

単一視点の深さ画像を検索要求とした3次元モデルの検索

Retrieving 3D models using single depth map from a viewpoint as a query

古屋 貴彦¹, 大淵 竜太郎²
Takahiko Furuya¹, Ryutarou Ohbuchi²

^{1,2} 山梨大学 〒400-8511 山梨県甲府市武田 4-3-11

^{1,2} University of Yamanashi, 4-3-11 Takeda, Kofu-shi, Yamanashi-ken, 400-8511, Japan

Mail: ¹g08mk025@yamanashi.ac.jp, ²ohbuchi@yamanashi.ac.jp

ABSTRACT

本論文では、単一、または少数の視点からレンジスキャナで得た3次元モデルを検索要求として用いた3次元モデルの検索について述べる。用いた手法は、我々の先行手法[1]に基づく。同手法では、検索要求をレンダリングして得た深さ画像と、データベース中の3次元モデルを多視点からレンダリングして得た深さ画像群とを比較し、形の類似する3次元モデルを検索する。個々の深さ画像から幾何変換に一定の不変性を持つ局所特徴の集合を抽出し、これをBag-of-Features法により物体あたり1つの特徴ベクトルに統合して3次元形状の類似比較を行う。しかしこの手法[1]のままでは、実物体の深さ画像に含まれる高周波のノイズや深さ値の欠損などのため、LoweのSIFT法の顕著点検出がうまく働かず、良い検索性能が得られなかった。そこで、我々は、顕著点検出を用いないランダムかつ密なサンプル配置と、低解像度画像により多くの特徴点を置く低周波強調サンプリングを組み合わせて用いる手法を提案する。実験的に評価した結果、提案手法は従来法に比べて高い検索性能を得ることが分かった。

1. はじめに

3次元モデルの用途が拡大し、その量が爆発的に増えるにつれ、その内容(形など)で3次元モデルを効率的に比較、検索、構造化、分類、等々する手法への要求が高まっている。既存の3次元モデル検索手法の多くは、ユーザが3次元モデルを検索要求(クエリ)としてシステムに与え、システムはクエリに類似したモデル群をデータベースから探して返す。しかし、たとえば考古学的遺物の検索、あるいはロボットの実世界におけるナビゲーションや実物体の把持、などの用途では、1つまたは少数の視点から得た深さ画像で3次元物体の形状を比較・検索する必要がある。本論文では、このような、部分視点(partial view)の深さ画像、あるいは深さ画像を元に作成した3次元モデル(の一部)を検索要求とする3次元形状の検索・比較の手法について検討する。

国際的な3次元モデルのコンテストShape REtrieval Contest (SHREC)が2006年から毎年開かれているが、2008年までの全ての部門では、検索要求は3次元モデルであった。2009年度に初めて、米国NISTのAfzal Godilらが主催し、部分視点からのレンジスキャンで得られた部分視点の3次元モデル(図1を参照)を検索要求とする部門 Partial 3D modelsが開かれた[6]。

参加したのは、我々のチームとPetros Darasらのチームの2チームで、そのいずれもが3次元モデルの見かけを比較する手法を採用した。つまり、データベース中の3次元モデルを多視点でレンダリングして得られた画像(深さ画像、シルエットなど)群と、与えられた部分視点3次元モデルをレンダリングして

得られた画像とを比較する手法である。

我々の手法は、我々が先に提案した、モデルの見かけに基づいた検索手法BF-SSIFT [1]に基づいている。BF-SIFTは、3次元モデルを、立体角上に均等配置した多視点(例えば42視点)からレンダリングして得た深さ画像群で比較する。同手法では、各画像に対しSIFT法 [5]を適用して一定の幾何変換不変性を備えた局所特徴群を抽出する。SIFT法はまず、多重解像度空間で明度勾配の大きい点を探索し「顕著」点とする。ついで、これらの顕著点(画像あたり30~300個)でSIFT特徴を抽出する。抽出されたSIFT特徴はその視点や画像上での位置を無視し、Bag-of-Features (BF)法により3次元モデルあたり1つの特徴ベクトルに統合する。3次元モデルあたり約1,300個程度の局所特徴それぞれは、まず代表ベクトル(視覚単語)にベクトル量子化され、ついでその視覚単語の頻度ヒストグラムが求められ、そのヒストグラムがそのモデルの特徴となる。評価の結果、この手法は、姿勢変化があるが比較的形状の単純な3次元モデルのデータベースMcGill Shape Benchmark (MSB) [3]において従来法を超える高い検索性能を得た。

我々は、SHREC 2009 Partial 3D modelsに参加するに当たり、このBF-SIFT法[1]を出発点として用いた。クエリは、1視点からレンジスキャナで得た深さ情報から生成した1視点の3次元メッシュモデルである(図1(a))。我々は当初、BF-SIFT法をほぼそのまま用いたが、その検索性能は悪かった。BF-SIFT法で用いた局所画像特徴であるSIFT法[5]が図1のような深さ画像の比較に適していなかったのである。その主な理由は以下の2つである。

- (1) 単一視点の3次元メッシュからレンダリングされた1枚の深さ画像から得られる顕著点数は平均数十個程度と少ない。これでは、検索要求の形状特徴が十分にとらえられない。
- (2) 検索要求からレンダリングされた深さ画像のなかで不必

