

回転運動を考慮した距離動画像の動き補償予測

Motion-Compensated Prediction of Range Image Sequence Considering Rotational Motion

福田光一
FUKUDA, Koichi

要旨

距離動画像のフレーム間の冗長性を低減する手法として、動き補償予測が一般に用いられている。フレーム間での動きの検出法としてはブロックマッチング法が広く利用されているが、検出しているのは剛体の並進運動である。本論文では、3次元空間における剛体の運動の拘束式をもとに推定した運動パラメータを利用することにより、回転運動を考慮した動き補償予測を行う方法を提案している。実験により提案手法の有効性を示している。

1 まえがき

複数のカメラ画像から任意の視点位置や視線方向の映像を生成する自由視点映像技術の開発が進められており、放送、教育、医療などの分野で応用が期待されている [1]。自由視点映像を生成する様々な手法が提案されているが、カメラ画像とともに奥行き情報を用いる Depth Image Based Rendering (DIBR) に関する研究が活発に行われている [2]。

奥行き情報はカメラ画像の各画素における被写体までの距離であり、ステレオマッチングなどにより複数の方向から撮影した画像をもとに推定される。最近では、ビデオレートで奥行き情報を取得するカメラも開発されている [3]。奥行き情報は距離に応じたグレースケール画像として扱うことができ、距離画像あるいはデプスマップ (depth map)、奥行き画像と呼ばれている [4]。図 1 に距離画像とカメラ画像の例を示している [5]。

自由視点画像を DIBR により生成するには複数のカメラ画像と距離画像を必要とするため、これらの伝送、記録にあたって圧縮符号化は必須の技術である。映像符号化の国際標準方式を検討する組織である MPEG (Moving Picture Expert Group) においても、奥行き情報の利用を含む自由視点テレビや 3 次元ビデオなどの符号化方式に関して活発に議論されている [6]。

これまでにカメラ画像と距離画像からなる映像データの符号化手法は多数提案されているが、その多くはフレーム内で隣接する画素間や時間方向での冗長性に注目した手法となっている。距離画像は奥行き情報からなる画像であることから、奥行き方向も含めた 3 次元の動き予測を行う手法 [7, 8] も提案されているが、並進運動のみを対象としており回転運動については考慮されていない。

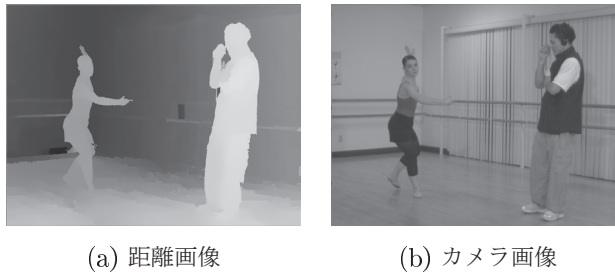


図1 距離画像とカメラ画像の例 ("Ballet Sequence" [5], 1024×768 pels/frame, 8 bits/pel)

そこで本論文では、距離動画中の物体の運動パラメータを推定し、回転運動も考慮した動き補償予測法を提案する。

2 回転運動を考慮した動き補償予測

3次元空間内で運動する剛体の運動パラメータの算出は、非線形な最適化問題として定式化されることが多いが、距離画像フレーム間で時間的な変化が少ない場合には動きの拘束式を導くことができ、線形連立方程式を解くことによって並進・回転運動のパラメータが得られる[9]。そこで提案法では、ブロックマッチング法により隣接する距離画像フレーム間で最も類似したブロックを探索し、マッチングした2つのブロック間で運動パラメータを推定する。最も類似したブロックと運動パラメータの推定値からブロックの奥行き値を予測する。

2.1 ブロックマッチング法による動き検出

時間的に連続して取得された距離画像では、隣接するフレーム間の冗長性が高くなっている[10]。この時間的な冗長性を低減することにより、距離動画を直接符号化する場合よりも少ない情報量での符号化が可能となることから、カメラ画像と同様にブロックマッチング法を利用した動き補償予測が行われている。

