台カリング, メタデータ, イメージベースドモデリング・レンダリング

## SIGMA Retriever: Design and Implementation of Retrieval Mechanism for Image-based Modeling and Rendering

KENTO YAMAZAKI<sup>1,a)</sup> GUAN SIKUN<sup>1</sup> AKIRA MATSUKI<sup>1</sup> ASAKO KIMURA<sup>1</sup> FUMIHISA SHIBATA<sup>1</sup>

Received: July 15, 2021, Accepted: November 26, 2021

**Abstract:** In this paper, we describe the design of a retrieval mechanism “SIGMA Retriever” for image-based modeling and rendering (IBMR), a content creation method using multi-view geometry. In the future, it is expected that numerous cameras will be installed everywhere to capture real-world images at all times. In order to utilize the captured images as open data, it is necessary to have a method to extract images suitable for the purpose of use. From these backgrounds, we have reviewed general IBMR methods and abstracted these processing flows. After that, we have designed “SIGMA Retriever” that systematically summarized retrieval methods suitable for IBMR methods, and implemented it in the SIGMA framework. We have confirmed that “SIGMA Retriever” can extract images necessary for generating 3D models and panoramic images, which are examples of IBMR, from a database of images taken of the clock tower of the old Kamakura station.

**Keywords:** multi view geometry, view frustum culling, metadata, image-based modeling and rendering

## 1. はじめに

いたるところに設置された監視カメラや自動車の車載カメラなどによって、実世界を観測した映像が常時記録される社会が訪れようとしている。近年活躍が期待されて

いる低空を安定して飛行できる UAV (Unmanned Aerial Vehicle) などの自律型移動体にもカメラは欠かせない。今後も我々の暮らす社会に存在するカメラの数は、増加の一途をたどるであろう (図 1)。

現状、これらのカメラから得られる映像はクローズドなデータとして、カメラの所有者などの特定の組織でのみ活用されている。しかし、近年ではオープンデータ化の機運が高まっている。国土交通省主導の 3D 都市モデルの整備・

<sup>1</sup> 立命館大学大学院情報理工学研究所  
Graduate School of Information Science and Engineering,  
Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

a) yamazaki@rm2c.ise.ritsumeai.ac.jp

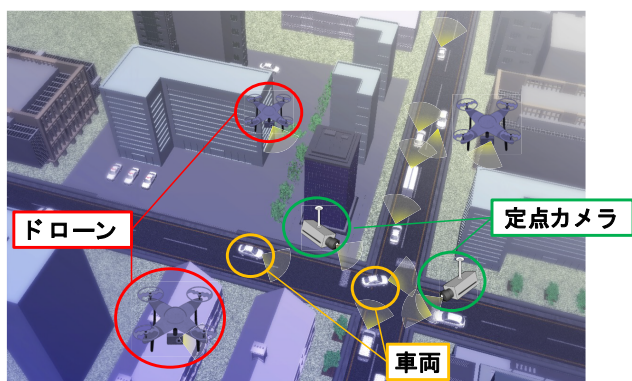


図 1 遍在したカメラのイメージ

Fig. 1 Image of omnipresent cameras.

オープンデータ化プロジェクト「PLATEAU」もその一例である [1].

我々は近い将来、街中に遍在するカメラから得られる映像をオープンデータ化する時代が来ると予想している。こうした時代を見据え、これらの映像をオープンデータとして収集・蓄積するアプリケーションフレームワーク SIGMA (Spatiotemporal Images with Generalized Management Architecture) の研究開発に着手した [2]。しかし、アプリケーションフレームワーク開発は広範な研究開発である。そこで本論文では SIGMA フレームワークの核となる機構の 1 つとして、データベース (Database : DB) に蓄積された画像から所望の画像を検索する仕組みについて議論する。

街中に遍在するカメラから得られる映像は、単独で使用するよりも、多視点の映像を組み合わせるほうが、利用の幅が広がる。多視点映像の利用方法には、物体認識など、様々なものがあるが、本研究では、多視点幾何 (Multi View Geometry : MVG) を用いたデジタルコンテンツ生成手法であるイメージベースドモデリング・レンダリング (Image-based Modeling and Rendering : IBMR) に焦点を絞ることとした [3].

IBMR では、複数視点の画像に基づいて 3D モデルを生成したり、広視野画像や仮想視点画像など撮影画像とは異なった風景を描画することが可能である。しかしながら、必ずしも入力画像の数に比例して、高精度なコンテンツが生成されるとは限らない。さらに、本研究で対象とする画像には、画像は監視カメラのように固定されたカメラで撮影した画像だけではなく、UAV や自動車などに搭載したカメラで撮影した画像も含むことを想定すると、たとえ、撮影位置が既知の画像を街中から大量に収集・蓄積できていたとしても、その中から利用目的に合った画像を得られなければ、利用者のニーズを満たすことはできない。

一般に、IBMR を利用する場面は千差万別であり、1 つの検索手法のみですべての利用目的に合致した画像を検索することは難しい。たとえば、Agarwal らは写真共有サイト Flickr で “Roma” を検索キーとして抽出した画像を

用いてローマの街並みの 3D モデルを作成した [4]。Flickr は 3D モデルを生成するために集めた画像がアップロードされているわけではないため、抽出した画像には 3D モデルの生成には不適切な画像も多く含む。そこで、Agarwal らは SfM (Structure from Motion) の処理を改善し、3D モデルの生成に適した画像のみを抽出している。このように、利用目的が一意であれば、それに対応する検索手法も一意に決めることができる。しかし、本研究のように 2 つの抽象的な条件「多種多様なカメラが遍在」「IBMR で利用」のみを制約とした場合、様々な利用目的に対して網羅的に検索できるような機能を準備することは困難である。したがって、少ない制約条件下で、広範な用途で有用な手法を提案するためには、利用目的や条件を抽象化する必要がある。そこで、本研究では、利用目的である IBMR について概観し、処理の流れを抽象化した上で、検索手法を体系的にまとめるとともに、検索キーの保存方法について検討することとした。こうして開発した検索機構を “SIGMA Retriever” と称する。

以降、2 章では関連研究について、3 章では本研究で設計した検索機構の詳細について述べ、4 章では検索機構の評価結果について考察する。続く 5 章において、まとめと今後の展望を述べる。

## 2. 関連研究

SfM や v-SLAM (Visual Simultaneous Localization and Mapping) [5] では現実空間にある物体から点群で表す 3D モデルを生成するだけでなく、各画像の位置姿勢を推定できる。3D モデルの生成例は机上の模型から、文献 [4] で例示されているようなローマの街並み、さらには航空機から地上に向けて撮影した画像を用いた数値表層モデルに至るまで [6]、様々な生成対象の大きさや、対象物までの距離が想定される。

映画『マトリックス』で有名になった Bullet time では、被写体を囲むようにカメラを配置し、仮想的に動かしたいカメラの位置にある映像を切り替えていくことで、被写体はスローモーションになるが、カメラは高速に動いているような映像を生成している。日産自動車のアラウンドビュー [7] では俯瞰映像を、Google Street View では広視野な画像を多視点カメラの映像から生成している。これらは仮想投影面上で、複数視点の画像を縫い合わせるように接合するため、Image/Photo Stitching と呼ばれている。この手法では各カメラの位置姿勢から任意の面に映像を投影し、映像の重複領域をブレンド処理することで、現実には存在しない映像を生成している [8]。Light Field Rendering では、3D モデルを生成せずに、多視点画像の光線を DB に蓄積することで、仮想カメラの光線を補間し、自由視点映像を生成している [9].

