

キーポイントマッチングにおける幾何学的の一貫性の評価法 -画像検索への応用-

大倉 有人¹ 和田 俊和¹

概要: Content-based Image Retrieval(CBIR) は、画像をクエリとして与えてデータベース内の類似する画像を検索する問題である。CBIR では、局所特徴から求められたコードブックを用いる方法がよく用いられているが、処理時間の上では、局所特徴そのものを用いた検索の方が有利である。この場合、幾何学的に無意味なマッチングが求められるが、これらは無視すべきである。そこで、複数の点の対応付けから RANSAC などを用いて剛体変換を求める手法が考えられるが、これも処理速度の低下をもたらす。本報告では、対応付けられた特徴点に対して幾何学的の一貫性のチェックを高速に行う方法 (Local feature hashing) を提案する。この手法では、画像間で局所特徴に基づいて対応付けられたキーポイントのペア組から、キーポイントの位置、スケール、オリエンテーションの情報を用いて相似変換を推定することで、幾何学的の一貫性のチェックを行う。この相似変換を全ての対応付けに対してを求め、パラメータ空間への投票を行い、最も投票数が多い変換の度数に基づいて幾何学的の一貫性を評価することで、相似変換のパラメータの推定精度を向上させ、画像検索の精度も向上させる。実験では約 1000 枚の画像を対象として実時間の画像検索が実現でき、提案手法により、誤った検索を抑制できることを確認した。

キーワード: 画像検索, 相似変換, パラメータ推定, 投票

A geometric consistency checking method for keypoint matching -Application to image retrieval-

OKURA YUTO¹ WADA TOSHIKAZU¹

Abstract: Content-based Image Retrieval(CBIR) is a problem of finding similar images to a given query image out of the image database. In CBIR, various methods using codebook obtained from local features are widely used. However the local feature based image retrieval without using codebook is much faster than this method. When we employ the codebook-less image retrieval method, we can examine the coordinate values of the corresponding points in the query and database images and geometrically meaningless matching should be ignored. For checking this geometric consistency, RANSAC based rigid body transformation estimation can be applied, but this method requires considerable amount of time for random sampling.

In this report, we propose Local Feature Hashing(LFH) that estimates similarity transformation between input and database images just from two corresponding keypoints. We perform voting to transformation parameters for all corresponded keypoints to obtain a reliable. If a salient peak is formed in the voting space after the voting to an image entry in the database, we can verify the retrieval, otherwise not. In the experiment, real-time image retrieval can be realized for the database consisting of 1000 images, and we confirmed that the method can suppress erroneous retrieval.

Keywords: Image retrieval, similarity transform, parameter estimation, voting

¹ 和歌山大学大学院システム工学研究科
Wakayama University Graduate school of systems engineering

1. はじめに

Content-based Image Retrieval(CBIR)とは画像をクエリとして与える画像検索問題であり、多くのアプリケーションで利用されている。例えば、ポスター等の展示物を撮影した画像を与え、催し物のホームページのURLを検索する仕組みなど、実世界への情報の埋め込みといった用途が考えられる。

CBIRでは、画像の色、テクスチャ、輪郭など様々の特徴を用いて検索を行うことができ、その中でもSIFT[1]やSURF[2]などのキーポイント検出手法で表現される局所特徴ベクトルが被写体の隠れ、回転、照明変化に対して頑健であり、高速かつ安定な画像検索に効果的であることが、多くの研究によって実証されている。局所特徴を用いた画像検索の手法は以下の2つのアプローチに大別される。

(1) Bag-of-Features(BoF) アプローチ

BoFアプローチでは、画像を1つの高次元かつスパースなBoFベクトルで表現する。BoFベクトルは画像から得られる局所特徴をクラスタリングし、クラスタ中心であるVisual Wordsごとの出現回数をカウントすることで生成される。BoFベクトルを用いた画像検索は、データベース画像から生成したBoFベクトルとクエリ画像から生成したBoFベクトルとの距離計算によって行われ、クエリ画像から作成したBoFベクトルに最も近い、BoFベクトルの画像が検索結果となる。

