

風景を撮影した動画における遮蔽物の除去

Diminished Reality for Landscape Video Sequence

武田 晃侑† 坂本 竜基†
Kosuke Takeda Ryuuki Sakamoto

1. はじめに

近年、インターネットにおける動画投稿サイトの利用者数が年々増加してきていることから判る通り、ビデオカメラは身近な機器となりつつある。このビデオカメラを用いて旅先で展望台から風景の動画を撮影するとすると広い範囲で風景が広がっているためすべてを収めるにはビデオカメラを横に動かしつつ撮影する必要がある。しかし、展望台は高い場所にあるため窓枠・柱・落下防止用の柵が設けられている事が多く、これらは動画に写り込んでしまう(図 1)。そこで、本研究では、このような場面において風景を撮影した動画を加工し、遮蔽物と遮蔽された風景に置き換える事であたかも遮蔽物が始めからなかったかのような映像を生成する手法を提案する。これにより、風景の画像に対する仮想的な解像度を向上させ、より壮観な情景の閲覧が期待できる。ただし、本研究における風景を撮影した映像はカメラをパンして撮影されたものではなく視点移動を伴っているという前提である。



図 1 遮蔽物を含む風景の様子

2. 関連研究

画像から物体を視覚的に除去する研究として、Image Inpainting がある。これは、本来、写真等を折り曲げた時に生じる細かい傷等の欠落領域に対して自動的に修復、補完を行う技術である。この手法は、非テキスト型 [1-2] とテキスト型 [3-7] に分類できる。両手法とも欠損領域の周辺ピクセルから補完に使用するテキストが計算される。これを欠落箇所の画素値と置き換える事で、画像の修復を行う。また、本研究のように、実在する物体を画像中から視覚的に除去する研究は Diminished Reality と呼ばれ、現実世界から不要な情報を減らす研究分野として、近年盛んに研究されてきている。Diminished Reality の実現は、カメラ位置の推定、除去対象物の検出・認識・追跡等、除去対象物に合成するテキストの取得等の技術を組み合わせている。よって Image Inpainting は、初期の Diminished Reality の研究にも用い

られてきた [8-9]。除去対象の領域に対して、擬似的に作成

したテキストを上書きし表示する。しかし、Image Inpainting は、あくまで画像を対象とするものであり動画等への適用は困難だと考える。それは、Image Inpainting を用いて作成されるテキストは、擬似的なものであり現実に存在するものではない点、また、動画に対して Image Inpainting を適用した場合は、フレーム毎に独立した補間が行われる点が要因である。これにより、動画としての一貫性が保たれず、不完全な補間結果となってしまう。また、この問題は、風景のように複数の物体を対象にした際に、より顕著に表れる。

近年の Diminished Reality の研究では、除去対象物によって遮蔽された領域を取得し、この領域に重畳する事で物体の除去を行う手法が用いられている。榎本らは、複数のハンディカメラを用いて遮蔽物の除去を行うシステムを提案した [10]。これは、複数台のカメラがそれぞれ主観カメラと遮蔽領域観測カメラの役割を担う事でカメラに写りこんだ遮蔽物とその被遮蔽領域を認識することで達成される。投影するカメラと遮蔽物、また被遮蔽領域との相対的な位置関係は AR マーカーを撮影環境に設置することで把握される。

lepetit らは、動画内にキーフレームを数枚設け、そのフレームの除去対象物を手動で指定したうえでキーフレーム間の画像を自動でセグメンテーションする手法を提案し [11]、その後、これを用いた遮蔽物の除去を行っている [12]。

3. 提案手法

3.1 処理の流れ

まず風景の動画の各フレームにおける対応点から仮想的なパノラマ画像を作成し、各フレームにおいて同じ 3 次元位置を映している画素位置同士を同定する。次に、同じ 3 次元位置の画素値を各フレームで観察し、通常の画素値をもつフレームと、それと大きく異なるフレームにクラスタリングする。このうち通常と大きく画素値が異なるフレームは、そのフレームにおいてその画素位置に遮蔽物が存在するものとして判断し、通常の画素値と入れ替える。これをすべての 3 次元位置に対して繰り返し、結果を元のフレームに書き戻すとあたかも遮蔽物がなかったかのような映像を生成することが可能となる。以下で、各処理について説明する。

†和歌山大学大学院システム工学研究科,
Graduate School of Faculty of Systems Engineering, Wakayama University

