

2 眼残像による移動物体の追尾

金吉雅人 †

三菱スペース・ソフトウェア (株)

動く 2 眼カメラによる移動体の追尾に対する新しい方式を提案する。2 台のカメラを近接設置し光学軸をほぼ平行固定した配置で、視差を利用した大域的検出処理と Blob 重心処理、色類似性尺度処理との併用を提案し、カメラ移動がある場合に移動体追尾が可能であることを示す。大域的検出は、差分強度の平均場検出と累積残像検出の、2 段階検出である。

Tracking of moving objects by 2 camera after image

Masahito Kaneyoshi

Mitsubishi Space Software, Inc.

We propose a new approach to the tracking of moving objects by two active camera. In the near affine configuration that two camera are set parallel like human eyes, it is shown for the new approach to detect and track a moving object in the situation of active camera, which is consist of global detectin algorithm, centroid processing of blob and color similarity measure processing. The global detection have two stages of detection: one is average mean field processing of difference intensity and the other is cumulative processing of after images.

1 まえがき

人間や肉食動物の目は、2 眼が近接しほぼ平行な光学軸であり視差による距離推定が可能であるが、どのようにして移動物体を認識するのであろうか？2 眼や多眼カメラによる物体検出は、ステレオ視として多くの研究がなされ、照明や影に影響されず、静止物体を検出し、物体 3D 復元が可能である [1, 2]。カメラ移動がある場合は、異なる視点からの画像間の特徴点追跡法が開発され、移動体の特徴点分離も開発されてきた [3, 4, 5, 6]。また、視差、奥行きに関しても多くの研究がなされてきた [9, 10, 11, 12]。しかし、形状が柔軟に変化する移動体では特徴点の画像間での全体的な対応付けが難しい、また、人間の目の残像効果や眼球運動（衝動性 saccade, 滑動性）の機能をコンピュータ・ビジョンとしてどの様にモデル化し模倣できるのであろうか？ここでは、人間や肉食動物の目を意識し、2 台のカメラを近接 (10cm 程度) 設置し光学軸をほぼ平

行固定した配置で、視差を利用した大域的検出処理と Blob 重心処理、色類似性重心処理との併用による追尾処理 [15] を提案し、カメラ移動がある場合に移動体追尾が可能であることを示す。

2 モデル化

2.1 大域的検出処理

(1) 初期設定：左右 2 眼画像の中にある遠方四角形領域 A に対して、2 眼 Gray 画像を残差逐次検定法 (SSDA) を行い最小の差分強度分布になるようにする。この時のスライド量を固定保存し、最小視差 d_A を得る。この結果、領域 A よりも手前にある物体は 2 カメラ画像の視差が d_A より大きな視差値になる。特定領域 A は、画面内の最遠方領域に初期設定する。(図 3,4 の例を参照)。これらの視差により生ずる Blob を、以下では検出に積極的に利用する。

(2) フレーム間差分処理：各フレームでは、初期設定スライド量により修正した左右 2 眼差分画像を

更に、フレーム間で差分画像 FD を取る。このフレーム間差分画像では、カメラ移動が緩やかな場合には、視差を解消した固定幅のエッジ画像となり、差分強度分布は小さなものとなる。(図 5,6 の例を参照)しかし、カメラ移動がある場合には、画面全体の視差の動きになる為、フレーム間差分画像 FD はエッジが画面全体に広がった画像となる。従って、画面 FD での差分強度分布は、広範囲に存在する。(図 7,8 の例を参照)この差分強度分布の違いを検出に利用する。

(3) 差分強度分布判定：フレーム間差分画面 FD を大域的に分割し、分割した画面セル内の平均強度 I を求める。セル内差分強度閾値 I_{cell} 、セル数閾値 N_{cell} を設定する。I が I_{cell} を超えるセル数を N とした時、移動体候補検出の判定条件とする：

- a) $N \geq N_{cell}$: 大域的検出処理をパス。
- b) $0 < N < N_{cell}$: 移動体候補登録。

移動体が画面内に存在する時には、画面 FD の分割セルの一部分の強度分布の変化となる。従って、多くのセルに広がった分布のカメラ移動の場合との違いとして検出する。(図 9,10 の例を参照)大きなカメラ移動がある場合、即ち、画面 FD の多くのセルに強度閾値を超えるものが存在する場合には、そのフレームでの移動体の検出処理を抑制(パス)する。

