

奥行き情報を用いた SIFT 特徴量の 高ロバスト化に関する一検討

後藤慎一[†] 青木輝勝[†]

物体認識の基盤技術として広く利用されている SIFT 特徴量は、スケール変化や回転変化に頑強な特徴量記述を目指したアルゴリズムであり、理論上は非常に高精度な画像マッチングの実現が可能である。しかし実際には、カメラと物体間の距離がわずか数メートル離れるだけでキーポイント数が激減し、その結果マッチング精度が著しく劣化するという脆弱性を持っている。従って、例えば車載カメラにおける自動運転等のアプリケーションを想定した場合には精度が不十分である。この問題を解決するため、本稿では“Distance-free”という概念を導入した新しい画像マッチング手法を提案する。

A study on high-robust SIFT features using depth information

SHINICHI GOTO[†] TERUMASA AOKI[†]

“SIFT features,” which is widely used as a fundamental technique of object recognition, is the algorithm which can describe image features invariant to image scale and rotation, and can provide high-robust matching between images in theory. In practice, however, SIFT has a weakness that its accuracy of matching falls down significantly when the distance between a camera and an object is changing only a few meters. For example, its performance is insufficient to apply for automatic driving using vehicle cameras. To resolve this problem, we propose a new image matching method using a concept of “Distance-free” in this paper.

1. はじめに

画像特徴量はコンピュータビジョンにおいて極めて重要な概念である。とりわけ画像の部分領域に注目する局所特徴量は、物体認識や画像検索など多くの分野において不可欠であるが、本研究ではその中でも精度が高いと言われる SIFT 特徴量[1][2]を対象とする。SIFT 特徴量は DoG や主曲率・サブピクセル推定の利用によりスケール変化に頑健であり、またキーポイント毎のオリエンテーションの割り当てにより回転に不変で、128 次元ベクトルによる最終的な記述を行うことで、理論上は非常に高精度な画像間マッチングが実現できる。しかし実際には、例えばカメラから物体までの距離の変化やこの空間上での回転変化、または光沢や陰影の影響・物体付近の背景の影響等、3次元空間上の環境変化に対しては影響を受けやすく、マッチング精度に問題が生じる。本稿ではこのうち、精度が距離に依存しやすいことに注目する。

3次元空間上においては物体が奥に存在すればするほど、その解像度の変化・画像ぼけの影響によってマッチング精度は大きく減衰してしまう。場合によってはわずか数メートル先の物体とテンプレート画像とのマッチングすら困難であるという脆弱性を示しており、例えば車載カメラによる自動交通標識認識[3]を考えればその安全性に不安が残るし、ロボットビジョンにおいても不十分な精度である。SIFT 特徴量は物体認識のうちでも全く同じ物体であるかどうかを言い当てる“特定物体認識”を主に対象としており、対象物体の一致・不一致の判断を的確に下すことが求められるにも関わらず、このようにカメラと物体間の距離に影響を受けやすいという欠点がある。本稿ではこの問題を解決するため、新たに画像の奥行き情報を利用することを提案し、距離に依存しない Distance-free である特徴量の記述を目標とする。

2. 従来研究と課題

SIFT(Scale Invariant Feature Transform)特徴量のアルゴリズムは以下の4段階に分類される。

1. 極値探索
2. キーポイントのローカライズ
3. オリエンテーションの割り当て
4. 特徴量記述

この4段階のうち、まず“極値探索”に関して述べ、残る3つのステップについても本稿での提案手法と関わりの深い部分のみ概説する。

[†] 東北大学 大学院情報科学研究科 / 東北大学 未来科学技術共同研究センター
Graduate School of Information Sciences (GSIS), TOHOKU University
/ New Industry Creation Hatchery Center (NICHe), TOHOKU University

2.1 極値探索

第1段階として、複数の平滑化画像と DoG 画像から成る空間毎に、画像中のキーポイント候補点とスケールを検出する。DoG 画像とは以下の DoG(Difference of Gaussian) 処理によって得られる画像を指す。