3.2 各フレーム間の対応関係の決定

まず、各フレームにおいて同じ3次元位置を映した箇所を決定し、仮想的なパノラマ画像を作成する。パノラマ画像の作成は Image Stitching と呼ばれ [13-14]様々な手法が提案されているが、風景の場合はホモグラフィが利用されることが多い。ホモグラフィとは、平面を写した画像間の変換行列であり対応点が4点以上存在する場合に求まる。そこで動画内の各フレームにおける特徴点抽出と対応付けからホモグラフィを計算し、仮想的なパノラマ画像と、そのパノラマ画像の位置、つまり3次元位置と各フレームにおける各画素位置との対応付けをおこなう。具体的には、一度ホモグラフィ H が求まれば以下のようにあるフレームにおける画素位置 (x,y) は他のフレームにおける画素位置 (x',y') に以下の式により変換することができる。

$$\lambda \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

ここで、 λ はスケールファクターである。

3.3 各画素位置のクラスタリング

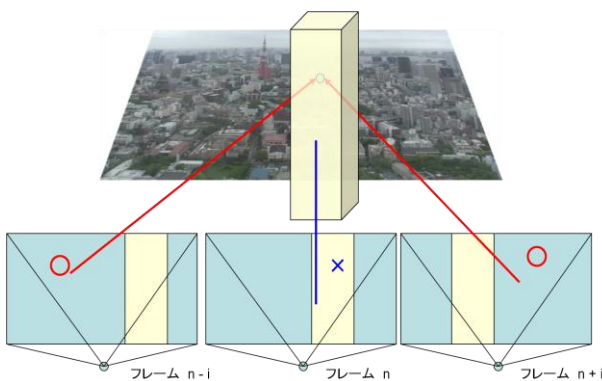


図2 各フレームと遮蔽物の関係

仮想的なパノラマ画像が作成できれば、パノラマ画像上の各ピクセルをすべてのフレームにおいて観察し、通常の画素値と遮蔽物が存在するときの画素値にクラスタリング可能である。これは図2に示した通り、同じ3次元位置を映そうとした画素値では、当然画素値も同一になるはずであるが、そのフレームにおいて遮蔽物が存在すればその画素値は異なるということを前提としている。また、遮蔽物はどの場所でも同じような画素値をもつという前提である。

図3 赤枠内の輝度値の推移 (右 フレーム番号 上: 1 中: 28 下: 39) 左のグラフは、同図右の同じパノラマ画像内の位置を示す赤枠の画素位置における各フレームでの輝度値の推移を示している。これによると、通常は200程度の輝度値を示しているが遮蔽物が通過時にはそれが大きく減少していることがわかる。これを2つにクラスタリ

ングすると各3次元位置における風景と遮蔽物のフレームに分割することが可能になる。なお、クラスタ間の距離が近い場合は、遮蔽物が通貨しなかったと判断する。

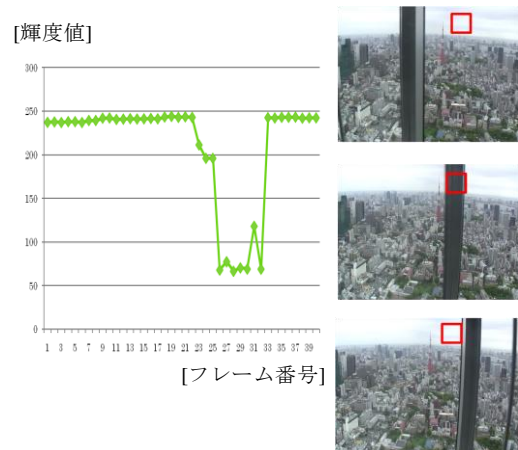


図3 赤枠内の輝度値の推移 (右 フレーム番号 上: 1 中: 28 下: 39)

4.2 被遮蔽領域の重畳

2つにクラスタリングされた結果から遮蔽物であると決定されたあるフレームにおける画素位置の画素値を、遮蔽物がないクラスタの画素値と入れ替えれば、遮蔽物の見た目上の除去ができる。どの画素値と入れ替えるかの決定は、一番近いフレームの画素値と入れ替える方法や、風景と判断されたクラスタの輝度値の平均と入れ替える方法が考えられる。次章の実験では、最も誤差が少なくなるであろう一番近いフレームの画素値と入れ替えた。

5. 実験

図5、6は、本手法を適用した動画のスクリーンショットである。どちらの動画もVGA、30fpsで撮影された動画である。図5は60枚、図6は50枚のフレーム数から成る。

図5は、風景と遮蔽物の輝度値の変化、つまりクラスタ間の距離が大きく比較的安定的に除去ができた例である。ただし、中段のフレームでは、右下、窓枠の一部が除去に失敗してしまっている。これは、仮想的なパノラマ画像における端は、すべてのフレームで撮影されているわけではないためクラスタリングに失敗していることに起因する。

