

# 動画像からの逐次運動抽出

九州職業能力開発大学校 飯 星 潤  
 九州芸術工科大学 岡 田 正 之  
 浦 濱 喜 一

Sequential Motion Extraction from an animation

Jun IIHOSHI, Masayuki OKADA, Kiichi URAHAMA

**要約** 動画像の圧縮方法の1つとして、動画像の中より複数の動きを含む画像を個々に抽出し、背景画面と動きのある物に分ける。そうすることにより背景と動物体のスタート地点およびアフィンパラメータを設定することにより動画像を復元できる。この方法は従来の動画像圧縮方法に比べ、高圧縮化が図れる。本論文は、動画像から複数の動きを抽出する方法を取り上げた。その手法として、ファジーハフ変換による関数回帰法に基づき、動画像から主要なアフィン動き領域を逐次に抽出する方法を提案する。

## はじめに

複数の動きを含む動画像から個々の動きを抽出するにはロバストな推定法が必要であり、代表的なロバスト推定法であるハフ変換もよく使われている。ハフ変換ではパラメータ空間での分布のモードを取り出すので、主要な動きから順に逐次に取り出すことができる。言い換えれば、画像内の各ピクセルが最も多く同一方向に移動しているピクセルを取り出すことをハフ変換を用いて行うことである。混合分布モデルのように同時に複数の動きを取り出す方法では動きの個数をMDL (Minimum Description Length) 法などの組み込んだ方法で決める必要がある<sup>(1)</sup>が、逐次抽出法では抽出の停止条件を与えるだけでよく、抽出された動きの主要順位も分かる。また動きを一つずつ求めていくので求解の困難度も軽減することができる。通常は少数の動き領域を抽出するだけで十分な場合が多く、特に背景画像を連結して作られるモザイクとかスプライトなどと呼ばれるパノラマ画像は通常、最も面積が広い動き領域だけを抽出すればよく、ハフ変換が適している。

本報告ではアフィン変換動きモデルによる大域的動き領域分割を考えるが、これにもハフ変換はよく用い

られている<sup>(2)</sup>。アフィン変換動きモデルによる大域的動き領域分割とは、各画素のアフィン変換のパラメータを求め、そのパラメータが適応する画素ごとに画像領域を分割していく方法である。しかしハフ変換には、量子化幅の設定やパラメータ数による探索量の増大などの問題点がある。計算量については、しきい値処理による逐次動き抽出による簡略法が提案されている<sup>(3)</sup>。一方、ハフ変換の種々の難点を解消する一方法として、ファジーハフ変換が提案されている<sup>(4)-(6)</sup>。そこで本報告では、ファジーハフ変換に基づいて動画像から主要な動き領域を逐次に抽出する方法を提案する。

## ファジーハフ変換による運動検出

多くの動き検出法で用いられているように、本報告でも動きをアフィン変換として表すことにする。すなわち場所  $(x, y)$  の動き  $(v_x, v_y)$  が

$$\begin{aligned} v_x &= (a - 1)x + by + c \\ v_y &= dx + (e - 1)y + f \end{aligned} \tag{1}$$

で表されるとし、係数  $(a, b, c, d, e, f)$  を動画像の二つのフレームから求める。これは1種の関数回帰

問題であるので、まず最初にファジーハフ変換による関数回帰法<sup>5)</sup>について述べる。

1 ファジーハフ変換による関数回帰

独立変数と従属変数の組のデータが  $n$  組  $(x_i, v_i) (i = 1, \dots, n)$  と与えられているとする。これに1次関数  $v = ax + b$  を当てはめることにし、データから  $a, b$  を求める。ある  $a, b$  の値を仮に考えるとこれは1本の直線を表す。データ  $(x_i, v_i)$  がこの直線に所属する度合いを