ブロックマッチング法による動き補償予測では、図2に示すように、対象となる距離画像フレーム（対象フレーム）を重ならない複数の正方ブロックに分割し、各ブロックについて直前の距離画像フレーム（参照フレーム）中から最も類似するブロックを検出する。本論文ではブロックの類似度の評価値として、以下の式により定義される差分絶対値和 (Sum of Absolute Difference, SAD) を用いる。

$$SAD(m_x, m_y) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} e_t(i, j) \quad (1)$$

$$e_t(i, j) = |f(x+i, y+j, t) - f(x+i+m_x, y+j+m_y, t-1)| \quad (2)$$

ここで、 $f(x, y, t)$ は時刻 t で取得された距離画像フレームの座標 (x, y) における奥行き値を表す。ブロックの大きさは $N \times N$ としている。 m_x, m_y はブロックの水平・垂直方向の移動量をそれぞれ表し、 $-(M-1) \leq m_x, m_y \leq (M-1)$ の範囲で変化させる。そして、 $SAD(m_x, m_y)$ が最小となるブロックを最も類似したブロックとみなし、そのときの移動量 (m_x, m_y) を動き

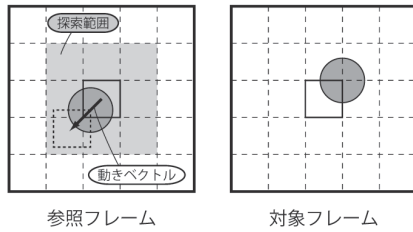


図2 ブロックマッチング法による動き検出

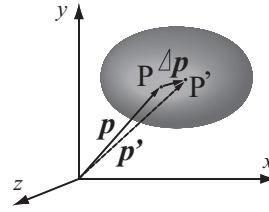


図3 3次元座標系での剛体の運動モデル

ベクトルとする。

マッチングしたブロックと動きベクトルを利用することで、対象フレームに対する予測フレームが作られる。しかし、検出された動きベクトルは水平・垂直方向の移動量であり、カメラ方向の移動量は求められていない。距離画像は奥行き値を記録していることから、 $SAD(m_x, m_y)$ が最小となる位置で参照ブロックをカメラ方向に移動し、奥行き方向の移動量を検出する3次元ブロックマッチング法 [7, 8] が提案されている。このような方法により3次元の並進運動パラメータを検出できるが、物体が移動する際にはある程度の回転運動も含まれていると考えられる。そこで、対象ブロックと探索範囲内で $SAD(m_x, m_y)$ を最小とする参照ブロックとから並進・回転運動パラメータを推定し、得られたパラメータを利用して予測ブロックを生成する。

2.2 並進・回転運動パラメータの推定

図3に示すように、 z 軸をカメラの方向と一致させた3次元座標系において、物体表面上の点 P の位置ベクトルを $\mathbf{p} = (x, y, z)^T$ とする。ここで、上付き添字 T は転置を表す。また、点 P の移動後の点を P' とし、点 P' の位置ベクトルを \mathbf{p}' 、変位ベクトルを $\Delta\mathbf{p}$ とする。

時刻 t における点 P の z 座標値を $z = f(x(t), y(t), t)$ とし、関数 f は時刻 t について微分可能とする。このとき、次式が導出される [9]。

$$\frac{dz}{dt} = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial f}{\partial t} \quad (3)$$

移動する物体は剛体と仮定していることから、その運動は並進と回転の合成として与えられる。並進速度ベクトルを \mathbf{v} 、原点を中心とする角速度ベクトルを $\boldsymbol{\omega}$ とすれば、物体表面上の点 P の速度を次式のように表すことができる。

$$\frac{d\mathbf{p}}{dt} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{p} \quad (4)$$

よって、式 (3) と式 (4) より次式が得られる。

$$(\mathbf{d} \times \mathbf{p}) \cdot \boldsymbol{\omega} - \mathbf{d} \cdot \mathbf{v} = \frac{\partial f}{\partial t} \quad (5)$$

ここで、 $\mathbf{d} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1 \right)^T$ である。ブロックマッチング法で検出されたブロックと対象ブロックの各点において式 (5) が成り立つと仮定し、各々のブロックで空間微分 $\partial f / \partial x, \partial f / \partial y$

