

EPIの直線構造に基づいた移動平均フィルタによる ノイズ除去を組み込んだロバストな視差推定手法

方便 剛^{1,a)} 高橋 桂太^{1,b)} 藤井 俊彰^{1,c)}

概要：光線空間（密な多視点画像）から得られる epipolar plane image (EPI) の直線構造を分析することで、高精度に視差を推定する手法 (EPI analysis) がある。この手法の課題は、light field camera など得た実写データの様なノイズを含んだ多視点画像を入力とした場合、EPI の直線構造が損なわれるため、精度が下がることである。そこで我々は、ノイズに対してロバストに視差を推定するために、EPI の直線構造を保持するノイズ除去手法を視差推定の過程に組み込むことにした。EPI の直線は多視点画像中のある物点に対応する画素で構成されているため、理想的には直線上に同じ画素値の画素が並んでいる。各直線の向きがわからないため、画素群が並ぶと想定される直線の向きを順番に変えつつデノイズすることで、EPI の直線構造を壊さずに、高精度なノイズ除去を可能にした。

Robust Disparity Estimation integrating Noise Removal according to Moving Average Filter preserving Line Structure of EPI

GOU HOUBEN^{1,a)} KEITA TAKAHASHI^{1,b)} TOSHIKI FUJII^{1,c)}

1. はじめに

2次元画像からの視差推定は、長い間研究者の興味を引き付けてきた。視差推定とは、画像上の被写体の奥行きを推定することであり、推定した視差から被写体空間の3次元再構成や自由視点映像の生成が可能になる。多視点画像から視差を推定する手法には、multi-view stereo (MVS) [1], [2], [3], [4], [5] や EPI analysis [6], [7], [8], [9], [10], [11], [12] がある。これらの手法の視差推定精度は同等であるが、MVS は計算コストが非常に高い。一方、EPI analysis は多視点画像が非常に密なときに、高速かつ高精度に視差を推定する。

近年の画像取得技術の発展 [13], [14], [15], [16], [17], [18] により、容易に密な多視点画像が取得できる light field camera [19] が開発された。そこで、light field camera から

取得した密な多視点画像に EPI analysis を適用して視差を求めることを考える。しかし、light field camera で得た実写データはノイズを含むため、これを入力とすると EPI の直線構造が損なわれ、視差推定精度が下がるという課題がある。そこで我々は、ノイズに対してロバストな視差推定を行うために、EPI の直線構造に基づいたノイズ除去手法を提案する。

2. EPI Analysis

図 1 左に示すような水平及び垂直方向に等間隔に配置したカメラにて撮影した多視点画像を考える。水平一列の画像群に着目し、それぞれの画像から同じ y 座標の画素列を抽出する。それらの画素列を積み上げると、図 1 右のような、横軸が画素の位置座標、縦軸が視点座標で構成される epipolar plane image (EPI) が得られる。EPI は多数の直線で構成されており、各直線は物点の軌跡である。また、各直線の傾きは各物点の視差に対応している。したがって、この直線の傾きを求めることが視差を推定することと等価となる。この概念に基づいて、Wanner と Goldluecke

¹ 名古屋大学 大学院工学研究科
Graduate School of Engineering, Nagoya University

a) houben@fujii.nuee.nagoya-u.ac.jp

b) keita.takahashi@nagoya-u.jp

c) fujii@nuee.nagoya-u.ac.jp

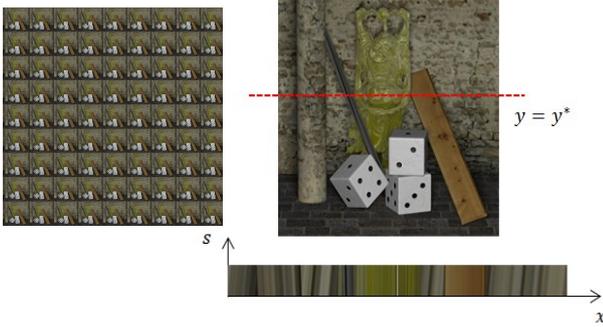


図 1: 多視点画像 (左) と EPI (右)

Fig. 1 Multi-view images (left) and EPI (right)

