

ライトフィールド情報を活用した特徴点の対応付けに関する検討

清水 政行[†] 川西 康友[†] 出口 大輔^{††} 井手 一郎[†] 村瀬 洋[†]

[†] 名古屋大学情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町

^{††} 名古屋大学情報連携統括本部 〒464-8601 愛知県名古屋市千種区不老町

あらまし 近年、ライトフィールドカメラが市場に導入されてライトフィールド情報を容易に得ることができるようになった。ライトフィールド情報からは、注目画素の鮮鋭度を算出しピントが最も合う焦点距離を推定することができる。本論文では SIFT 特徴量の対応付け問題において、焦点距離を活用して信頼度の低い特徴点を削除し対応付け精度を向上させる手法を提案する。本手法により、誤対応の原因となる、テクスチャが少ない特徴点を削除することができると考えられる。提案手法の有効性を確認するために対応付けの精度を確認した。その結果、対応付けの総数、正対応付け数、Precision のいずれも向上することを示した。

キーワード ライトフィールド, SIFT, 特徴量, マッチング

A study on keypoint matching with light field information

Masayuki SHIMIZU[†], Yasutomo KAWANISHI[†], Daisuke DEGUCHI^{††}, Ichiro IDE[†], and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan

^{††} Information Strategy Office, Nagoya University Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan

Abstract Recently, it is easier to obtain light field data because light field camera is commercially available. From light field data, we can use contrast-based measure to find an optimal focal length at each pixel. We propose a new method to eliminate lower confident keypoints from the conventional SIFT keypoints with an optical focal length. As a result, our proposed method improve number of all matching keypoints, correct matches, and precision.

Key words light field, SIFT, SIFT feature, keypoints matching

1. はじめに

カメラ画像を用いた認識技術の向上は目覚ましく、パターン認識だけでなく、カメラモーションの推定や地図を生成するような幾何パラメータを推定する研究も盛んである。特に SIFT 特徴量 [1] は Bag of Keypoints [2] や SLAM [3] まで幅広く活用されており、また SIFT そのものについても数多くの改良手法が提案されている [4] [5]。一方でカメラについてはライトフィールドカメラが市場に導入され注目され始めている。[6] [7] [8]。そこで本稿では SIFT 特徴を基にライトフィールド情報を利用して特徴点对応付けの精度を向上する手法と結果を報告する。本提案手法の全体フローを図 1. に示す。まず、ライトフィールド情報から焦点距離方向 (奥行き) を含めた 3 次元の画像データを取得する (2. 節)。以降はここで得られた 3 次元画像に対して処理を行なう。得られた 3 次元画像から鮮鋭度を用いてピントが最も合う焦点距離を各画素毎に算出する。ここでピントが

最も合う焦点距離を算出できなかった画素はテクスチャが不鮮明で特徴点として相応しくないため、特徴点から削除する (4. 節)。SIFT 特徴点の抽出と特徴量演算はすべての画像に対し処理を行なう (4. 節)。最終的に 3 次元の特徴点を 2 次元に統合し、特徴点对応付けの精度について検討を行った (6. 章)。

2. ライトフィールド

ライトフィールド (光線空間) とは場に対する光線情報を意味しライトフィールドカメラは光線情報を取得できるカメラである。初期のライトフィールドカメラは、多数のカメラを縦横に並べたカメラアレイによる取得方法により実現された。一方、Ng 等は撮像素子の前面にマイクロレンズアレイを配置したより実用的なライトフィールドカメラを製品化した。本研究ではライトフィールド情報を取得するカメラとして Lytro 社の Lytro illum [9] を使用する。ライトフィールドカメラは主レンズとマイクロレンズを介して画像センサにより撮像される。図

