

画像検索技術を用いた建築データベースシステム開発に向けた試行

○中林 拓馬*¹ 神山 一馬*²
黒澤 紀之*² 平沢 岳人*³

キーワード：建築写真 画像検索 データベース

1. はじめに

計算機技術の発達やデジタルカメラの普及により、従来のキーワードをクエリにした検索システムのほかに、画像をクエリにした検索システムの実用化がなされ始めている。キーワード検索では、ユーザが望む回答を得るためには適切なキーワードを入力する必要があるため、ある程度の予備知識が必要となるが、画像検索では画像さえあれば求める情報にたどり着くことができる。建築分野においても、検索すべきキーワードが不明な建築写真のみからその情報を引き出すことができるということは、非常に有用であると考えられる。

筆者らはこれまでの研究で、多重解像度ウェブレット変換を利用して類似画像検索を行う **imgseek** というソフトウェアを利用し、予め構築した建築写真データベースを対象に画像検索を行うシステムを試作し、その実用性を検証した。この検証ではある程度の正答率は得られたものの、類似度が一位の画像が正解画像であるケースは少なく、写真に写る建築物の情報を得るためには検索結果上位の画像群からユーザが正解画像を選択する必要がある。

本研究では、画像をクエリとした建築情報検索システムの検索精度をより高めて、最終段においてユーザ自身が検索結果から正解画像を選択する作業を極力排除することを目的とする。これを実現するため、本稿では画像検索に有用であると考えられる **SIFT** 特徴量・カラーヒストグラムなどの技術について検証を行い、これらの結果について考察を行う。

2. 既往研究

画像内の物体を認識することを目的とした研究は情報処理分野を中心に活発に取り組まれている。本研究は、この中でも特定物体認識と呼ばれる物体認識を目的とした研究であるが、本道¹⁾らは、これに用いられる様々な局所特徴量の精度比較結果を報告している。この他、岩村²⁾らは画像内の特徴点がどのように配置されているかを利用した画像検索手法について検証を行なっている。

また、Google 画像検索のように、本研究と同様に画像をクエリとした検索サービスも提供され始めている。この検索では Web 上に存在する膨大な画像やテキストを利用した検索を行うことができる。しかしながら、いくつかの建築写真を対象に検索を行ったところ、ほぼ同一の写真が



図1 本研究で用いた画像例

Web 上に存在しない限り特定物体の認識は難しく、多少の視点変更でも認識精度に影響が生じる結果であった。

建築写真は建築そのものが壁・柱・窓などといったように共通の構造を持つ上、それぞれに異なった周辺背景を持つため、画像からの特定が難しいことが予想される。本研究では建築情報検索システムの作成のため建築写真に範囲を限定し、複数の画像検索アルゴリズムを用いて特定物体認識の成功率を検証した。

3. 研究手法

3. 1. 概要

本研究では、大規模な写真投稿サイトである **Flickr** から 1000 枚の建築写真を収集し、独自の画像データセットを作成することで建築写真データベースを構築した。これらの一部を図1に示す。写真の収集においては、一般物体認識などで利用されている大規模画像データセットの **Caltech-256**^{注1)}などと同様、各建築物の典型的視点 (**canonical view**) で撮影された写真を採用した。典型的視点とは、一般的には物体の斜め前方からの視点とされ、物体を同定しやすい視点である。建築物を対象にした場合にも、例えば建築雑誌に見られる視点など、それぞれの建築に印象的かつ典型的視点があると考えられる。不明な建築の情報を検索しようとした場合、クエリとなる画像もこのような画像が多いことが予想されるため、本研究では典型的視点を集めたデータセットが適当であると考えた。建築においてはそれぞれの建築作品によって典型的視点が異なると考えられるため、**Flickr** 上において頻繁に見られた構図^{注2)}を典型的な見え方と設定することとした。

また、建築写真データベースは階層型として構築した。

建築家名/作品名/画像という階層を作ることにより、検索して得られた画像データから建築の名称・建築家名を得ることができる。本研究では、31人の建築家、100個の建築の写真データを集めた。なお、各建築物につき写真を10枚と定めることでデータの偏りにも注意した。

