勾配画像処理に基づく動画中の流体部分抽出

Extraction of Fluid Dynamic Texture using Gradient-domain Image Processing

岡部誠†

安生健一草

尾内理紀夫*

Makoto OKABE[†],

Ken ANJYO[‡]

and

Rikio ONAI*

†* 電気通信大学

†* The University of Electro-Communications

†科学技術振興機構さきがけ †JST PRESTO

‡ オー・エル・エム・デジタル ‡ OLM Digital, Inc.

E-mail: †m.o@acm.org, ‡anjyo@olm.co.jp, *onai@onailab.com

はじめに 1

映像製作の現場において流体動画データベースは爆発, 炎,煙等を含むシーンを作るための重要な資源である.物理 的な流体シミュレーションも有効なツールであるが,専門 家でないと使いこなせないという問題もあるので,流体動 画を組み合わせて高画質なシーンを合成する手法は合理的 で人気がある. 他方, そのような流体動画素材の多くが特別 なスタジオで撮影されている. 例えば, 爆発を撮影する際, 後で合成することを考えて単一色の幕を背景とし[1], 更 に適切なライティングで撮影される必要がある. そのよう な特別なスタジオの準備は高価だが,一方でYouTube 1等 の動画共有サービスに目を向けると使えそうな流体動画が 数多く存在する. このような流体動画は,しかし,単一色 の幕でなく複雑な背景を持っていることが多く,しばしば, そのまま合成に使うことは難しい. そこで, 我々は流体を 背景から分離してシーン製作に使うことを考える.

単一の動かないカメラで撮影された動画から流体のみを 抽出する問題はill-pose な問題として良く知られている. こ の問題への既存のアプローチは大きく次の3つがある.1 つ目の簡単で便利な方法は Bayesian matting に基づくも のである [2]. 結果のアルファマットは美しいが,背景が 既知で,かつ,流体自身は単一色で構成されなければなら ない. 他に霧に特化した手法 [3] もあるが, 我々は背景が 未知でも動き、もっと一般的な流体を扱える手法を開発し たい.

2つ目のアプローチは動画を半透明な層に分割する技術 である. 流体も半透明な物体なのでこのアプローチが使え る可能性がある. "layer information exchange" はこの問 題を上手く解いていて,透明なガラス越しに見えるものと ガラスに映ったものを分離できる [4]. しかし, Ghanem ら [5] の報告では,この手法では流体動画は分離できない. 理由は大きく2つあり、1つはこの手法は繰り返しのある 動きが入力ビデオに現れることを条件にしていることだ. 流体の動きに繰り返しはない. もう1つの理由は,この手 法が intrinsic image decomposition [6] に基づいているこ

とで,これでは流体動画の背景の推定は困難だ.

3つ目のアプローチは流体の物理的シミュレーションに 基づく手法である [5]. 流体動画を Navier-Stokes 微分方程 式と隠れマルコフモデル (HMM) でモデリングする. この 手法は今までの方法で分離できない煙や炎を上手に分離で きるが,この手法にも3つの制限があり,我々はこれらを 取り除きたい. 1つ目は流体動画中の全てのピクセルで背 景が1度はしっかり見える必要があることだ. 現実の流体 には濃くて背景が一度も見えない箇所は存在しうる.2つ 目は前景は単一色のみでなくテクスチャも扱えるものの、 時間変化が許されないことだが、ほとんどの流体の色は刻 一刻と変化する. 3つ目の制限は計算時間だ.流体シミュ レーションが原因でフレーム毎に数分の計算時間が掛かる. これではフル HD の動画を多く処理したい場合など, 現実 的には使い難い.

他の関連技術としてはセグメンテーションとマット推定 の組み合わせがある [7,8]. 背景差分等のセグメンテーショ ン手法で2値のマスクをフレーム毎に計算し[9,10,11],そ れを元に「前景」「背景」「未定」の領域を表現する trimap を生成してアルファマットを推定する. この手法は人や動 物の髪の毛等,不透明な物体の境界部分の半透明な領域に 対しては上手く動く. しかし, 流体は半透明の領域が多す ぎて trimap を作ることが難しく, 故にこのアプローチで アルファマットを生成することは難しい.