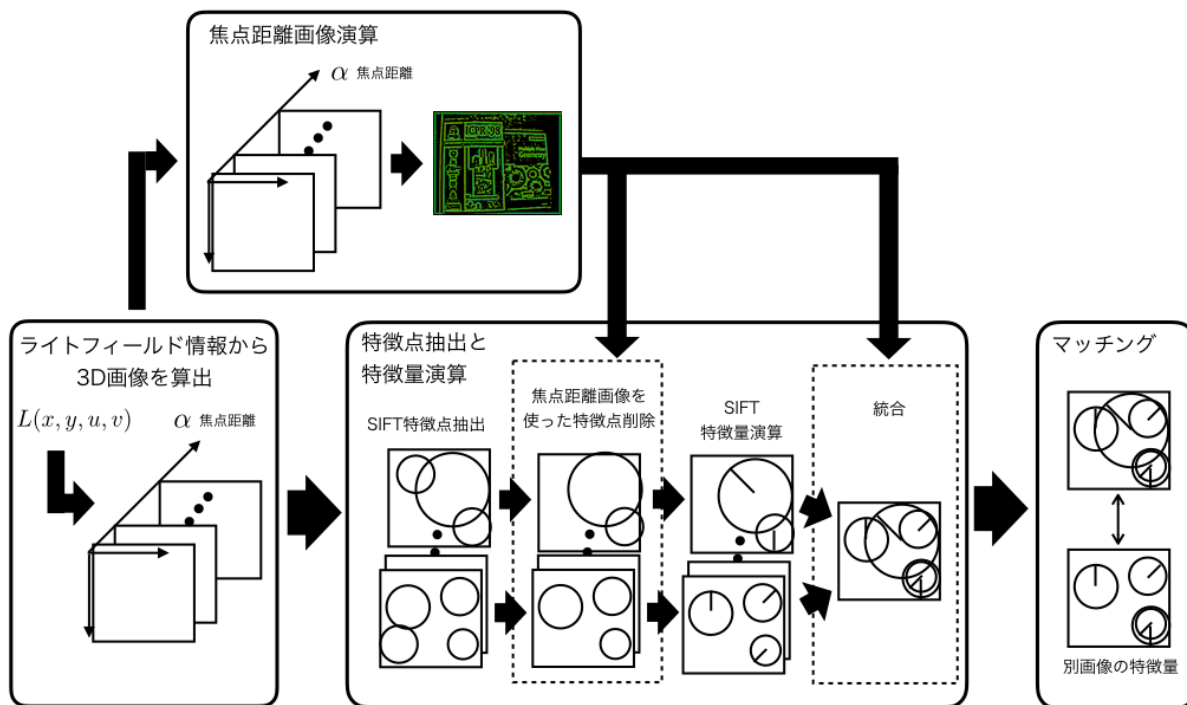


図1 ライトフィールド情報を活用した特徴量マッチングの全体図
 Fig.1 Overview of keypoints matching with light field information

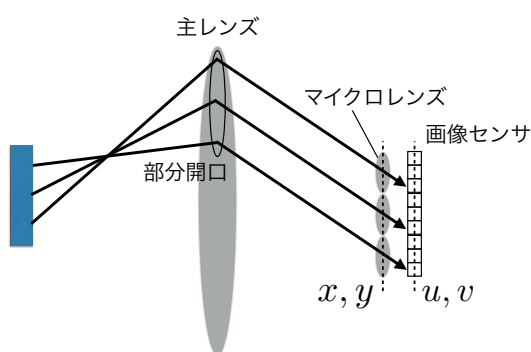


図2 ライトフィールド情報の取得原理
 Fig.2 Principle of recording light field information

2. にライトフィールドが取得される原理を示す。物体からの光線は部分開口を通り各マイクロレンズの同じ座標で記録される。この原理に基づき、マクロレンズ面での座標 x, y と画像センサ面の座標 u, v を取得することで光の方向を含めた4次元のライトフィールド情報 $L(x, y, u, v)$ を得ることができる。例えば、式(1)は u, v が0の画素を集めてできた部分開口画像となる。

$$I(x, y) = L(x, y, 0, 0) \quad (1)$$

ライトフィールド情報を活用することで任意の焦点距離の画像を生成することができる。任意の焦点距離の画像、すなわちリフォーカス画像は以下の式(2)で得ることができる[6]。ここで αF は任意の焦点距離、 F はベースとなるカメラの焦点距離