$$d(a, b) = w_i e^{-\rho(ax_i + b - v_i)^2} \tag{2}$$

とする。  $\rho$  は適当な正定数である。  $w_i$  はこのデータの重みであり、今  $w_i = 1$  である。この  $d_i$  を、データ  $(x_i, v_i)$  がこの  $a, b$  の値に所属するメンバーシップ値と呼ぶ。すると自分に所属するデータの数が最も多い  $a, b$  の値は

$$\max_{a, b} \sum_{i=1}^n d(a, b) \tag{3}$$

で求まる。これを  $a_1, b_1$  とするとこれが、与えられた  $n$  個のデータから抽出される最も主要な直線を表す。

次に2番めに主要な直線は、1番めの直線に所属するデータを取り除いた残りで最も主要な直線である。データ  $(x_i, v_i)$  が1番めの直線に所属しない割合は  $1 - d(a_1, b_1)$  であるから、式(2)で  $w_i = 1 - d(a_1, b_1)$  とおいて再び式(3)で  $a, b$  を求めれば2番めに主要な直線が求まる。同様に3番めでは、

$$w_i = [1 - d(a_1, b_1)][1 - d(a_2, b_2)]$$

とおけばよい。式(3)は通常、複数の局所最適解を持つのでアニーリングが必要である<sup>6)</sup>。

以上がファジーハフ変換による関数回帰法であるが、同じことはデータの表現法を少し変えても可能である。  $x$  の変域を  $L$  区間に分割し、同様に  $v$  も  $M$  区間に分割して、  $(x, v)$  空間を格子状に分割し、格子の各セルの中心点  $(x_j, v_j); (j = 1, \dots, LM)$  全てをデータとする。そして各セルについて、そのセル内に上記の個のデータが存在するかどうか調べ、セル内のデータの個数をそのセルの重み  $w_j$  とする。こうすると式(3)は

$$\max_{a, b} \sum_{j=1}^{LM} w_j e^{-\rho(ax_j + b - v_j)^2} \tag{4}$$

と等価(但し量子化の誤差が生じるが)である。次節の運動抽出ではこのようなデータ表現法を用いる。式(2)では  $v_i$  は観測値で  $x$  の1価関数であるが、式(4)の  $v_j$

は観測値ではなく  $x$  について多価であるという違いに注意されたい。観測値は式(4)では重み  $w_j$  で表される。

式(4)の解  $a, b$  は、式(4)の  $a$  による偏微分と  $b$  による偏微分とがともに0になる点であるから、適当な初期値から出発して

$$\begin{aligned} a &= \frac{\sum w_j (v_j - b) x_j e^{-\rho(ax_j + b - v_j)^2}}{\sum w_j x_j^2 e^{-\rho(ax_j + b - v_j)^2}} \\ b &= \frac{\sum w_j (v_j - ax_j) e^{-\rho(ax_j + b - v_j)^2}}{\sum w_j e^{-\rho(ax_j + b - v_j)^2}} \end{aligned} \tag{5}$$

が収束するまで反復計算して求める。アニーリングでは  $\rho$  を最初小さな値にして式(5)で  $a, b$  を求め、  $\rho$  を徐々に大きくしながら  $a, b$  を追跡し、  $a, b$  が変化し