画像をクエリとした検索システムを構築するため、SIFT特徴量・カラーヒストグラム・Bag of Featuresを複合的に利用することで、精度の高い検索システムの開発に取り組んだ。本稿では、それぞれの特徴量を用い、画像間の類似度を算出し、実用性を評価する。

3. 2. SIFT特徴量

SIFTとはScale Invariant Feature Transformの略称で、Lowe³⁾によって考案された特徴量である。これを用いることで画像内の局所特徴量を128次元のベクトルとして表すことができる。SIFT特徴量は、回転・スケール変化に不変な特徴量であり、照明変化にも頑健であるという特徴を持つ。また、小規模な視点位置・角度の変化であれば特徴量にあまり影響が出ないため、同一の物体を異なる位置から撮影した写真を比較する場合にも、それぞれの特徴点を同一であると関係づけられる。建築固有の典型的視点から撮影した複数の写真間での対応付けも十分に可能であると考えられる。

以上の特徴から、本研究ではSIFT特徴量を画像検索に用いることとした。SIFT特徴量の検出は、計算コストが大きく、大規模なデータセットの特徴量を全て抽出する処理は数時間単位の時間が必要となる。そのため、GPU^{注3)}を利用して処理の高速化を実現したSiftGPU^{注4)}を用いた。また、画像セットとして収集した画像の総画素数はそれぞれに異なるが、画素数の差が特徴点数量の差につながるのには望ましくない。アスペクト比の違いにより画素数に多少の差は生じるが、画像の長辺が1,000ピクセルとなるようにダウンサンプリングすることで画素数の影響を軽減した。図2に建築写真に対してSIFT特徴量を描画した例を、表1に抽出された点数とその処理時間を示す。

3. 3. カラーヒストグラム

SIFT特徴量は画像の色に関する情報を持たないため、本研究では画像のカラーヒストグラムを用いて色の類似度を比較することとした。ヒストグラムを作成するカラーモデルにはいくつか考えられるが、予備実験においてRGBとHSVの2つのカラーモデルのヒストグラムで画像検索を行った結果、RGBを用いた場合の正解率が高い結果であった。

一般的なデジタル写真の色数は、 $256 \times 256 \times 256 = 16,777,216$ 色であるが、これと同じ数のビン数を持つヒストグラムを作成するのは非常に計算コストが大きく、現実的でない。加えて、天候など撮影環境の影響が現れやすいヒストグラムになってしまうことが予想された。これらの問題を軽減するため、64色に減色してヒストグラムを作

表1 特徴点数と処理時間

	SIFT特徴点数	処理時間(ms)
特徴点数最小画像	975	141
特徴点数最大画像	12,595	85
全画像	4,184,401	113,292

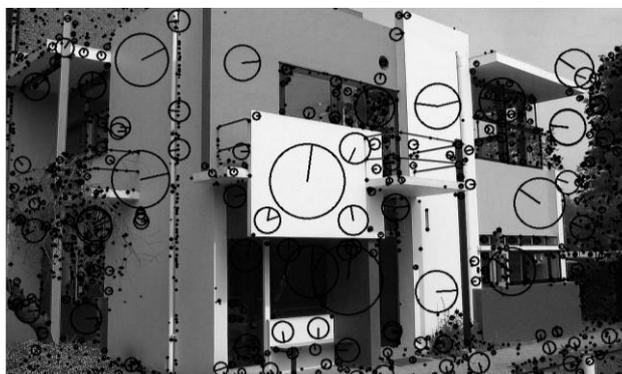


図2 SIFT特徴量を描画した例

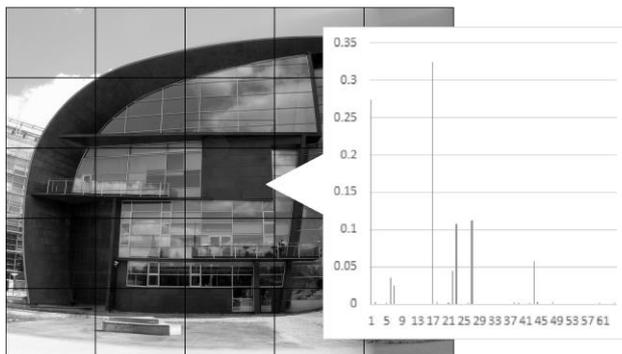


図3 グリッド分割した画像とグリッドのヒストグラム例

成した。

また、予備実験として、画像全体から作成したヒストグラム、画像を等間隔にグリッド分割して作成したヒストグラムによる画像検索の比較も行った。1×1、3×3、4×4、5×5、10×10の計5種類の実験を行ったうち、5×5のグリッドで作成した場合が最も正答率が高い結果であった。