IBMR では、カメラ位置だけではなく、撮影時刻につい

て考慮しているような場合もある。たとえば、映像の中から任意の物体を視覚的に消去する隠消現実感 (Diminished Reality : DR) では、他視点の映像を用いて、消去対象の背景映像を生成し、この背景映像を物体に重畳するが [10], 隠背景に写っているものが非静止物の場合、撮影時刻が異なる画像から生成した隠背景では不都合が生じる場合がある。また、撮影時刻の異なる画像群から街並みの時間的な変化を推定して 3D モデル化する 4D Cities Project もある [11].

ここで述べたように、一口に IBMR といっても、その目的や用途は大きく異なっており、そこで必要とされる画像についても多様であることが分かる。

### 3. SIGMA Retriever の設計

#### 3.1 処理の流れの抽象化

2章で列挙した様々な手法をもとに IBMR の処理の流れを抽象化すると、図 2 に示す 4つのステップ「撮影」「メタデータの推定」「実世界の計測」「描画」に分けられる。

「撮影」とは、カメラで実世界を撮影することである。このとき、撮影画像には GNSS で取得したカメラの位置、撮影日時や焦点距離など、撮影時のカメラの状態を表すメタデータである Exif (Exchangeable image file format) タグが付与される場合がある。

「メタデータの推定」とは、カメラの位置姿勢や内部パラメータなどを推定することである。本研究では、チェスボードなどによって「撮影」前にキャリブレーションしたものや、Exif タグのようにセンサが付加したものを計測値と定義し、このステップでは、画像や計測値を初期値として使用することで、高精度なメタデータを推定する。撮影時のメタデータが高精度であれば、推定値を計算しないこともある。たとえば、SfM において、カメラキャリブレーション済みのカメラを用いる場合は、内部パラメータは計算対象にならない。

「実世界の計測」とは、カメラ位置と画像を基に実世界を 3次元的に計測することである。SfM や v-SLAM では、計測結果を点群で表すが、この点群が SIFT などの特徴点に基づく場合、疎な点群と呼ばれる。より正確な実世界の情報である密な点群やメッシュを生成するためには、Patch Match Stereo [12] など、目的に合わせた多視点ステレオ (Multi View Stereo : MVS) が必要となる。さらに、SfM 中の一処理であるバンドル調整では、カメラの焦点距離や位置姿勢とともに点群の座標値を最適化するため、「メタデータの推定」と「実世界の計測」は両方向の矢印としている。

「描画」では、利用者に提示するため、計測した実世界を描画面に投影することである。ただし、Image/Photo Stitching などは、実世界の計測自体はしていないため、「メタデータの推定」から直接「描画」に進む矢印も追加した。

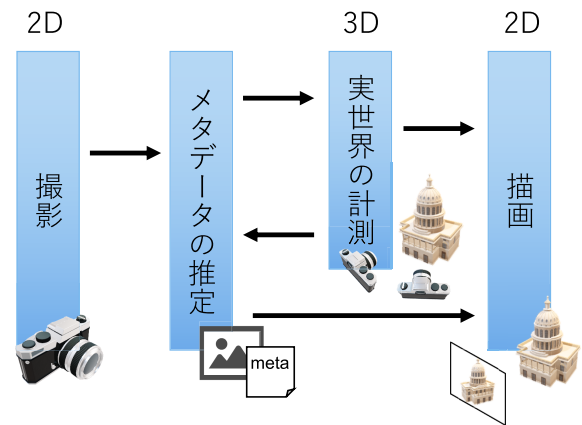


図 2 抽象化した IBMR の処理の流れ  
Fig. 2 Flowchart of abstracted IBMR.

#### 3.2 設計方針

3.1 節で整理した処理の流れから、IBMR には、「メタデータの推定」「実世界の計測」「描画」の 3つのタイミングで画像の入力が必要となる。これらの画像を多くの画像の中から抽出するためには検索条件が必要となるが、画像そのものを検索キーとして DB から検索することはできないため、画像を撮影したカメラのメタデータはすでに存在することが前提となる。

「実世界の計測」では計測対象を撮影している画像が抽出対象であり、「描画」では描画対象を撮影している画像が抽出対象である。「実世界の計測」は対象の実世界 (被写体) が写っているかを画像中から認識する必要がある。認識する被写体が事前に決まっていれば、アップロードされた画像に検出した物体をラベル付けし、メタデータに含むことができるかもしれない。しかし、広範な用途が想定される本研究の検索において、検索に合わせたラベル付けを設計することは難しい。そこで、実世界において、対象となる領域を指定し、その対象が撮影されている可能性があるかを、カメラの位置姿勢や画角などを基に計算することで判断する。

大量に蓄積した全画像に対して検索すると、たとえば、神奈川県鎌倉駅前を対象領域に指定した場合、どこで撮影した画像であったとしても、鎌倉駅前の方向を向いていれば、抽出される可能性があるため、全画像を対象にすることは現実的ではない。したがって、検索対象となる画像は空間上で制限を設ける必要がある。

本研究が想定しているカメラは街中に遍在しているだけでなく、時間とともに移動する場合がある。したがって、画像を検索する際には、検索対象の画像が空間軸上だけでなく、時間軸上にも撮影時のカメラが遍在することを考慮する必要がある。手順としては、時空間軸に基づく絞り込みを行ったうえで、カメラ画角などの内部パラメータを用いて検索する。「描画」も同様に、描画対象が写っているかを内部パラメータを利用して検索する。「メタデータの



推定」は「実世界の計測」「描画」で使用する画像のメタデータを抽出することが必要条件であるが、十分な条件は一意に決められない。そのため、時空間軸を利用して検索するしかない。

ここまでの方針において必要なデータは、カメラの位置姿勢、撮影時刻、焦点距離、主点、解像度、歪み係数である。本研究では、時空間軸を表すカメラの位置姿勢、撮影時刻を外部パラメータ、焦点距離、主点、解像度、歪み係数を内部パラメータと定義し、これらのデータをメタデータとして、DBに画像とともに蓄積する。

### 3.3 検索手法の提案

DBに蓄積した画像を抽出するためには、画像そのものをIDなどで指定か、画像のメタデータを基に検索する手法がある。街中に遍在するカメラの画像を検索するには、外部パラメータをキーとして検索するのが一般的な手法である。しかし、外部パラメータを利用して検索できるのは、カメラの位置や撮影時刻であり、この手法で検索した画像中に任意の被写体が撮影しているとは限らない。たとえば、画像中の被写体は、カメラの位置から数m離れた位置に存在することが一般的である。3.2節で述べたとおり、外部パラメータで検索対象を絞ったうえで、任意の領域を撮影しているかを、画角や焦点距離などの内部パラメータに基づいて検索する。

任意の領域とは「実世界の計測」「描画」から、「空間で指定できる領域」「任意の視点から見た領域」の2つから定義できる。これらの領域を検索するために、一般的な検索手法に限定せず、目的に合致する手法を選択することが重要である。「空間で指定できる領域」は一般的に、空間軸で多面体を指定できるため、その多面体が各画像を撮影したカメラで撮影できるかを判定するには、視錐台カリングを用いる手法を提案する（視錐台カリングに基づく検索）。しかし、各画像に対して視錐台カリングを用いると計算量が大きい。また、指定した多面体が遠くにある場合は、計算量を削減するため、多面体を点で代替する「注視点に基づく検索」を提案する。「任意の視点から見た領域」は仮想カメラと検索対象のカメラとのエビポラ幾何に基づいて検索する。

本節では、図3に示すとおり、「外部パラメータをキーとした検索手法」と「内部パラメータをキーとした検索手法（視錐台カリングに基づく検索、注視点に基づく検索、エビポラ幾何に基づく検索）」について述べる。

