

HSヒストグラムを用いた動画像中の対象物体トラッキング Target Object Tracking in Moving Pictures using HS Color Histogram

山口 鉄平*
Teppey YAMAGUCHI

鈴木 寿†
Hisashi SUZUKI

1. まえがき

興味ある物体を任意の画像中から分離、抽出する技術として、物体抽出あるいはセグメンテーションと呼ばれる技術が存在する。その中でも、ユーザによって入力された情報を初期値として、対象物体を追跡する半自動トラッキング技術は、MPEG4のオブジェクトベース符号化等の基礎的な分野から、監視カメラからの特定人物抽出等の応用的な分野まで、幅広い分野において利用されている。これらの技術は以前より広く研究されており、浜崎らの研究 [1] では、3次元 Watershed 法を用いた高精度な対象物体トラッキングが可能となっている。しかし、Watershed 法は初期情報として、対象物体の詳細な形状指定を必要とし、対象物体の形状が複雑である場合には、詳細な形状指定に多くの時間と労力がかかる。

我々は従来、色相 (Hue) と彩度 (Saturation) の2次元ヒストグラム (以下、HSヒストグラムと呼ぶ) を用いた領域成長法 [2] を提案し、静止画像から画素単位の分解能で明度の変化にロバストな対象物体の抽出を実現した。一般的に対象物体は、照明の影響によって色の明るさは複雑であるが、色そのものは単純な組み合わせによって構成される。そのため、対象物体を構成する色そのものは矩形などによって容易に指定できる。文献 [2] では、HSヒストグラムを用いることによって、明度の変化だけでなく、初期情報入力の簡易化や複雑な対象物体形状への対応を可能にしている。本稿では、動画像中の対象物体をトラッキングする手法として、HSヒストグラムを用いた領域成長法を拡張し、HSヒストグラムを用いた初期情報入力の容易なトラッキング手法を開発したので報告する。

2. 提案手法

2.1 概要

提案手法は、対象物体の初期情報として、入力動画像の任意フレームの矩形領域をユーザが設定し、それをHSヒストグラムへ変換する。その後、入力動画像の各画素を画素の色が持つ度数へ置換し、矩形領域から対象物体の領域を成長させ、最終的に対象物体のトラッキングを実現する。以下に領域成長処理および処理で用いる閾値の設定指針、精度向上のための工夫としての成長方向の限定について具体的に述べる。

2.2 領域成長処理

提案手法では、対象物体が時空間にまとまって存在することを利用する。入力動画像の各画素を画素の色が持つHSヒストグラム上の度数へ置換し、置換後のデータの時空間方向に領域成長処理をおこない対象物体をトラッキングする。

領域成長は度数が0より大きく、時空間方向の $3 \times 3 \times 3$ 近傍に対象物体が存在する画素において、近傍の状態に応じた判定をおこない、その判定を成長点がなくなるまで繰り返す。状態判定として、近傍に対象物体として抽出した画素または度数閾値 T_p 以上の画素が何割存在するか計算し、密集度 T_j を超える際、注目画素を対象物体と

して追加する。 $f(x, y, t)$ を近傍画素の状態として、

$$f(x, y, t) = \begin{cases} 1 & ((x, y, t) \in R \text{ または } F(x, y, t) \geq T_p) \\ 0 & ((x, y, t) \notin R \text{ かつ } F(x, y, t) < T_p) \end{cases} \quad (1)$$

と定義する。 R は対象物体の画素集合であり、 $F(x, y, t)$ は画素 (x, y, t) の度数である。注目画素 (i, j, n) が抽出領域である場合、

$$\sum_{x=i-1}^{i+1} \sum_{y=j-1}^{j+1} \sum_{t=n-1}^{n+1} \frac{f(x, y, t)}{27} > T_j \quad (2)$$

が成立する。

なお、初期情報として与えられる入力動画像の任意フレームの矩形領域を、領域成長をおこなう初期対象物体の画素として用いる。この判定は、対象物体と色は同じであり離れた位置に孤立している画素や、対象物体と隣接して色異なる画素を背景とみなし、対象物体の抽出を実現する。