図 1: 提案手法の概要.

我々は既存手法よりも仮定が少なく計算量の軽い手法を

¹http://www.youtube.com/

提案する. 単一の固定カメラで撮影した流体動画を入力すると,本手法は背景画像を推定し,前景及びアルファマットを推定する. 背景画像は動きのないピクセルに着目し,それらを全フレームを通して重み付き平均することで得る(図. 1-b). 密度の濃い流体については背景色の推定も行う.

背景を推定した後は、ユーザがパラメータを手動で操作しラフな初期アルファマットを作る。それをコンピュータが受け取り自動的にアルファマットを精製する。コンピュータがアルファマットをイチから完璧に推定することは難しいので、初期アルファマットをユーザに作ってもらい、一方、初期アルファマット上のノイズを手作業で取り除くのは難しいので、そこはコンピュータが担当するという作業分担が本手法のアイデアである。

初期アルファマットは背景差分に基づいて得る (図. 1-c). 煙の領域をラフに得ることはできるものの,背景のレンガ模様が載っている. この模様を取り除いて精製したアルファマット (図. 1-d) を得て,同時に前景画像 (図. 1-e) を推定する. この時,勾配画像処理を用いる. 各フレームについて,横方向,縦方向に勾配画像を得て,煙に無関係な,つまり,背景に強く関係する勾配成分を取り除き,それを積分することでアルファマットを精製する. この勾配成分を取り除く過程が我々の技術的貢献である. 計算時間は 640×360 の解像度でフレーム毎に数秒である. アルファマットも既存手法に比べて同程度か,もしくはより高品質な結果が得られている. 得られた前景画像及びアルファマットを用いて緑の背景と合成したのが図. 1-f である.

2 提案手法

流体動画のi番目のフレームは標準的なアルファブレンディングで表現される.

$$I_i(\vec{x}) = F_i(\vec{x})\alpha_i(\vec{x}) + B(\vec{x})(1 - \alpha_i(\vec{x})), \tag{1}$$

において \vec{x} はピクセル位置 , I_i , F_i , B, α_i はフレーム , 前景 , 背景 , アルファマットの画像である. 固定カメラを仮定するので , B は単一画像である. 以下 , ピクセル値の範囲は 0 から 1(255 ではない) とする.

2.1 背景推定

背景は動きの少ないピクセルに着目することで推定する. 動きはオプティカルフローで計算し速度場を得る. ここでは OFLib を用いている [12]. このライブラリは GPU を用いており, 筆者の環境では 640×480 の解像度でフレームあたり 0.1 秒程度の計算速度である. 背景画像 B(x) は次 の重み付き平均で計算される.

$$B(\vec{x}) = \frac{\sum_{i} I_i(\vec{x}) W_i(\vec{x})}{\sum_{i} W_i(\vec{x})}, \qquad (2)$$

$$W_i(\vec{x}) = e^{-\beta|V_i(\vec{x})|},\tag{3}$$

において V_i は i 番目と i+1 番目のフレーム間の速度場, $V_i(\vec{x})$ は \vec{x} における速度, W_i は重み, β は 40 に設定されている.図.2 に煙の動画における背景推定の過程を示す.

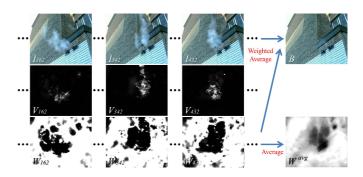


図 2: 背景推定.