#### 3.3.1 外部パラメータをキーとした検索手法

この検索手法でキーとする外部パラメータは3.2節で定義した時空間軸である。具体的な手法としては、各座標軸における任意の領域を指定し検索する。領域は、各座標軸の中心座標値  $(x, y, z, t)$  と範囲  $(a, b, c, d)$  を指定する。その検索方法の空間軸は立方体となるため、範囲は立方体の

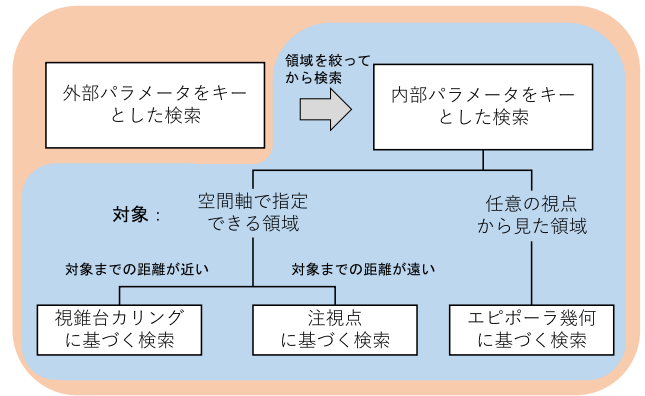


図3 提案手法の分類

Fig. 3 Classification of proposed method.

1辺の長さとなる。また、中心座標値からの距離に応じた領域も指定可能とし、検索値をそのままに次の式(1)が成り立つ検索方法も提案する。式(1)の場合、空間軸は楕円球となる。

$$\frac{x^2}{\left(\frac{a}{2}\right)^2} + \frac{y^2}{\left(\frac{b}{2}\right)^2} + \frac{z^2}{\left(\frac{c}{2}\right)^2} \leq 1 \quad (1)$$

指定できる領域は1つだけではなく、複数指定でき、それらは積集合や和集合として扱うことも可能である。

時間軸は空間軸とは異なり周期性があることから、直線的な側面と円環的な側面を合わせ持つ。たとえばUNIX時間のように、任意の時点からの経過時間で表すシステム時刻は直線的である。一方、「正午からの前後1時間のみ」は円環的な表現である。検索においても、円環の一部を指定したい場合もあるため、周期性を満たす検索値も使用可能とした。

#### 3.3.2 内部パラメータをキーとした検索手法

##### 【視錐台カリングに基づく検索】

3.3節で述べたとおり、本手法は視錐台カリングを用いる。視錐台カリングは、CGI (Computer Generated Imagery) 生成時に視錐台内に3Dモデルが含まれているかを判定することで、描画処理を減らす手法である。しかし、視錐台カリングは計算量が多いことから、多面体に軸平行境界ボックス (Axis-Aligned Bounding Box: AABB) を用いて計算量を削減することが多い。3Dモデルを囲むAABBを指定し、視錐台の各面とAABBとの関係を計算する[13]。図4(a)に示すとおり、AABBの各頂点から、視錐台の任意の面の法線ベクトルの向きに最も近い対角線を形成する2頂点を選ぶ(視錐台カリングにおいて、面の法線ベクトルは視錐台の内側を向いている方が正)。任意の面から選択した2頂点への法線ベクトルの大きさを計算し、大きい方をp-vertex、小さい方をn-vertexと称する。各面に対するp-vertexがすべて正の場合は視錐台内にAABBが含まれていると判定できる。ただし、本検索の対象はカメラ画像であるた



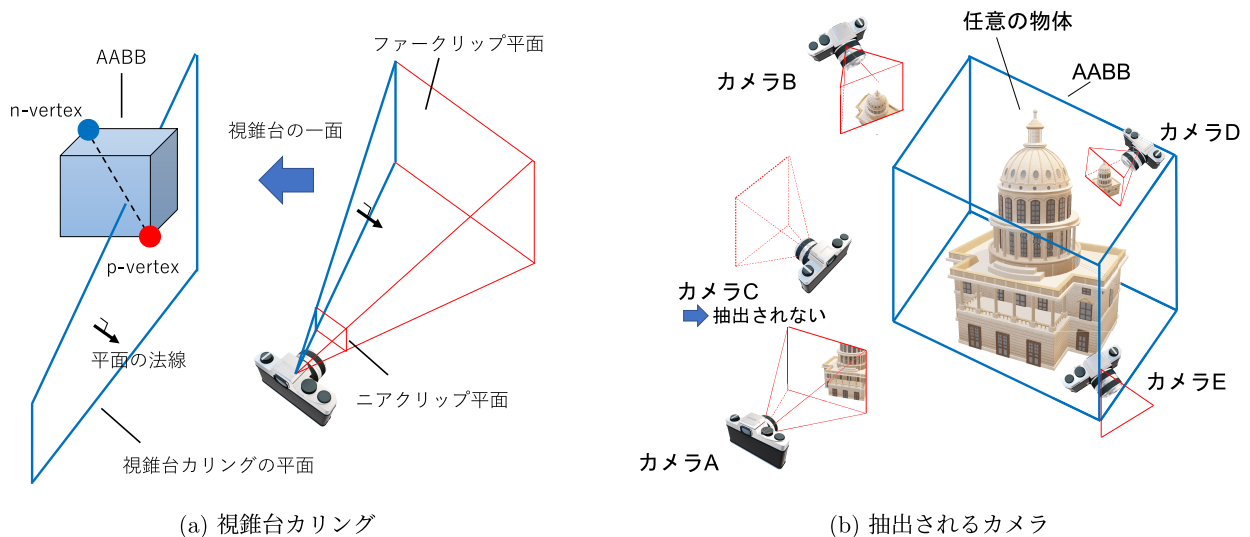


図 4 視錐台カリングに基づく検索  
Fig. 4 Retrieval by viewpoint frustum.

め、CGI 描画におけるファークリップ面を設定することができない。同様にニアクリップ面はカメラの視点に近い。これらから、AABB の頂点と、視錐台のニアクリップ面やファークリップ面との関係は計算しないことで、計算量も減らすことができる。したがってカメラ 1 台に対する計算量  $s$  は、式 (2) ( $\alpha$  = 対角線と面の法線とのなす角の計算,  $\delta$  = p-vertex から面までの法線の大きさの計算) となる。つまり、カメラ  $n$  台におけるオーダはカメラ台数に比例することから  $O(n)$  となる。

$$s = 4 \times (4\alpha + \delta) \quad (2)$$

図 4 (b) に示すとおり、カメラが 5 台 (カメラ A~E) あるときは、各カメラの視錐台と AABB との関係計算することで、カメラ C は抽出されることが分かる。また、少しでも撮影画像内に AABB が含まれると検索条件にあてはまるため、図 4 (b) に示すとおり、任意の物体を向いていないカメラ E の画像であっても抽出する。さらに、遠方の物体を検索する場合、AABB が相対的に小さくなる (図 5)。また遠方であるために対象となるカメラの総数も多いことが想定される。そこで、AABB よりも簡易的な手法として注視点を利用した手法も提案する。

【注視点に基づく検索】

空間軸上に注視点を指定し、各カメラが注視点を撮影可能かを判断する。これは次の 2 条件で判断可能である。

- (1) 注視点はカメラの光軸方向において正の値
- (2) 注視点をカメラ画像に投影可能

検索対象のカメラが位置  $t$ 、姿勢  $R$ 、焦点距離  $f$ 、主点  $(c_x, c_y)$  の場合、注視点の世界座標系の座標値  $(X_w, Y_w, Z_w)$  は式 (3) を用いてカメラ座標系の座標値  $(X_c, Y_c, Z_c)$  に変換する。このとき  $Z_c$  が正の値であれば条件 (1) を満たす。また、カメラ座標系の座標値  $(X_c, Y_c, Z_c)$  は式 (4) を