(2) Number-of-Match アプローチ

Number-of-Matchによる画像検索は、クエリ画像から抽出された局所特徴(クエリ特徴)とデータベース画像から抽出された局所特徴(インデックス特徴)の間で最近傍探索を行い、最も距離が近いものを対応付ける。そして、対応付けられたインデックス特徴をもつ画像に対して投票を行い、最も投票数が多い画像が検索結果となる手法である。図1にNumber-of-Matchアプローチの概念図を示す。

これら2つのアプローチを比較すると、BoFアプローチは、BoFベクトルがスパースであるため、BoFベクトル間の距離計算を行う際に、あるVisualWordを持つベクトルとそうでないベクトルでは、ベクトル間の距離が大きくなり、クエリ画像に似た画像を早い段階で絞り込めるため、大規模な画像検索問題に向く。ただし、クラスタリングなどの事前処理を必要とするため、計算コストが大きくなるといった問題がある。それに対して、Number-of-Matchアプローチは、クラスタリングなどの事前処理が不要であるため、計算コストが少なく高速である。ただし、データベースの画像の量が増えると、その分だけインデックス特徴の量も増えるため、大規模な画像検索に適用することは難しいといった問題がある。以上のことから、動画を

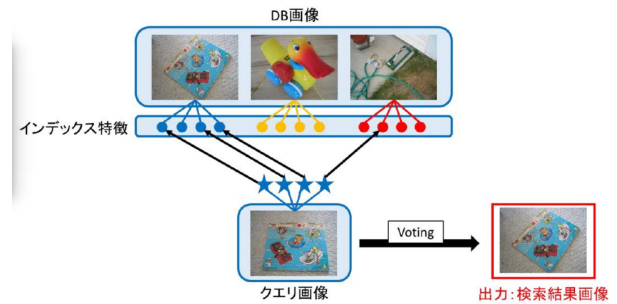


図1 Number-of-Match アプローチの概念図
Fig. 1 Illustration of Number-of-Match approach

連続してクエリとして与えて、似た画像を15fps程度のスループットで検索する、リアルタイム処理を行う画像検索システムでは、Number-of-Matchアプローチを用いた方が有利である。本研究ではこのアプローチでの画像検索法の改良に取り組む。

前述の通り、Number-of-Matchアプローチでは、画像から抽出されたキーポイントの対応付けに基づいて投票が行われるが、異なる画像間でのキーポイントの対応付けの中には、画像間での相似変換、アフィン変換、あるいはHomography変換などの剛体変換に合致しない対応付けが存在する可能性がある。これは、キーポイントの対応付けが、画像から抽出された局所特徴ベクトル間の距離に基づいて行われるからである。そのため、誤った対応付けに基づく投票が行われてしまい、クエリ画像と全く似ていない画像が検索されるという問題がある。

この問題を解決するためには、画像間の剛体変換を推定し、キーポイント同士の対応付けに幾何学的一貫性があるかどうかをチェックする必要がある。このとき、画像間の剛体変換を推定する方法として、キーポイント間の対応付けから、RANSACを用いて剛体変換を推定する方法がある。しかし、RANSACではノイズが存在するサンプルからパラメータを推定するために、多数回のランダムサンプリングが必要になる。このランダムサンプリングに時間がかかるため、リアルタイムの画像検索システムにはRANSACを用いることはできない。

本報告では、画像間に対応付けられた1組のキーポイントからRANSACを用いずに、キーポイントのスケール、オリエンテーションを利用することによって、画像間の剛体変換を推定し、剛体変換に従う対応付けを求める方法を提案する。また、提案手法を実際の画像検索システムに実装し、速度や投票数にどのような影響を与えるかということについても検討する。