ガウス関数 $G(x, y, \sigma)$ と入力画像 $I(u, v)$ を畳み込むことで得られる平滑化画像 $L(u, v, \sigma)$ に対し、ガウス半径 σ を変化させた前後でのその差分により生成される画像を DoG 画像 $D(u, v, \sigma)$ と定義する。これらはそれぞれ以下の式で求められる。

$$L(u, v, \sigma) = G(x, y, \sigma) * I(u, v) \quad (1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi \sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2)$$

$$D(u, v, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(u, v) \quad (3)$$

$$= L(u, v, k\sigma) - L(u, v, \sigma)$$

輝度値の勾配が急峻な部分ほど平滑化前後での変化量も大きくなる。従って図1に示すように、その平滑化画像で差分をとるという処理を σ_0 から k 倍ずつ大きくした異なるスケール (ガウス半径と同義) 間で行った時、DoG の値が大きくなる σ においてはスケールの変化領域に画像中の急峻な部分の情報を多く含んでいると考えることができる。その極値を検出することで画像中の特徴となるキーポイント候補点を決定することが可能で、この検出では3枚の DoG 画像一組を利用し、注目画素とその周辺26近傍との比較を行う。この処理を σ の小さい DoG 画像から順に全画素に関して行う。この際、スケールと DoG の出力値は比例することが分かっており、これによりスケール変化に頑健な特徴量の記述が可能となる[4]。

なお、平滑化を行う際にガウス半径を増加し続けると、フィルタのサイズも並行して増大するため、計算量が膨大になる。このコストを低減するために SIFT ではダウンサンプリングを行う。

初期値 σ_0 のスケールで得られた平滑化画像 $L_1(\sigma_0)$ に対し、 σ_0 を $k = \sqrt{2}$ 倍ずつ順に増加させ、丁度 $2\sigma_0$ による平滑化画像 $L_1(2\sigma_0)$ を得た場合を考える。 $L_1(2\sigma_0)$ を $1/2$ のサイズ $L_2(\sigma_0)$ にダウンサンプリングとしたとき、

$$L_1(2\sigma_0) \approx L_2(\sigma_0) \quad (4)$$

であることが証明されている。これは、 $1/2$ にダウンサンプリングした画像を $1/2$ のガウス半径で平滑化を行うことで、元の平滑化画像と同様の結果が得られることを示しており、結果生ずる DoG 画像、および検出される極値も同様となる。すなわち、ダウンサンプリングすることで計算コストを大きく減らした状態での極値の検索が可能となる。ここで、図1のように同サイズの各処理空間単位を1オクターブと呼んでいる。

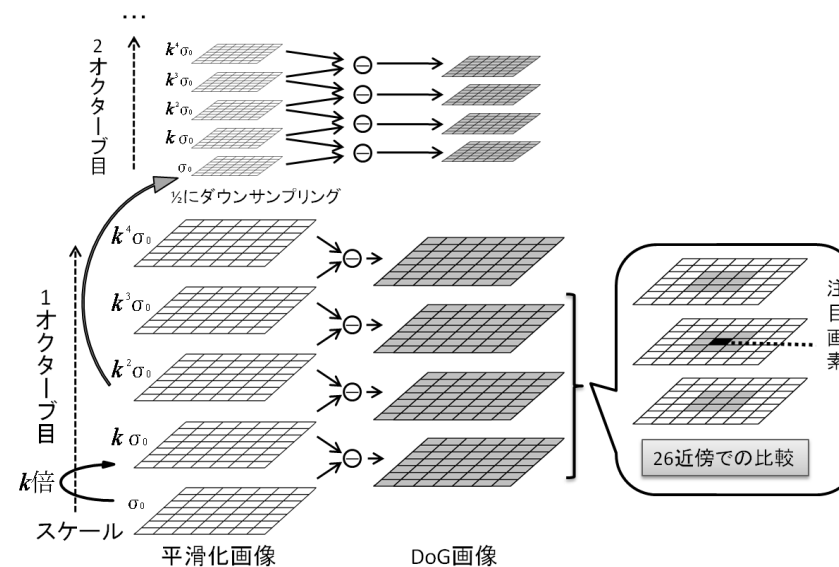


図1 DoG(Difference of Gaussian)処理

2.2 その後の3ステップ

極値探索によって決定した候補点のうち、マッチングに向かない点を手順2のローカライズで削除する。まず候補点のサブピクセル推定を行い、推定座標において DoG としての出力値が小さい点、および主曲率の算出により画像のエッジ上に存在すると考えられる点を省き、最終的なキーポイントを決定する。