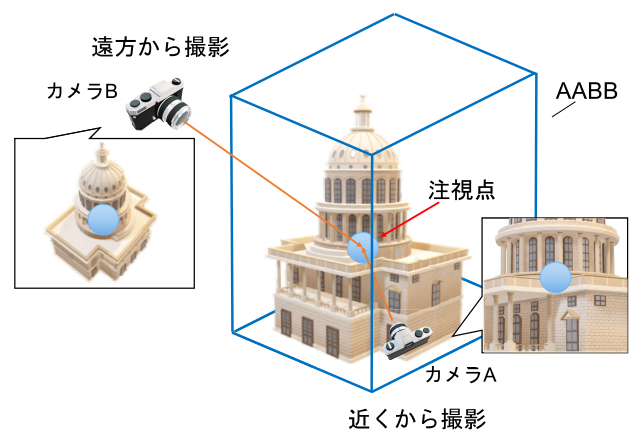


図 5 被観測物とカメラとの距離  
Fig. 5 Distance between observed object and camera.

用いて、カメラ画像座標系の座標値  $(u, v)$  に透視投影する ( $Z_c = 1$  の面に投影されるように  $h$  を用いて正規化する)。この  $(u, v)$  が 0 以上かつ撮影画像の横および縦の画素数未満の値であれば条件 (2) を満たす。図 6 に示すとおり、両条件を満たすのはカメラ C のみである。

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = \begin{pmatrix} R_{3 \times 3} & t_{3 \times 1} \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (3)$$

$$h \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \quad (4)$$

「注視点に基づく検索」も「視錐台カリングに基づく手法」同様、カメラ  $n$  台におけるオーダは  $O(n)$  であるが、式 (3) は乗算が 12 回であり、式 (4) は 0 を考慮すると乗算が 5 回、除算が 2 回であるため、カメラ 1 台に対して、「注視点に基づく手法」は乗数 17 回と除算 2 回で計算が可能で

ある. 式 (2) では  $\alpha$  を 16 回,  $\beta$  を 4 回計算するため, 「注視点に基づく検索」は「視錐台カリングに基づく手法」より計算量が少ないことが分かる. 本論文では, 「注視点に基づく検索」は「視錐台カリングに基づく検索」よりも指定領域が遠方のときに使用することを提案しているが, 遠方に具体的な数値的基準はなく, 指定した領域やカメラの位置, 総数などの要因によって最適な検索手法が変化するため, 利用目的に合わせて決める必要がある.

【エピポーラ幾何に基づく検索】

設定した仮想カメラの撮影領域を検索対象の各カメラが撮影しているかを調べる. 図 7 (a) に示すカメラ A, C のように, 仮想カメラの撮影領域と検索対象の任意のカメラの撮影領域に共通部分がある場合は検索条件を満たす. この検索条件は 2 カメラの撮影領域における共通部分の有無と見換えることができる.

図 7 (b) に示すとおり, お互いのカメラが撮影領域内にある場合は, 2 カメラの撮影領域に共有部分があるといえる. 具体的にはどちらかのカメラの視点をもう一方のカメラの画像に投影できるか (エピポールの有無) を調べる. 両方のカメラの視点が投影できるかではなく, どちらか一方でも条件を満たせば共有部分があると判断できる. 図 7 (a)

ではカメラ C がこの条件を満たす.

しかし任意のカメラの撮影領域内に, もう一方のカメラがない場合もある. この場合は, 共通部分の有無についてはエピポーラ線を用いて調べる. もし共通部分がある場合は, どちらか一方のカメラの画像の各画素から Ray-casting のように光線を飛ばし, もう一方の画像に, この光線を投影し, できた線 (エピポーラ線) が 1 本でもあれば, 共通部分があるといえる. しかし, 全画像に対して処理するには, 計算量が大きい. そこでカメラ画像の四隅の光線のみを飛ばし, エピポーラ線の有無によって判断する. しかし, 四隅の光線だけでは図 7 (c) に示すとおり, エピポーラ線が生成できない場合がある. これは, 画角の大きい方のカメラから小さい方のカメラに光線を飛ばしたことが原因であるため, 画角の小さい方から光線を飛ばすことで判定が可能となる. 図 7 (a) ではカメラ A がこの条件を満たす.

前述したとおり, 「視点の投影」は計算量が少ないため, 「視点の投影」を優先して用いる. 「エピポーラ幾何に基づく検索手法」のフローチャートは図 8 に示す.

3.4 リソース DB の構築

SIGMA フレームワークに “SIGMA Retriever” を実装するために関係型データベースを用いて 5 つのテーブルを持つリソース DB を構築した (図 9). 3.2 節で述べた外部パラメータは Geometry テーブルに入れた. 3.1 節で述べたとおりメタデータは計測値と推定値が必要であるため, Geometry テーブルでは両値を蓄積できるようにした. 内部パラメータは Resource テーブルに入れた. Resource テーブルにおける内部パラメータは歪み係数を含むが, 歪み係数は, 通常のレンズと魚眼レンズでは補正方法が異なる [14], [15]. 本研究では多機種のカメラを想定しているため, レンズも様々なものが想定される. 歪み係数を使用するアルゴリズムを明確にするため, 補正手法の種類もリソース DB に蓄積した. Resource のパスは画像が保存されている補助記憶装置のファイルパスである.

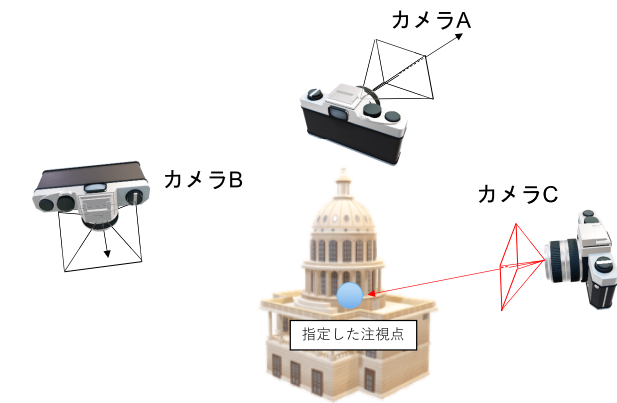


図 6 注視点に基づく検索  
Fig. 6 Retrieval based on gaze of view.

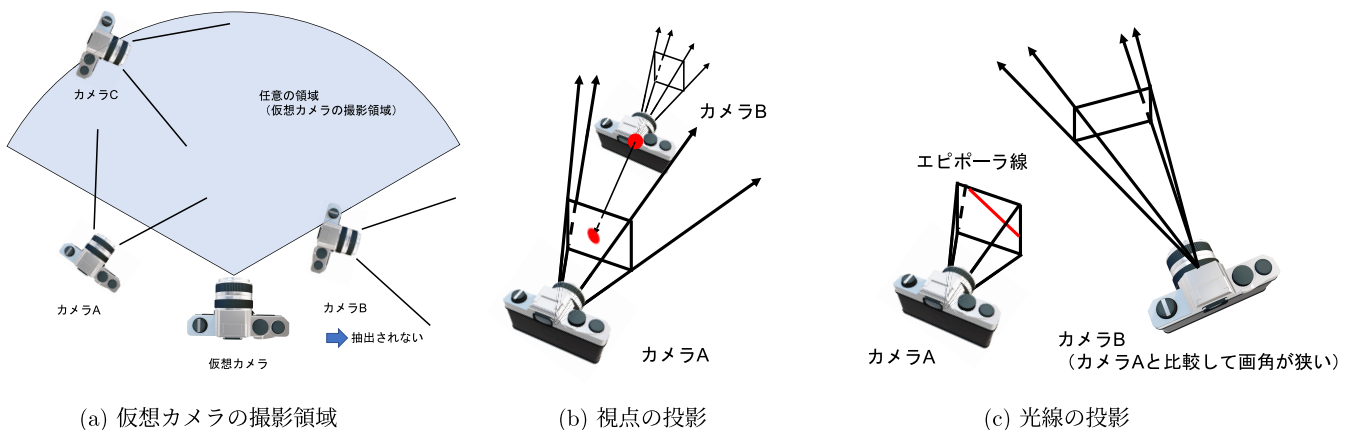


図 7 エピポーラ幾何に基づく検索  
Fig. 7 Retrieval based on epipolar geometry.