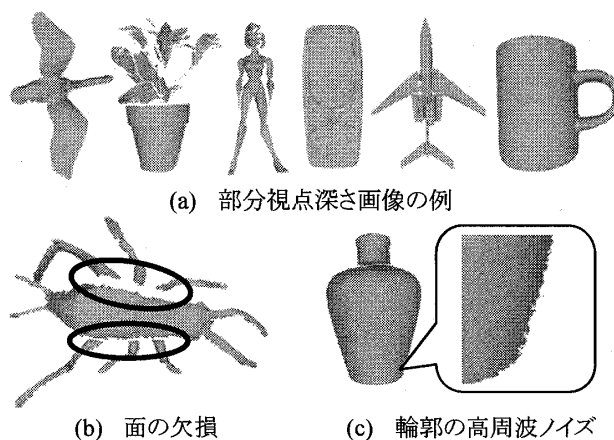


図1. SHREC 2009 Partial Model Retrieval トラックの検索要求の例(a). 部分視点モデルにはノイズが含まれる(b), (c).

要な部分に特徴点が配置され、したがって不必要な(検索性能を下げる)局所特徴が抽出されている。

前者は単純に検索要求の深さ画像が1枚で、SIFTがその画像に見出す顕著点数が限られていることによる。後者は、レンジスキャナから得られる3次元メッシュモデルに含まれる種々のノイズが主な原因である。レーザー光が反射しない部分に面の欠損(図1(b))があり、たとえば割れ目が生ずることがある(例えば、昆虫の胴と脚の繋ぎ目)。また、メッシュの輪郭部に高周波のノイズ(図1(c))が含まれる。SIFT法の顕著点検出は、スケールにかかわらず、画像の明度値の勾配に反応するため、輪郭のノイズや面の欠損などに生ずる明度勾配の高い部分に数多くの特徴が抽出される。反面、重要と思われる検索要求モデルの全体的形状をとらえる特徴点はほとんどなくなる。図2(a)と(d)はBF-SSIFT法による特徴サンプル点であるが、メッシュの割れ目やメッシュの輪郭にあるノイズに顕著点が配置され、サンプル点の配置が偏っているのが分かる。

そこで、我々は、SHREC2009 Partial 3D Modelsトラック [6]に参加する際に、SIFTの顕著点検出を停止し、強制的に格子状のサンプル点を配置してSIFT特徴を計算するBF Grid-SIFT (BF-GSIFT)を考案しこれを用いた。図2(b)と(e)はBF-GSIFT法による顕著点である。当然であるが、BF-SSIFT法(例:図2(a), (d))ではメッシュの割れ目付近に多く配置されていた顕著点がBF-GSIFT法(例:図2(b), (e))無くなった。BF-GSIFT法の検索性能はSHREC 2009 Partial 3D modelsトラック参加手法の中で最も良かったが、それでも平均検索性能 First Tier=22.5%と、3次元モデル全体を検索要求とする多くの手法と比べて低めの数値であった。

本論文では、単一(少数)視点からの深さ画像を検索要求とした場合の検索性能の向上を目指し、あらたな手法を提案する。中心的アイデアは、(1)低周波画像に重点を置く低周波強調サンプリングと、(2)多数のサンプル点の、ランダムかつ密な配置、の2つである。

前者は、SIFT法がその過程で生成する画像の多重解像度ピラミッドのなかで、低解像度(大スケール)の画像に対し、他のスケールの画像にくらべて多くの特徴サンプル点を配置するものである。面の欠損や輪郭のノイズ等、スキャンによる副作用が多重解像度ピラミッドの高解像度(高周波数)画像に多く

含まれる。そのため、低解像度画像からの特徴数を増加させるようなサンプリングを行うことでこれらのノイズの影響を低減できると期待される。また、ランダムかつ密に多数の特徴抽出点を配置する手法は、全体形状検索で高い検索性能を示した我々自身の先行手法BF-DSIFT [2]のものである。BF-DSIFTでは図2(c), (f)のようにランダムに、かつ密に特徴点を配置し、部分視点モデルの多様な特徴をバランスよく捉えることができると期待される。

評価実験の結果、提案手法である低周波強調サンプリングを行ったBF-DSIFTの検索性能は、Partial 3D Modelsトラックで最も良かった我々自身のBF-GSIFTと比較し、First Tierで約15%高いことが分かった。

2. 関連研究

3次元モデルの形状類似検索に、モデルの見かけの類似性を用いる手法では、大抵、多視点からレンダリングした画像(2値画像、あるいは深さ画像)から何らかの特徴量を抽出し、比較する。画像から抽出する特徴は、大域特徴と局所特徴の2種類に分類することができる。

大域特徴は、画像あたり1個抽出される。ChenらのLFD [7]では、10視点よりレンダリングした2値画像から、Fourier特徴とZernikeモーメントの2種の大域特徴を抽出し、比較する。LFDは、飛行機や家具等、多様なモデルのデータベースにおいて高い検索性能を示したが、全ての視点において、クエリと検索対象モデルの視点のマッチングを行う必要があったため、3次元モデル対間の比較の計算量が問題である。(LFDを用いた、2次元スケッチによる検索ウェブページが公開されている [10].) DarasらのCMVD [6]は、Chenらが用いたFourier特徴とZernikeモーメントに加え、Krawtchoukモーメントの計3種の大域特徴を抽出し、統合するもので、LFDに比較して高い検索性能を実現した。SHREC2009 Partial 3D modelsトラックに参加した2チームのうちの1つがCMVDで、もう一方が我々のBF-GSIFT(と、参考のために加えたBF-SSIFT)である。