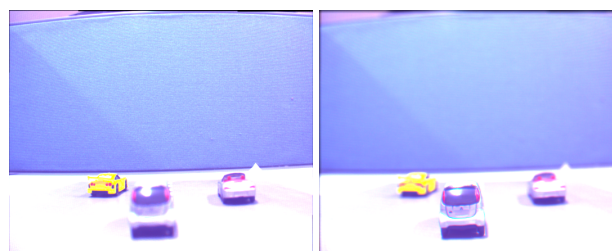


図3 リフォーカス画像
 Fig.3 Refocus images

を表す。

$$I_\alpha(x, y) = \iint L\left(x + \left(1 - \frac{1}{\alpha}\right)u, y + \left(1 - \frac{1}{\alpha}\right)v, u, v\right) dudv \quad (2)$$

式(2)は積分による連続系で表現しているが実際の演算では離散的に部分開口画像を加算する。式2の演算により得られた任意の焦点画像を図3に示す。左の画像は遠い焦点距離の画像、右の画像は近い焦点距離の画像を生成した結果を示す。画像左奥の黄色い車両と中央下の銀の車両を見ると焦点距離が変化した画像が生成できているのがわかる。

以降、本稿では α の値を変えながら式2により得られる焦点距離の異なる3次元画像で特徴量マッチングの検討を行った。

3. 鮮鋭度の算出とピントが最も合う焦点距離の推定

ここでは各画素毎にピントが最も合う焦点距離を求める。画

像は焦点が合った際に最も鮮鋭度が高くなるため、焦点距離を求めるには、2. 節で得られた 3 次元画像から、各画素近傍で最も鮮鋭度が高くなる焦点距離を求めれば良い [10]. 鮮鋭度 D_α はガウシアンフィルタを施した画像と元画像の差分画像 I' を求めた後、注目画素を中心とした窓領域内の画素値を加算することで計算する.

$$I'_\alpha(x, y) = I_\alpha(x, y) - G(x', y', \sigma) * I_\alpha(x, y) \quad (3)$$

$$D_\alpha(x, y) = \frac{1}{|W_D|} \sum_{W_D} |I'_\alpha(x, y)| \quad (4)$$

ここで $G(x', y', \sigma)$ はガウス関数, W_D は窓の幅を表す. 鮮鋭度 D_α が最大値となる焦点距離 α_D が最もピントが合った焦点距離となる.

$$\alpha_D(x, y) = \arg \max_{\alpha} D_\alpha(x, y) \quad (5)$$

図 4 に本手法で算出した焦点距離画像を示す. 画像中の色がピントが最も合った焦点距離を示しており、赤が近い距離、青が遠い距離を表している. 黒は鮮鋭度の最大値が見つからなかった画素を表している. 鮮鋭度の最大値が見つからないとは、いずれのリフォーカス画像でもテクスチャが不鮮明な画素、もしくはテクスチャがない物体表面である可能性が高いことを示している. このような画素では特徴量の精度も低くなる為、後述の特徴点抽出処理において特徴点から削除する. 特徴点の削除については節で更に詳しく説明する.

4. 特徴点と特徴量

ライトフィールドを用いた特徴点候補と特徴量の算出について述べる. 本手法では焦点距離が異なる画像毎すべてに SIFT [1] を適用し特徴点と各特徴点の 128 次元特徴量を計算する. 図 5 に焦点距離が異なる 3D 画像に SIFT を適応した結果を示す. 上段は SIFT による従来手法, 下段は提案手法により削除した後の特徴点抽出の結果を示している. 左右の画像は焦点距離の異なる画像を示しており、右に行くほど焦点距離が遠い画像になっている. また、赤い円の中心が抽出された特徴点を、円の大きさがスケールを表している. 上段の従来手法を見ると、ピントが合っていないボケた画像や背景のようにテクスチャが少ない領域においても特徴点が多く抽出されている. 一方で 3. 節で述べた焦点距離が算出できない画素をテクスチャが少なく特徴点として信頼度が低いとし削除する本手法ではボケた画像やテクスチャが少なく背景部分の特徴点を削除することができている. 次に図 6 にリフォーカス画像毎で抽出された特徴点の数を示す. 横軸は図 5 と同様の並びで画像番号を示す. このグラフから本提案手法と従来の SIFT 特徴の数は共にピントが合わない画像ほど少なくなることがわかる. しかし、提案手法の方が特徴点の減少の傾向がより強く現れた. これは提案手法の方が鮮明な画像の特徴点を残しつつ、信頼度の低い特徴点を削除できていることを示している.