2.3 閾値の設定指針

度数閾値 T_p と密集度 T_j の設定指針を考える。 T_p は初期情報からトラッキング対象物体の情報のみを抽出するための閾値であり、 T_j はトラッキング対象物体の形状に対する敏感さを調整する閾値である。初期情報がトラッキング対象物体の画素で構成される場合には、 T_p を小さい値にするとトラッキングの誤差が小さくなる。対象物体の形状に細かい部分が存在しない場合には、 T_j を小さい値にするとトラッキングの誤差が小さくなる。本稿では、トラッキングの誤差は次のように定める。 $O_h(x, y, t)$ 、 $O_a(x, y, t)$ を、それぞれ正しいトラッキング処理および提案手法によって生成した t フレームの画素 (x, y) の状態として、

$$O_{\{h,a\}}(x, y, t) = \begin{cases} 1 & ((x, y) \in \text{対象物体}) \\ 0 & ((x, y) \notin \text{対象物体}) \end{cases} \quad (3)$$

$$D(x, y, t) = \begin{cases} 1 & (O_h(x, y, t) \neq O_a(x, y, t)) \\ 0 & (O_h(x, y, t) = O_a(x, y, t)) \end{cases} \quad (4)$$

と定義する。そして、トラッキング誤差 E を

$$E = \frac{1}{N} \sum_{x,y,t} \frac{D(x, y, t)}{O_h(x, y, t)} \quad (5)$$

とする。

T_p 、 T_j の変化によるトラッキング誤差の変化を実際の動画像を用いて示す。3節の実験で用いる動画像“Foreman”と“Rolling”に対して、 T_p 、 T_j を変化させた際のトラッキング誤差の変化を図1、図2に示す。図では、横軸は T_p 、 T_j であり、縦軸はトラッキング誤差である。動画像“Foreman”と“Rolling”において、それぞれの初期情報は対象物体のみの情報である。また、対象物体の形状には細かい部分があまり存在しない。そのため、閾値 T_p 、 T_j が小さい値の場合には、トラッキング誤差が小さいが、閾値が大きい値の場合には、対象物体が抽出されずトラッキング誤差が大きくなる。

* 中央大学大学院理工学研究科

† 中央大学理工学部

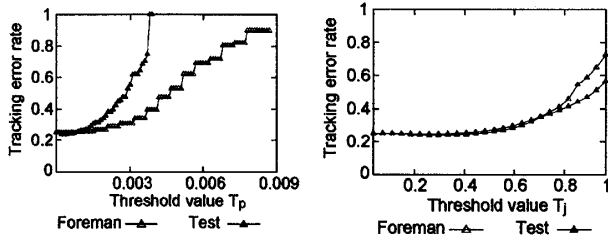


図1 T_p による変化 図2 T_j による変化

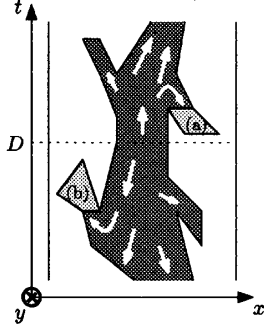


図3 成長方向の限定

表1 実験用動画像の諸元

	Foreman	Rolling
Frame rate [fps]	30	30
Picture size [pixel]	176 × 144	360 × 240
Frame number [frame]	30	25

2.4 成長方向の限定

領域成長によって、初期情報を含むフレームの過去および未来においても、対象物体をトラッキングすることができる。図3は、初期情報を含むフレームDから過去および未来への領域成長の過程を示しており、色つきの領域が対象物体として得られた結果を示している。初期情報を含むフレームより未来において、既存の対象物体とは全く別の位置(図3の位置(a))に対象物体が出現することはない。また、初期情報を含むフレームより過去において、既存の対象物体とは全く別の位置(図3の位置(b))で対象物体が消滅することはない。すなわち、単純な領域成長では、対象物体ではない領域が成長し、トラッキング誤差が大きくなる。そこで、トラッキング精度向上のために、領域成長の成長に方向性を持たせ、初期情報を含むフレームより未来においては過去へ逆戻りする成長を禁止し、初期情報を含むフレームより過去においては未来へ逆戻りする成長を禁止する。

