Vol.2016-CVIM-200 No.41
2016/1/22

Video Completion via Depth Estimation by Motion Analysis for Outdoor Omni-directional View
カメラの動き解析に基づいたシーンの奥行き推定による屋外全方位映像補完

CARLOS MORALES[†1]　小野 晋太郎[†1]　岡本 泰英[†1]
MENANDRO ROXAS[†1]　大石 岳史[†1]　池内 克史[†2]

概要: ビデオ映像の補完では、不要な領域を除去し、除去した領域を整合性が取れるように埋める必要がある。屋外映像の場合、フレーム間の時間・空間的距離が大きくなることから、多数のフレームに渡って広い領域を補完しなければならない。そのためより大きな範囲の画像と 3 次元情報が必要となり、通常のオプティカルフロー推定に基づく映像補完手法を適用することは難しい。そこで本論文では、全方位映像を対象映像とした映像補完手法を提案する。本手法では、まず、カメラの動きから各点の画像上の動きを推定してシーン全体の奥行きを推定する。そして奥行きマップを用いて、参照フレームから対象フレームにカラー情報を伝搬させることで補完を行う.

Keywords: 映像補完, 奥行き推定, 動き解析, 全方位動画

1. はじめに

映像補完はコンピュータビジョン技術のひとつであり，動画像列から不要な物体を削除し，その領域を前後および全体のフレーム画像間の視覚的に整合性を保ったまま自然に補完する技術のことである（図 1）. 視覚的な整合性とは，幾何学的，光学的，時間的という 3 要素からなる．幾何学的整合性を保つためには，背景領域上での欠落領域が空間的に不連続にならないようにする必要がある．光学的整合性においては欠落領域とそうでない領域間のテクスチャ，色，陰影の対応付けが必要となる．そして時間的整合性では欠落領域内部もしくは外部の画素の動きをフレーム間で調整することが必要となる.

欠落領域を埋め，かつ視覚的に整合性が保たれた動画像列を出力するために用いられる基本的なアプローチとしては，連続した画像間でその領域を追跡することで欠落領域を埋める方法である．しかしながらこうした基本的な手法では，大きな領域もしくは多数のフレームに渡って存在する欠落領域を補完することは難しい.

動画像の欠落領域を埋める一般的な手法は映像修復法（video inpainting）と映像補完法（video completion）の 2 種類に分類される.

映像修復法は小規模の欠落領域を埋めるために広く使われており，形状伝播を使った静止画修復の手法を用いて行われる．この手法は静止画修復法に対して時間という 1 次元を加えたものであることから，静止画修復法の拡張としても捉えることができる．この手法では一般的にフレーム間での平滑化と連続性を定式化した変動問題もしくは微分方程式を解くことによって欠落領域の色情報を復元する（[1], [2], [3], [4], [5]）. ただし，フレーム間で局所的に不連続な変化があった場合，不自然な修復結果（ゴースト）を生成してしまい，整合性が失われてしまう．この現象は欠落領域や，カメラの動き，光源変化が激しい場合には特に顕著となる．こうした問題を解決するため，Partwardhan らはカメラと背景は静止，前景は動的であるという条件の下で，映像修復法に対して優先度付きの空間補完を行う手法[6]を提案している．また Cheung らは静止背景にたいしては単純な背景移動と変則的な画像修復を施し，動きのある前景物体に対しては動的に補完を行うという複合的な手法を提案している[7].

これに対して映像補完法は，テクスチャの合成などを利用することで，より大きな欠落部を埋めることが可能な手法である．静止画の補完手法では，テクスチャを既知の領域から取り出し，事例ベースのテクスチャ合成を行うことで, 欠落領域に相当する新たなテクスチャを生成している．Drori らが提案するパッチベースの補完手法[8]では，既知の画像からテクスチャを多数サンプリングし，対象画像における欠落領域に隣接する領域に対してその中から最も整合性の高いものを選択して補完する方法を用いている．また Criminisi らはエッジ強度や画素の信頼性などの既知の幾何的情報をもとに割り当てた優先度を利用したテクスチャ合成手法を提案している[9].

