

解像度方向への変分ベイズ情報伝搬による動画からの 3 次元復元 3D Recovery from Image Sequences based on Variational Bayesian Information Propagation in Resolution Direction

油山 耀也[†] 田川 憲男[†]
Teruya Aburayama Norio Tagawa

1. はじめに

動画からの 3 次元情報復元は、コンピュータビジョン研究の開始当初から検討されてきた基礎的な課題である。本稿では、連続する 2 フレームに対する密なデプス復元処理に焦点を当てる。デプスの多重解像度階層モデルを用いることでデプス自由度を抑えながら、エイリアス問題を避け、デプス不連続に対処可能な方法を提案する。

ここでのエイリアス問題とは、画像上での速度場（オプティカルフロー：OF）が輝度パターンを一次近似するには大きすぎて、偽の小さな動きが検出されてデプスが大きく推定されてしまうものである。そこで、画像を多重解像度分解し、低解像度から順次、デプスを復元して次の高解像度処理に伝搬させることで、エイリアスを回避する。各解像度において、デプス一定と仮定する領域のサイズを解像度に応じて小さくしながら、階層間のデプス遷移を確率モデルで表現することで、デプス自由度が削減され[1]、安定な復元が可能になる。デプス不連続にも自然に対応できる。

2. 方法

2.1 投影モデルとカメラ運動

透視投影モデルを用い、レンズ中心を原点とする座標系をとる。Z = 1 に画像面をおき、3 次元空間中の点(X, Y, Z)の画像上の投影点を(x, y)とする。カメラの並進ベクトルと回転ベクトルをそれぞれ $\mathbf{u} = (u_x, u_y, u_z)$, $\mathbf{r} = (r_x, r_y, r_z)$ とする。画像輝度値とその時空間微分をそれぞれ、 f, f_t, f_x, f_y とする。このとき、運動前後の輝度値不変性を一次近似する勾配方程式

$$\mathbf{f}_t = -f_x v_x - f_y v_y \quad (1)$$

が画素毎に成り立つ。 $\mathbf{v} = (v_x, v_y)$ は OF であり、撮影対象が剛体であれば、shallowness $d = 1/Z$ を使って次式のように表現できる。

$$v_x = xy r_x - (1 + x^2) r_y + y r_z - (u_x - x u_z) d \quad (2)$$

$$v_y = (1 + y^2) r_x - xy r_y - x r_z - (u_y - y u_z) d \quad (3)$$

これらを式(1)に代入することで、剛体勾配方程式が定義できる。

$$\mathbf{f}_t = -\mathbf{f}_r^T \mathbf{r} - \mathbf{f}_u^T \mathbf{u} d \quad (4)$$

2.2 階層的デプスモデル

画像輝度値の空間波長に対して OF が大きい場合、式(1)、(4)の一次近似では偽の OF が検出される危険があり、デプス復元の誤差に繋がる。これを避けるために、空間波長の長い低解像度画像で得られた OF を用いてワーブ処理を行うことで、エイリアスの生じにくい f を計算して式(4)の左辺とすることができる。