局所特徴は、画像あたり複数個の特徴を抽出する。部分から特徴を抽出するため、局所特徴で比較する手法はモデルの姿勢変化に対して頑強という利点がある。しかし、モデルあたり n 個の局所特徴群を全対で比較すると、2モデル間の比較回数が $O(n^2)$ となり、比較の計算コストが高い。

我々の先行研究であるBF-SIFT [1]では、Bag-of-Features (BF)法により、複数の局所特徴群を、ベクトル量子化された視覚単語の頻度ヒストグラム (BF特徴)に統合することで、比較の計算量を $O(1)$ に低減した。局所特徴は、予め学習した代表ベクトル群(コードブック)に対して最近傍探索を行うことで、最も近い代表ベクトルに量子化される。局所特徴には、スケール変化や画像面内の回転に対して不変な局所視覚特徴 SIFT [5]を用いた。SIFT特徴は、多重解像度ピラミッドを構成する複数の画像上にサンプリングした特徴点から抽出される。

BF-SSIFT [1]はもともとのSIFT法の顕著点検出を用い、顕著度の高い点で少数の特徴を計算する「疎な」サンプリング方法である(図2(a), (d))。深さ画像あたり平均数十個の特徴が抽出されるため、特徴抽出やベクトル量子化の計算量は小さいが、顕著点検出に失敗すると、モデルの特徴を捉えられず検索性能が悪くなる。これに対し、BF-GSIFTとBF-DSIFT [2]は「密な」サンプリング方法で、特徴点を、BF-GSIFTでは格子状(図2(b), (e))、BF-DSIFTではランダムかつ密に(図2(c), (f))、それぞれ配置する。深さ画像あたりの約300個の特徴が抽出されるため、計算量が大きくなるが、WuらによるGPUによ

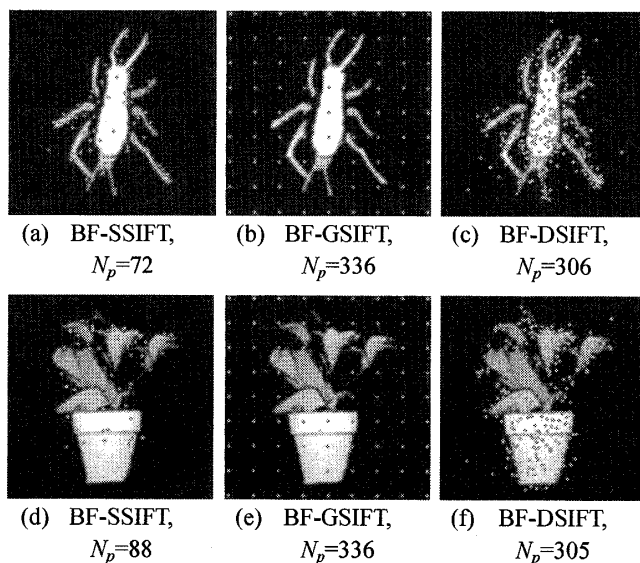


図2. サンプリング法による特徴点の差異。(N_p は画像あたりの特徴点数。)

るSIFT特徴抽出 SiftGPU [8]や, Geurtsらによる木構造ベクトル量子化器 ERC-Tree [9]の利用により高速化が成された. 評価実験の結果, BF-DSIFTは, 多様なモデルのデータベース Princeton Shape Benchmark (PSB) [3] において, LFDより約10%高い検索性能を示した. 姿勢変化のあるデータベース MSB [4]では, 局所特徴を用いたBF-SSIFTとBF-DSIFTのいずれも, その検索性能がLFD法よりも約20%高かった.

3. 提案手法

提案手法の流れを図3に示す. まず, クエリとなる部分視点モデルは1視点, 検索対象モデルは多視点からレンダリングし, 深さ画像を得る. 各深さ画像上に, ランダムかつ密に多数の特徴点を配置し, 各特徴点から SIFT 特徴を抽出することで, 視点毎の局所特徴群を得る. 次に, 予め SIFT 特徴空間を学習して得たコードブック (代表ベクトル群)を用いて, 局所特徴をベクトル量子化し, 視点毎の視覚単語群を得る. 各視覚単語の出現頻度を1つのヒストグラムに統合し, 視点毎のBF特徴ベクトルを得る. 最後に, クエリモデルから計算した1つのBF特徴と, 検索対象モデルから計算した視点数分のBF特徴群を比較することで, モデル間の相違度を求める. 以下, 各処理について具体的に説明する.