以上から、本稿では写真を64色に減色した上で、図3のように5×5のグリッドのRGBカラーヒストグラムを作成することとした。

3. 4. Bag of Features

Bag of Featuresとは、Csurka⁴⁾らにより考案され、一般物体認識のためによく用いられる手法であり、画像をVisual Wordsと呼ばれる特徴の集合であると見なす。画像を一つの特徴ベクトルで表現する事ができるため、計算コストを軽減することが可能であり、また、画像分類への有効性が先行研究⁵⁾で報告されている。本稿では以下のような手順で行った。

1. 全画像からSIFT特徴量を抽出する。
2. 全SIFT特徴量をK-Means法によりクラスタリン

グする。

3. 各クラスタの重心位置を Visual Words として設定する。
4. 全画像の特徴点を Visual Words に置き換え、ヒストグラムを作成する。

手順2においてクラスタ数を設定する必要があり、予備実験として24、48、128、500、1000個にクラスタリングして画像検索を行った。結果として、1000クラスタに分類した場合が最も精度が良かったため、これを K-Means のパラメータとして採用することとした。

4. 検証結果と考察

4. 1. SIFT 特徴量による画像検索の検証

SIFT 特徴量を利用した画像検索では、クエリ画像とデータベース画像間で対応付けがなされた特徴点の総数により、検索結果を順位付けすることとした。対応付け成立の可否の判断方法は、クエリのある特徴点とデータベース画像の全特徴点で総当りに(1)式の計算を行い、その内の最短距離と次点距離を用いた(2)式が成立すれば最短距離である2点が対応点と定めた。なお、本稿ではrの値を0.8と設定している。この計算をクエリ画像の全特徴点とデータベースの全画像の全特徴点に対して行うため、膨大な計算量ではあるが、SiftGPU の高速な処理により1クエリにつき1分程度で全探索を終了する。

$$d = \sqrt{\sum_{n=1}^{128} (q_n - d_n)^2} \quad \dots(1)$$

但し、d:特徴点間距離、 q_n :クエリ画像の特徴点のn番目の要素、 d_n :データベース画像の特徴点のn番目の要素

$$d_1 \leq d_2r \quad \dots(2)$$

但し、 d_1 :最短特徴点間距離、 d_2 :2番目に短い特徴点間距離、r:任意の倍率

以上の処理により、建築画像データベースに含まれる画像1000枚を全てクエリにし、1000回の検索実験を行った。結果を図5に示す。1位に正解が選ばれる確率は73.9%で上位10位に正解が選ばれている確率は86.5%であり、順位が下がるに従って正答数も少なくなるという結果であった。

順位と正答率に関係性があることから、対応付けの総数が多ければ多いほど正答である可能性が高いといえる。1000回の実験において、50組以上の対応付けが成立した検索結果は合計で1040枚あり、これらの正答率は99.4%であった。また、正解が含まれていない場合の上位10位の対応付け数は平坦な勾配であるのに対し、正解が含まれている場合の対応付け数はほとんどの場合に図5のように急勾配になっていた。この特性は正答率を上げるうえで非常に重要であろう。