Geometry の ID は、一意なものにするとともに、近傍探索が高速となる ID を生成する。具体的には、計測や推定誤差があったとしても位置・時間の値がすべて同じ数値になるとは考えられないため、ID は位置・時間を基に生成したバイナリーコードを Base32 を用いて文字にした。このとき、位置は GeoHash を参考に MGRS (Military Grid Reference System) の各グリッドを南北/東西を交互に分割した 2 分木を基にバイナリーコードを生成する手

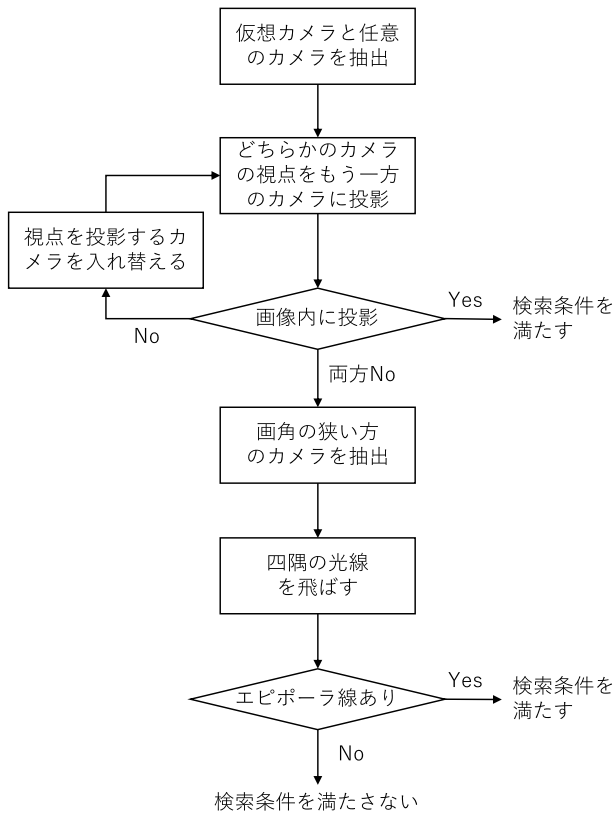


図 8 エピポーラ幾何に基づく検索手法のフローチャート

Fig. 8 Flowchart of retrieval based on epipolar geometry.

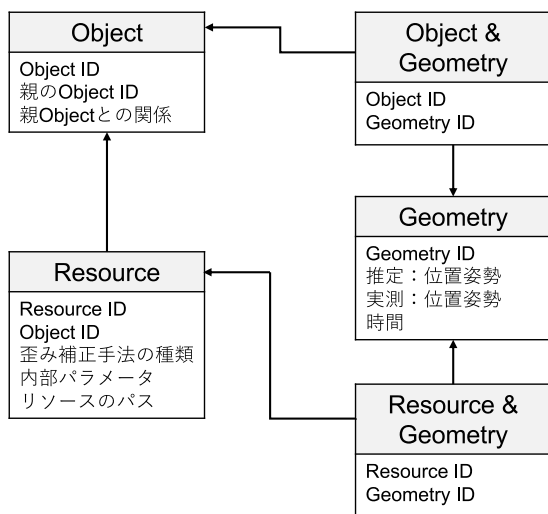


図 9 テーブルの関係

Fig. 9 Relation of the table.

法 [16], [17] を用いて高速化した。また、3.2 節で述べた「内部パラメータをキーとした検索」は Geometry の値のみで検索する。

## 4. 評価と考察

### 4.1 評価の概要

本章では、提案した検索手法の動作確認および評価を実施した。4.4 節では「外部パラメータをキーとした検索」の動作確認および、動作確認に対して考察した。4.5 節では「内部パラメータをキーとした検索」を評価した。3.3 節で述べたとおり、「内部パラメータをキーとした検索」は「外部パラメータをキーとした検索」に条件を追加して使用する。そのため、「外部パラメータをキーとした検索」を評価基準とし、これと比較することで、「内部パラメータをキーとした検索」によって生じた変化を考察した。評価 1 では「注視点に基づく検索」の比較評価し、評価 2 では「視錐台カリングに基づく検索」の比較評価、評価 4 では「エピポーラ幾何に基づく検索」の比較評価した。また、評価 3 では、本評価で準備した環境において、「視錐台カリングに基づく検索」「注視点に基づく検索」を比較し、比較に対して考察した (図 10)。

### 4.2 評価のための準備

3.4 節で構築したリソース DB に画像を登録するため、鎌倉駅周辺を撮影した。使用したカメラは 4 台 (カメラ 1 : Canon Kiss M (EF-M15-45 IS STM), カメラ 2 : SONY XPERIA X Compact メインカメラ, カメラ 3 : SONY Cyber-shot DSC-WX350, カメラ 4 : GoPro HERO3) で

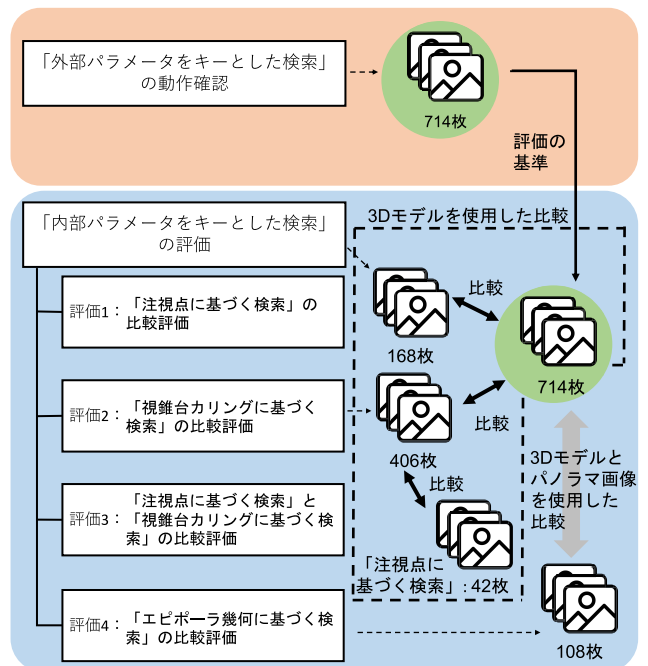
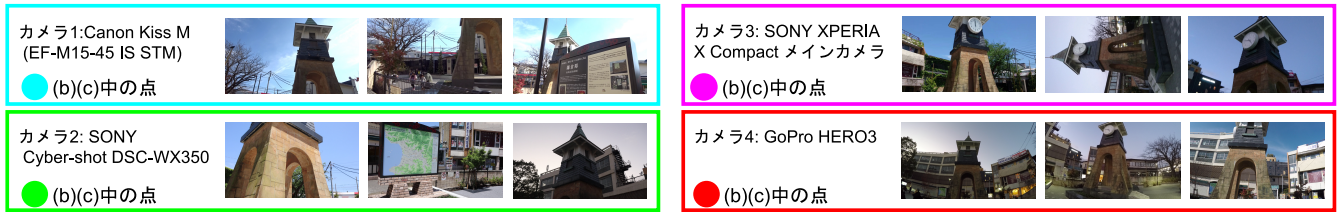


