

セグメンテーション情報を利用した画像修整手法の検討

Image Completion with Supplemental Segmented Image

大久保 英彦†
Hidehiko Okubo

藤井 真人†
Mahito Fujii

1. はじめに

映像制作の分野では、その加工や合成のニーズに対して、要求されるスキルや処理コストなどの課題が未だ多く存在している。我々はそのような画像・映像合成を、より簡便に行えるような要素技術の考案と新しい制作プラットフォームの開発をテーマとしている。その柱となる要素が、映像素材撮影の際に取得される撮影カメラの状態などの情報、あるいは別に与えられる撮影対象の奥行きや領域情報などを利用して映像制作の高度・効率化を図ることである。

本研究は、難易度の高い合成処理の1つの要素として、画像の指定領域を他の領域の画像情報を利用して補填する画像修整処理に関して検討をおこなったものである。処理対象の画像と対応した付加的なセグメンテーション情報を利用することにより、これまで難しいとされてきた大きな領域で、かつ構造的な対象領域に対する修整処理の品質向上を図った。そして良好な処理結果を得ることができたので報告する。

2. Space-Time Completion of Video

従来の代表的な修整手法としては、修整の適用に制限が少なく、動画への適用も可能な Wexler らの Space-Time Completion of Video [1]がある。これは、ビデオシーケンス中の時空間の欠損部分を時空間方向ともに補間する手法である。この欠損部分は、データセットと呼ぶ素材シーケンスからサンプルされた時空間パッチによって埋められる。その際に欠損部分の内部および周辺部に配置されたすべてのパッチ間のグローバルな時空間整合性の評価をおこなうことで、つじつまのあった動画シーケンス生成を可能にする。後述する提案手法はこの手法をベースに改良を加えたものである。そこで、以下に従来手法の概要を記述する。なおここでは静止画への適用を前提とし、時間方向の議論は省略する。

2.1. 視覚的コヒーレンス

修整対象の画像を S 、修整に利用する素材となるデータセット画像を \mathcal{D} とする。それぞれの画像内のピクセルを p, q としたとき、 W_p, V_q は p, q を中心とした小さな k ピクセルの空間パッチ (たとえば $k=7 \times 7=49$) をあらわすものとする。もし S に属するそれぞれのパッチが \mathcal{D} 中のパッチのどこかに見つけることができた場合、画像 S は \mathcal{D} に対して視覚的コヒーレンスを持つ、と定義する。一般には $\mathcal{D} = S \setminus \mathcal{H}$ 、つまり「穴」となる部分 \mathcal{H} の外側の領域が穴を埋める素材として使用されるケースが多い。

今、入力とする画像 S 中の修整対象領域である穴を $\mathcal{H} \subset S$ とし、これが修整結果である新しいデータ \mathcal{H}^* で補完されると考える。その結果、処理後の結果 S^* は、

リファレンスであるデータセット \mathcal{D} に対して同等の視覚的コヒーレンスを持つはずである。そのために、以下の目的関数を最大にする画像 S^* を見つけるようにする。

$$\text{Coherence}(S^* | \mathcal{D}) = \prod_{p \in S^*} \max_{q \in \mathcal{D}} \text{sim}(W_p, V_q) \quad (1)$$

ここで、

$$\text{sim}(W_p, V_q) = e^{-\frac{d(W_p, V_q)}{2\sigma^2}} \quad (2)$$

$$d(W_p, V_q) = \sum_{(x,y)} \left\| \mathbf{W}_p(x,y) - \mathbf{V}_q(x,y) \right\|^2 \quad (3)$$

なお $\mathbf{W}_p(x,y), \mathbf{V}_q(x,y)$ は、 p, q を中心とするパッチに含まれるすべてのピクセルの色情報のベクトルである。すなわち、 $d(W_p, V_q)$ はパッチ間の色情報の差の二乗和で局所的特徴の距離を意味し、そこから算出される $\text{sim}(W_p, V_q)$ はパッチ間の類似性の尺度を表すことになる。そしてこの尺度の導入によって、繰り返し処理で局所的なパッチ間で類似性の高いパッチを探索しつつ、画像全体の整合性を高めることが可能になる。

2.2. 最適化と反復処理

最適化のための入力となる要素は、画像 S と欠損部分 $\mathcal{H} \subset S$ である。そして式(1)を満足させるように欠損部分のすべてのピクセル値を決めていく。そのピクセル値は、空間上の点 $p \in S$ に対応する各パッチからの色 $c(R, G, B)$ と信頼値 α_p から以下のように決定される。まず α_p は、既知の点 $p \in S \setminus \mathcal{H}$ で高い信頼値をもつようにする一方で、穴の点 $p \in \mathcal{H}$ の信頼値は低くなるように設定する。

$$\alpha_p = \gamma^{-\text{dist}} \quad (4)$$

ここで dist は、穴ピクセルを含む対象パッチの中心ピクセルの、欠損領域境界への最短距離である。これを利用することにより、欠損領域の内部になるほどそのパッチの信頼性を下げる効果をもたせることが可能になる。

式(1)を満たすべく選択されたパッチ群から、ピクセル p に関する k 個すべてのパッチ $W_p^1, W_p^2, \dots, W_p^k$ それぞれの信頼値 α_p^i ($i=1, 2, \dots, k$)、そして各パッチにおける p の位置に対応した候補ピクセル \tilde{p}^i の色 c^i とから、以下のようピクセル値 c_p を算出する。

$$c_p = \frac{\sum_i w_p^i \cdot c^i}{\sum_i w_p^i} \quad (5)$$

ここで、

$$w_p^i = \alpha_p^i \cdot \text{sim}(W_p^i, V_q) \quad (6)$$

である。実際には、式(5)のような単純な加重平均値ではなく、可変バンド幅の mean shift アルゴリズム[2]を使って、よりロバストなピクセル値推定を行うほうが望ましい。