手順3では各キーポイントの近傍の局所領域をキーポイントが得られたスケールに応じて決定し、この領域内において平滑化画像での勾配強度・勾配方向を算出し、 360° 方向を量子化した36方向ヒストグラムを作成する。この最大値の80%以上のピーク値となった方向をそのキーポイントのオリエンテーション (方向) と決定し、これにより回転に不変な特徴量が記述できる。なお、勾配強度の算出の際はスケールの1.5倍のガウス円により重み付けを行っている。

手順3で求めたオリエンテーション方向にキーポイント周辺領域の勾配情報の座標軸を回転することで正規化を行った上で手順4を行う。キーポイントが持つスケールを半径とした円領域を 4×4 のブロックに分割し、スケールに応じた重み付けを利用し、ブロックごと8方向の勾配強度に関するヒストグラムを作成する。この結果、 $4 \times 4 \times 8 = 128$ 次元の特徴ベクトルが記述される。最終的に記述された各ベクトルに関

してユークリッド距離の最小値を探索することで、SIFT 特徴量間のマッチングが可能となる。

2.3 SIFT 特徴量の課題

DoG の利用や高次元での記述のために計算量が莫大であることその他、そもそものマッチング精度に影響を与えてしまう要因として以下が挙げられる。

- ① 対象物体付近の背景
- ② 3次元空間上での物体の向き
- ③ 画像ぼけ
- ④ 物体の奥行き距離

本稿が対象とするのはこのうちの④である。SIFT は理論上、すなわち、画像情報がアナログ波形であると想定した場合には物体サイズの変化に頑健であるにも関わらず、デジタル画像を対象とした場合にはサイズが異なるだけの2つの画像同士の特徴量が完全には一致しない。画像幾何学上、奥行きが n 倍の物体は画像上でその一辺の長さが $1/2$ (面積は $1/n^2$) となるため、この不一致は用途 (車載カメラ等) によっては大きな問題となる。次章以降で、これを改善する提案手法に関して説明していく。

3. 奥行き情報を用いた SIFT 特徴量

3.1 Distance-free の概念

既存の SIFT 特徴量を用いた画像間のマッチングは物体とカメラ間の距離が離れるほどその精度が劣化する (図 2 左)。本稿の提案手法は新たに奥行き情報を付与することで、図 2 右に示すように距離に依存しないマッチングを達成できる Distance-free である特徴量の記述を目標とする。本章ではその実現方式を提案する。

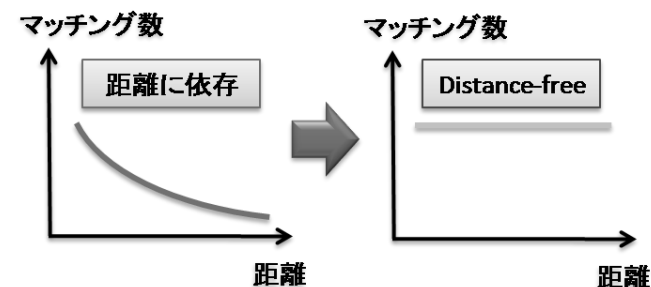


図 2 Distance-free の概念

3.2 ガウス半径による補正

SIFT アルゴリズムの問題点として、平滑化画像の生成やオリエンテーション割り当てに使用しているガウス半径の初期値が常に一定であることが挙げられる。物体のサイズが把握できていればそれに応じてこの値は変更すべきであり、例えば図 3 のようにある画像の $1/2$ の画像 (奥行き距離 2 倍の画像) でのガウス半径の初期値を $1/2$ にする。これによって元画像でキーポイントが検出される低スケール空間に対応する空間を $1/2$ の画像の場合にも考慮することができ、元画像の場合に近い特徴量を記述することができる。