2. 関連研究

画像検索の分野では、D.LoweによるSIFT[1]や、H.BayらのSURF[2]などの局所特徴がよく用いられている。中

でもコードブックを用いた画像検索の手法は、J.Sivic らの Vido Google[3] をはじめ、数多く提案されている。また、これらの手法を高速化する方法として、k-means クラスタリングを再帰的に適用することで得られる階層的コードブック (Vocabulary tree) を用いた手法 [4] も提案されている。しかし、コードブックを用いる手法では、画像をヒストグラムとして表現することにより、画像の空間的な情報が失われる問題がある。その問題に対して、Shen らの研究 [5] では、局所化されたコードブックを用いることで、コードブックの欠点である、空間的な情報の欠落を解消した画像検索を行っている。

一方で、局所特徴そのものを用いて画像検索を行う手法も提案されている。湯浅らの研究 [6] では、Multiple Instance Learning などの分野で用いられる、共通性を表す指標である Diverse Density を用いて、データベース画像の局所特徴の削減を行うことで画像検索の精度を向上させている。

本報告では、局所特徴そのものを用いた画像検索において、データベース画像とクエリ画像から得られるキーポイントの対応付けが、幾何学的に正しいかどうかを評価することによって、画像検索の精度向上を目指す。また、この手法は画像検索システムに適用しても、リアルタイムでの検索が実現できるよう、実行速度の面での評価も行う。

3. Local Feature Hashing

本章では、キーポイントの対応付けから RANSAC を用いずに画像間の剛体変換を推定し、誤った対応付けを削減する方法について説明する。図 2 のようにデータベースに存在する画像に対して、回転・拡大された画像がクエリとして与えられたとき、データベース画像上の点 $P_1 = [x_1 \ y_1]^T$ で検出されたキーポイント $K_1 = (P_1, \cos\theta_1, \sin\theta_1, s_1)$ と、クエリ画像上の点 $P_2 = [x_2 \ y_2]^T$ で検出されたキーポイント $K_2 = (P_2, \cos\theta_2, \sin\theta_2, s_2)$ が対応づいているとする。 θ_1, θ_2 はキーポイントのオリエンテーション、 s_1, s_2 はスケールであり、画像中の C_1 はデータベース画像の中心である。ここで 2 つの画像間の剛体変換 T を相似変換で表すと、キーポイント K_1 と K_2 の関係は

$$\begin{aligned} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} &= \Delta s \begin{bmatrix} \cos\Delta\theta & -\sin\Delta\theta & \Delta x \\ \sin\Delta\theta & \cos\Delta\theta & \Delta y \\ 0 & 0 & 1/\Delta s \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \\ &= T \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \end{aligned} \quad (1)$$

と表すことができる。行列 T の $\Delta x, \Delta y$ は平行移動、 Δs はスケールの変化、 $\Delta\theta$ は画像の回転を表している。

このとき、行列 T の未知パラメータは $\Delta s, \Delta\theta, \Delta x, \Delta y$ の 4 つであるため、通常であれば 1 組の対応点からではこれら全てのパラメータを求めることはできない。そこで、

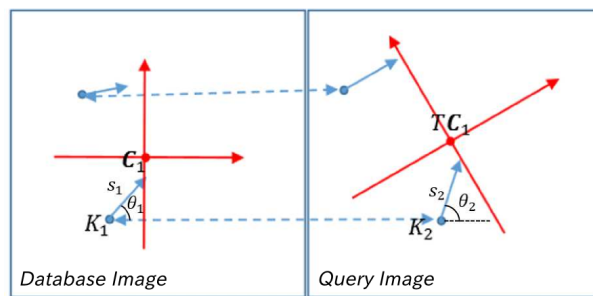


図 2 DB 画像とクエリ画像間のキーポイント対応け
Fig. 2 Corresponded Keypoints between DB and Query image