フレーム I_i を第 1 行に ,対応する速度場 V_i の大きさを示す画像を第 2 行に ,重み W_i の画像を第 3 行に示す.直感的には ,速度場の大きさ $|V_i(x)|$ が小さいと重み $W_i(x)$ が高い確率で背景色を与えるという考え方である.図.2 の一番右は推定された背景 B と ,M をフレーム数とした時の重みの平均 $W^{avg}=\sum_i^M W_i/M$ である.B にいくらか煙が残っており,その領域で W^{avg} は黒く,つまり小さな値を持つことが分かる.

流体が濃い場所では真の背景色を得ることが難しい. 図. 3 の窓の周辺では煙が非常に濃く(図. 3-a), 煙が推定された背景に多く残っている(図. 3-b の黄色い楕円). この煙の色は後々背景差分の計算に悪影響なので取り除きたい. W^{avg} を用いて流体の濃さを見積もるため,2つのピクセル位置の集合を考える. パラメータ γ を導入し, $\Phi = \left\{ \vec{\phi} | W^{avg}(\vec{\phi}) < \gamma \right\}$ と $\Psi = \left\{ \vec{\psi} | \gamma < W^{avg}(\vec{\psi}) \right\}$ の2つの集合だ. Φ は図. 3-c の黒い領域, Ψ は明るい領域に相当する. $\vec{\phi}$ における煙の色を取り除くため, $\vec{\phi}$ に最も近いピクセル位置 $\psi^{ne\vec{a}rest} \in \Psi$ から色 $B(\psi^{ne\vec{a}rest})$ を $B(\vec{\phi})$ にコピーする. 結果は図. 3-d に示す. 煙の色が消え,図. 3-b よりも良い背景として使える.

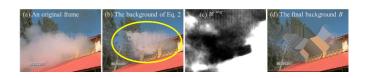


図 3: 流体密度 W^{avg} を考慮した背景推定.

2.2 初期アルファマットの推定

推定した背景 B を使って, I_i と B の差分により初期アルファマット A_i^{init} を計算する. I_i と B の各ピクセルは赤,緑,青チャンネルから成る 3 次元ベクトルとし, A_i^{init} の各ピクセルは輝度のスカラー値を持つ.初期アルファマット A_i^{init} はシグモイド関数で次のように計算する.

$$A_i^{init}(\vec{x}) = \frac{1}{1 + e^{-g(|I_i(\vec{x}) - B(\vec{x})| - t)}},$$
 (4)

において,g と t は曲線の傾きとオフセットを決めるパラメータである.図.4-a は g と t を 1.4 と 0.18 に設定して得られる.差分を計算する前にガウスぼかしを I_i と B に掛けてノイズを除去する.ガウスカーネルの直径は 5 とした.図.4-b と c は得られた初期アルファマットである.背景のレンガ模様が出てしまっている.

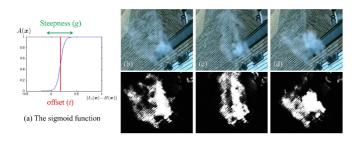


図 4: シグモイド関数と初期アルファマット.

2.3 アルファマットの精製

ユーザが作った初期アルファマットは大まかな流体領域を与えるが背景のテクスチャ模様の影響を受けてしまう.これを取り除くため勾配画像処理を用いる.図.5に概要を示す.初期アルファマット(図.5-a)の横縦方向の勾配画像を計算する(図.5-b and c).これらの勾配のうち,煙に関する勾配のみ残し,背景のレンガ模様に関する勾配を弱める(図.5-d and e).最後に処理された勾配画像を積分して最終的に精製されたアルファマットを得る(図.5-f).勾

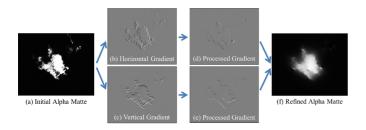


図 5: アルファマット精製の概要.