3. 実験

2種類の動画像に対して、提案手法を適用した結果を示す。対象動画像として、“Foreman”および独自に撮影した“Rolling”を用いた。動画像それぞれのフレームレートおよび画像サイズ、フレーム数を表1に示す。“Foreman”は撮影対象の人物と背景の動きが異なる30フレームを選択した。“Rolling”はトラッキング対象物体をいすの上にのせ、いすを反時計回りに回転させた動画像であり、物体の見た目が大きく変化する。

“Foreman”では、トラッキング対象物体を人物の顔とし、初期情報は0フレームの(88, 72)を中心とした横43[pixel]、縦45[pixel]の矩形部分を利用した(図4の(a)の白枠内)。“Rolling”では、トラッキング対象物体をピンク色のぬいぐるみとし、初期情報は0フレームの(106, 114)を中心とした横112[pixel]、縦63[pixel]の矩形部分を利用した(図5の(c)の白枠内)。なお、閾値は、 $T_p = 0.00001$, $T_j = 3/27$ とした。

表2 トラッキング誤差の数値評価

	Foreman	Rolling
Tracking error	0.248419	0.280047



(a) $t = 0$ (処理前) (b) $t = 0$ (処理後) (c) $t = 29$ (処理前) (d) $t = 29$ (処理後)

図4 “Foreman”に対するトラッキング処理



(a) $t = 0$ (処理前) (b) $t = 0$ (処理後) (c) $t = 24$ (処理前) (d) $t = 24$ (処理後)

図5 “Rolling”に対するトラッキング処理

“Foreman”に対するトラッキング処理の結果として、0フレームと29フレームの処理前を図4の(a), (c)に示し、その処理後を(b), (d)に示す。同様に、“Rolling”に対するトラッキング処理の結果として、0フレームと24フレームの処理前を図5の(a), (c)に示し、その処理後を(b), (d)に示す。また、対象物体を各フレーム上において手動で切り出した画像を用いて、2節で定義したトラッキング誤差を数値的に求めた。その結果を表2に示す。

提案手法によるトラッキングの結果は、表2や図4や図5で明らかのように、対象物体の端や物体内部の欠損などにトラッキング誤差が存在するが、極めて簡単な初期情報からでも対象物体の概要や動きが得られるなど比較的良好な結果となった。誤差の要因としては、対象物体の移動幅が大きく、端の部分では近傍に既抽出領域が存在しないため、対象物体として判定されず誤差が発生すると考えられる。

4. まとめ

本稿では、動画像中の対象物体をトラッキングする手法として、HSヒストグラムを用いた領域成長法を時空間へ拡張し、HSヒストグラムを用いた初期情報入力容易なトラッキング手法を提案した。提案手法でトラッキングをおこなった動画像と手動でおこなった動画像との比較をおこない、提案手法のトラッキング精度を示した。今後の課題として、グローバル動き補償などを採用し、さらなるトラッキング精度の向上をおこなう予定である。

5. 謝辞

本研究は文部科学省の21世紀COEプログラム「電子社会の信頼性向上と情報セキュリティ」プロジェクトによる援助を受けた。

参考文献

[1] 浜崎 唯史, 吉田 俊之, “3次元 Multivalued Watershed 法を用いた動画像中の対象物体トラッキング,” 映像情報メディア学会誌, vol.58, no.5, pp.695-704, May, 2004.
 [2] 山口 鉄平, 水谷 啓, 鈴木 寿, “HSヒストグラムを利用した領域抽出法,” 第4回情報科学技術フォーラム講演論文集 (FIT2005), 3, pp.153-154, Sep. 2005.