一方，静止画補完法を用いても大きな欠落部の正確な幾何的情報を求めることは難しく，動画像列にそのまま適用することは困難である．この問題に対して様々なアプローチが提案されており，例えば Jia らはテクスチャ合成処理を主眼においた手法を提案している[10]．この手法ではまず運動する対象物体を追跡することにより，欠落領域の境界として尤もらしい画素部分を特定する．そして色の類似性を利用して欠落領域に尤もらしいテクスチャ片を選択して合成することで補完を行う．Wexler らは時空間でグローバル最適化と 3 次元パッチを利用したノンパラメトリックなサンプリングを行うことで欠落領域を補完する手法を提案している[11]．白鳥らは動き場の変化を利用した映像補完法を提案している[12]．この手法では，まず映像のいくつかの部分から局所的な動きの時空間パッチをサンプリングし，欠落領域に対して求められた動き場を使って，欠落領域周辺から色の情報を伝播させることで補完を行っている．Roxas らは，大規模な欠落領域の補完にも対応した，時空間最適化による補完手法を提案している[13]．この手法ではオプティカルフローの最小化と色伝播を同時に反復最適化することにより補完結果を求めている.

†1 東京大学
†2 Microsoft Research Asia

Vol.2016-CVIM-200 No.41
2016/1/22

本論文においては，映像中の欠落領域は追跡できるものと仮定し，静的な背景シーン中における時空間的に大規模な欠落領域を補完する問題の解決に取り組むものとする．特に，全方位カメラを使って撮影された動画像列を対象として，撮影された全方位画像列内の各点の動きを解析することにより映像補完を行っている．本手法では，まずカメラの動きから，複数全方位画像フレーム中にある各画素の動きを定式化する．フレーム画像の深度情報は，カメラの動きとそれに基づいて推定された画素の動きから求めることができる．求められた深度マップはフレーム画像内のエッジを失わないように平滑化処理を施す．このようにして求めた深度情報を用いることで，キーフレームの色情報を各フレームに伝播させていき，最終的に欠落領域の補完を行う．

本論文は以下のように 5 つの章で構成される．2 章では提案手法のうち，カメラの動きをもとにした各点の画像中の動きの定式化について説明する．3 章では提案手法の映像補完手順について説明する．そして 4 章において提案手法の実験と評価，5 章にて考察と結論を述べる．

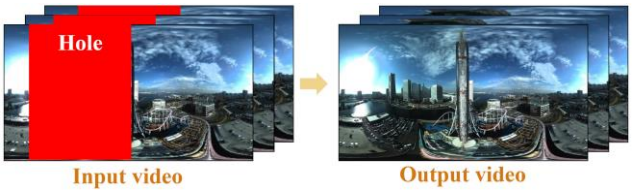

図 1. 映像補完技術の概要

2. 動き解析

2.1 カメラ投影モデル

世界座標系上で全方位画像中の画素の3次元点をマッピングするにあたっては，図2で示すようなカメラが単位球の中心に配置されたカメラ投影モデルを利用することができる．3次元世界座標上における点Pは，時間tにおいてカメラ投影モデル上の点pに投影することができるものとする．そのとき点pは，図3で示すような円筒投影した全方位画像フレームeに対して以下の式によって投影することができる．

$$x_t = \begin{cases} -\frac{w}{2\pi}\phi_t & ;\phi_t < 0 \\ w - \frac{w}{2\pi}\phi_t & ;\phi_t \geq 0 \end{cases}, \qquad (1)$$

$$y_t = \frac{h}{\pi}\theta_t$$

ここでxとyは円筒投影された全方位画像上における対象画素の2次元座標を表し，wとhは全方位フレーム画像の幅と高さを画素単位で表す．またθとϕはそれぞれ天頂角と方位角を表す．