は構造テンソル分析を用いた視差推定手法を開発した [20]. EPI $l(x, s)$ 上の構造テンソルは以下のように定義される.

$$J(x, s) = \begin{bmatrix} G * (l_x l_x) & G * (l_x l_s) \\ G * (l_x l_s) & G * (l_s l_s) \end{bmatrix} \quad (1)$$

ここで, $G*$ はガウスフィルタカーネルを用いた畳み込みであり, l_x と l_s はそれぞれ EPI 上の x 方向と s 方向の勾配を表す. 行列 J を勾配ベクトルの共分散行列とみなし, J の固有ベクトルを求めることによって, 直線の傾き $\theta(x, s)$ とその信頼度 $c(x, s)$ が得られる.

$$\theta(x, s) = \frac{1}{2} \arctan \left(\frac{2J_{12}(x, s)}{J_{11}(x, s) - J_{22}(x, s)} \right) \quad (2)$$

$$c(x, s) = \frac{\sqrt{(J_{11}(x, s) - J_{22}(x, s))^2 + 4J_{12}^2(x, s)}}{J_{11}(x, s) + J_{22}(x, s)} \quad (3)$$

ここで, $c(x, s)$ が大きいほど, 推定した傾き $\theta(x, s)$ が信頼できる. また, 視差 $d(x, s)$ は以下のようにして得られる.

$$d(x, s) = \tan \theta(x, s) \quad (4)$$

EPI analysis を水平方向と垂直方向の 2 方向に対して実行し, それらの結果を画素単位で信頼度に基づいて統合することにより, ある視点 (s^*, t^*) の視差マップを取得する.

$$d^{s^* t^*}(x, y) = \begin{cases} d^{t^* y}(x, s^*) & c^{t^* y}(x, s^*) \geq c^{s^* x}(y, t^*) \\ d^{s^* x}(y, t^*) & \text{otherwise} \end{cases} \quad (5)$$

3. 提案手法

提案手法の目的は, ノイズのある多視点画像から正確な視差推定を行うことである. ノイズを含む場合, EPI の直線構造が不明慮になるため, ノイズを除去する必要がある.

EPI 上では, 同じ物点に対応する画素が直線状に並んでいる. そのため, EPI 上にて, その直線状に並んだ画素群に対して平均値をとると, 最も高精度なノイズ除去ができると考えられる. ある画素に注目したとき, その物点に対応する直線に沿おうとすると, その直線の傾き, すなわち

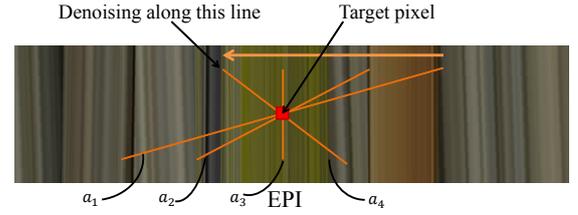


図 2: 提案手法の図解

Fig. 2 Illustration of proposed method

注目画素の持つ視差が必要である. すなわち, ノイズの除去と視差推定は表裏一体の関係となっている.

そこで我々は, 図 2 のように, 平均値をとる際に沿う直線の傾き a_n を順番に変化させることにした. EPI 上の全画素に対し, a_n に沿ったデノイズングを行う. デノイズング後の EPI にて EPI analysis を行い, 視差と信頼度を求める. y, t を固定した EPI において, 傾き a_n に沿ったデノイズング後に求めた視差を $d^{t^* y^*}(x, s; a_n)$, 信頼度を $c^{t^* y^*}(x, s; a_n)$ とする. ここまでの処理を a_n を変化させながら繰り返し, 各画素において, 最終的に最も信頼度が高いときの視差を採用する.

$$a(x, s) = \arg \max_{a_n} c^{t^* y^*}(x, s; a_n) \quad (6)$$

$$d^{t^* y^*}(x, s) = d^{t^* y^*}(x, s; a(x, s)) \quad (7)$$

4. 実験

実験では, 以下に記す 5 通りのフローにて視差推定精度を測定した.