各キーポイントのオリエンテーションとスケールを利用して、 $\Delta\theta = \theta_2 - \theta_1, \Delta s = s_2/s_1$ とし、 $\Delta x, \Delta y$ は (1) 式を変形して、

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} - \Delta s \begin{bmatrix} \cos\Delta\theta & -\sin\Delta\theta \\ \sin\Delta\theta & \cos\Delta\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \quad (2)$$

と求めることができる。

ただし、1 組の対応点から相似変換を求めても、それが画像間の相似変換として正しいということは保証されない。そこで、全ての対応点から相似変換を計算し、式 (2) を用いて $\Delta x, \Delta y$ の 2 次元の投票空間に対して投票を行い、最も多くの対応組からの投票によって支持される相似変換を画像間の相似変換であるとみなす。さらにこのとき、最も投票数が多い bin に対して投票を行った対応点は、幾何学的に正しい対応付けであるとみなすことができる。画像検索を行う際には、この方法で正しいとみなした対応付けのみに基づいて投票を行うことで、検索精度を向上させることができる。と考える。

本来、投票によってパラメータを推定する場合は、 $\Delta s, \Delta\theta$ も含めた 4 つのパラメータに対しての投票を行うべきであるが、4 次元空間に対する投票となるため、メモリの消費量増えることや、最も投票の多い bin を発見するのに時間がかかるという問題が考えられる。4 次元の投票空間でピークが現れているならば、投票を 2 次元の投票空間に射影してもピーク場所は変わらないため、本手法では $\Delta x, \Delta y$ の 2 次元の空間での投票を行う。これにより、幾何学的に正しい対応付けを、全ての対応点間のパラメータの計算と 2 次元空間への投票という、少ない計算コストで求めることができるため、リアルタイムシステムへの組み込みが可能となっている。

4. 実験

提案手法による幾何学的の一貫性の評価が画像検索において有効であるかを確認するために、実験を行った。実験は提案手法によって、幾何学的に正しい対応付けを求めることができるかの確認と、実際に画像検索システムに実装した際に検索結果にどのような影響を及ぼすかを確認した。



図 3 Graffiti 画像セット
 Fig. 3 Graffiti image set

4.1 幾何学的一貫性の評価法に関する実験

Local Feature Hashing による幾何学的一貫性のチェックによって、幾何学的に正しいキーポイントの対応付けを求めることができるか、確認するための実験を行った。実験は Affine Covariant Regions Dataset[7] の Graffiti 画像セット (図 3) で、壁と正対して撮影された画像 (図 3.a) と、別の視点から撮影された画像 (図 3.b-図 3.d) とのマッチングを行い、投票の分布と、投票によって求められた対応付けが、幾何学的に正しい対応付けであるか確認を行った。キーポイント検出のアルゴリズムには SIFT を用い、投票空間 $\Delta x, \Delta y$ の bin サイズは 10×10 としている。

マッチングを行った画像間のすべての対応付けから $\Delta x, \Delta y$ を求め、投票を行った際の投票マップを図 4 に示す。img1 と img2 の対応付けでは投票のマップを見ると、投票空間内のある bin に投票が集中しピークが出現していることが確認できる。また、このピークの bin に対して投票を行った対応付けが、正しい対応付けであるかを、画像セットに付属している homography 行列を用いて確認したところ、すべての対応付けが正しいことが確認できた。次に img1 と img3 の対応付けでは、img2 との対応付けのときと比較すると、ピークの bin に対する投票数が少なくなっているが、依然として投票空間にピークが現れ、ピークの bin に対して投票を行った対応付けがすべて正しい対応付けであることが確認できた。最後に img1 と img4 の対応付けでは、投票が投票空間全体に散っており、局所的なピークも見られないため、剛体変換の推定及び幾何学的一貫性のチェックは行えなかった。

これらの実験の結果から、提案手法が画像間の剛体変換として相似変換を想定していながら、視点変化が発生しているような、画像間でのキーポイント対応付けでも幾何学的一貫性の評価が行うことができ、正しい対応付けも求め