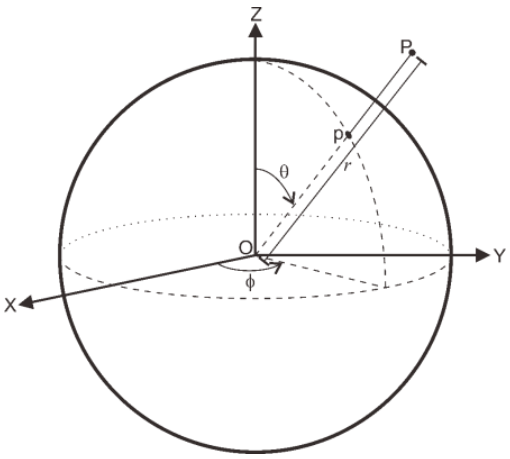

図 2. 単位球を用いたカメラ投影モデル．

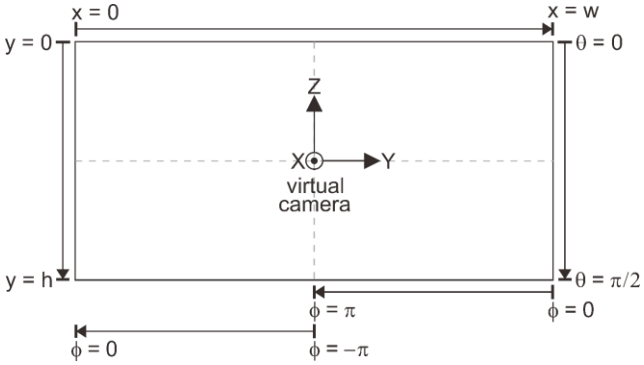

図 3. 単位球カメラモデルの円筒座標への投影．

2.2 カメラの運動に基づく画素の動き

静的なシーンの背景において，円筒座標上での画素の動きは図3のように表され，これはカメラの動きによってのみ決められる．図2と式1よりそのようなカメラの動きは時間tにおけるθ_t，ϕ_t，r_tによって求めることができる．しかし，画素の動きをモデリングするにあたって，時間によって変化するr_tを利用して補完を行うのは，画素の3次元点の時間tにおけるカメラからの相対的な奥行き情報をその都度求める必要があり，現実的でない．そこで提案手法では，画素の動きを$\{\theta_t, \phi_t, r_t\}$を使って定式化するのではなく，$\theta_t$と$\phi_t$を使って時間に対して不変（time-invariant）な奥行きsを次のような式により定義し，これを用いて画素の位置を定式化する．

$$x_{t'} = \begin{cases} -\frac{w}{2\pi}\phi_{t'}(x_t, y_t, t'-t, s) & ;\phi_{t'} < 0 \\ w - \frac{w}{2\pi}\phi_{t'}(x_t, y_t, t'-t, s); & \phi_{t'} \geq 0 \end{cases}, \qquad (2)$$

$$y_{t'} = \frac{h}{\pi}\theta_{t'}(x_t, y_t, t'-t, s)$$

ここでsは3次元点Pの，図4で示すような仮想カメラ位置に対する奥行き値を表す．

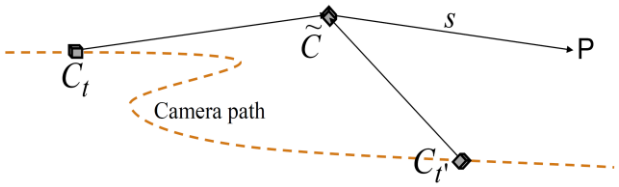

図 4. 固定仮想カメラ位置からの奥行き値

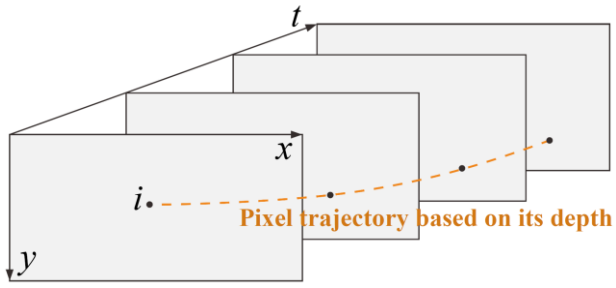

図 5. 全方位画像フレーム間での画素の軌跡

3. 映像補完パイプライン

本論文において我々は，前節で説明したようなカメラの運動を利用した画素の動き解析によって映像の補完を行う．図5で表現されるように，映像中のフレーム画像間での画素の軌跡を推定し，キーフレーム画像上での既知のカラー情報を欠落部のある対象フレームへ伝播させることで補完を行う．画素の軌跡はその奥行き値に依存するため，これを行うにはまず奥行き値を求める必要がある．提案する映像補完手法のパイプラインは1）奥行きマップの推定，2）奥行きマップの平滑化，3）カラー情報の伝播，という3段階で実行される．本節ではそれぞれの処理について詳細を説明する．