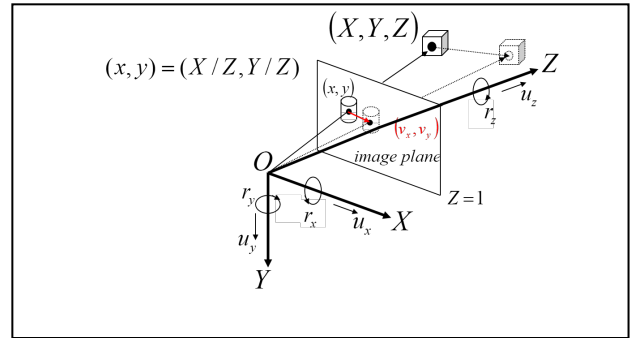


図 1 カメラモデル

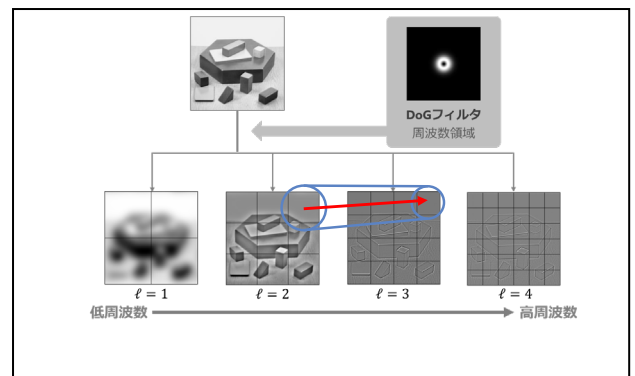


図 2 多重解像度復元処理のイメージ

一方、デプス不連続の復元を考慮し、本研究では局所的にデプスを一定と仮定する局所最適化の枠組みを採用する。デプス不連続を考慮する従来法では、ラインプロセスやセグメンテーションを用いる陽な不連続のモデリングが用いられており、処理が複雑になる。それに対し、低解像度ではデプス一定の領域を大きくし、解像度が高くなるにつれて小さくしていくことで、不連続の復元が自然に実現できる。すなわち、解像度が高まるにつれて不連続がデプス一定領域の境界として表現されやすくなる。解像度に応じてこの領域を狭めることは、冗長性の観点からも適切である。

上記の方法では、解像度間でデプス復元情報を伝搬させなければならない。階層的なデプスモデルとして、解像度間の遷移関数が必要となる。ここでは簡易なモデルとして、低解像度（下位層）のデプスの線形補間に正規性の揺らぎ $\epsilon_p^{(\ell+1)}$ が加わって上位層が生成される確率モデルを導入し、その揺らぎの分散が各層で独立であるとする。

$$d^{(\ell+1)} = U^{(\ell)} d^{(\ell)} + \epsilon_p^{(\ell+1)} \quad (5)$$

l は層番号を表し、大きくなるほど高解像度を意味する。 U は線形補間演算子である。本来はノンパラメトリックな自由度を持つデプスを、この階層的なモデルによってパラメ

トリックに扱うことで、実質的な自由度を下げられる。式(5)によって次層のデプスの事前確率を計算して復元に利用できるため、安定な結果が期待できる。図2には、DoG (Difference of Gaussians) フィルタによる解像度分解と階層的デプスモデルの連携の様子を示す。

2.3 変分ベイズ推論による情報伝搬アルゴリズム

各層におけるすべての確率変数の同時事後確率は、観測量 $f_t^{(l)}$ の条件付き確率と未知量 $d^{(l)}$, $m^{(l)}$ の事後分布の積で書ける。

$$\begin{aligned} p(\mathbf{d}^{(l)}, \mathbf{m}^{(l)}, \mathbf{f}_t^{(l)} | D^{(l-1)}, \boldsymbol{\theta}^{(l)}) \\ = p_f(\mathbf{f}_t^{(l)} | \mathbf{d}^{(l)}, \mathbf{m}^{(l)}, \boldsymbol{\theta}^{(l)}) p_{dm}(\mathbf{d}^{(l)}, \mathbf{m}^{(l)} | D^{(l-1)}, \boldsymbol{\theta}^{(l)}) \end{aligned} \quad (6)$$

ここで、 $\mathbf{m} = (\mathbf{u}, \mathbf{r})$ はカメラの運動パラメータであり、全層に共通のパラメータとなる。この事後確率をそのまま計算するのは困難であるため、変分ベイズ法を適用する。 $d^{(l)}$ と $m^{(l)}$ の相関を平均場近似によって除去して、それぞれの事後分布の積で表現する。

$$\begin{aligned} p_{dm}(\mathbf{d}^{(l)}, \mathbf{m}^{(l)} | D^{(l-1)}, \boldsymbol{\theta}^{(l)}) \\ \approx q_d(\mathbf{d}^{(l)} | D^{(l-1)}, \boldsymbol{\theta}^{(l)}) q_m(\mathbf{m}^{(l)} | D^{(l-1)}, \boldsymbol{\theta}^{(l)}) \end{aligned} \quad (7)$$

変分自由エネルギーの最大化原理 [2] により、 $\ln q_d(\mathbf{d}^{(l)} | D^{(l)}, \boldsymbol{\theta}^{(l)})$ と $\ln q_m(\mathbf{m}^{(l)} | D^{(l)}, \boldsymbol{\theta}^{(l)})$ を具体的に書き下すことができる。具体的な数式は省略するが、両式を微分してゼロとおいて得られる方程式は、それぞれの式がもう一方の変数を定数として含むため、両者の解を交互に更新し、収束解を求めればよい。 $f_t^{(l)}$ の観測雑音と $d^{(l)}$ の層間の遷移、並びに $\{d\}$ と \mathbf{m} を正規分布でモデル化することで、平均と分散を求めることに帰着される。なお、雑音の分散と階層的デプスモデルに含まれる揺らぎの分散が確定的な未知数として存在するため、それらも含めた計算はEMアルゴリズムによることとなる。