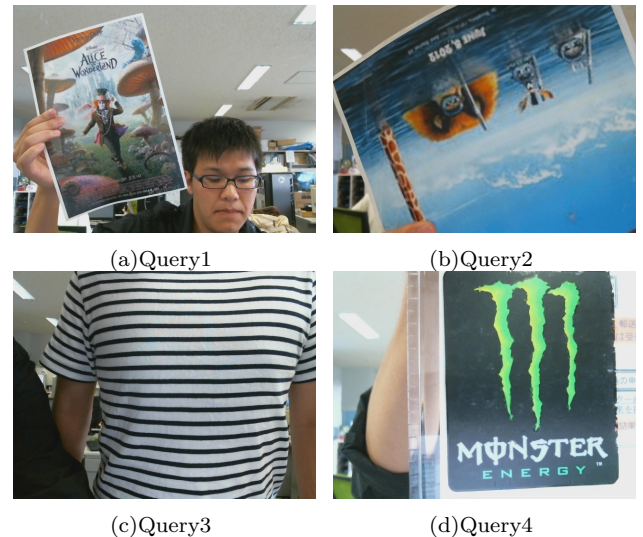


図 5 実験で用いたクエリ画像
 Fig. 5 Query images

ることができた。しかし、視点変化の激しい画像間の対応付けでは、幾何学的一貫性の評価が行えないという結果になった。

4.2 画像検索システムへの実装

画像検索システムに提案手法による幾何学的一貫性のチェックを実装して実験を行った。画像検索実験の手順は以下の通り。

- (1) クエリ画像からキーポイントを検出
 - (2) クエリ特徴とインデックス特徴を最近傍探索によって対応付け、対応づいたインデックス特徴を持つ DB 画像に対して投票を行う
 - (3) 投票数が多い上位 5 枚の画像を検索候補とし、それらに対して幾何学的一貫性のチェックを行い、正しい対応付けのみに基づく投票を行う
 - (4) 最も投票数が多い画像を検索結果とする
- (3) の部分で、データベースの画像の中からある程度候補を絞っているが、これは提案手法を全てのデータベース画像との対応付けに関して行くと、システムの実行速度の低下を招く可能性があるからである。

キーポイント検出のアルゴリズムには SURF を高速化した整数化 SURF[8] を使用し、データベースに保存されている画像の枚数は 1000 枚、幾何学的一貫性のチェックを行う際の、投票空間の bin サイズは 10×10 としている。また、データベース画像に対する投票数が少ない場合、クエリ画像と似ていると判断することは妥当ではないので、ある一定の投票数に満たなければ、似た画像がデータベースに存在しないと判定する閾値を設ける必要がある。この実験ではその閾値を 4 と設定している。