3.1. 深さ画像レンダリング

クエリとなる部分視点モデルを1視点から, 検索対象モデルを N_i 個の視点からレンダリングする. 本手法では, 3次元モデルを囲む80面体の頂点 ($N_i = 42$)からレンダリングし, 42個の深さ画像を得る (画像サイズは 256×256). 部分視点モデルは,

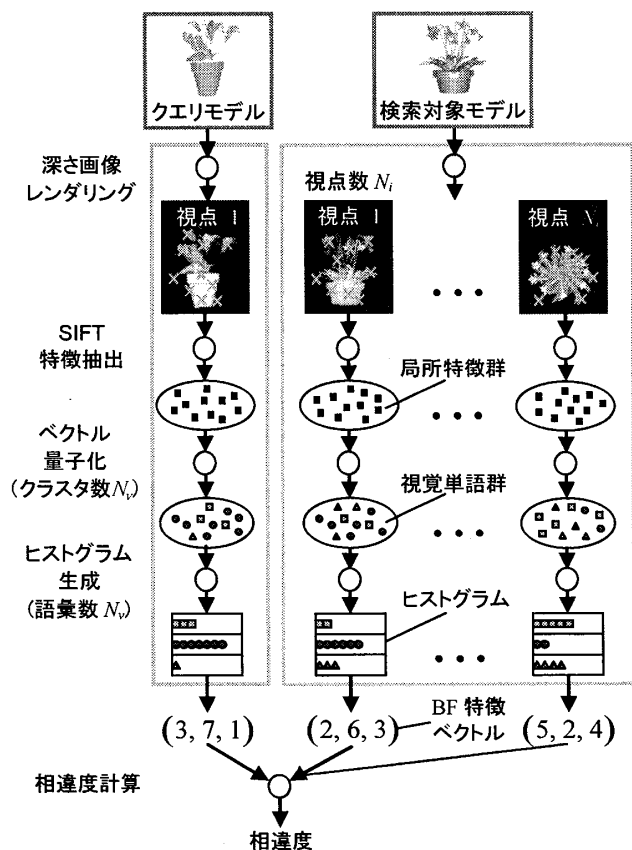


図3. 提案手法の流れ図. 深さ画像から抽出した局所特徴群は, Bag-of-Features (BF)法により, 視点毎に1つの特徴ベクトルに統合される.

1024×1024 でレンダリングし, アンチエイリアシングを施し, 256×256 にサイズを落とすことで, スキャンで発生したノイズ (特に, メッシュの輪郭部に生じたノイズ)を軽減させる.

3.2. 局所特徴抽出

レンダリングして得た深さ画像から, SIFT 特徴を抽出する. SIFT 特徴 [5]は, スケール変化, 回転に対して不変性を持つ128次元の局所視覚特徴である. 多重解像度表現を用いることで, 画像の大域的特徴, 局所特徴の双方を捉えることができる. 抽出方法には, 特徴点検出を行う SSIFT (図2(a), (d)), 格子状に特徴点を配置する BF-GSIFT (図2(b), (e)), ランダムかつ密に配置する BF-DSIFT (図2(c), (f))の3つを用いる. いずれも Wu らの SiftGPU [8]により, 高速に特徴抽出を行う.

BF-SSIFT は, 多重解像度ピラミッドの各画像上に顕著点 (人間が目目するような点)を複数検出する. 特徴点数は, 画像により異なる. 見かけに高周波成分を多く含む画像 (例えば, 小さな葉を多く持つ植物)からは, 数百個の特徴が抽出される. 対して, 高周波成分の少ない画像 (例えば, 平面の多いドア)からは, 数十個の特徴が抽出される.

BF-GSIFT は, 等間隔に特徴点を配置する. 図2(b), (e)では, 全解像度の特徴点が重なっているため, 同位置から異なるスケールの SIFT 特徴が複数抽出される.

BF-DSIFT は, ランダムかつ密に特徴点を配置する. 多重解像度ピラミッドの各解像度画像において, 輝度値が0でない画素群からランダムに特徴点をサンプリングする. 我々の先行研究 [2]では, 各解像度のサンプリング密度が均等になるように, 特徴点数を決定していた.

しかし, 本手法で扱う部分視点モデルには, 面の欠損や輪郭のノイズ等, スキャンによる副作用が含まれる. これらの副作用は, 多重解像度ピラミッドの高解像度画像に多く含まれるため, 低解像度画像の特徴数を増加させるような, 重み付きサンプリングを行うことで, 副作用による影響の低減が期待できる. 図4の例では, 256×256 の高解像度画像と比べて, 128×128 や, 64×64 の画像では, 脚と胴の接続部に発生した面の欠損が, 平滑化によって目立たなくなっている. 提案手法では, これら低解像度画像のサンプル数に重みを付ける.

具体的には, パラメータ w を設け, ある解像度 (サイズ $N \times N$)の, 次の解像度 (サイズ $N/2 \times N/2$)の特徴点数が, $1/w$ 倍となるようにサンプリングする. $w = 4.0$ を均等サンプリングとし, $w > 4.0$ で高周波強調サンプリング, $w < 4.0$ で低周波強調サンプリングとなる. 多重解像度ピラミッドは同サイズの画像を複数含むが, 同サイズの画像の特徴点数は同じとする.

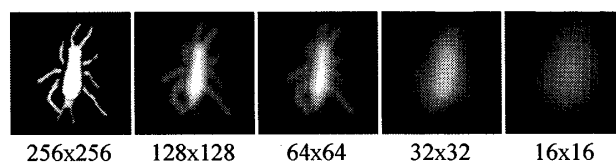


図4. 多重解像度画像の例 ($N \times N$ は画像サイズ).

3.3. ベクトル量子化, ヒストグラム生成

視点毎に抽出した局所特徴群を, ベクトル量子化して視覚単語群に変換し, 視覚単語の頻度ヒストグラム (BF 特徴)を視点毎に生成する.