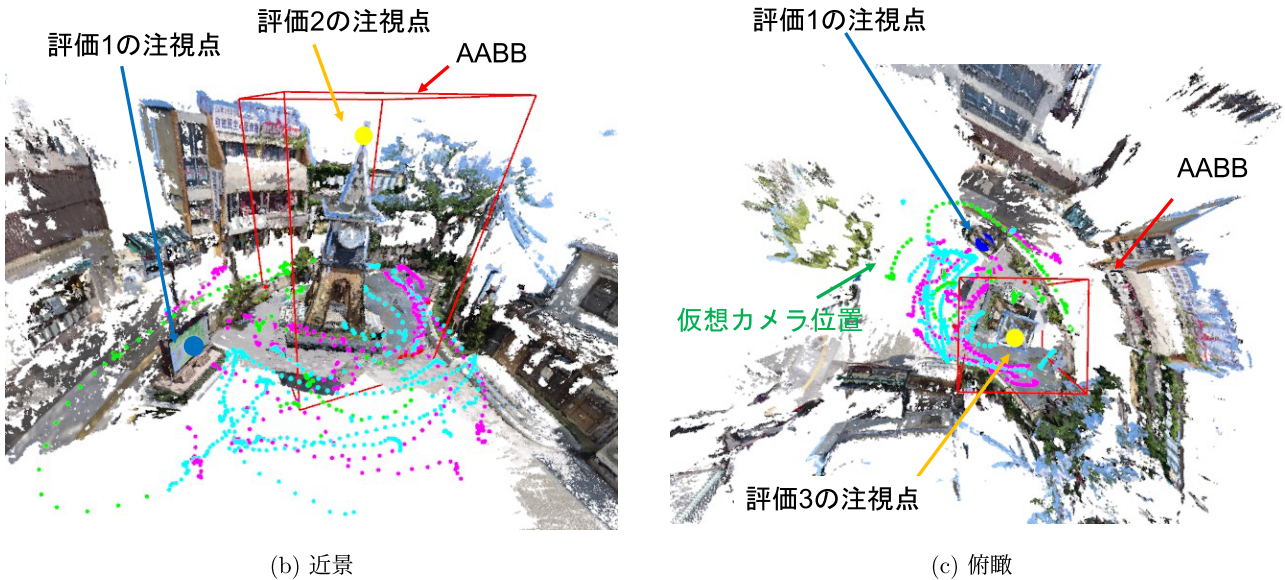
図 10 評価の種類

Fig. 10 Type of evaluation.





(a) 撮影した画像の一例



(b) 近景

(c) 俯瞰

図 11 時計台と周辺の三次元再構成 (プロットした点はカメラの撮影位置を示す)

Fig. 11 3D reconstruction clock tower.

表 1 撮影時期と撮影枚数

Table 1 Date taken and number of pictures.

	5月昼	12月昼	12月夕方
カメラ 1 (青)	227	30	114
カメラ 2 (緑)	324	163	55
カメラ 3 (紫)	-	117	92
カメラ 4 (赤)	-	12	26
計	551	322	287

あり、カメラ 4 は魚眼レンズである。カメラ 1, 2, 3 は  $1,920 \times 1,080$  [px] で撮影し、カメラ 4 は  $2,704 \times 1,520$  [px] で撮影した。また、撮影時期と撮影枚数は表 1 に、各カメラで撮影した画像の一部を図 11 (a) に示す。

リソース DB に計 1,160 枚の画像を登録した後、OpenMVG の SfM を用いて画像の位置姿勢を推定し、リソース DB の位置姿勢の推定値を更新した [18], [19]。その結果、モーションブレンダーなどが原因で推定できなかったものを除く、1,105 枚の位置姿勢をリソース DB の推定値に蓄積した。これらの画像と推定した位置姿勢を基に鎌倉駅旧駅舎時計台および、その周辺を三次元再構成した結果を図 11 (b), (c) に示す。図 11 (b), (c) 中でプロットした色は撮影位置を表しており、色は図 11 (a) に示したカメラの色に合わせてある。評価 1 の注視点の時計台近くの案内図

上に青丸で表し、評価 3 の視点は時計台頂上に黄色い丸で表し、評価 4 で指定した仮想カメラの位置を緑丸で表した。さらに、評価 2, 3 で使用する AABB もあわせて描画した。

### 4.3 評価方法

比較方法は、2 章の例であげた 3D モデルの生成や、パノラマ画像の生成を用いる。

3D モデルは、提案する検索手法によって抽出した画像を SfM によりカメラの位置姿勢を計算し、疎な点群を構成した後、Patch Match Stereo をベースに密な点群を構成することで生成する。このとき、リソース DB からは、画像に対応するメタデータを抽出している。SfM のバンドル調整時は、このメタデータを使うが、内部パラメータは固定し、位置姿勢の推定値は初期値として最適化した。そのため、SfM に用いる画像群が異なると計算結果である位置姿勢の値と点群に誤差が生じる。

3D モデルを構成する点群自体の品質を評価するには、Ground truth との比較が必要になる [20]。しかし、本研究では、3D モデルの品質向上するための手法を提案しているわけではなく、検索手法を提案している。つまり、本評価では、生成した 3D モデルは比較評価するためのものであって、それ自体が評価の対象ではないため、評価基準を

設定し、評価対象との比較して評価した。目標は、提案手法を基に抽出した画像から生成した点群が評価基準と同じ点群になっていることである。さらに、「内部パラメータをキーとした検索」の結果どうしも比較する。比較評価する項目は点群の数とするが、単純に数だけ比較した場合、点群の一部が密になったり、一部が欠損したりするなど、3Dモデルの見た目が異なっても、数値上の変化は見られない。そこで、見た目に変化が現れる数値を基に比較するために、点の密度を一定にした。点の密度は、三次元グリッドを1辺が10cmの正立方体とし、点群処理ライブラリ Point Cloud Library の Voxel Grid フィルタを用いたダウンサンプリングによって一定にしている [21]。以後、比較する点群は、このダウンサンプリングしたものとする。

評価4では、抽出した画像からパノラマ画像を生成し、仮想カメラが想定した視野のパノラマ画像になっているかを、点群とパノラマ画像とを比較し評価した。

#### 4.4 「外部パラメータをキーとした検索」の動作確認

本節では、時空間軸として図12に示すアプリ座標系上の赤丸（時計台の中央）と正午を中心に、範囲は、 $x$ 軸と $y$ 軸を10m、 $z$ 軸6m、時間1時間を指定して検索した。検索した結果、714枚が抽出されたことを確認した（図12は本論文のために地理院地図とOpenStreetMap [22]を加工して作成した図であるが、本フレームワークではこれらの地図は使用していない）。時間軸を検索値に使用できることから、太陽高度などによる陰影の条件をそろえることが可能なため、特徴点の検出・対応付け精度を向上させることもできる。

#### 4.5 「内部パラメータをキーとした検索」の評価

本評価では、4.4節で使用した「外部パラメータをキーとした検索」に条件を追加して検索している。つまり714枚から「内部パラメータをキーとした検索」により抽出した画像を基に評価した。