- (1) EPI analysis from noise-free input: ノイズを乗せていない多視点画像から EPI analysis にて視差推定を行う. このフローが EPI analysis の性能の上限を示す.
- (2) EPI analysis from noisy input: ノイズの乗った多視点画像から EPI analysis にて視差推定を行う.
- (3) NLF on spatial domain + EPI analysis: 画像それぞれに画像ドメイン上で non-local means filter (NLF) をかけ, ノイズを除去した多視点画像から EPI analysis により視差推定を行う. NLF のカーネルサイズは 3×3 , 探索範囲は 5×5 とした.
- (4) NLF on EPI domain + EPI analysis: EPI それぞれに EPI ドメイン上で NLF をかけ, ノイズを除去した EPI から視差推定を行う. NLF のカーネルサイズは 3×3 , 探索範囲は 5×5 とした.
- (5) Proposed method: 3 章にて述べた手法.

フロー 3, 4 では, ノイズ除去の精度が良いとされている NLF を用いた.

4.1 CG シーケンス

ノイズに対するロバスト性を定量的に調べるため, 正解

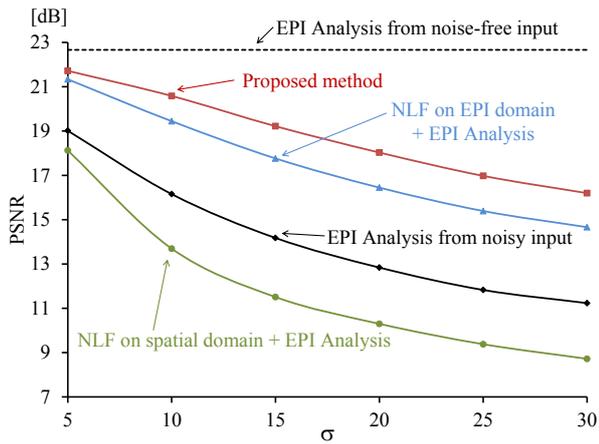


図 3: ノイズ強度と PSNR の関係
Fig. 3 Noise intensity and PSNR

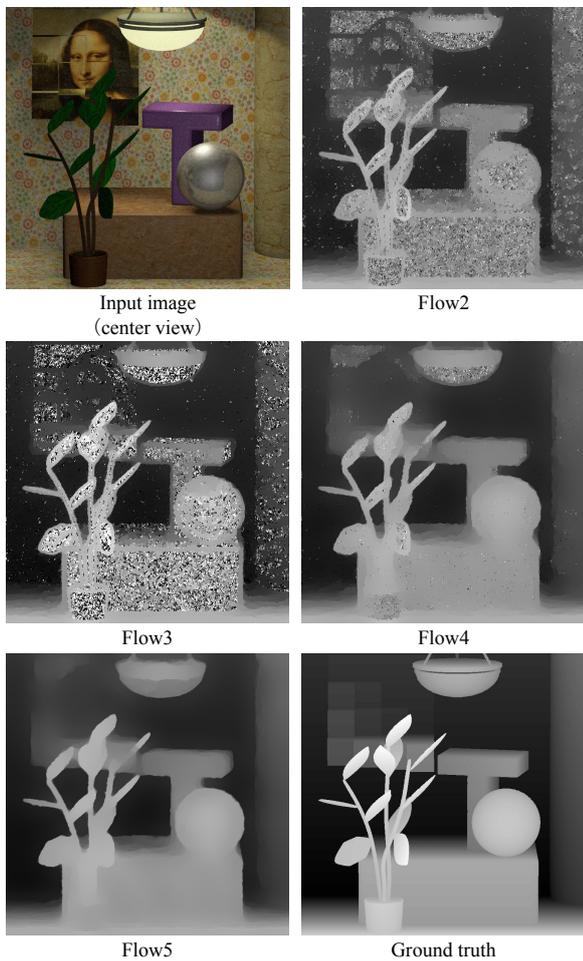


図 4: CG 入力画像と視差マップ
Fig. 4 CG input image and depth maps

視差マップを持った CG の多視点画像に人工的にノイズを加え、各フローを適用した。ここでは、 9×9 視点、 768×768 画素の CG シーケンス “monasRoom” [21] に MATLAB にて平均 0、標準偏差 σ のガウシアンノイズを加えたものを用いた。この標準偏差 σ を変化させながら、先に記した 5 通りのフローにて視差推定精度を測定した。フロー 5 にお