3.1 奥行きマップの推定

この処理において，まず入力画像列よりキーフレームを含むN枚の画像フレームを取り出す．そしてキーフレームにおける画素iの奥行き値を推定する．これを行うために，RGBのカラー情報の変化が画素の軌跡間で最小となるような奥行き値sを以下のエネルギー関数を使い求める．

$$\hat{s}_i = \arg\min_{s_i \in \Re} \int [\sum_{c\in\{r,g,b\}} (I_c(x(t,s_i), y(t,s_i)) - \bar{I}_c)^2] ds_i \quad (3)$$

ここでIは画素のRGB値を表し，$\bar{I}$はNフレーム中の平均のRGB値を表す．

適切な最適化アルゴリズムを用いることによりこの最小化問題を解くことができる．しかし，式3のエネルギー関数の最小化を解く際には局所解に陥る場合も多い．そうした場合に対応するため，我々は奥行き値をNs個に離散化させ，どの値が大域解であるかの評価を行うという離散的アプローチを行う．これは単純なアプローチではあるが，正確な画素の軌跡を求めるのに十分な奥行き値を求めることが可能である．

3.2 奥行きマップの平滑化

式3の最小化問題と離散的アプローチにより求められた奥行きマップにはその手法の特性上，ノイズとなるような誤差のある奥行き値も多く含む．そのようなノイズ的な奥行き値を排除するために，奥行きマップに対してエッジ情報を保ちつつ平滑化処理を施す．奥行きマップの平滑化には，[14]で提案された手法をベースとした，次の最小化問題を解くことにより行う．：

$$\hat{m}_i = \arg\min_{m_i} \sum_i \left((m_i - \hat{s}_i)^2 + \lambda \left(a_{x,i}(L) \left(\frac{\partial m}{\partial x} \right)_i^2 + a_{y,i}(L) \left(\frac{\partial m}{\partial y} \right)_i^2 \right) \right) \quad (4)$$

ここでa_xとa_yは平滑化係数であり，L は推定奥行き値の自然対数とする．式4は[14]で提案された同様のアルゴリズムで求められる．

3.3 カラー情報の伝播

平滑化された奥行きマップが得られれば，式1と式2を使うことでそれぞれの画素の軌跡を求めることができる．そのため画素の軌跡をたどることで，カラー情報を持つフレーム画像から，欠落部の対象画素へとカラー情報を伝播させることができ，最終的に映像の補完をすることが可能となる．

4. 実験結果

本節においては，本論文で提案した手法による奥行き推定結果と，映像補完結果の評価を行う．これらの実験は，提案手法を Matlab 上で実装し，OS は Windows7，CPU に Core i7 2.93GHz，RAM として 16GB を搭載したコンピュータ上において行った．

4.1 奥行きマップ推定結果

まず提案手法をシミュレーション画像列に対して適用し奥行き推定を行った．シミュレーションデータとしてバーチャル飛鳥京プロジェクト[15]の3次元モデルを使用した．このモデルは飛鳥時代の都を再現したもので数百m四方の範囲に当時の建物が立ち並んだ小規模な都市モデルである．このシミュレーションにあたっては，図6で示すような観覧車が3次元モデルの地表上にあると仮定し，カメラはそのゴンドラに取り付けられて回転しながら撮影するものとして入力動画像列の作成を行った．時間tからt'におけるカメラ運動は以下のように定式化される．

$$\begin{cases} T_X = -R(\cos\alpha_t - \cos\alpha_{t'}) \\ T_Y = R(\sin\alpha_t - \sin\alpha_{t'}) \\ \Omega_Z = \alpha_t - \alpha_{t'} \end{cases} \quad (5)$$