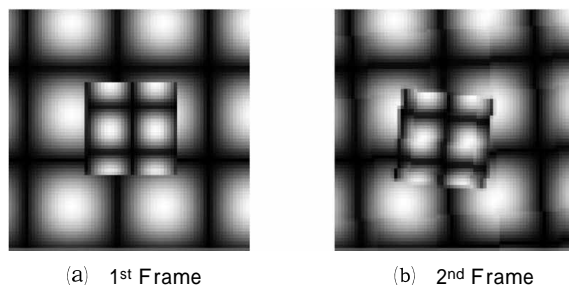


図1 動画像例

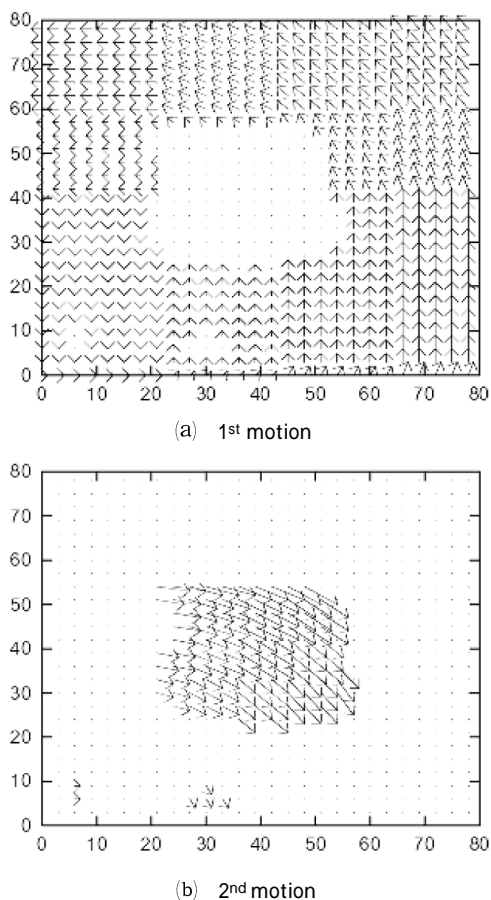


図2 抽出された動き

なくなったところで  $p$  を止め、そのときの  $a, b$  を式(4)の解とする。式(5)は一般に大域収束性は保証されないが、局所的には収束し、アニーリングでは一つ前の  $p$  での収束値が次のステップでの初期値になるので局所的に収束すればよく、ここでの実験では全て収束した。

## 2 逐次運動抽出

2枚のフレームから動きを求めるとし、画素  $(x, y)$  のカラー値が第1フレームでは  $I(x, y)$ 、第2フレームでは  $I(x, y)$  であるとする。また画素の幅(高さも同じとする)を  $h$  とすると  $v_x = ih, v_y = jh; (i, j = -m, -m-1, \dots, m-1, m)$  を動きのデータとする。すなわち全ての画素  $(0 \leq x \leq Nx, 0 \leq y \leq Ny)$  でこれら  $(2m+1)^2$  個の動きデータがある。よってデータの総数は  $N_x N_y (2m+1)^2$  である。これらの動きデータの重み(式(4)のを  $w_j$ ) を

$$w_{x,y,i,j} = e^{-q p [I(x+ih, y+jh) - I(x,y)]^2} \quad (6)$$

とする。 $q$  は動きの確率分布の分散(の逆数)に対応する正定数であり、ここでは適当に設定する。すると式(4)はこの場合

$$\max_{a,b,c,d,e,f} \sum_{x=0}^{N_x} \sum_{y=0}^{N_y} \sum_{i=-m}^m \sum_{j=-m}^m w_{x,y,i,j} e^{-p \{ [a \cdot 1]x + by + c - ih \}^2 - p [dx + (e-1)y + f - jh]^2} \quad (7)$$

となる。これにより最も主要な動きのアフィンパラメータが求まる。式(7)は式(5)と同様な反復法で解く。また2番め、3番めの動きの求め方も - 1節と同じであるから省略する。アニーリングでは前記の説明と同じく  $p$  を変化させる。これは文献(6)で説明したように  $p$  がエントロピーの係数すなわち温度の逆数であるからである。収束性についても文献(6)を参照されたい。 $q$  を変化させても同様な効果を持たせられる可能性もあるが、それは厳密にはアニーリングとは異なる。ここでは  $q$  は固定したままとする。このようにして何個かの動きを求めた後、各画素がどの動きに所属するかを求める。まず、求まった各アフィンパラメータについて各画素  $(x, y)$  での動き  $(v_x, v_y)$  が式(1)で求まる。これを量子化すると、各画素での各動きの重みが式(6)で求まる。まず最初にこの重みが全ての動きで、あるしきい値  $(\epsilon < 1)$  よりも小さい画素はどの動きにも属さないとする。残りの画素については、重みが最大の動きに属すとする。