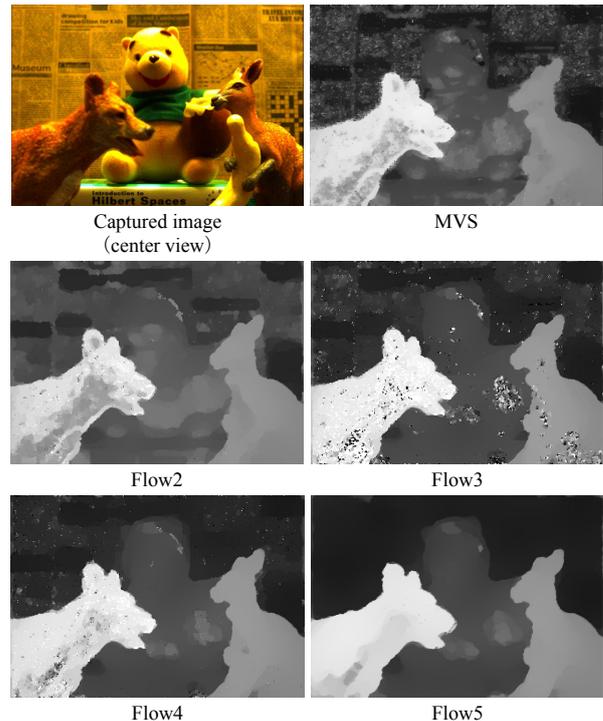


図 5: 実写入力画像と視差マップ
Fig. 5 Captured input image and depth maps

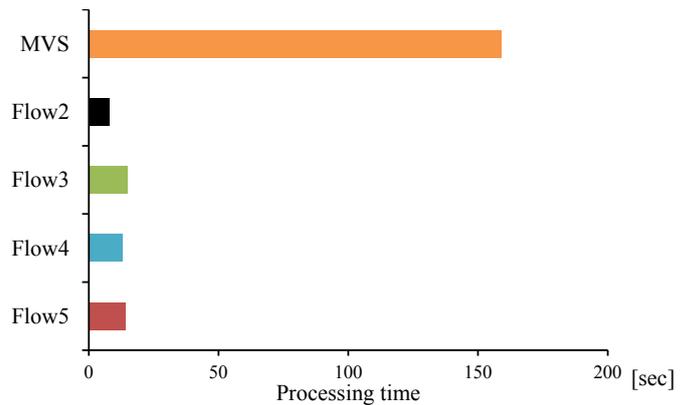


図 6: ノイズ除去と視差推定にかかる処理時間
Fig. 6 Processing time of denoising and disparity estimation

いて、 a_n は視差にして -2.0 – 2.0 画素の範囲を 0.01 刻みで変化させた。

図 3 にノイズ強度と視差推定精度の関係を示す。精度は推定された視差マップと正解視差マップとの PSNR 値とした。提案手法がノイズに対して最もロバストに視差を推定できていることがわかる。また、図 4 には $\sigma = 10$ のノイズを乗せた多視点画像から各フローにて求めた視差マップを示す。視覚的にも提案手法の視差推定精度が高いことがわかる。

4.2 実写画像

本研究の最終目的である、light field camera で撮影した実写の多視点画像からの正確な視差推定を試みた。図 5

には, Lytro Illum で撮影した 9×9 視点の実写データに MVS, フロー 2, 3, 4, 5 をそれぞれ適用して得た視差マップを示す. フロー 5 において, a_n は視差にして -2.0 – 2.0 画素の範囲を 1.0 刻みで変化させた. 図 6 にはそれぞれのフローに要した処理時間を示す. 使用した CPU のプロセッサは Intel Xeon (R), OS は ubuntu 14.04 LTS, メモリは 16GB である. また, GPU は GeForce GTX 660 Ti を使用した. 提案手法が, MVS と比較して高速であるという EPI analysis の特徴を保持しつつ, 視差推定精度を大きく向上させていることがわかる.