ここでRは観覧車の半径を表し，αはゴンドラの回転角度を表す．ゴンドラの角速度は一定とする．本論文で提案したtime-invariantな奥行き推定アプローチを適用するため，仮想のカメラ位置は観覧車の中心に設定され，各画素の奥行き値はそこからの距離として求められる．推定結果の評価にあたっては，我々は$S_{error} = (S_{test} - S_{gound\ truth})/S_{test}$,という誤差関数を用いた．ここでそれぞれの$S$は奥行き値を表し，$S_{test}$が本手法により求めた平滑化済みの奥行き値である．この評価の結果を図7に示す．この結果から求められた奥行き

値と真値の間には大きな誤差はなく，十分に正確な値が求められているといえる．

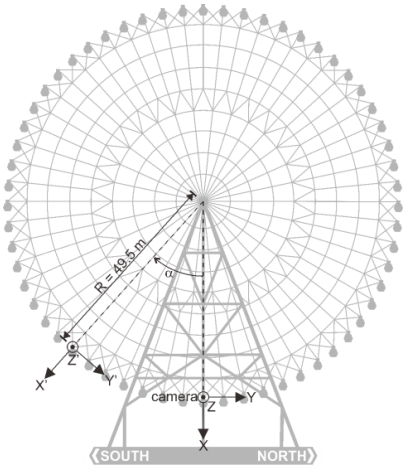

図 6. カメラの動きの設定

4.2 映像補完結果

次に提案手法をシミュレーションによる画像列と現実の観覧車（コスモクロック21，横浜）で撮影した画像列に対して適用し，フレーム画像の補完結果について評価を行う．シミュレーション画像列と実画像列の複数の対象フレームに欠落部を設け，キーフレームからの補完を行う．現実の観覧車でのカメラ運動も式5で定式化することができる．同様の手順により奥行き推定を行うため，仮想的なカメラ位置を観覧車の中心とし，そこからの奥行き値を求める．映像補完手法の評価のため，我々はRoxasらの手法との比較を行った．結果の評価には単純な誤差評価関数$I_{error} = I_{gound\ truth} - I_{test}$を用いた．ここで$I$は画素の輝度値を表し，$I_{test}$は我々の手法で得られた結果，およびRoxasらの手法で得られた結果となる．得られた補完結果は図8と図9に示す．

5. 結論

本論文において我々は全方位動画列中における時間的，空間的に大規模な欠落部を補完する手法の提案を行った．提案手法ではまずフレーム画像の奥行きマップの推定を行う．そうして推定されたキーフレーム画像の奥行きマップを利用して画素の軌跡を求めることで，キーフレーム画像から欠落部の画素へとカラー情報を伝播させ，映像補完を行うことが可能となる．実験により，推定された奥行き値は映像補完をするために十分な精度が得られることが示された．また映像補完の結果に関しても，欠落部内の復元結果は，Roxas らの提案する映像補完手法[13]に比べて正確にもとめられたという結果が得られた．

提案手法の課題として，第一には補完結果内に欠落部が残ってしまう場合があることが挙げられる．これは実験で仮定したカメラの動きが急激な回転運動であったことが原因のひとつに考えられる．カメラが並進運動のみであれば画素も線型的な動きを仮定することができ，映像再生時間の逆方向へカラー情報を伝播させる処理などが容易となる．しかし，回転運動を考慮した場合その処理はより複雑になり，結果として補完できない画素が発生することが考えられる．本論文における実験では順方向のカラーの伝播のみを用いているが，キーフレームから補完対象フレームが離れるほど補完結果が悪くなることが考えられる．もう一つの課題としては，映像補完手法全般に共通の問題ではあるが，大きなオクルージョンがあった場合には対応できないことである．時空間的に大きなオクルージョン領域が発生した場合，当然画素の欠落部の補完や奥行きの推定は非常に難しいものとなる．将来的には内挿補間手法や映像修復手法と組み合わせることにより，この問題を解決していきたい．