ベクトル量子化には, 予め学習したコードブック (代表ベクトル群)を用いる. コードブックは, データベースの全モデルから

抽出した局所特徴群を、特徴空間上でクラスタリングすることで得られる。クラスタリング後、各クラスターの重心を代表ベクトル(視覚単語)として決定する。コードブックに入力された局所特徴は、代表ベクトル群に対して最近傍探索を行うことで、最も近い代表ベクトルに量子化される。

本手法では、ランダム化決定木の1つERC-Tree [9]を、コードブック学習とベクトル量子化に用いる。コードブック学習では、ERC-Treeは、1つのクラスターの局所特徴数が S_{min} 以下になるまで、ランダムに、かつ再帰的に特徴空間を分割することで、木構造のコードブックを学習する。 S_{min} がクラスター数 N_p (コードブックの語彙数)を決定する。ベクトル量子化では、入力された局所特徴は、コードブックを根から葉へ辿る(近似最近傍探索をする)ことで、視覚単語にベクトル量子化される。

我々の先行研究 [2]では、一般的に用いられるk-meansによるコードブック学習+線形探索によるベクトル量子化と比較し、検索性能を維持しつつ、高速化することに成功した。

ベクトル量子化によって得た視覚単語群は、視点毎にそれぞれ、視覚単語の頻度を要素に持つヒストグラム(BF特徴)に統合される。BF特徴の次元数は、語彙数 N_p に等しい。

3.4. 相違度計算

部分視点モデルのBF特徴1個と、検索対象モデルの $N_i = 42$ 個のBF特徴群との比較を行う。部分視点モデルと検索対象モデルのBF特徴の組み合わせは、 N_i 通りある。各組み合わせについてBF特徴間の相違度を計算し、最も小さい値をモデル間の相違度とする。BF特徴間の相違度尺度には、先行研究と同様、Kullback-Leibler Divergence (KLD)を用いる。

4. 実験と結果

提案手法の有用性を調べるため、以下の実験を行った。

- 実験 1. SSIFT, GSIFT, DSIFT の検索性能比較 (4.1 節)
- 実験 2. 重み付き局所特徴サンプリングの効果 (4.2 節)
- 実験 3. コードブックの語彙数と検索性能の関係 (4.3 節)
- 実験 4. 既存手法との検索性能比較 (4.4 節)

全ての実験に共通する実験条件を以下に記す。検索対象モデルから深さ画像をレンダリングする際の視点数は $N_i = 42$ とした。コードブック学習にはERC-Treeクラスタリングを用い、全検索対象モデルから抽出した局所特徴群からダウンサンプリングした500,000個の局所特徴群を学習サンプルとした。

評価用ベンチマークには、SHREC2009 Partial 3D modelsトラック [6]を用いた。クエリは、モデルを3次元スキャナで1方向から取り込んだ、20個の部分視点モデルである(図1)。検索対象モデルは、植物や家具、飛行機等、40個のカテゴリに分類された720個の3次元モデルである。

検索性能の評価指標には、First Tierを用いた。これは、クエリQが属するカテゴリのモデル数をqとした場合、検索結果上位q個中の正解モデルの割合である。クエリ毎に求めたFirst Tierの平均値を、ベンチマーク全体の検索性能とする。

4.1. SSIFT, GSIFT, DSIFT の検索性能の比較

部分視点による検索に適した局所特徴サンプリングを知るため、BF-SSIFT, BF-GSIFT, BF-DSIFTの3手法について画像あたりの特徴点数 N_p と検索性能First Tierの関係を調べる。

BF-SSIFTの画像あたりの特徴点数 N_p は、顕著点検出アルゴリズムによって決まるため、本実験では変化させない。一方で、BF-GSIFTとBF-DSIFTは、画像あたりの特徴点数 N_p を自

由に変化させることができる。本実験では、 N_p を100 ~ 3,600程度まで変化させ、検索性能の変化を調べる。各解像度のサンプリング密度は均等とする。

図5に、BF-SSIFT, BF-GSIFT, BF-DSIFTの、画像あたりの特徴点数 N_p と検索性能First Tierの関係を示す。検索性能の高い順に、BF-DSIFT, BF-GSIFT, BF-SSIFTとなった。

BF-SSIFTは3つのサンプリング手法の中で最も検索性能が低くFirst Tier=9.2%であった。検索性能が低い原因として次の2点が考えられる。1つ目は、クエリからの特徴抽出失敗である。スキャン時に生じた面の欠損や、輪郭部のノイズ上に特徴点が検出され、不要な特徴が抽出された可能性がある。2つ目は、特徴点数の少なさである。画像あたりの特徴点数が平均 $N_p = 66$ と少数だったため、BF特徴がモデルの特徴を記述し切れなかったと推測される。

BF-GSIFTは、画像あたりの特徴数 N_p の増加と共に検索性能が向上し、 $N_p = 3,700$ でFirst Tier=27.2%を示した。画像あたりの特徴数が増加すると、格子点の間隔が狭くなるため、検索性能がDSIFTに近づく。しかし、同位置に異なるスケールの特徴点が複数個配置されるため、多様な特徴が得られず、DSIFTの検索性能には及ばなかった。