最終的に、3 次元画像で得られた特徴点からピントが最も合った焦点距離 α_D の特徴点のみを有効として統合することで 2 次

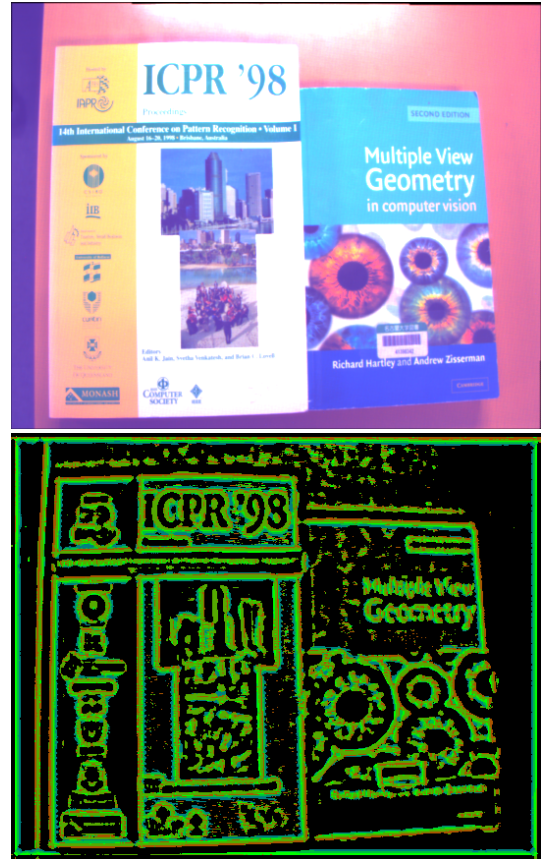


図 4 焦点距離画像 (上段: 元画像)

Fig. 4 Focal length estimation (upper: original image)

元座標の特徴点を得ることができる. 図 7 に本手法で得られた特徴点と従来手法として SIFT 特徴点抽出の結果を示す. 従来手法の入力画像は全く同じ条件の通常カメラ画像を使う必要があるが、全く同じ条件で撮像することはできない. そこで本論文ではマイクロレンズ中心の部分開口画像 $L(x, y, 0, 0)$ を通常のピンホールカメラとし、SIFT 特徴量を算出した. 最終的な特徴点の数は従来の SIFT 特徴量 1,285 個に対し、提案手法は 1,594 個となった. 焦点距離情報を活用しテクスチャの少ない特徴点を削除しつつ、3 次元画像から特徴点を抽出することで従来手法より多くの特徴点を抽出することができた. また、図 7 から、狙い通り特徴点としては信頼度が低いと考えられるテクスチャの少ない画素や背景のエッジ部分の特徴点を多く削除できていることが見て取れる.

5. 特徴点の対応付け

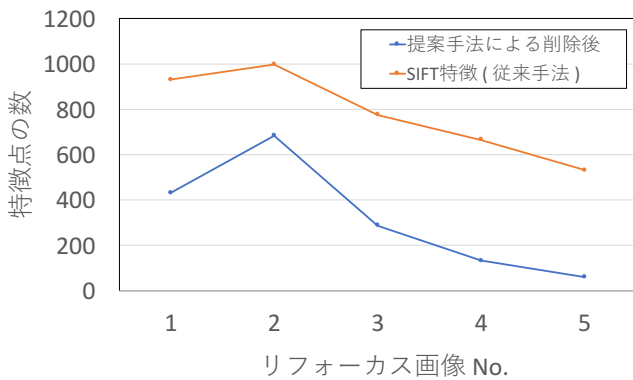
特徴点の対応付け処理では、L-2 ノルム

$$d_2 = \sqrt{\sum_{i=0}^{128} (\nu_i^{I_1\alpha} - \nu_i^{I_2\alpha'})^2} \quad (6)$$