4. 2. カラーヒストグラムによる画像検索の検証

カラーヒストグラムを利用した画像検索では、クエリ

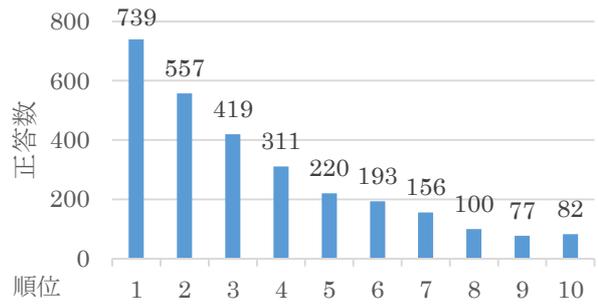


図4 SIFT 特徴量による画像検索結果

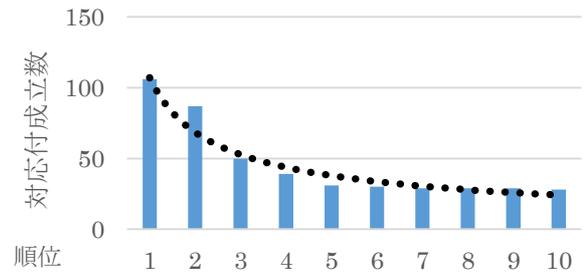


図5 正解が存在するときの順位別対応付け成立数

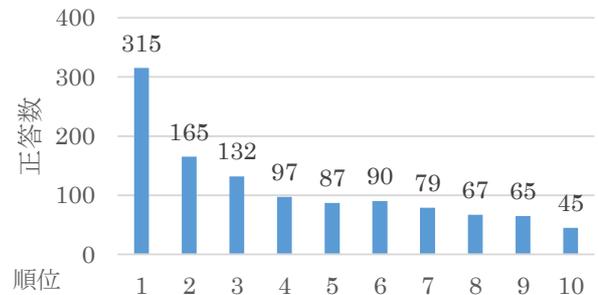


図6 カラーヒストグラムによる画像検索結果

画像とデータベース画像間のヒストグラムの距離が最も近いものから順に順位付けすることとした。なお、ヒストグラムはグリッドごとに正規化しており、グリッドごとに(3)式で算出された Bhattacharyya 距離の全グリッドの総和を画像間の距離としている。

$$d = \sqrt{1 - \frac{1}{\sqrt{\sum_{n=1}^{64} h_n \sum_{n=1}^{64} h'_n}}} \sum_{n=1}^{64} \sqrt{h_n h'_n} \quad \dots(3)$$

但し、d:ヒストグラム間距離、 h_n :クエリ画像のヒストグラムのn番目のビンの値、 h'_n :データベース画像のヒストグラムのn番目のビンの値

以上の処理を用いて、SIFT 特徴量による画像検証と同様の検証を行った。結果を図6に示す。1位に正解が選ばれる確率は31.5%で上位10位に正解が選ばれている確率は59.4%であった。1位から3位程度までは距離が短いほど順位が下がるに従って正答数も少なくなるという結果ではあったが、SIFT 特徴量の検索結果と比較すると、正答数の勾配は小さかった。

結果を概観すると、色情報に依存していることが要因なのか、建築により得手不得手が出やすい検索方法であ

った。視覚的に特徴的な色を持つ建築作品であれば、検索精度は上がるが、例えば白や灰色がファサードに多く見られるような、目立った色味のない建築作品では十分な検索精度は得られなかった。また、正解・不正解に関わらず上位10位のヒストグラム間距離にあまり大きな差は生じなかった。しかしながら、同一の色味の写真を抽出できるため、誤回答の除去フィルターなどとしての利用価値は十分にあると考えられる。

4. 3. Bag of Features による画像検索の検証

Bag of Features を利用した画像検索でもカラーヒストグラムと同様、クエリ画像とデータベース画像間の Visual Words のヒストグラムの距離が最も近いものから順位付けにすることとした。なお、K-Means 法によるクラスタリングの際には、クエリ画像は未知であるため、本検証では予め、各建築の写真群から1枚ずつクエリ画像を設定し、クエリ画像を除去した900枚の画像セットでクラスタリングを行い、Visual Words を得た。また、クエリ画像群のクラスタリングは、各画像のすべての特徴点において Visual Words との最近傍探索をすることでやっている。

本検証では、ヒストグラムはグリッドごとに正規化しており、ヒストグラム間の距離の算出には(4)式のインターセクトを用いた。

$$d = \sum_{n=1}^{1000} \min(h_n, h'_n) \quad \dots(4)$$

但し、 d :ヒストグラム間の距離、 h_n :クエリ画像のヒストグラムの n 番目のビンの値、 h'_n :データベース画像のヒストグラムの n 番目のビンの値