そうして得られたピクセル値で S^* を更新し、それを基に新たな最近似パッチを探索し、ピクセル値を更新するという繰り返し処理を行うことで S^* を改善していく。

† 日本放送協会 放送技術研究所



図1. 入力素材



図3. 修整結果：リファレンス手法（左）／提案手法（右）

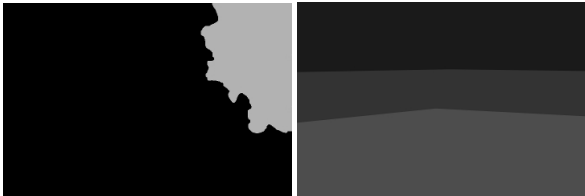


図2. 修整対象領域（左）とセグメンテーション分布（右）

なお上記のながれについては、この上位に解像度ピラミッドを利用することによる複数スケールの反復プロセスを導入する。それによって、グローバルな整合性と収束速度の向上を図ることが可能になる。

3. 提案手法

上記の従来手法は自然画や幾何模様など、さまざまな素材を対象に適用してもおおむね有効に動作するが、特に建造物などの構造化された領域を対象とし、かつ大きな欠損領域をもつ場合は、欠損部分の構造的関係が接続される保証はなく、仮にパラメータ調整を行ったとしても、実際には視覚的に望ましい修整結果を得ることはきわめて難しい。そこで本研究では、修整領域の構造的な情報を与えるセグメンテーション画像を利用して画像修整を行う手法を提案する。全体的な処理フローは踏襲しつつ、パッチ探索プロセスとピクセル値決定プロセスにおいて、以下に提案する処理を行うことで処理結果の品質向上をはかる。

3.1. パッチ探索領域の特定

付加情報としてのセグメンテーション情報は、ここでは図2右のように手動で描画した大まかな分布情報を利用する。そのセグメンテーション領域分布を $\{C_i\}_{i=1,2,\dots,N}$ とし、修整対象領域 \mathcal{H} の周辺部分を $\delta\mathcal{H}$ とすると、この $\delta\mathcal{H}$ を部分に持つセグメントの集合領域 $\{C_u\}_{u=1,2,\dots,M}$ が得られる。これを式(1)の最大の入力データセットとすることにより、隣接した周辺領域のパッチにデータセットが限定され、コヒーレンス評価時に発生する望ましくないパッチ選択を避けることが可能になる。

3.2. ピクセル値の決定

ピクセル値決定プロセスでは、最近似パッチ群から得られる各候補ピクセルのうち、対象ピクセルと同じセグメントに属する場合にそのピクセル値が高い寄与を与えるようにする。具体的には、従来手法における式(5)によるピクセル値決定のプロセスにおいて、利用する重み付け w_p^i を式(6)から式(7)のように変更する。

$$w_p^i = \alpha_p^i \cdot b_p^i \cdot \text{sim}(W_p^i, V_q) \quad (7)$$

ここで、

$$b_p^i = \begin{cases} 1 & (u=v \text{ for } p \in C_u, \tilde{p}^i \in C_v) \\ \beta & (u \neq v \text{ for } p \in C_u, \tilde{p}^i \in C_v) \end{cases} \quad (8)$$

である。すなわち b_p^i は所属セグメントを評価し、各パッチからの候補ピクセルが p の属するセグメントと異なる場合に $\beta (<1)$ のペナルティを与える。これにより、ピクセル値決定時に同一セグメントからの寄与を高められる。

4. 実験

本実験でのセグメンテーション情報は、大まかな分布として手動で描画した分布画像を利用し、また「穴」となる修整対象領域も手動で描画した(図2)。そして従来手法と提案手法による修整処理を適用し結果比較を行った。図1に元素材を、図3に従来手法と提案手法による修整処理結果をそれぞれ示す。

5. 考察

提案手法は、セグメンテーション情報により対象領域の構造的な連続性が補強され、大きな領域を対象とした修整処理においても視覚的に整合のとれた結果を得ることができた。実際には処理コンテンツに応じて重み付けなどのパラメータを調整する必要があるが、逆にそれを積極的に利用することで、処理結果の演出的なコントロールが可能になると考えている。

6. まとめ

本手法においては、セグメンテーション情報を用いることにより、パッチベース画像修整の処理結果が改善可能であることを示した。有効なセグメンテーション情報の生成手段としては、特定対象物のオブジェクト形状情報を利用して撮影カメラのパラメータを割り出すモデルベースカメラトラッキングから得られるCGオブジェクトID分布なども利用できる。本来セグメンテーション利用に関しては、静止画よりも領域指定が困難である動画への適用を主たる目的としており、今後は本手法とあわせた修整処理全体の動画適用を進め、制作ツールの1つとして利用できる汎用性の高い処理手法の構築をめざしていく。

文献

- [1] Y. Wexler, E. Shechtman, and M. Irani, "Space-Time Completion of Video", In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), March 2007.
- [2] D. Comaniciu, P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis", IEEE Trans. Pattern Analysis Machine Intell., Vol. 24, No. 5, 603-619, 2002