が最小となる特徴点の組を対応付けた. 本提案手法で得られた対応付け結果を図 8 に示す. 比較のために提案手法だけでなく、従来手法で得られた SIFT 特徴量の対応付けと 3 次元画像から得られた SIFT 特徴量を削除せず全ての特徴点をマッチングする手法の 3 手法の結果を合わせて示した. ここでも従来手



図 5 リフォーカス画像毎の SIFT 特徴点 (上段 従来手法, 下段 提案手法)
 Fig.5 SIFT keypoints in each refocus image (upper conventional method, lower our proposed method)



画像 No は図 5 の画像と対応している, 左から 1, ..., 5 となる

図 6 リフォーカス画像毎の特徴点の数

Fig. 6 The number of keypoints in each refocus image

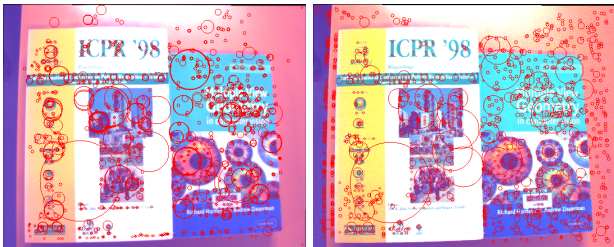


図 7 提案手法により得られた特徴量 (左) と SIFT 特徴量 (右)
 Fig.7 Keypoints by our proposed method (left) and SIFT keypoints (right)

法の入力画像はマイクロレンズ中心の部分開口画像 $L(x, y, 0, 0)$ とした. また全対応点を表示すると数が多すぎるため, L-2 ノルムが 0.03 以下の対応点のみを図示する. 線の色は特徴点对応付けの精度を表し, 赤線は 5 画素以下で対応付けできた特徴点, 黒線は 10 画素以上離れた対応点, 青線は 5-10 画素以内で対応付けできた特徴点を示している. 図 8 を見ると本提案手法が最も対応付けを表す線の角度と長さのばらつきが少ないことがわかる. また総当たりのマッチング結果は L-2 ノルムを 0.03 以下と絞り込んだにも関わらず多数のマッチングが得られているが, 線のばらつきは大きく, 本提案手法が信頼度が低い特徴点を削

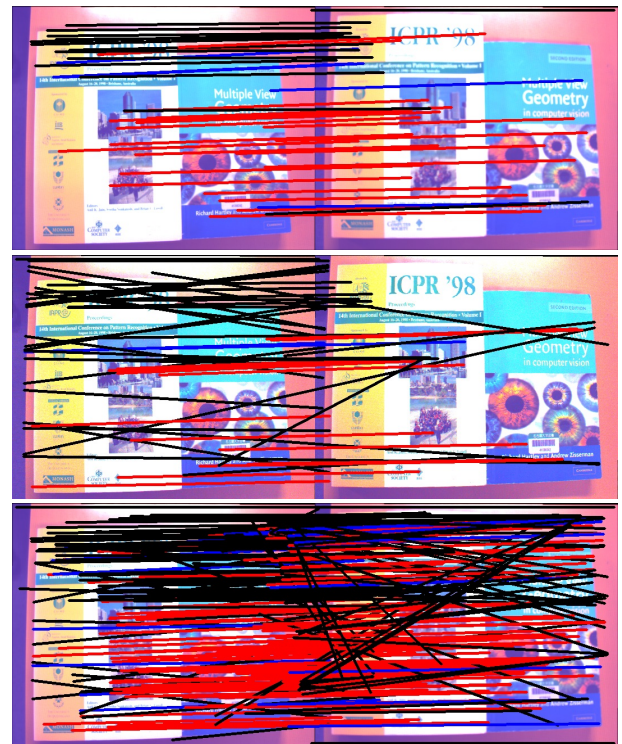


図 8 本提案手法 (上), 従来手法 (中) と総当たり手法 (下) の特徴量マッチング結果 (L-2 ノルム 0.03 以下のみ) 赤線 5 画素以下で対応付けできた特徴点, 黒線 10 画素以上離れた対応点, 青線 5-10 画素以内の対応点

Fig. 8 Keypoint matching result of our proposed method(upper), All matching method(middle) and SIFT(lower) (only keypoints with less than 0.03 L-2 norm) red line: lower than 5 pixels match, black line: more than 10 pixels match, blue line: 5-10 pixels match

除して効率良くマッチング精度を向上できていることが見て取れる.