参考文献

[1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer Graphics and Interactive techniques*, July 2000.
[2] A. Levin, A. Zomet, and Y. Weiss. Learning how to inpaint from global image statistics. In *Proceedings of the 9th International Conference on Computer Vision*, 2003.
[3] T. Chan and J. Shen. Variational image inpainting. *Communications on Pure and Applied Mathematics*, 2005.
[4] X. Shao, Z. Liu, and H. Li. An image inpainting approach based on the poisson equation. In *Proceedings of the 2nd International Conference on Document Image Analysis for Libraries*, 2006.
[5] J. Dobrosotskaya and A. Bertozzi. A wavelet-laplace variational technique for image deconvolution and inpainting. *IEEE Transactions on Image Processing*, 2008.
[6] K. Patwardhan, G. Sapiro, and M. Bertalmio. Video inpainting of occluding and occluded objects. In Proceedings of *IEEE International Conference on Image Processing*, 2005.
[7] S. Cheung, J. Zhao, and M. Venkatesh. Efficient object-based video inpainting. *IEEE International Conference on Image Processing*, 2006.
[8] I. Drori, D. Cohen-Or, and H. Yeshurum. Fragment-based image completion. *ACM Transactions on graphics*, 2003.
[9] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 2004.
[10] Y. Jia, S. Hu, and R. Martin. Video completion using tracking and fragment merging. *The Visual Computer*, vol. 21, pp. 601-611, 2005.
[11] Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *IEEE Transactions on Pattern analysis and machine Intelligence*, 2007.
[12] T. Shiratori, Y. Matsushita, X. Tang, and S. Kang. Video completion by motion field transfer. *IEEE conference on Computer Vision and Pattern recognition*, 2006.
[13] M. Roxas, S. Takaaki, and K. Ikeuchi. Video completion via spatio-temporally consistent motion inpainting. *IPSJ Transactions on Computer Vision and Applications*, 2014.
[14] Z. Farbman, R. Fatal, D Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics (TOG)*, 2008.
[15] Virtual Asukakyo Project: http://www.cvl.iis.u-tokyo.ac.jp/research/virtual-asukakyo/

謝辞 本論文における成果の一部は日本文部科学省の NET プロジェクトの支援を受け行われた．また，観覧車からの全方位画像列の撮影は泉陽興業株式会社の協力の下，横浜のコスモクロック 21 にて行われた．

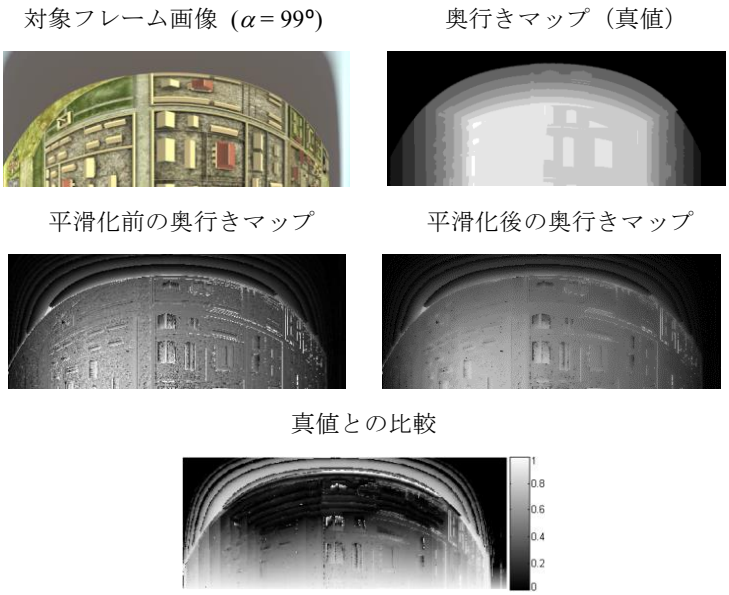

図 7. 奥行き値の比較.