BF-DSIFTは、画像あたりの特徴数 N_p の増加と共に検索性能が向上し、 $N_p = 1,200$, First Tier=32.5%付近で飽和した。密に、かつランダムに特徴点を配置することで、BF-SSIFTとBF-GSIFTで捉えられなかった特徴が抽出され、3手法中で最も検索性能が高くなった。本データベースについては、サンプリング手法がBF-DSIFT、画像あたりの特徴数を $N_p = 1,200$ が最も良い検索性能を与えるようである。

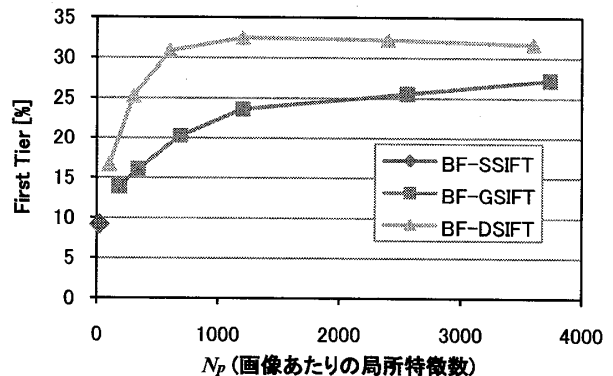


図5. 画像あたりの局所特徴数と、検索性能の関係。

4.2. 重み付き局所特徴サンプリングの効果

SHREC2009 Partial 3D modelsトラックの部分視点モデルには、面の欠損や輪郭のギザギザなど、レンジスキャンによるノイズが含まれる。本節では、これらの副作用に対し、多重解像度空間で重み付きサンプリングを行う効果を調べる。局所特徴には、4.1節において最も良い検索性能を示したBF-DSIFTを用いる。画像あたりの特徴数は $N_p = 1,200$ とする。3.2節で定義したサンプリングの重みを決定するパラメータ w を、1.0~12.0まで変化させる。

図6に、重み付きサンプリングを行った際の、検索性能の変化を示す。低周波に重みを付けるにつれ、検索性能が向上し、 $w = 1.5$ で検索性能がピークとなり、均等サンプリングと比較して検索性能が5%向上した(First Tier=37.2%)。 $w = 1.5$ より小さい値では、均等サンプリング時より検索性能が数%低くなった。一方、高周波重点サンプリングと均等サンプリングの検索性能

差は1%以内と、ほとんど影響が無かった。

低周波強調サンプリングによって、検索性能が最高で5%向上した理由として、次の2点が挙げられる。1つ目は、部分視点に含まれる、面の欠損やノイズによる影響の軽減である。多重解像度ピラミッドにおける低解像度画像では、平滑化によって、面の欠損や輪郭部のノイズが目立たなくなっているため、これら低周波画像からの局所特徴数を増加させることで、モデルの特徴を正しく捉えたと推測される。2つ目は、大域的な特徴の増加である。低解像度画像からの SIFT 特徴は、モデルの大域的特徴を捉える。実験に用いた Partial 3D models トラックには、カップやボトルなど、詳細部の少ないクエリも含まれるため、大域的な特徴が増加することで、検索性能が改善したと考える。ただし、低周波を強調し過ぎると、局所的な特徴が不足し、検索性能が低下した。

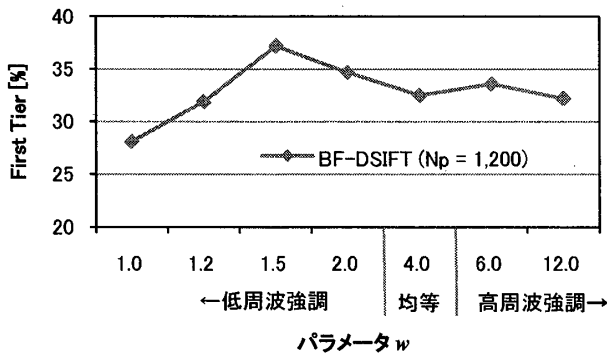


図6. 重み付きサンプリングによる検索性能の変化。

4.3. コードブックの語彙数と検索性能の関係

本節では、4.1 節と 4.2 節の実験で用いた手法における、コードブックの語彙数 N_v と検索性能 First Tier の関係を調べる。実験には、次の4手法を用いた (N_p は、画像あたりの平均局所特徴数を指す)。

- BF-SSIFT ($N_p = 66$)
- BF-GSIFT ($N_p = 1,200$)
- BF-DSIFT ($N_p = 1,200$, 均等 $w = 4.0$)
- BF-DSIFT ($N_p = 1,200$, 低周波強調 $w = 1.5$)

図7に、コードブックの語彙数と、検索性能の関係を示す。ERC-Tree は、ランダム木であるため、語彙数と検索性能の曲線上の点には、本質的にばらつきがあることに注意して欲しい。また、我々の先行研究 [2] で得た知見として、特徴数が多くなるほど、また、多様な特徴が多く抽出されるほど、必要なコードブックの語彙数も大きくなるのが分かっている。