ガウス半径による補正は理に適ったものではあるが、SIFT アルゴリズムの各手順において、以下の箇所の変更が必要となってしまう。

- ① 極値探索 - 26 近傍の定義
- ② オリエンテーションの割り当て - 近傍領域
- ③ オリエンテーションの割り当て - 勾配強度への重み
- ④ 特徴量記述 - 円領域の半径値
- ⑤ 特徴量記述 - 勾配強度への重み

3.3 画像サイズによる補正

前節で述べた手法は実装がやや複雑であることから、この手法と理論上同様の結果が期待でき、かつ実装が容易な手法を提案する。前節の例を再び用いると、図 3 のように $1/2$ の画像を何らかの方法によりあらかじめ定義した補正サイズに直す (図 3 の例の場合は 2 倍することになる) ことで、低スケール空間に対応する空間を考慮することができ、省かれていた特徴量を記述することが可能となる。

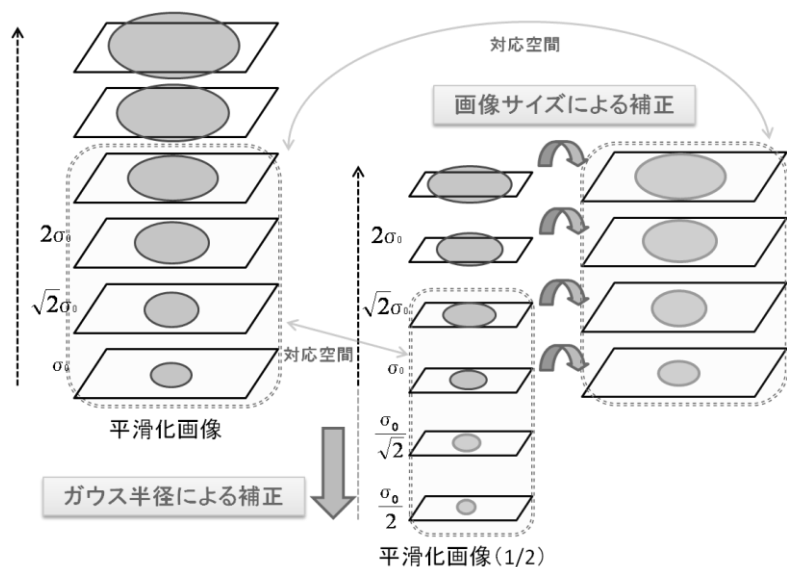


図 3 ガウス半径による補正と画像サイズによる補正

3.4 デプスマップを利用した分割記述の流れ

画像サイズによる補正を行うにあたり、カメラと物体間の距離情報が得られれば、奥行き距離と画像上のサイズの相似関係に基づき予め定義された補正サイズを容易に得ることができる。そこで本稿ではデプスマップにより奥行き情報を取得し、補正された特徴量を以下の手順に従って記述することを提案する。

まず2枚のステレオ画像を取得し、グラフカットの原理を組み込んだ上で左右視差を利用したステレオマッチングによってデプスマップを生成する[5]。デプスマップの明度値と距離情報の関係は予め定式化しておき、この明度値に応じて画像の領域を分割して考える。各々に関して適当な補正を施した特徴量を元画像の輝度値に基づいて記述し、全体で和をとり最終的な特徴量とする。例としてデプスマップを均等に4分割した場合の処理の流れを図4に示す。この場合は視覚化しやすいように4分割しているが、実装上256分割することも可能である。