##### 4.5.1 評価1：「注視点に基づく検索」の比較評価

本評価では、時計台近くの案内図を三次元再構成するために、図11に示した青丸の座標値を注視点とし、「注視点に基づく検索」を行った。検索した結果、168枚を抽出した。抽出した画像を用いて三次元再構成した結果、案内図が再構成されていることが分かる（図13）。714枚の画像により再構成した案内図の点群（評価基準の点群）は1,826点であり、評価1の検索結果における画像により再構成した案内図の点群は1,737点であった。三次元再構成に使用した画像は23.5%であったが、点群は95.1%であったため3Dモデルの品質はほぼ同等と考えられる。

一方で、図13から時計台で遮蔽された画像も抽出していることも分かる。筆者が目視で数えたところ16枚が時計台の遮蔽により案内図を視認できなかった。遮蔽も考慮

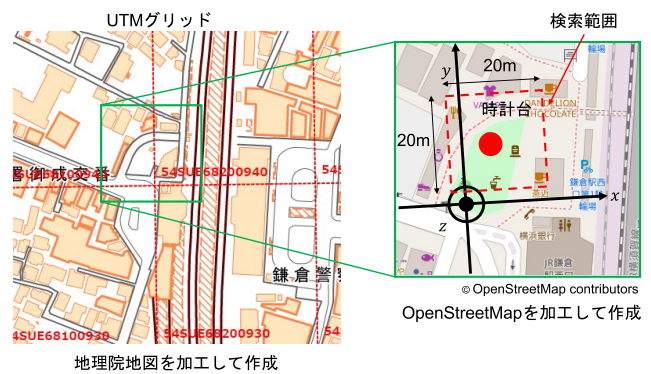


図12 検索範囲

Fig. 12 Retrieval range.

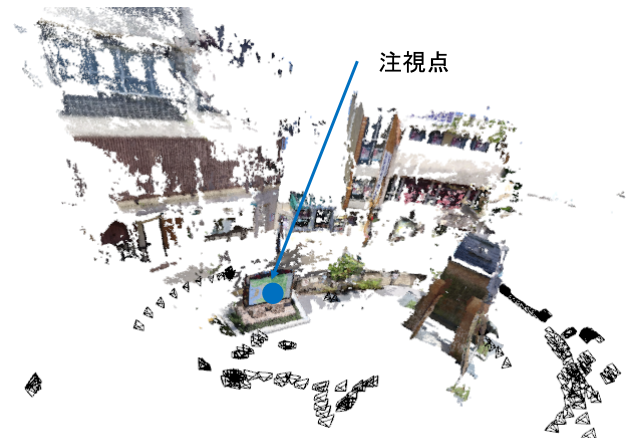


図13 案内図（評価1の結果）

Fig. 13 Map (Result of evaluation 1).

するためには、画像中の物体を認識する必要がある。

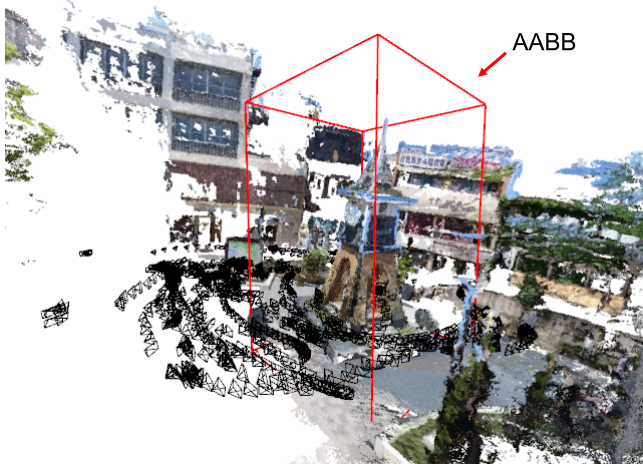
##### 4.5.2 評価2：「視錐台カリリングに基づく検索」の比較評価

本評価では、時計台を三次元再構成するために、時計台を囲むようにAABBの頂点8つを指定し（図11に示した赤枠）、「視錐台カリリングに基づく検索」を行った結果、615枚を抽出した。抽出した画像を用いて三次元再構成した結果を図14(a)に示す。この再構成した点群は16,035点であり、714枚の画像により再構成した時計台の点群（評価基準の点群）16,102点と比較しても99.6%の点群を維持できている。99.6%になった理由は、同じAABB内であってもバンドル調整により点群の座標値が評価基準時と異なるため、ダウンサンプリング時に異なる点が削減された可能性があるため、3Dモデルの品質は同等と考えられる。

##### 4.5.3 評価3：「注視点に基づく検索」と「視錐台カリリングに基づく検索」の比較評価

本評価では、「注視点に基づく検索」と「視錐台カリリングに基づく検索」を比較するために、あえて三次元再構成結果が悪くなりそうな点を注視点の座標値に指定することで、三次元再構成の結果に影響が出ることを示す。たとえば、図11に示した黄色い丸の座標値を注視点に設定し42枚の画像を抽出した。この画像から三次元再構成した結果





(a) 視錐台カリングに基づく検索の結果



(b) 注視点に基づく検索の結果

図 14 時計台 (評価 2, 3 の結果)

Fig. 14 Clock tower (Result of evaluation 2, 3).



(a) 仮想カメラから見た点群



(b) パノラマ画像

図 15 「エピポーラ幾何に基づく検索」の評価結果

Fig. 15 Result of evaluation “retrieval based on epipolar geometry”.

を図 14 (b) に示す。時計台の「とんがり帽子」の先端を注視点に指定し「注視点に基づく検索」し、42 枚を抽出した。これらの画像を用いて三次元再構成した結果を図 14 (b) に示す。本評価における点群は 5,376 点であり、評価 2 で、「視錐台カリングに基づく検索」した結果と比較すると、33.4% になっており、品質が低下していることが分かる。評価 1 では品質良く再構成できた「注視点に基づく検索」であるが、時計台は案内図より大きく、撮影位置からも多いため悪い結果となった。注視点として選ぶ場所や DB に蓄積された画像、再構成対象の大きさや距離によって三次元再構成の結果は大きく影響を受けていることが分かる。

一方、三次元再構成の品質だけを比較した場合、「視錐台カリングに基づく検索」が適しているように見える。しかし、検索により抽出した画像は DB からダウンロードすることを考えると、画像のデータ量は少ない方が好ましいが、視錐台カリングに基づく AABB に入っているカメラは、時計台を向いていなくても抽出される。時計台を三次元再構成する目的を満たすには、画像中に時計台が少し

でも写っている必要があるが、筆者が目視で数えたところ、この目的に合っていない画像が 103 枚あった。これはカメラの位置が AABB 内であるが、被写体を向いていない画像である。これを改善するためには、「注視点に基づく検索」などの他の検索手法と組み合わせが必要になる。

#### 4.5.4 評価 4: 「エピポーラ幾何に基づく検索」の比較評価

図 11 (b) の位置から水平画角 90°, 解像度 2K の仮想カメラを設定し「エピポーラ幾何に基づく検索」をした結果、108 枚を抽出した (図 15 (a))。この 108 枚から、文献 [8] の手法を基に、この 108 枚中の 10 枚を使用し、生成したパノラマ画像を図 15 (b) に示す。

仮想カメラから見た景色がパノラマ画像において生成されていることが分かる。「エピポーラ幾何に基づく検索」では仮想カメラの視錐台内を撮影しているカメラ画像はすべて検索されるため、本来ならば、仮想カメラから見た景色よりも広視野な画像が生成される。図 15 に示すとおり、案内図よりも左側までパノラマ画像が生成された。パノラマ画像を生成するとき、必要に応じてトリミングするな