なお、従来の動き抽出では文献(1)のように  $I_1, I_2$  の

確率分布に基づいて動きを推定しており、その評価式は本論文で言えば式(2)を代入した式(3)に対応する。それに対し式(7)は式(4)に対応する。

## 実験

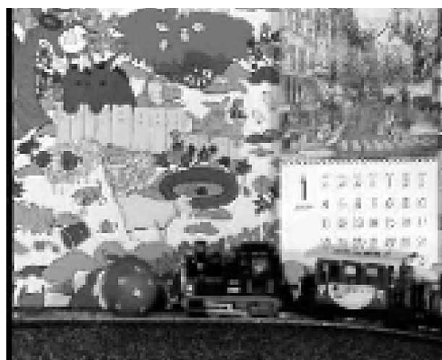
まず最初に図1のような人工的な画像から動きを二つ抽出した。正しい動きは、背景上で中央の四角が  $a = 10, b = 0.05, c = -1.0, d = -0.05, e = 1.0, f = 0.0$  である。図2に1回め(左)と2回め(右)に取り出された動きを示す。ノイズが少し生じているがほぼ正しく領域分割できている。アフィンパラメータはそれぞれ  $a = 1.0, b = -0.1, c = 0.0, d = 0.1, e = 1.0, f = -1.0$  と  $a = 1.0, b = -0.1, c = -0.3, d = 0.11, e = 1.0, f = -0.92$  であり、ほぼ正しい値が得られた。式(6)のは  $0.05$  とし、動きの範囲  $m$  は  $5$  とした。 $p$  のアニーリングの最終値は  $0.1$  とした。 $q$  は前にも述べたように動きの分散の逆数に対応し、データによるがここでの実験ではどの例でも  $0.05$  あたりの値でほぼ正しい結果が得られた。なお  $q$  が小さすぎると領域が融合しセグメンテーションがあいまいになり、大きすぎるとノイズに敏感になり細かな誤り領域が生じてくる。 $p$  の最終値  $0.1$  は前にも述べたように  $a, \dots, f$  の変化が十分小さくなる値であり、これ以上大きくしても  $a, \dots, f$  は変化しない。

次に図3に示す“Mobile and calendar”の中の2つのフレームから3つの動きを抽出した。取り出された動きを図4に示す。所々ノイズがあり、また右方のカレンダーの下部分の動きが1回めの背景の動きに組み込まれて抽出されているなど不完全な点もある(この例については同様な傾向が他の方法でも見られる<sup>(7)</sup>) が、ほぼ良好に運動が抽出されている。

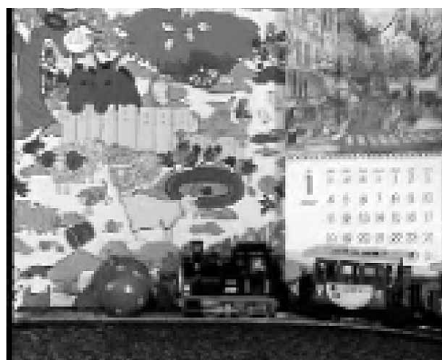
Sawhney<sup>(1)</sup>らの実験では逐次抽出法はいくつかの例で誤ったので彼らは同時抽出法を採用したが、本論文では彼らとは異なる評価関数を使って逐次抽出法でもある程度良好な結果を得ることができた。文献(1)では分散も求めているので信頼性の高い結果が得られているが、逐次法でも評価法の改良によりある程度信頼性の高い結果を得ることができると思われる。

## おわりに

ファジーハフ変換に基づいて動画画像から主要な動き領域を逐次に抽出する方法を提案して、ほぼ良好に動きが抽出できることを実験で確かめた。同時抽出法<sup>(1)</sup>



(a) 1st Frame



(b) 2nd Frame

図3 Mobile and calendar

との性能比較については更に多くの実験が必要である。

多くの運動抽出法と同様に本方法でもテクスチャの少ない様な領域では運動の推定があいまいになり、正しい動き検出が難しい。画素の色による領域抽出も組合せれば運動の検出の性能が向上するものと思われる。次の課題として、実際の動画像圧縮に適応させた実例を上げ、本方法の有用性について検討する。