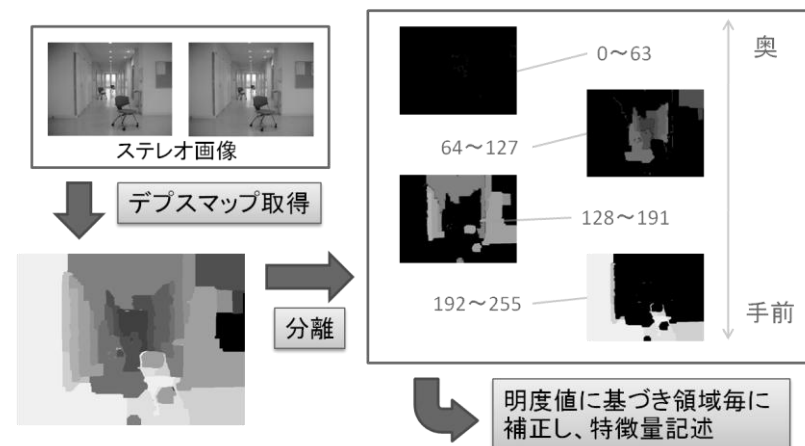


図 4 分割記述の流れ

4. 提案方式の評価

4.1 実験目的

デプスマップの生成技術は細部まで見るとその精度に難があるため、本実験では画像中において奥行き情報が表れやすい位置に対象物体を配置し、その明度値から距離情報を取得する。この距離情報を用いて画像サイズを予め定義されたサイズに補正し、用意しておいたテンプレート画像とのマッチングを行う。これを奥行き距離の異なる様々な画像に適用することで、Distance-free となるマッチングの実現が可能であることを実証する。

4.2 撮影条件

実験に利用した画像の撮影の諸条件を以下に示す。

- 撮影対象とした物体は A, B の 2 種類 (図 6 に結果と共に示す)
- $d = 2.8, 4.2, 5.6, 7.0, 8.4, 9.8, 11.2$ [m] の場合に関して撮影
- 対象物体の垂直方向に対する角度が $0^\circ, 30^\circ$ の 2 つの場合に関して撮影

ここで d はカメラと対象物体間の距離であり、 d がある値の α 倍となったとき、画像中での物体のサイズは $1/\alpha$ 倍となる。すなわち、撮影条件は $d = 2.8$ [m] を基準点として、物体のサイズが 1 倍, $2.8/4.2(=2/3)$ 倍, $2.8/5.6(=1/2)$ 倍, ... となるよう調整を行ったものである。

4.3 実験方法と結果

提案手法に基づき奥行き情報を利用した SIFT 特徴量を記述し、通常の SIFT 特徴量の場合の結果と比較し評価した。手順は以下の通りである。なお計算量削減のため、事前に全ての撮影画像を 640×480 サイズに変換した。

- ① 画像上で対象物体付近を切り抜く
- ② デプスマップを利用し対象物体の明度値を決定(面積割合が最大のものとする)
- ③ 明度値からカメラと対象物体間の距離 x を計算し推定
- ④ $d = 2.8[m]$ 時を基準とし、③で求めた距離を利用し画像サイズを補正
- ⑤ SIFT 特徴量の記述
- ⑥ テンプレート画像とのマッチング

基準とした $d = 2.8[m]$ 時にテンプレート画像とのマッチングが可能なのは事前に確認済みである。補正においては、距離が $x[m]$ であると推定された物体は基準画像中での物体の $(2.8/x)$ 倍であると予想されるから、サイズを $(x/2.8)$ 倍すればよい。

手順①から⑥までを d が $4.8[m]$ 以上の画像に対して実行した場合の結果(距離に対するキーポイント数およびマッチング数)を、既存の SIFT 特徴量による結果ととものに図 6, 図 7 に示す。実線が角度 0° 、点線が角度 30° の場合である。