どの処理が求められるが検索としては、十分な画像を抽出できたといえる。

## 5. むすび

本論文では、様々な視点かつ様々な機種のカメラから収集した大量の画像を IBMR に用いるために検索機構を設計したことについて述べた。一言で IBMR といっても様々な手法があり、利用目的・利用手法に応じて必要な画像も変わってくる。このような状況下で、利用者のニーズを満たすには、IBMR の処理の流れを整理し、抽出を所望する画像を明確化する必要があった。

本研究では、SfM などの IBMR の手法を概観し、抽象化した処理の流れを基にカメラの外部パラメータを検索キーにする手法や、視錐台カリング、エピソード幾何など、一般的には検索に使用しない既存技術を応用した検索手法を提案した。提案した検索手法は検索機構 “SIGMA Retriever” の機能として SIGMA フレームワークに実装し、利用目的・利用手法に合った画像を抽出できることを確認した。しかし、評価において、視錐台カリングやエピソード幾何に基づいた検索手法で抽出した画像は必要な画像を含んでいたものの、提案した手法では不必要な画像も含んでいた。本研究は利用目的・利用手法を狭めることなく、IBMR の広範なものに対応することを目標にしていたため、提案手法によって必要条件は満たした。一方で、提案手法のままでは不必要な画像も含む可能性がある。4.5.1 項でも述べたとおり、画像だけでは、より高度な検索手法を提案することが難しい。そこで、リソース DB に蓄積した画像に紐づくメタデータだけではなく、画像から生成した 3D モデルも活用できるようにリソース DB を拡張するとともに、活用方法を検討する。

今後は、さらなる利用目的・利用手法を視野に検索手法を検討するとともに、IBMR だけではなく、物体検出などの異なるアプローチから検索手法を検討する。さらに、画像の収集や配信手法などを検討し、SIGMA フレームワークの機能拡充を目指す。

謝辞 本研究の一部は、科研費：基盤研究 (B) 課題番号 21H03487 による。

## 参考文献

- [1] 国土交通省：Project “PLATEAU”，入手先 (<https://www.mlit.go.jp/plateau/>)。)
- [2] 山崎賢人，关 斯琨，松木 輝，木村朝子，柴田史久：SIGMA フレームワークにおける時空間映像データの管理手法，情報処理学会研究報告デジタルコンテンツクリエーション (DCC)，Vol.2020-DCC-26，No.9，pp.1–6 (2020)。
- [3] 佐藤洋一：イメージベースドモデリング，計測と制御，Vol.41，No.9，pp.633–638 (2002)。
- [4] Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S.M. and Szeliski, R.: Building rome in a day, *Comm. ACM*, Vol.54, No.10, pp.105–112

- (2011).
- [5] Taketomi, T., Uchiyama, H. and Ikeda, S.: Visual SLAM algorithms: A survey from 2010 to 2016, *IPSS Trans. Computer Vision and Applications*, Vol.9, No.1, pp.1–11 (2017).
- [6] 中野一也：解説：Structure from Motion (SfM) 第三回 SfM による三次元形状復元ソフトウェアについて，写真測量とリモートセンシング，Vol.55, No.5, pp.321–324 (2016)。
- [7] 酒井和彦：世界初アラウンドビューモニター，自動車技術 = Journal of Society of Automotive Engineers of Japan, Vol.62, No.3, pp.100–101 (2008)。
- [8] Brown, M. and Lowe, D.G.: Automatic panoramic image stitching using invariant features, *International Journal of Computer Vision (IJCV)*, Vol.74, No.1, pp.59–73 (2007)。
- [9] Gortler, S.J., Grzeszczuk, R., Szeliski, R. and Cohen, M.F.: The lumigraph, *Proc. 23rd Annual Conference on Computer Graphics and Interactive Techniques*, pp.43–54 (1996)。
- [10] Mori, S., Ikeda, S. and Saito, H.: A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects, *IPSS Trans. Computer Vision and Applications*, Vol.9, No.1, pp.1–14 (2017)。
- [11] Schindler, G., Dellaert, F. and Kang, S.B.: Inferring temporal order of images from 3D structure, *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–7, IEEE (2007)。
- [12] Bleyer, M., Rhemann, C. and Rother, C.: PatchMatch Stereo - Stereo Matching with Slanted Support Windows, *BMVC* (2011)。
- [13] Assarsson, U. and Moller, T.: Optimized view frustum culling algorithms for bounding boxes, *Journal of Graphics Tools*, Vol.5, No.1, pp.9–22 (2000)。
- [14] Zhang, Z.: A flexible new technique for camera calibration, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.22, No.11, pp.1330–1334 (2000)。
- [15] Kannala, J. and Brandt, S.S.: A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.28, No.8, pp.1335–1340 (2006)。
- [16] Morton, G.M.: A computer oriented geodetic data base and a new technique in file sequencing (1966)。
- [17] 小川輝樹，伊藤嘉博，中村嘉隆，高橋 修，白石 陽：空間統計処理のための Geohash を用いたデータベースシステムの提案と評価 (2012)。
- [18] Moulon, P., Monasse, P. and Marlet, R.: Adaptive structure from motion with a contrario model estimation, *Asian Conference on Computer Vision*, pp.257–270, Springer (2012)。
- [19] Moulon, P., Monasse, P., Marlet, R. and Others: OpenMVG, available from (<https://github.com/openMVG/openMVG>)。)
- [20] Seitz, S.M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, Vol.1, pp.519–528, IEEE (2006)。
- [21] Rusu, R.B. and Cousins, S.: 3D is here: Point cloud library (PCL), *2011 IEEE International Conference on Robotics and Automation*, pp.1–4, IEEE (2011)。
- [22] OpenStreetMap Foundation: OpenStreetMap, available from (<https://www.openstreetmap.org>)。)



山崎 賢人 (正会員)

2013年立命館大学情報理工学部卒業。2015年同大学大学院情報理工学研究科博士前期課程修了。同年三菱電機(株)入社。同社情報技術総合研究所にて複合現実感、映像合成等の研究開発に従事。現在、立命館大学大学院情報理工学研究科博士後期課程在学中。日本バーチャルリアリティ学会会員。



关 斯琨

2019年立命館大学情報理工学部卒業。2021年同大学大学院情報理工学研究科博士前期課程修了。複合現実感に関するシステムの研究開発に従事。



松木 輝

2020年立命館大学情報理工学部卒業。現在、同大学大学院情報理工学研究科博士前期課程在学中。複合現実感に関するシステムの研究開発に従事。



木村 朝子 (正会員)

1996年大阪大学基礎工学部卒業。1998年同大学大学院基礎工学研究科博士前期課程修了。同大学助手、立命館大学理工学部助教授、科学技術振興機構さきがけ研究員等を経て、2009年立命館大学情報理工学部准教授。現在、同教授。博士(工学)。実世界指向インタフェース、バーチャルリアリティ、複合現実感、多感覚知覚の研究に従事。ヒューマンインタフェース学会、日本バーチャルリアリティ学会、ACM、IEEE等の会員。日本バーチャルリアリティ学会学術奨励賞・論文賞、ヒューマンインタフェース学会論文賞、情報処理学会山下記念研究賞等を受賞。



柴田 史久 (正会員)

1996年大阪大学大学院基礎工学研究科博士前期課程修了。1999年同研究科博士後期課程修了。大阪大学産業科学研究科助手を経て、2003年立命館大学理工学部助教授。同大学情報理工学部准教授を経て、2013年より同教授。博士(工学)。モバイルコンピューティング、複合現実感等の研究に従事。IEEE、日本バーチャルリアリティ学会等の会員。日本バーチャルリアリティ学会複合現実感研究委員会委員長。日本バーチャルリアリティ学会学術奨励賞・論文賞、ヒューマンインタフェース学会論文賞を受賞。