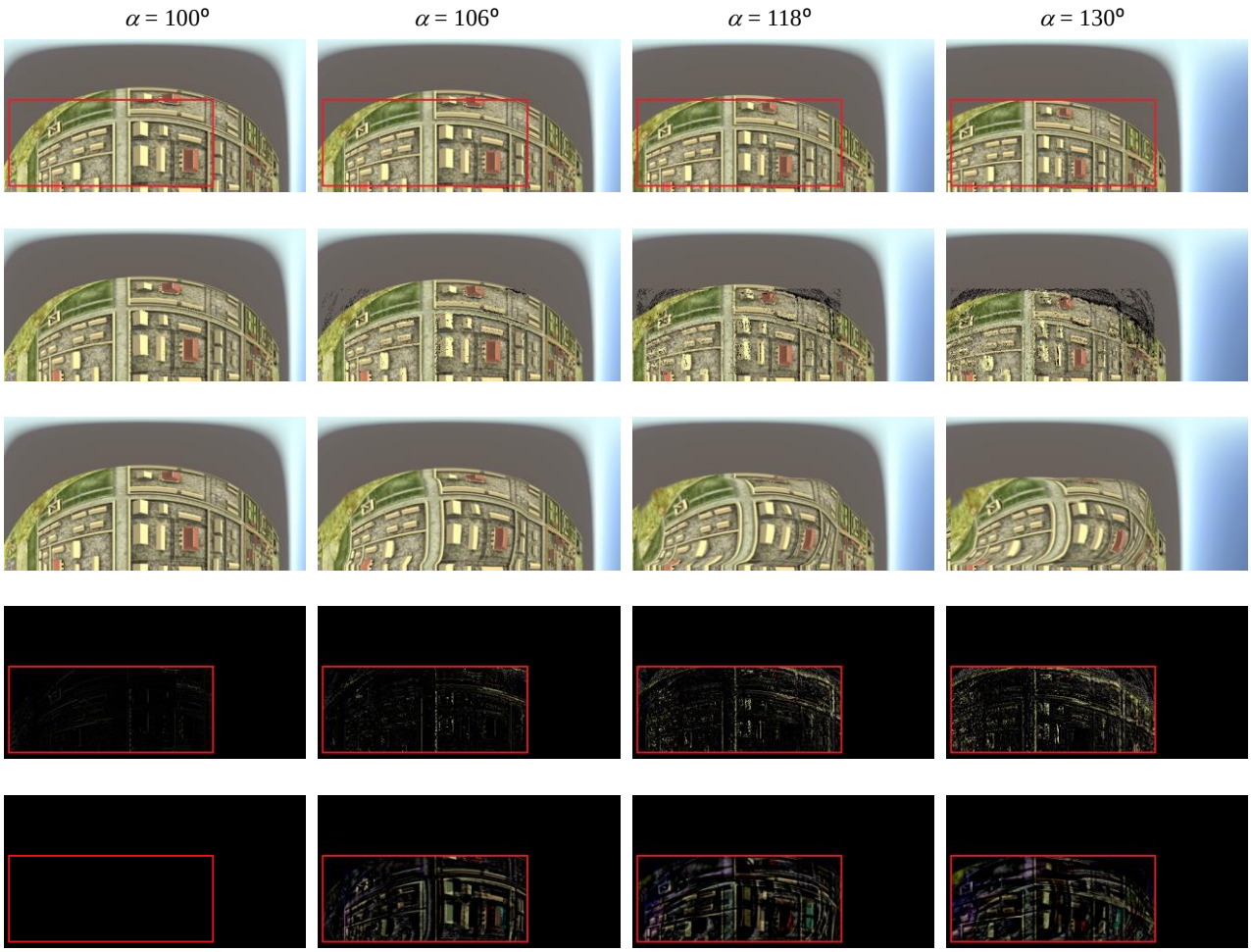

図 8. シミュレーションモデルにおける映像補完の結果。キーフレーム画像と奥行きマップには図 7 のものを用いた。赤の矩形領域が欠落部であると仮定する。1 行目:真値、2 行目:提案手法による補完結果、3 行目:Roxas らの手法による結果、4 行目:真値と提案手法の誤差評価結果、5 行目:真値と Roxas らの手法との誤差評価結果