5. むすび

Light field camera で撮影した実写の多視点画像から正確に視差を推定するために, EPI の直線構造に基づいたデノイズングを EPI analysis に組み込んだ手法を提案した. 4.2 章に示すように, 提案手法は処理時間の増加を抑えつつ, ノイズを含んだ多視点画像からの高精度な視差推定を可能とした. 将来的には, 提案手法に Sheard EPI analysis と Multi-scale EPI analysis[23], [24] とを組み合わせ, EPI analysis の弱点である低テクスチャ領域や大きな視差範囲を含んだ実写の多視点画像からも高精度に視差が推定できるようにしたい.

参考文献

- [1] M. Okutomi and T. Kaneda: “A Multiple-Baseline Stereo,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 4, pp. 353–363, 1993.
- [2] S. Kang, R. Szeliski and J. Chai: “Handling Occlusions in Dense Multi-View Stereo,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 103–110, 2001.
- [3] C. Strecha, R. Fransens and L. V. Gool: “Combined Depth and Outlier Estimation in Multi-View Stereo,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2394–2401, 2006.
- [4] S. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski: “A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 519–528, 2006.
- [5] N. Campbell, G. Vogiatzis, C. Hernandez and R. Cipolla: “Using Multiple Hypotheses to Improve Depth-Maps for Multi-View Stereo,” *Proc. European Conference on Computer Vision*, pp. 766–779, 2008.
- [6] R. Bolles, H. Baker and D. Marimont: “Epipolar-plane image analysis: An approach to determining structure from motion,” *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.
- [7] A. Criminisi, S. Kang, R. Szeliski and P. Anandan: “Extracting layers and analyzing their specular properties using epipolar-plane-image analysis,” *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 51–85, 2005.
- [8] J. Berent and P. Dragotti: “Segmentation of epipolar-plane image volumes with occlusion and disocclusion competition,” *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 182–185, 2006.
- [9] S. Wanner and B. Goldluecke: “Globally consistent depth labeling of 4d light fields,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 41–48, 2012.
- [10] M. Diebold and B. Goldluecke: “Epipolar plane image refocusing for improved depth estimation and occlusion handling,” *Annual Workshop on Vision, Modeling and Visualization*, pp. 145–152, 2013.
- [11] J. Li, M. Lu and Z. Li: “Continuous depth map reconstruction from light fields,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3257–3265, 2015.
- [12] I. Tosic and K. Berkner: “Light field scale-depth space transform for dense depth estimation,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 435–442, 2014.
- [13] C. Zhang and T. Chen: “A self-reconfigurable camera array,” *Proc. ACM SIGGRAPH 2004 Sketches*, p. 151, 2004.
- [14] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz and P. Hanrahan: “Light field photography with a hand-held plenoptic camera,” *Stanford Tech Report*, vol. 2, pp. 1–11, 2005.
- [15] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz and M. Levoy: “High performance imaging using large camera arrays,” *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, 2005.
- [16] Y. Taguchi, K. Takahashi and T. Naemura: “Real-time all-in-focus video-based rendering using a network camera array,” *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pp. 241–244, 2008.
- [17] K. Marwah, G. Wetzstein, Y. Bando and R. Raskar: “Compressive light field photography using overcomplete dictionaries and optimized projections,” *ACM Trans. Graph.*, vol. 32, no. 4, 2013.
- [18] H. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. Tai and I. Kweon: “Accurate depth map estimation from a lenslet light field camera,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1547–1555, 2015.
- [19] Lytro: <http://www.lytro.com/>.
- [20] S. Wanner and B. Goldluecke: “Variational light field analysis for disparity estimation and super-resolution,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, 2014.
- [21] <http://klimt.iwr.uni-heidelberg.de/HCI/Research/LightField/vmv-benchmark-2013.php>.
- [22] B. Goldluecke: “Cocolib/light field suite.” <http://sourceforge.net/p/cocolib/home/Home/>, 2016.
- [23] T. Suzuki, K. Takahashi and T. Fujii: “Disparity estimation from light fields using sheared epi analysis,” *Proc. IEEE International Conference on Image Processing*, pp. 1444–1448, 2016.
- [24] 鈴木貴博, 高橋桂太, 藤井俊彰: “多重解像度 EPI 分析法による高精度デプス推定”, *IEICE Technical Report*, vol. 116, no. 36, pp.5–10, 2016.