クエリ画像として 4 枚の画像 (Query1 - Query4) を与え、似た画像の検索を行い、幾何学的一貫性のチェックを

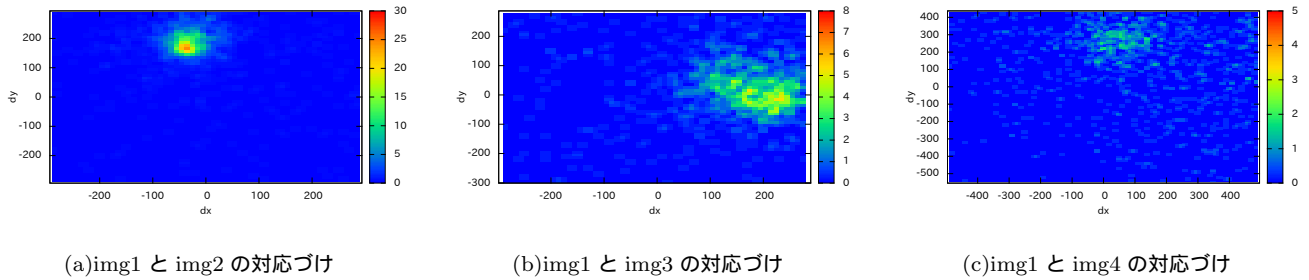


図 4 画像間の対応点による投票マップ

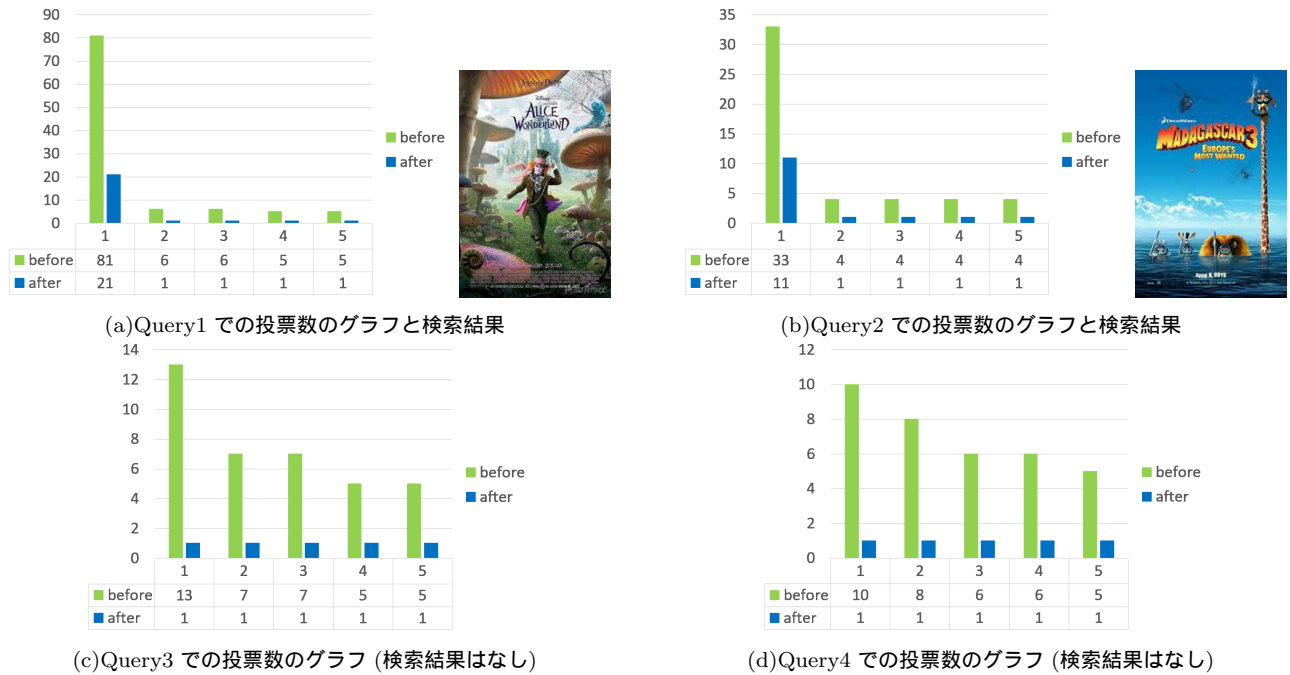


図 6 幾何学的一貫性のチェックを行う前後の投票数と検索結果

行う前後での検索候補の画像に対する投票数と、画像検索の結果に注目した。Query1, Query2 はクエリ画像中にデータベース内に登録されている画像が写っているものとなっており、Query3, Query4 はデータベースに登録されていない画像が写っている画像となっている。また、検索システムの動作速度は提案手法実装前と変わらず、15fps ほどのスループットで検索できていることを確認した。