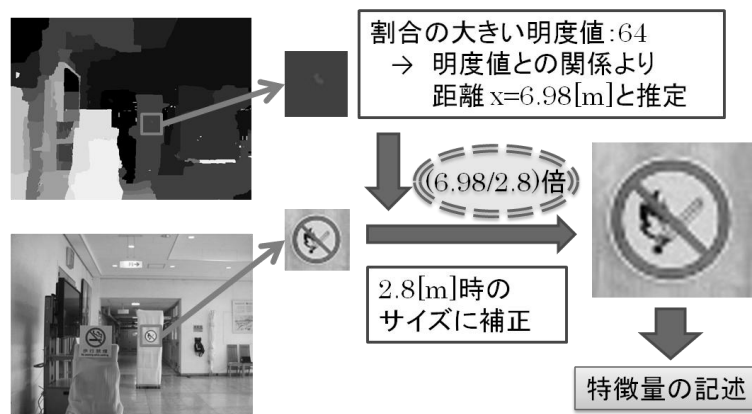


図 5 評価実験の流れ

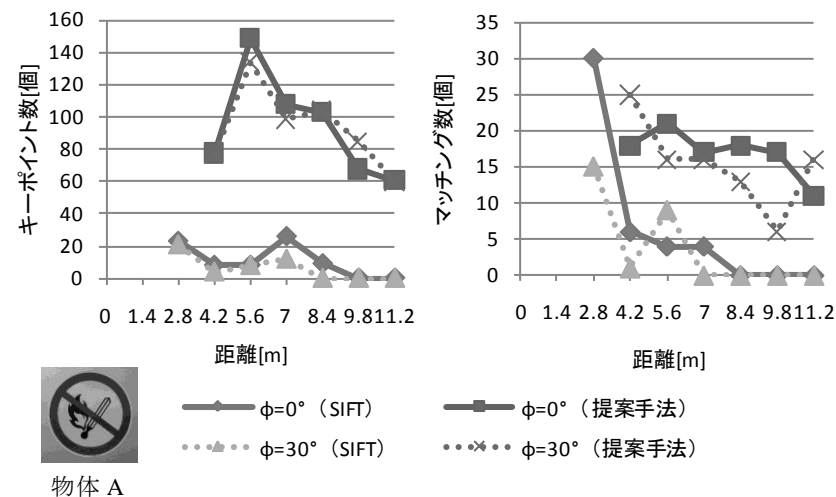


図 6 物体 A での結果

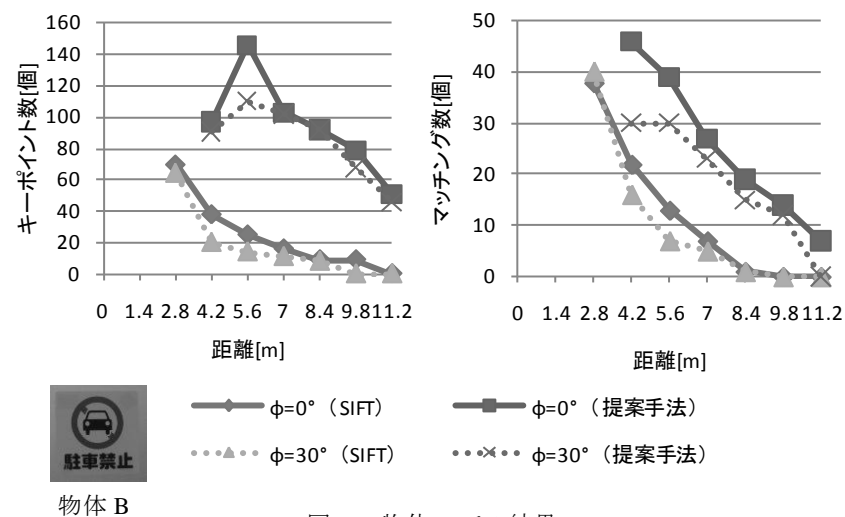


図 7 物体 B での結果

4.4 結果に対する考察

キーポイント数・マッチング数ともに、全ての場合に関して提案手法において既存手法以上の場合を上回る値が得られた。キーポイント数は数と性能が比例するわけではないが、ある程度の個数が存在しないとマッチングの信頼性が揺らぎかねないことから、奥行き距離に依存せず一定数のキーポイントが抽出できること(Distance-free 化)は重要である。図 6, 図 7 より、物体 A,B の両方でキーポイント数についてある程度の Distance-free 化が実現できていることが読み取れる。