配画像は横縦方向に隣接するピクセル間で差を計算して得

る. 積分は制約付き最小二乗法で定式化する.

$$\underset{a_1, \dots, a_N}{\operatorname{argmin}} E = \sum_{p} \sum_{q \in \nu(p)} (a_p - a_q - g_{pq})^2 + \sum_{p} \lambda_p (a_p - a_p^{init})^2,$$
(5)

において, $\{a_1,...,a_N\}$ は精製されるアルファマットの全ピクセル値,p はピクセル位置, $\nu(p)$ は横縦方向に p に隣接するピクセル位置の集合, g_{pq} は p と q の間の勾配で図. 5-d と e に相当する. λ_p は p における制約の重み, a_p^{init} は初期アルファマット A_i^{init} の p における値である.このエネルギー関数の最小化は,preconditioner を用いたポアソン方程式の解法アルゴリズムを GPU 実装したものを用いる [13]. Conjugate Gradient 法の繰り返し回数は 80 回に設定した.精製されたアルファマットを得るためには,各ピクセルにおける勾配 g_{pq} と制約の重み λ_p を適切に与える必要がある.以下,それぞれ説明する.

2.3.1 勾配と制約の設定

図. 5-d と e に示す勾配画像 g_{pq} を得るアイデアは A_i^{init} の みに見られる勾配を残し A_i^{init} と背景 B に共通して見える勾配は消す,というものだ. このアイデアは Background Cut の "contrast attenuation" と類似である [10]. Background Cut では,あるピクセル位置で B の勾配の値が大きい時, A_i^{init} の勾配を消すことで A_i^{init} のみに存在する勾配だけを残し他を削除する. しかし,勾配の大きさのみに基づくこの方法は,我々の目的には不十分である. 例えば,レンガ模様は多くの場所で勾配が大きいので,この方法では背景のみならず煙を表現する勾配も消してしまうためだ. より注意深く勾配を選択するため,本手法は大きさのみでなくテクスチャ模様の類似度を A_i^{init} と B の間で観察する.

テクスチャ模様を 5×5のパッチを用いて記述する (図. 6-a). パッチの各ピクセルで勾配方向と大きさを見てヒストグラムを作る. 8 方向を見つつ 180 度異なる方向は同一視し,4次元のヒストグラムを考える. 各ビンは各勾配の大きさの合計値が入る. ヒストグラムは全ビンの合計が1となるように正規化される.

各ピクセル位置で A_i^{init} と B の間でヒストグラムをユークリッド距離で比較する. 図. 6-d は結果で輝度値がユークリッド距離を表現している. g_{pq}^{init} を A_i^{init} の勾配として次のように処理をする.

$$g_{pq} = g_{pq}^{init} (1 - e^{-d_p^2/K^2}),$$
 (6)

において, d_p は p におけるユークリッド距離(図. 6-d),K は今回は 0.01 に設定する.処理した勾配 g_{pq} は図. 5-d と e に相当する.レンガのテクスチャ模様が勾配画像と最終結果のアルファマットから取り除かれていることが分かる.

制約は初期アルファマットの値 a_p^{init} が 0.8 より大きい , もしくは 0.02 より小さい場合のみ λ_p を 0.01 に設定し , それ以外では λ_p はゼロとした.

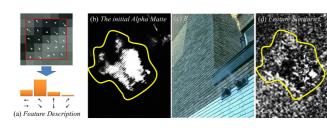


図 6: (a) 勾配を 5×5 のパッチで計算し 4 次元のヒストグラムを作ってテクスチャ類似度を計算する.(b) 初期アルファマット A_i^{init} .(c) 背景 B.(d) A_i^{init} と B の間のヒストグラムのユークリッド距離. 黄色い境界は対応する領域を示す. 黒い値が高い類似性を示す. 黒い場所では勾配がより除去される.

2.4 前景推定

前景の推定もアルファマットの精製と同様の手法で行う. 横縦方向の I_i の勾配を計算し , 式. 6 で勾配選択を行い , 最 後に積分して前景画像を得る. 処理は赤 , 緑 , 青チャンネ ルで独立に制約付き最小二乗法を解く事となる. 式. 5 で , a_i を前景の各チャンネルの輝度値とし , a_p^{init} を I_i の輝度値とする. 勾配選択は式. 6 を用いる. この時 g_{pq}^{init} は I_i の 勾配 , d_p は I_i と B の間のヒストグラムのユークリッド距離に基づくテクスチャ類似度である. 制約は a_p^{init} が 0.8 より大きい場合のみ λ_p を 0.01 に設定する. 図. 7 は結果の前景画像である.