幾何学的一貫性のチェックを行う前後での、各検索候補の画像に対する投票数を図 6 のグラフで示している。Query1 をクエリとして与えて画像検索を行った結果、正解の画像に対する投票数が、幾何学的一貫性のチェックを行う前に比べて減少してはいるが、検索候補の中で最も投票数が多いため、正しい検索が行われていることが確認できた。Query2 に関しても同様に、幾何学的一貫性チェックの後に投票数が減少したが、正しい検索が行われていることが確認できた。Query3, Query4 に関しては、幾何学的一貫性のチェックを行う前では、検索候補の画像の投票数が閾値を上回っているため、データベースに存在しない画像をクエリとしているにも関わらず、データベース内の画

像が検索されてしまっている。ところが、幾何学的一貫性のチェックを行うことで、全ての検索候補の画像に対する投票数が減少し、投票数が閾値より低くなったため、クエリ画像に似た画像がデータベースに存在しないと判定できた。

実験の結果から、提案手法実装前と比べて、データベースに存在する画像と似た画像を与えたときの検索結果に変わりはない。データベースに似た画像が存在しない画像をクエリとして与えたとき、全く関係のない画像を検索結果として返すことなく、クエリ画像と似た画像がデータベースに存在しないと判定することが出来たことから、提案手法により誤った検索を抑止できることが確認できた。

5. おわりに

本報告では、画像間のキーポイントの対応付けから RANSAC を用いず、キーポイントのスケール、オリエンテーションの情報を利用し高速に相似変換を求め、求めた相似変換のパラメータに対して投票を行うことで、幾何学的に正しい対応付けを求める方法を提案した。

提案手法では、画像間の剛体変換として相似変換を想定していながらも、ある程度視点変化のある画像間でも正しいキーポイント対応付けを求めることができたが、視点変化の激しい画像間での対応付けではうまくいかなかった。この問題の解決策として、データベースの画像に Homography 変換をかけた画像を作成しその画像もデータベースに保存しておくことで、視点変化の激しい画像間でも幾何学的に正しい対応付けを求めることができると考えられる。

画像検索においては、処理速度は提案手法実装後も動作速度に変化はなく、15fps 程度のスループットで動作することが確認できた。検索精度の面では、データベースに格納されている画像が写っているクエリ画像を与えた場合、正しい検索が行われ、データベースに存在しない画像をクエリとして与えた場合の、誤検索を抑えることができた。

今後は提案手法による幾何学的の一貫性のチェックの結果から、クエリ画像中のどの領域に正しい対応付けが存在するかを求めることで、クエリ画像中のどの部分に、データベース内の画像と似た画像が存在する領域を特定する方法について検討する。これが実現できれば、背景から現れる不要なキーポイントを除外し、画像検索に必要なキーポイントのみを用いた検索が可能になる。

参考文献

- [1] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110 (2004).
- [2] Bay, H., Tuytelaars, T. and Van Gool, L.: Surf: Speeded up robust features, *European conference on computer vision*, Springer, pp. 404–417 (2006).
- [3] Sivic, J. and Zisserman, A.: Video Google: A text retrieval approach to object matching in videos, *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, IEEE, pp. 1470–1477 (2003).
- [4] Nister, D. and Stewenius, H.: Scalable recognition with a vocabulary tree, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2, IEEE, pp. 2161–2168 (2006).
- [5] Shen, X., Lin, Z., Brandt, J., Avidan, S. and Wu, Y.: Object retrieval and localization with spatially-constrained similarity measure and k-NN re-ranking, *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3013–3020 (2012).
- [6] Yuasa, K. and Wada, T.: Keypoint Reduction for Smart Image Retrieval, *Multimedia (ISM), 2013 IEEE International Symposium on*, IEEE, pp. 351–358 (2013).
- [7] : <http://www.robots.ox.ac.uk/vgg/data/data-aff.html>.
- [8] 吉岡勇太, 和田俊和: FPGA を用いた SURF の実時間計算法, *Technical report of IEICE. HIP*, Vol. 109, No. 471, pp. 241–246 (オンライン), 入手先 (<http://ci.nii.ac.jp/naid/110008002595/en/>) (2010).