一方マッチング数に関しては、特に物体 A は 4.2[m]以降のマッチング数がある程度安定傾向にあり、Distance-free である特徴量が記述できていることが読み取れる。これに対し物体 B では既存の SIFT に比べ高精度の結果を示したものの、距離に比例してマッチング点数が減少する結果となった。物体 A, 物体 B の結果に差が生じている原因としては画像の密度と色が考えられる。画像の密度に関しては、物体 B は文字情報を含んでおり、サイズが大きい場合はこの付近でも多くのマッチングを行っている。しかし対象領域が小さくなるにつれてサイズ補正の際のぼけの度合いが強くなり、特にこの密な部分はその影響をより強く受け、正しい演算が困難になってしまったと考えられる。一方、画像の色に関しては、SIFT のアルゴリズムでは全処理をグレースケールへの変換後に行っているため、画像内で差分の出にくい物体 B ではマッチング精度が劣化してしまったと考えられる。これらを解決するため、画像の密度や色を予め判定し、高密度な場合やグレースケール変換時に差分が小さい場合の前処理を工夫することにより、精度向上が期待できると考えられる。

また、物体の角度を 30°傾けることにより精度が多少不安定になった。更に大きい角度変化、若しくは物体自身に奥行きがある場合には精度がより低下してしまうため、例えば ASIFT[6]の利用を検討することが必要である。

5. まとめと今後の課題

本稿では、SIFT 特徴量の精度が対象物体までの距離が離れることで劣化してしまうという問題に対し、新たにデプスマップから取得される奥行き情報を利用し、Distance-free である特徴量の記述手法を提案した。実験では画像中の対象物体に注目し、デプスマップからその物体付近の明度値を読み取り、推定される距離を利用して画像に補正を加えた上で特徴量を記述した。テンプレート画像とのマッチング精度に関して既存の SIFT 特徴量による結果と比較し、その有用性を評価した。改善点は残されているものの Distance-free の実現可能性を確認でき、既存手法を上回る精度結果を得ることに成功した。

但し今回採用した画像サイズによる補正は、同時に走査が必要な画素数が増加して

しまうため、処理領域が画像全体になった場合に元々長い SIFT アルゴリズムの演算時間が更に膨大になってしまうことが危惧される。従って今後の課題として、まずガウス半径による補正手法の理論を細部まで確立させる必要があり、そのために 3.2 節において挙げた各項目を検討する必要がある。また本稿ではカメラと物体間の距離が 2.8[m]の際を基準としたが、この補正目標の決定手法を一般化する必要もあるだろう。

更に、画像全体を対象として明度値に応じて領域毎に処理を行うアルゴリズムを実装するために、精度の高いデプスマップをステレオ画像から得る技術も必要である。画像全体での記述を考えた場合の莫大なキーポイント数も問題となることが予想されるため、キーポイントのローカライズの際の条件を変更するか、若しくは鮮鋭化等の新たな処理によりキーポイント数を削減することが必要となる。

なお、今回は条件としてステレオ画像が必須なため、画像検索等への応用は困難であるかもしれない。しかし 2 眼カメラが用意できる環境下、例えば車載カメラへの利用を考えた場合利用価値は高く、車椅子やロボットに組み込むことで障害物判定にも利用可能である。デプスマップの精度も徐々に改善されつつあり、今後コンピュータビジョンという分野の発展に貢献していくことが期待される。

参考文献

- 1) David G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints, Proc. of International Journal of Computer Vision (IJCV), 60(2), pp. 91-110 (2004).
- 2) 藤吉弘亘: Gradient ベースの特徴抽出 - SIFT と HOG -, 情報処理学会 研究報告 CVIM 160, pp. 211-224 (2007).
- 3) 高木雅成, 藤吉弘亘: SIFT 特徴量を用いた交通道路標識認識, 第 13 回画像センシングシンポジウム SSI07, LD2-06 (2007).
- 4) T. Lindeberg: Feature Detection with Automatic Scale Selection, Int'l J. Computer Vision, vol.30, no.2, pp. 79-116 (1998).
- 5) M. Bleyer, M. Gelautz: Graph-based surface reconstruction from stereo pairs using image segmentation, SPIE Symposium on Electronic Imaging (Videometrics VIII), vol.5665, pp. 288-299 (2005).
- 6) Jean-michel Morel, Guoshen Yu: ASIFT: A New Framework for Fully Affine Invariant Image Comparison, SIAM Journal on Imaging Sciences, Volume 2 Issue 2 (2009).