6. 評価結果

ここでは最終的な評価として特徴点对応付けの精度について検討を行なう. 10 画素以内で対応付けできた特徴点を正対応点と定義し, 対応点数と正対応点数, 誤対応点数, Precision (=

表 1 提案手法と従来手法のマッチング数 (正対応の定義: 10 画素以内)

Table 1 The number of matches of our proposed method and conventional method (Correct match: within 10 pixels)

	all matches	correct matches	wrong matches	precision
SIFT with light field (our proposed)	630	309	321	0.49
SIFT (conventional method)	524	212	312	0.40
All matching SIFT with light field	2,459	979	1,480	0.40

表 2 提案手法と従来手法のマッチング数 (正対応の定義: 5 画素以下)

Table 2 The number of matches of our proposed method and conventional method (Correct match: less than 5 pixels)

	all matches	correct matches	wrong matches	precision
SIFT with light field (our proposed)	630	283	347	0.45
SIFT (conventional method)	524	183	341	0.35
All matching SIFT with light field	2,459	874	1,585	0.36

正対応点 / 全対応点数) を集計した結果を表 1 に示す. 更に厳しい条件として, 5 画素以下で対応付けできた特徴点を正対応点と定義した場合の結果を表 2 に示す. 表 1, 表 2 から本提案手法により対応点数, 正対応点数, Precision の項目で性能向上することが確認できた. また総当たり対応付け手法は対応点数は増えたが, 対応付けの精度は従来手法と変わらない結果になった. これら結果から本提案手法が対応点数, 対応付け精度を共に向上させる効果があることが示された.

7. ま と め

本論文でライトフィールド情報を用いた特徴点の抽出と対応付けの手法を提案した. 各画素の鮮鋭度から最もピントが合った焦点距離を推定し, 得られた焦点距離を用いて特徴点の絞り込み, 対応付けを行なう. 本提案手法は従来手法と比べて正しい対応点数の増加と対応付け精度を向上させることが確認できた. 今後は検討結果をもとにライトフィールド情報を更に活用し, より精度の高い特徴量算出の手法やマッチング手法を検討する. 加えてカメラモーションの推定等のアプリケーションへの応用も検討したい.

謝 辞

本研究を遂行するにあたり, 多くの皆さまにご指導及びご協力いただきましたことに心より感謝申し上げます. 本研究の一部は, 科学研究費補助金の助成を受けたものです.

文 献

- [1] D. Lowe, "Distinctive image features from scaleinvariant keypoints," *International Journal of Computer Vision (IJCV)*, 60(2), pp. 91-110, 2004.
- [2] G. Csurka, C.R. Dance, L. Fan, and C. Bray, "Visual categorization with bags of keypoints," *Proc. of the 8th European Conference on Computer Vision (ECCV)*, pp. 1-22, 2004.
- [3] Raúl Mur-Artal, J. M. M. Montiel, and Juan D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, Volume 31, Issue 5, Oct. 2015
- [4] Y. Ke, R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *Proc. of IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), pp. 511-517, 2004.

- [5] W. Cheung and G. Hamarneh, "N-dimensional scale invariant feature transform for matching medical images," *Proc. of IEEE International Symposium on Biomedical Imaging (ISBI)*, pp.720-723, 2007.
- [6] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Stanford University Computer Science Tech Report CSTR 2005-02*, April 2005.
- [7] R. Ng, "Digital light field photography," *Ph.D thesis, Stanford University*, July 2006.
- [8] 蚊野浩, "ライトフィールドカメラ Lytro の動作原理とアルゴリズム," 第 127 回微小光学研究会, Vol.31, No.1, pp.17-22, 2013.
- [9] <http://www.lytro.com/>
- [10] Michael W. Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi, "Depth from Combining Defocus and Correspondence Using Light-Field Cameras," *Proc. of the 14th International Conference on Computer Vision*, Pages 673-680, December 01 - 08, 2013