図6は、高層ビルから街の風景を撮影した動画であり、風景と遮蔽物の画素値が比較的近い例である。ここでも大域的には遮蔽物の除去に成功している。しかし、やはり一部では除去が不十分な箇所も散見される。

6. おわりに

本論文では、風景を撮影した動画において遮蔽物を除去する手法について述べた。また、本手法での遮蔽物の除去の例を示した。結果、多くのフレームにおいて撮影されている3次元位置においては問題なく除去できているが、そうでない箇所では除去に失敗することもわかった。また、風景と遮蔽物の輝度値の差が僅かなピクセルでも完全な除去はできなかった。今後、これらの箇所に対する改善手法が望まれる。

参考文献

- [1] Bertalmio, M. Sapiro, G. Casells, V.; Image Inpainting; . In Proc. ACM Conf. Comp. Graphics (SIGGRAPH), pages 417–424, 2000.
- [2] Chan, T.F. Shen, J. ; Nontexture inpainting by curvature-driven diffusions : Journal of Visual Communication and Image Representation, 2001.
- [3] Criminisi, A. Perez, P. Toyama, K. ; Object removal by exemplar-based inpainting. : IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003.
- [4] Hays, J. Efros, A. A. ; Scene Completion Using Millions of Photographs ; Computer Graphics Proceedings, Annual Conference Series, 2007.
- [5] Wei, L.Y. Levoy, M. ; Fast texture synthesis using tree-structured vector quantization ; Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 2000.
- [6] 河合 紀彦, 佐藤 智和, 横矢 直和 ; パターン類似度に基づくエネルギー最小化による画像修復 ; 電子情報通信学会技術研究報告, PRMU2006-163, pp. 13-18, 2006.
- [7] 日詰 遼子, 藤代一成, 山口 泰 ; 特徴に基づく絵画風画像の補完 ; 情報処理学会全国大会講演論文集, pp175-176
- [8] Mann, S. Fung, J. Videorbits on eye tap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene ; International Symposium on Mixed Reality, 2001.
- [9] Herling, J. Broll, W. ; Advanced self-contained object removal for realizing real-time Diminished Reality in unconstrained environments ; The 2010 IEEE International Symposium on Mixed and Augmented Reality , 2010.
- [10] 榎本 暁人, 斎藤 英雄 ; 複数のハンディカメラを利用した Diminished Reality ; 映像メディア学会誌, pp901-908, 2007.
- [11] Lepetit, V. Berger, M.O. A semi-automatic method for resolving occlusion in augmented reality ; IEEE Conference on Computer Vision and Pattern Recognition2000, pp225-230, 2000.
- [12] Lepetit, V. Berger, M.O. ; An intuitive tool for outlining objects in video sequences: applications to augmented and diminished reality ; ISMR2001, pp159-160, 2001.
- [13] Zhang, Z. ; A Flexible New Technique for Camera Calibration ; IEEE TPAMI, 22(11) , pp1330-1334, 2000.
- [14] Faugeras, O. Luong, Q.T. Papadopoulo, T. ; The geometry of multiple images ; MIT Press, Cambridge, 2001.