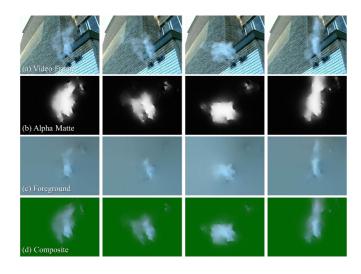


図 7: (a) 動画フレーム.(b) 精製されたアルファマット.(c) 推定された前景.(d) 緑の背景と合成した結果.

3 結果と議論

提案手法を炎,煙,爆発の流体動画に適用した.動画素 材は DynTex データベース [14] と Getty Images ²から収 集した. 各動画は流体が静止したカメラの前を流れるものだ. 動画中の流体は疎な部分と密な部分の両方を持っており, 背景は複雑なテクスチャ模様を持っている. 各結果の情報は表 1 に示す. 流体の抽出結果及び, 他の背景との合成結果, Adobe After Effects の背景差分との比較は添付のビデオをご覧頂きたい. 実験には Intel(R) Core(TM) i7 1.87 GHz CPU と NVIDIA Quadro FX 3800M GPU を積んだノートブック PC を用いた.

システムは,まず数分掛けて背景を推定する.次にユーザはパラメータ γ とtとgを操作し初期アルファマットを作る.満足のいく初期アルファマットを得たら,コンピュータが自動的に全フレームについてアルファマットを精製する.

表 1 に示すように $\gamma=0$, t=0.18 , g=1.4 が良いデフォルト値であることが分かった。 $\mathrm{Smoke2}$ では γ も調整した、煙の濃い場所があり背景を注意深く作る必要があったためだ (図. 3)。 $\mathrm{Smoke3}$ でも異なるパラメータを使ったが、これは結果的に失敗例となった。3.1 章で述べる。

図.7,図.8と図.10に成功例を示す.図.7では煙が比較的疎であり背景が上手く推定できた(図.1-b).煙が少し残るが,特に最終的な煙の抽出には問題ない.図.8では初期アルファマットと精製されたアルファマットを比較している.赤,黄色,青の円内で本手法が背景の屋根のタイル等の模様を上手く削除できていることが分かる.

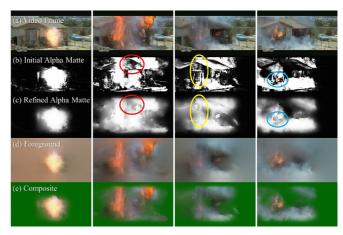


図 8: Explosion2 の結果. 赤, 黄色, 青の円内で本手法が 背景由来の模様を消せている.

3.1 手法の限界と将来課題

不要な勾配を弱めるというアプローチなので,前景がぼけてしまう傾向がある.図.8にこの現象が見える.コントラストとテクスチャ模様が弱まりぼけている.これは勾配選択手法の改善の余地を示唆している.他の解決策は,後処理でコントラストを上げたり,ヒストグラムマッチングで見た目を復元することが考えられる.

本手法の成否は初期アルファマットが上手く作れるかど うかにかかっている. そのため推定された背景が前景と似

²http://www.gettyimages.com

流体動画	フレーム数	解像度	γ	t	g	時間
Smoke1	883	352×288	0	0.18	1.4	2.92
Smoke2	359	640×360	0.035	0.18	1.4	7.42
Smoke3	441	480×360	0	0.05	2.5	6.25
Fire1	503	640×360	0	0.18	1.4	2.19
Explosion1	209	480×360	0	0.18	1.4	6.31
Explosion2	947	640×360	0	0.18	1.4	5.0
Explosion3	687	640×360	0	0.18	1.4	5.45
Explosion4	609	640×360	0	0.18	1.4	6.0

表 1: 各流体動画について,フレーム数,解像度,パラメータ γ ,t,g,フレームあたりの計算時間(秒)を示す.