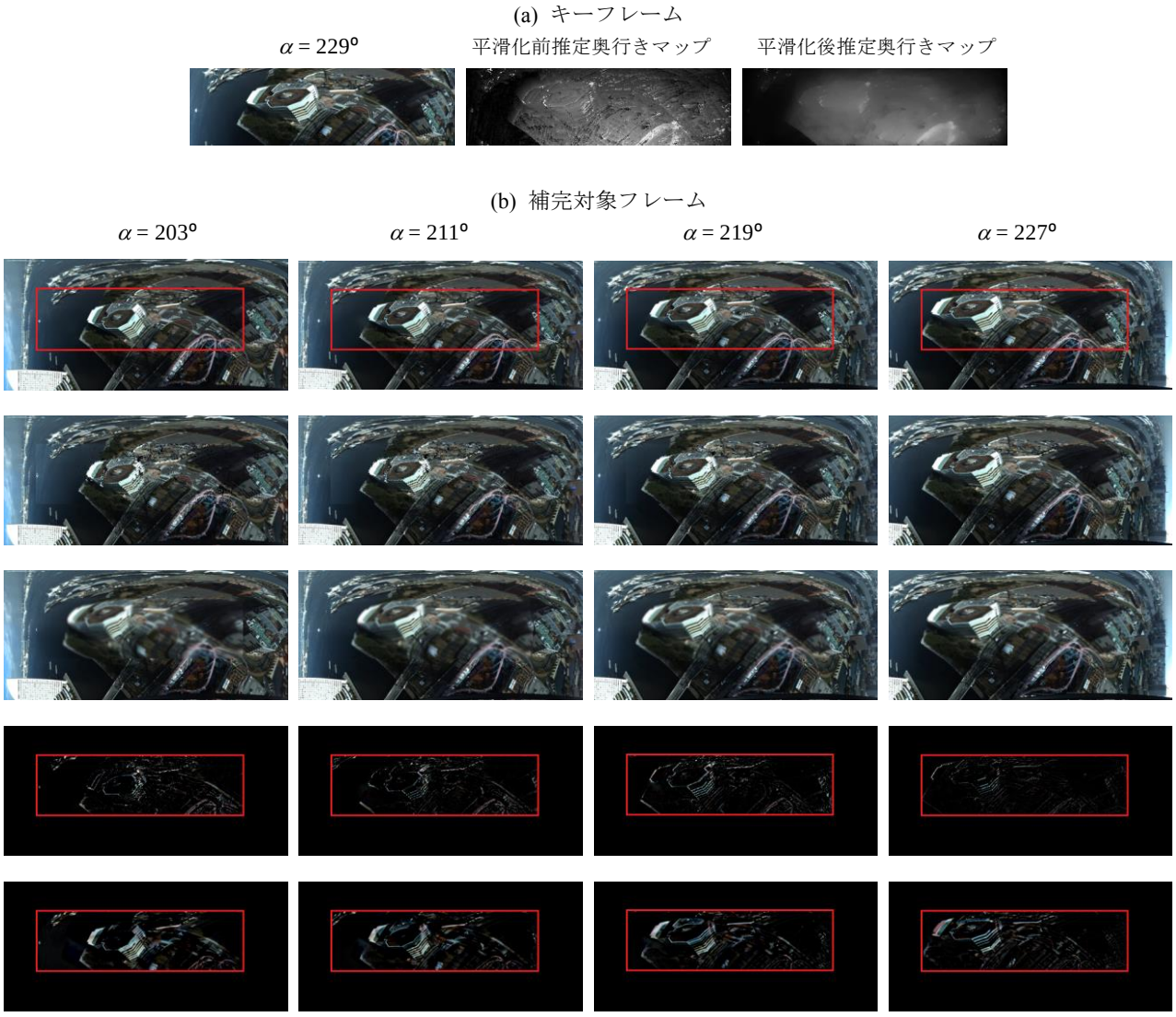

図 9. 実画像に対する映像補完結果 (a) キーフレーム画像とその平滑化前推定奥行きマップおよび平滑化後の奥行きマップ (b) 補完対象とするフレーム。赤の矩形内を欠落部と仮定する。1 行目: 真値画像、2 行目: 提案手法による補完結果、3 行目: Roxas らの手法による結果、4 行目: 真値と提案手法との誤差評価結果、5 行目: 真値と Roxas らの手法との誤差評価結果