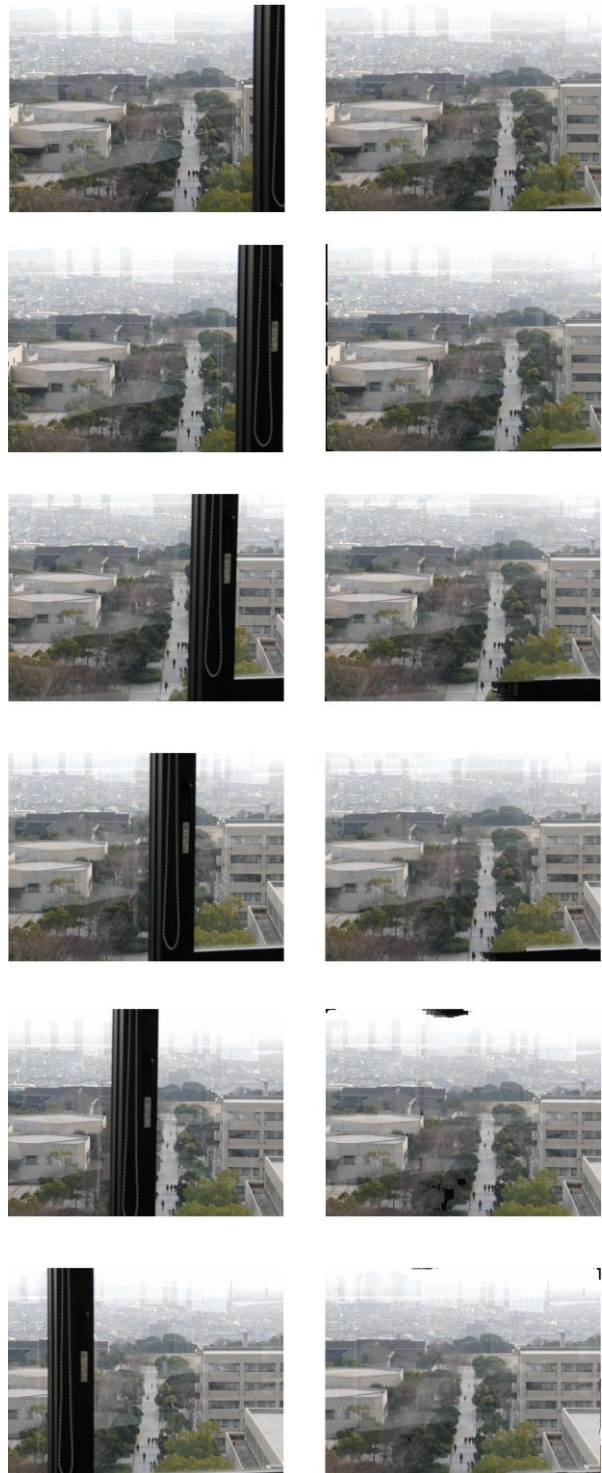


図4 左：入力映像 右：提案手法

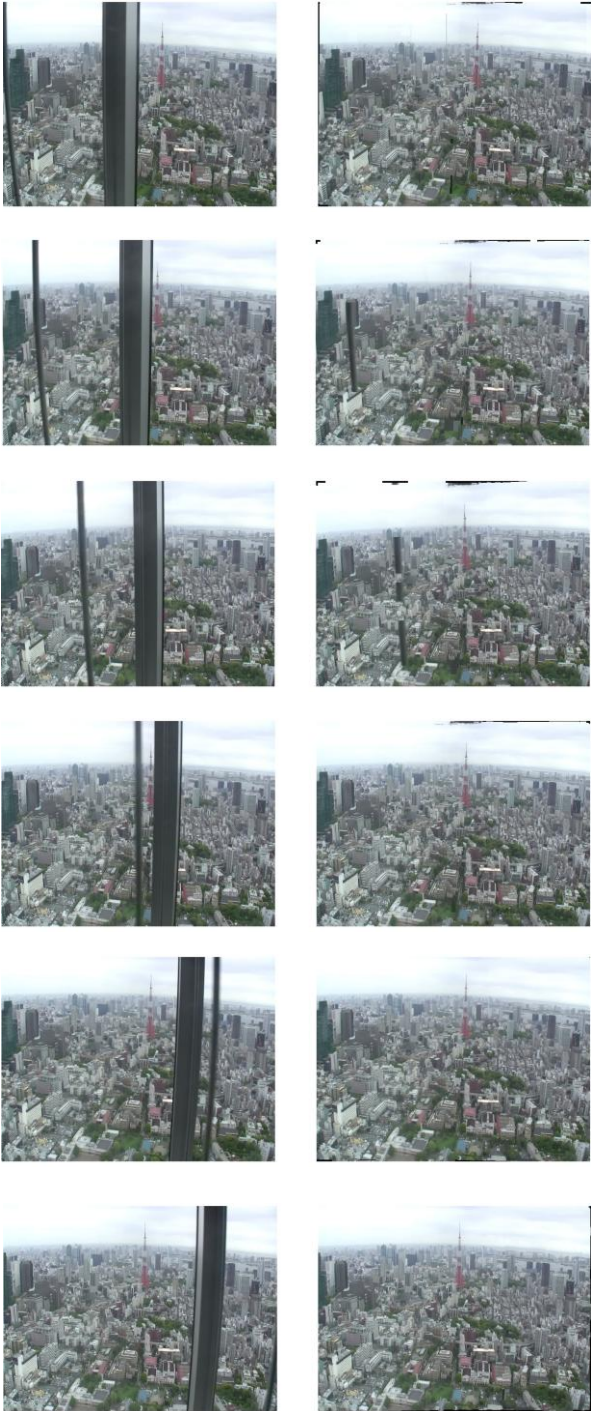


图5 左：入力映像

右：提案手法