BF-SSIFT は、画像あたりの局所特徴数が $N_p = 66$ と少ないため、小さな語彙数で検索性能のピークを持ち、語彙数の増加と共に検索性能が低下した。BF-GSIFT は語彙数 $N_v = 10,000$ 付近で検索性能のピークを持ち、その後で検索性能が低下した。BF-DSIFT (均等, 低周波強調サンプリング) は、語彙数の増加と共に検索性能が向上し、語彙数 $N_v = 10,000$ 付近で飽和した。飽和後の検索性能の変動は数%あるが、これはランダム木を用いたためである。飽和した理由として、ランダムかつ密に局所特徴を抽出することで、SSIFT や GSIFT では抽出できなかった多様な特徴により特徴空間が張る部分空間が広がったこと、が考えられる。また、語彙数 $N_v = 10,000$ 以上の BF 特徴は、多くの要素が 0 の疎なヒストグラムとなる。そのため、語彙数が増加しても頻度が 0 になるビンが増えるだけ

で相違度への影響は小さく、検索性能が横這いになった。

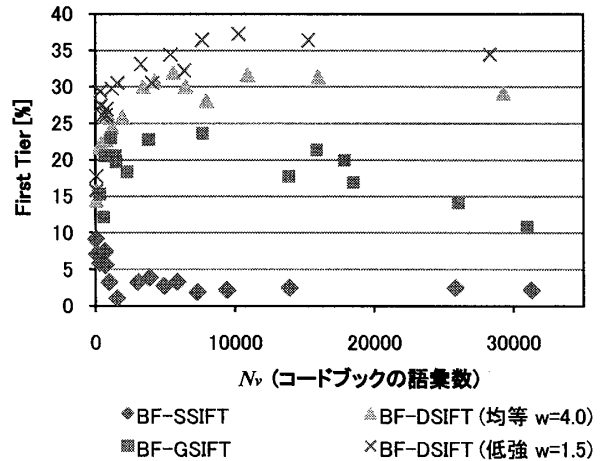


図7. コードブックの語彙数と、検索性能の関係

4.4. 既存手法との検索性能比較

提案手法と、既存手法 (SHREC2009 Partial 3D models トラックの参加手法) の比較を行う。比較対象には2手法を用い、1つ目は、我々の BF-GSIFT [6] で、高解像度画像からは SSIFT 特徴、低解像度画像からは GSIFT 特徴を抽出し、視点毎に BF 特徴に統合する。2つ目は、Daras らの CMVD [6] で、2値画像から3種の大域特徴 (フーリエ特徴と2種のモーメント特徴) を抽出する。提案手法との区別のため、これら2つの参加手法を BF-GSIFT (SH09), CMVD (SH09) と表記する。

図8に各手法の Recall-Precision プロットを、図9に検索結果例を、表1に First Tier の値を示す。Recall-Precision 曲線は右上に位置するほど高い検索性能を示すため、低周波強調サンプリングをした BF-DSIFT の検索性能が最も高い。表1において、参加手法である CMVD の FT=21.7%, BF-GSIFT の FT=22.5% に対し、低周波強調サンプリングをした BF-DSIFT は、FT=37.2% と、約15%高い検索性能を示した。

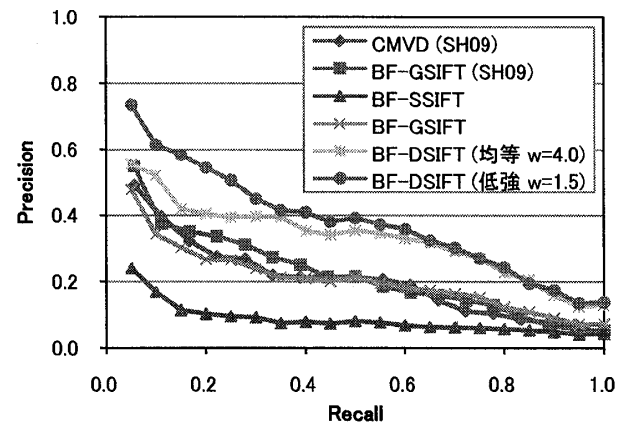


図8. 各手法の Recall-Precision 曲線

表1. 各手法の検索性能 FirstTier [%]

手法	FirstTier [%]
CMVD (SH09)	21.7
BF-GSIFT (SH09)	22.5
BF-SSIFT	9.2
BF-GSIFT	23.6
BF-DSIFT (均等 $w = 4.0$)	32.5
BF-DSIFT (低強 $w = 1.5$)	37.2

5. まとめと今後の課題

本論文では、部分視点モデルをクエリとする、3次元モデルの検索手法を提案した。レンダリングして得た深さ画像から、モデルの見かけに基づく局所特徴を、密に、かつランダムに抽出し、Bag-of-Features法で1つの特徴ベクトルに統合し、モデル間の比較を行った。部分視点モデルは、スキャンの結果、レーザー光の当たらない部分の面の欠損や、モデルの輪郭部のノイズが発生する。これらの副作用の影響を軽減させるため、低解像度画像からの特徴数に重みを付けることで、従来法と比較し、検索性能が約15%向上した。

今後の課題は、さらなる検索性能の向上である。例えば、本手法で用いた特徴とは別種の局所特徴、もしくは大域特徴と組み合わせた際の検索性能の評価を行うことが挙げられる。