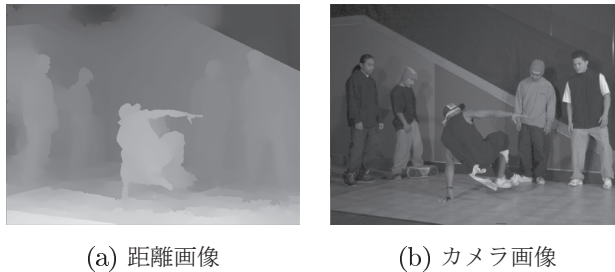


図4 "Breakdancing Sequence" [5] (1024×768 pels/frame, 8 bits/pel, 100 frames, 15 fps)

と時間微分 $\partial f / \partial t$ を算出することができれば、最小二乗法により運動パラメータ \mathbf{v} と $\boldsymbol{\omega}$ を推定することが可能である。運動パラメータの推定値と式 (4) よりフレーム間の変位ベクトル $\Delta \mathbf{p}$ が得られることから、ブロックマッチング法により検出されたブロックを $\Delta \mathbf{p}$ だけ変位させることにより予測ブロックが得られる。

3 実験

提案手法の有効性を検討するため、図1に示した"Ballet Sequence"[5]と図4に示す"Breakdancing Sequence"[5]を用いて実験を行った。いずれもフレームサイズ 1024×768 画素 (8 bits/pel), フレーム数 100 (15 フレーム/秒) である。各フレームは大きさ 16×16 画素 ($N=16$) のブロックに分割して処理を行った。ブロックマッチング法を用いた動き検出では、探索範囲 ± 16 画素 ($M=16$), 探索精度 1 画素とした。参照フレームにはオリジナルのフレームを用いている。また、ブロックの空間微分は Sobel オペレータを利用して算出し、時間微分は 3×3 近傍の平均値のフレーム間差分により求めている [11]。

図5に、各手法により得られた"Ballet Sequence" 第9距離画像フレームの動き補償予測結果を示している。3次元ブロックマッチング法では、奥行き方向の探索範囲を ± 16 とした。提案法を用いることで、他の方法に比べて良好な動き補償予測フレームが生成されており、オリジナル対象フレームとの平均二乗誤差 (Mean Square Error, MSE) も改善されている。

図6には動き補償予測フレームごとの MSE を示している。提案法を用いることで予測性能が向上していることが確認できる。

4 むすび

本論文では、回転運動を考慮した距離動画の動き補償予測手法を提案した。ブロックマッチング法により検出されたブロックをもとに、並進・回転運動パラメータを推定している。マッチングしたブロックと運動パラメータの推定値から、対象ブロックの奥行き値を予測した。本手法を用いて動き補償予測を行うことにより、ブロックマッチング法および3次元ブロックマッチング法を用いた場合に比べ、予測性能が改善されることを確認した。

今後は、復号された参照フレームを用いた場合の予測性能や、符号量の削減効果について検討する予定である。

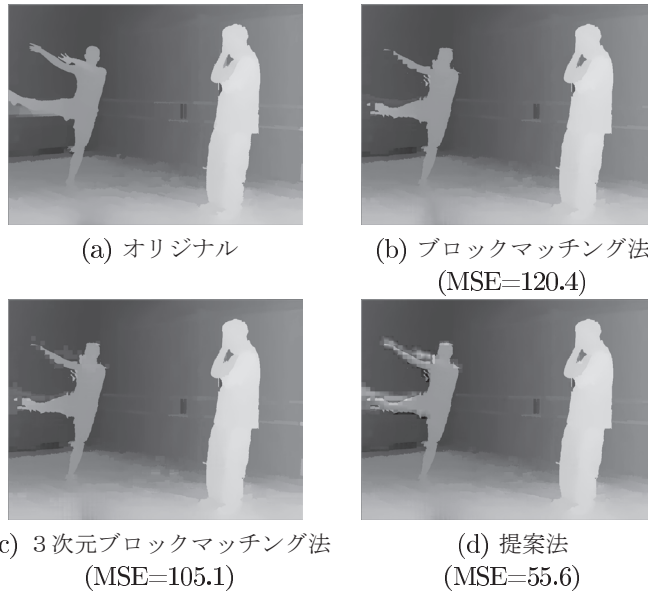
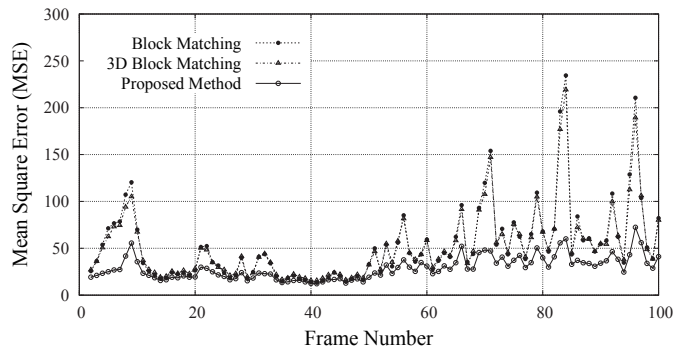
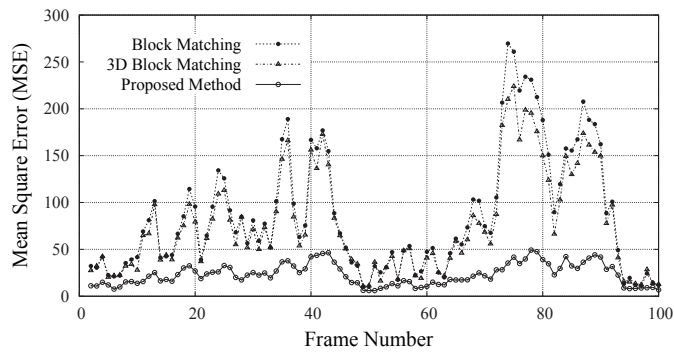