たような色を持っている場合は上手くいかない. 図. 9 にこの失敗例を示す. まず,雲という濃すぎる流体のせいで背景 (図. 9-b) が上手く推定されない. そして,色差分に基づいて作られる初期アルファマットには望ましくない穴ができた (図. 9-c). 本手法はこの穴を埋めるべく勾配選択を行ったが,多少は穴がぼけただけで望ましいアルファマットは得られない (図. 9-d). この問題の解決策として色のみでなく流体の動きの分析を初期アルファマットの計算に使うことを考えている.



図 9: I_i と B の色差分があまりないために作られる穴

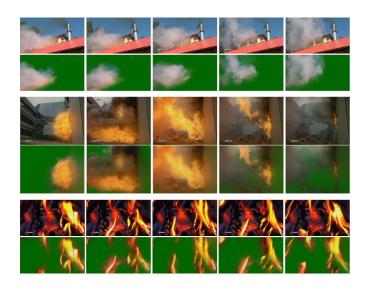


図 10: Smoke2, Explosion1, Fire1の結果.

参考文献

- Smith, A.R., Blinn, J.F. Blue screen matting. In Proc. SIGGRAPH '96. pp. 259–268, 1996.
- [2] Chuang, Y.Y., Agarwala, A., Curless, B., Salesin, D.H., Szeliski, R. Video matting of complex scenes. In Proc of SIGGRAPH 2002, pp. 243–248, 2002.
- [3] Zhang, J., Li, L., Zhang, Y., Yang, G., Cao, X., Sun, J. Video dehazing with spatial and temporal coherence. Vis. Comput. 27, pp. 749–757, 2011.
- [4] Sarel, B., Irani, M. Separating transparent layers through layer information exchange. In ECCV 2004, pp. 328–341, 2004.
- [5] Ghanem, B., Ahuja, N. Extracting a fluid dynamic texture and the background from video. In Proc. of CVPR '08, pp. 1–8, 2008.
- [6] Weiss, Y. Deriving intrinsic images from image sequences. In Proc. of Int'l Conf. on Computer Vision '01, pp. 68–75, 2001.
- [7] Gong, M., Wang, L., Yang, R., Yang, Y.H. Real-time video matting using multi-channel poisson equations. In Proc. of Graphics Interface 2010, pp. 89–96, 2010.
- [8] Sarim, M., Hilton, A., Guillemaut, J.Y., Kim, H., Takai, T. Multiple view wide-baseline trimap propagation for natural video matting. In Proc. of CVMP, pp. 82–91, 2010.
- [9] Criminisi, A., Cross, G., Blake, A., Kolmogorov, V. Bilayer segmentation of live video. In Proc. of CVPR '06, pp. 53–60, 2006.
- [10] Sun, J., Zhang, W., Tang, X., Shum, H.Y. Background cut. In Europ. Conf. on Computer Vision, pp. 628–641, 2006.
- [11] Bai, X., Wang, J., Sapiro, G. Dynamic color flow: a motion-adaptive color model for object segmentation in video. In Europ. Conf. on Computer Vision, pp. 617–630, 2010
- [12] Zach, C., Pock, T., Bischof, H. A duality based approach for realtime tv-l1 optical flow. In Pattern Recognition (Proc. DAGM), pp. 214–223, 2007.
- [13] Szeliski, R. Locally adapted hierarchical basis preconditioning. In SIGGRAPH 2006, pp. 1135–1143, 2006.
- [14] Péteri, R., Fazekas, S., Huiskes, M.J. DynTex: a Comprehensive Database of Dynamic Textures. In Pattern Recognition Letters. http://projects.cwi.nl/dyntex/.