以上を各層で行い、 $\{d\}$ と \mathbf{m} の事後分布が次層の事前分布として伝搬される。なお、今回は第一層には無情報分布 (十分に大きな分散を持つ正規分布) を用いている。

3. 性能評価

3.1 評価条件

1枚目の画像とデプスマップ (図3) を設定し、 $\mathbf{u} = (0.2, -0.2, -0.1)$, $\mathbf{r} = (-0.01, 0.0, 0.01)$ と定めて2枚目の画像を生成した。2枚の画像には、最大輝度値の3%の一樣乱数雑音を加えた。DoG フィルタで2枚の画像をそれぞれ4枚に分解して処理した (図4)。層毎のデプス一定のブロックサイズは、32x32 (ブロック数 8x8), 16x16 (16x16), 8x8 (32x32), 8x8 (32x32) である。

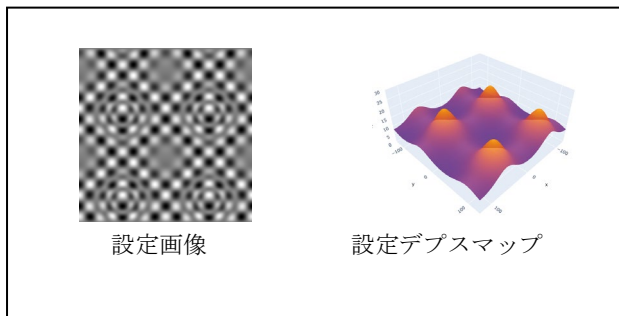


図3 評価用人工データ例

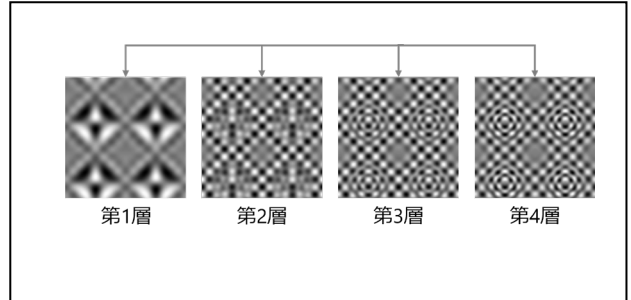


図4 多重解像度分解の結果

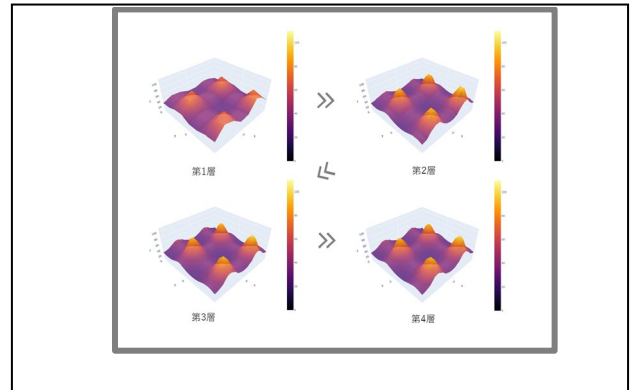


図5 デプス復元結果

3.2 評価結果

図5に、各層でのデプス復元結果を示す。RMSEは真値に対して12%程度 (1層目では20%) であった。共通パラメータである \mathbf{u} のRMSEは5%程度と非常に高精度であった。不連続性の強いデプスも、不連続を保持して同程度の精度で復元できることを確認した。デプスモデルを用いないMLEでは、概形すらも復元できず、本手法で画像ワーブを用いなかった場合は、デプスの低い周波数成分は復元できるものの、高周波成分はランダム雑音相当であった。

4. おわりに

デプスに階層的なモデルを仮定することで、低解像度画像からの復元結果を高解像度画像処理に伝えることが可能となり、そのことは画像ワーブによるエイリアス回避にも繋がることを確認した。その結果、デプス不連続に強い復元が実現できた。今後は、撮影画像に対する性能の評価を行う予定である。さらに、深層ネットワーク [3] による経験の学習と、本法のような物理モデルに基づく方法との統合は、今後の重要な課題である。数式で表現しきれない部分を経験で補うことが期待される。

参考文献

- [1] N. Tagawa, M. Yang: "On computing three-dimensional camera motion from optical flow detected in two consecutive frames," Proc. of Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 931-942. INSTICC (2023).
- [2] 中島伸一, "変分ベイズ学習", 講談社 (2016).
- [3] V. Arampatzakis, G. Pavlidis, N. Mitianoudis, N. Papamarkos: Monocular depth estimation: A thorough review. IEEE Trans..PAMI, Vol. 46, No.4, pp.2396-2414 (2024).