図5 動き補償予測画像の比較 ("Ballet Sequence" 第9フレーム, 1024×768 pels)



(a) "Ballet Sequence"



(b) "Breakdancing Sequence"

図6 フレームごとの動き補償予測誤差

参考文献

- [1] I.Ahmad : Multi-View Video: Get Ready for Next-Generation Television, IEEE Distributed Systems Online, Vol.8, No.3, art.no.0703-o3006. (2007)
- [2] P.Ndjiki-Nya, M.Koppel, D.Doshkov, H.Lakshman, P.Merkle, K.Muller, T.Wiegand : Depth Image-Based Rendering With Advanced Texture Synthesis for 3-D Video, IEEE Transactions on Multimedia, Vol.13, No.3, Jun., pp.453 – 465. (2011)
- [3] 川人 祥二 : 3次元レンジファインダの最新動向, 映像情報メディア学会誌, Vol.62, No.3, pp.313 – 316. (2008)
- [4] 井口征士, 佐藤宏介 : 三次元画像計測, 昭晃堂. (1990)
- [5] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski : High-quality video view interpolation using a layered representation, ACM SIGGRAPH and ACM Trans. on Graphics, Los Angeles, CA, Aug. pp. 600–608. (2004)
- [6] 木全英明 : 3次元映像符号化の国際標準化動向と研究開発 – 多視点映像と自由視点映像 –, 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM) , vol.2008, No.27(2008-CVIM-162) , pp.243 – 248. (2008)
- [7] Yu-Cheng Fan, Shu-Fen Wu, and Bing-Lian Lin : Three-Dimensional Depth Map Motion Estimation and Compensation for 3D Video Compression, IEEE Trans. Magnetics, Vol.47, No.3, Mar., pp.691 – 695. (2011)
- [8] B.Kamolrat, W.A.C.Fernando, M.Mrak, and A.Kondoz : 3D Motion Estimation for Depth Image Coding in 3D Video Coding, IEEE Trans. Consumer Electron., Vol.55, No.2, May, pp.824 – 830. (2009)
- [9] Berthold K. P. Horn : Rigid Body Motion from Range Image Sequences, CVGIP: Image Understanding, Vol.53, No.1, Jan., pp.1 – 13. (1991)
- [10] Han Oh and Yo-Sung Ho : H.264-based depth map sequence coding using motion information of corresponding texture video, Advances in Image and Video Technology, Lecture Notes in Computer Science, Volume 4319/2006, pp. 898–907, Springer. (2006)
- [11] 山本正信, ピエール・ボランジャー, アンジェロ・ベラルディン, マーク・リュウ, ジャック・ドメイ : 距離動画像を用いた非剛体運動パラメータの直接推定法, 情処学論, Vol.32, No.9, Sep., pp.1129 – 1141. (1991)