以上の処理を用いて、各建築作品につき1枚の画像を無作為に抽出し、計100枚のクエリ画像をもとに検索実験を行った。結果を図7に示す。1位に正解が選ばれる確率は49%で上位10位に正解が選ばれている確率は68%であった。正答数の傾向としては、カラーヒストグラムと同様、1位から3位以外では順位と正答数の関連性は見られない。

結果を概観すると、カラーヒストグラムと比較しても、より得手不得手が出やすい傾向があることが分かった。特に、得意な画像においては上位10位の中に正解画像が9枚全て含まれることも多く、上位10位以内に2枚以上の正解画像が含まれている場合には1位の画像の正答率が非常に高かった。この傾向も画像検索システムの精度向上に活用できると見込まれる。

また、SIFT 特徴量の対応付けは非常に計算コストの大きいものであったが、Bag of Features では Visual Words によるクラスタリングから検索までを通して非常に高速に検索することが可能であった。カラーヒストグラムも同様であるが、Bag of Features では画像を1つのヒストグラムとして表すため、被検索対象群の増加が計算コストに与える影響がグリッド分割したカラーヒストグラムよりも少ない。SIFT の前段階処理として用いることで、計算コストの軽減にも利用価値が期待できそうである。

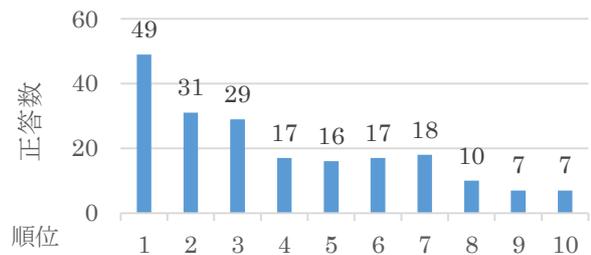


図7 Bag of Features による画像検索結果

5. まとめ

本研究では、ユーザが検索結果から正解画像を選択するという作業を必要としない高精度な建築情報検索システムを開発することを目的とし、画像をクエリとした検索技術の開発に取り組んでいる。本稿では、このうち、このような検索方法に有用であると考えられる SIFT 特徴量・カラーヒストグラム・Bag of Features の3つの技術について検証を行い、その結果について報告した。

それぞれの結果から、SIFT 特徴量を用いた検索が、最も正答率が高い検索方法であることがわかった。また、それぞれの手法による検索結果の傾向を確認し、これらを複合することで、より高い正答率を得ることが可能であると期待できる結果が得られた。

謝辞

本研究は、特別研究員奨励費 24005895 によって行われたものである。

【参考文献】

- 1) 本道貴行, 黄瀬浩一, 大規模画像認識のための局所特徴量の性能比較, 画像の認識・理解シンポジウム(MIRU2008)論文集, p.550, 2008
- 2) 岩村雅一, 中居友弘, 黄瀬浩一, 特徴点の配置に基づく画像検索手法の解析 —Geometric Hashing と LLAH の比較と解析—, 電子情報通信学会論文誌 D, J93-D, 4, pp.494-501, 2010
- 3) Lowe, D. G.: Distinctive image features from scale-invariant keypoints, IJCV, Vol.60, No.2, pp.91-110, 2004
- 4) Csurka, G., Dance, C.R., Fan, L., Willamowski, J. and Bray, C.: Visual Categorization with Bags of Keypoints, ECCV International Workshop on SLCV, 2004
- 5) 永橋知行, 伊原有仁, 藤吉弘亘, "画像分類における Bag-of-features による識別に有効な特徴量の傾向", 情報処理学会研究報告 CVIM169, No.3, pp.1-8, 2009.

【注釈】

- 1) 約3万枚、256カテゴリの画像を収録した画像データセット。
http://vision.caltech.edu/Image_Datasets/Caltech256/
- 2) なお、正確な分析は行っていないが、建築においても一般的な典型的視点と同様に斜め前方からの写真が多かった。
- 3) GPU(Graphics Processing Unit)とはPCの画像処理を担当するパーツである。
- 4) Changchang により公開されているオープンソースライブラリである。
<http://cs.unc.edu/~ccwu/siftgpu/>

- *1 千葉大学大学院工学研究科 日本学術振興会特別研究員 DC
 *2 千葉大学大学院工学研究科
 *3 千葉大学大学院工学研究科 准教授