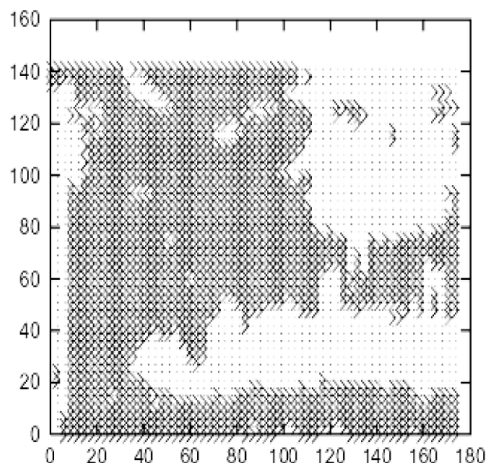
[参考文献]

(1) H. S. Sawhney and S. Ayer, "Compact representations of videos through dominant and multiple motion estimation," IEEE Trans. Patt. Anal. Mach. Intell., 18, 8, pp. 814-830, 1996.

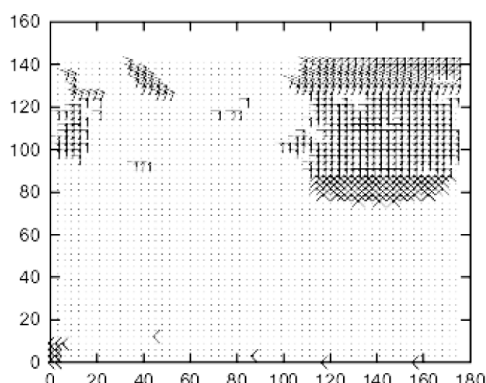
(2) 大綱亮麿, 佐藤真樹, 相澤清晴, 羽鳥光俊, "不連続性を考慮したアフィン動きパラメータの推定", 信学論, J 79-D-II, 7, pp. 1195-1203, 1996.

(3) 小松隆, 斎藤隆弘, "段階的な大域的動き領域分割アルゴリズムの一検討", 信学総大, D-12-203, p. 410, 1997.

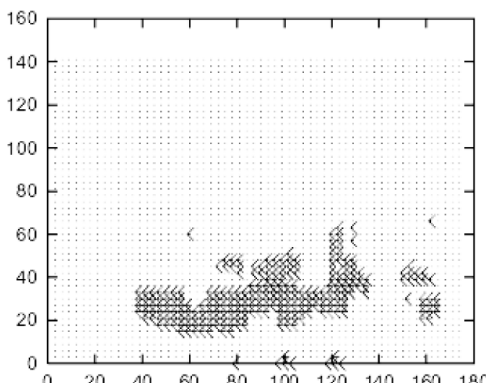
(4) J. H. Han, L.T. Koczy and T. Poston, "Fuzzy Hough transform," Patt. Recog. Lett., 15, pp. 649-658, 1994.



(a) 1st motion



(b) 2nd motion



(c) 3rd motion

図4 抽出された動き

(5) 岡田正之, 半田美恵, 松永浩之, 浦浜喜一, "ファジーハフ変換による関数回帰", 映情学誌, 51, 1, pp. 1899-1905, 1997.

(6) 岡田正之, 半田美恵, 浦浜喜一, "アニーリングファジーハフ変換によるロバスト回帰", 信学論, J 80-D-II, 9, pp. 2316-2323, 1997.

(7) Y. Altunbasak, P. E. Eren and A. M. Tekalp, "Region-based parametric motion segmentation using color information," Graph. Models Imag. Proc.

ess., 60, 1, pp.13-23, 1998.

- (8) Y. Altunbasak, P. E. Eren and A. M. Tekalp, "Region-based parametric motion segmentation using color information," Graph. Models Imag. Process., 60, 1, pp. 13-23, 1998.