参考文献

- [1] Ryutarou Ohbuchi, Kunio Osada, Takahiko Furuya, Tomohisa Banno, Salient local visual features for shape-based 3D model retrieval, *Proc. IEEE Shape Modeling International (SMI) 2008*, pp. 93-102, (2008).
- [2] Takahiko Furuya, Ryutarou Ohbuchi, Dense Sampling and Fast Encoding for 3D Model Retrieval Using Bag-of-Visual Features, *accepted, Proc. ACM International Conference on Image and Video Retrieval 2009 (CIVR 2009)*, July 8-10, (2009)
- [3] *Princeton Shape Benchmark*, <http://shape.cs.princeton.edu/benchmark/>
- [4] *McGill 3D Shape Benchmark*, <http://www.cim.mcgill.ca/~shape/benchMark/>
- [5] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *Int'l Journal of Computer Vision*, **60**(2), (2004)
- [6] A. Axenopoulos, P. Daras, H. Dutagaci, T. Furuya, A. Godil, R. Ohbuchi, SHREC 2009 - Shape Retrieval Contest of Partial 3D Models, *Eurographics Workshop on 3D object retrieval*, (2009)
- [7] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, M. Ouh-Young: On Visual Similarity Based 3D Model Retrieval, *Computer Graphics Forum*, **22**(3), pp. 223-232, (2003).
- [8] C. Wu, SiftGPU: A GPU Implementation of David Lowe's Scale Invariant Feature Transform (SIFT), <http://cs.unc.edu/~ccwu/siftgpu/>
- [9] P. Guerts, D. Ernst, L. Wehenkel, Extremely randomized trees, *Machine Learning*, **36**(1), 3-42, (2006)
- [10] 3D Model Retrieval Homepage, <http://3d.csie.org/~dynamic/3DRetrieval/>

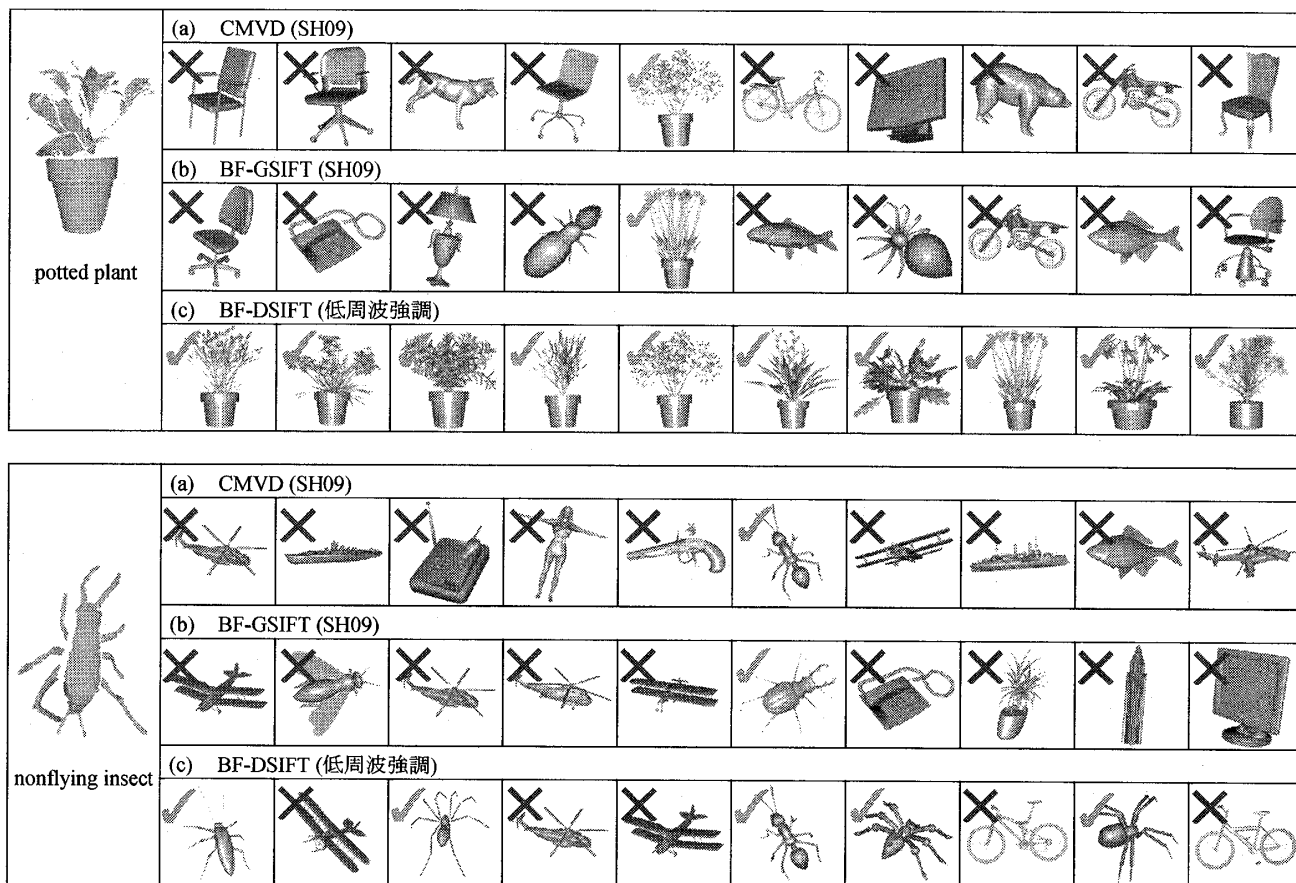


図9. 部分視点モデルをクエリとする検索の検索結果例 (低周波強調サンプリングした DSIFT は、他手法よりも良い結果を返すのが分かる).