(4) 連続フレーム検定処理：連続累積回数の閾値 D_{th} を設定し、連続移動体候補検出回数 D の検出処理を行う：

- a) $D \geq D_{th}$: 移動体検出とする。
- b) $D < D_{th}$: 移動体候補登録を抹消。

登録された移動体候補が $D < D_{th}$ 条件ならば、outlier として、その候補の登録も抹消する。

(5) 大域的検出処理の出力は、移動体セル番号である。

この大域的検出処理は、連続フレーム数 D_{th} 期間を使っているが、フレーム間隔内、即ち、人の目の残像時間内(30msec 程度)で終了していることが最適であり、人の目の残像、眼球運動の saccade 及び saccade 抑制の一部機能に相当しているのではない

かと考察される。[16, 17]

2.2 追尾処理

(1) 大域的処理で検出した各フレームの画面 FD の移動体セル近傍に対して、Blob 重心位置検出処理を行う。

(2) その Blob 重心位置を基準に、小窓を得て、その小窓の色比率ヒストグラム検出処理を行う。また、Blob 重心位置を基準に、頭部検出処理を行う。これらのデータを追尾目標データとして登録する。

(3) フレーム間の異なる移動体セル間では、セル位置近傍相関、色類似性相関を行って追尾目標の一致を検証する。一致しない場合は新規追尾目標データとして登録する。

(4) 追尾データの確立している移動体目標は、大域的検出処理でセル検出が無いフレームでも、予測目標位置を算出して、色類似性処理、頭部検出処理及び、局所的な特徴点対応処理から目標位置を算出・更新する。[15]

この追尾処理は、滑動性眼球運動の機能に相当しているのではないかと考察される [16, 17]。

2.3 色類似性相関処理

色類似性相関については、2通りの方式：Color Histogram Intersection (CHI) 色類似性、Bhattacharyya 係数 (BC) による色類似性、で実施し比較する。

(1) CHI による色類似性

CHI による2つの密度分布 P_u 、 Q_u 間の色類似性は次で定義される：[13]

$$CHI[p(y), q] = \frac{\sum_{u=1}^m \text{Min}(P_u(y), Q_u)}{\sum_{u=1}^m (Q_u)}$$

この CHI 分布を、追尾枠内の重心を基準にした小窓内部で計算する。追尾開始時点の色ヒストグラム分布を Q_u として記憶し、以降のフレームでの小窓内部の色ヒストグラム分布を $P_u(y)$ として、色類似性を検定する。色は比率カラー (r,g,b) と Bin サイズ 16 を用いて計算する。

(2) BC による色類似性

色類似性の尺度は、色密度分布に基づいている。目

標の基準色を表現する特性 z が一つの密度関数 q_z を持つとし、一方、新たなフレームで位置 x に中心を持つ目標候補が $p_z(x)$ に従って分布した特性を持つと仮定する。その時、 B 係数に従って2つの密度間の類似性分布を計算する、その一般的な形式は以下で定義される [14] :

$$B(y) \quad B[p(y), q] = \int \sqrt{p_z(y)q_z} dz$$

類似性尺度 (1) 式は 任意の密度に対して正しく、目標のスケールに対して 不変である。サンプルデータから (1) 式を計算する為に、密度 p と q の計算では、カラーヒストグラムから導いた簡易密度を次の様に定義する :

(a) 基準目標の離散色密度 $Q = Q_{u=1, \dots, m}$ (正規化条件付) は、目標の追尾開始時点での小窓から得られる基準 m -bin 比率カラーヒストグラムとする。

(b) 新ビデオフレームでの目標候補の色密度 $P(x) = (P_{u(x)})_{u=1, \dots, m}$ (正規化条件付き) は、追尾 grouping box 内で頭部座標と重心座標とから得られる小窓の中で m -bin 比率カラーヒストグラムを計算して得られる。

B 係数分布のサンプル値計算は 次で与えられる ;

$$B(x) \quad B[P(x), Q] = \sum_{u=1}^m \sqrt{P_{u(x)} Q_u}$$

この式の幾何学的な意味は、 m 次元、単位ベクトル間の角度のコサインである。 B 係数 は次式で Bhattacharya 距離 R_B と関係付けられる :

$(x) = \exp(-R_B(x))$, 2つの分布の非類似性距離 R_B を最小にすることは、 B 計数 を最大化することである [14]。

3 実験

3.1 移動2眼カメラによる大域的な検出処理

簡易型2眼 Web カメラによる映像を図 1,2 に示す。2カメラは 10cm に設置し、ほぼ平行に光学軸を向け、カメラ校正無し、で撮影した。

図 3 の画像内中央の事務所キャビネット領域 A にて、差分 Gray 画像の強度和が最小になるようにピクセル単位でマッチングをとった後の、2眼差分



図 1: 左カメラ画像

図 2: 右カメラ画像

Gray 画像を図 3 に示す。領域 A(四角枠)の視差が最小になっている。領域 A より手前にある物体は、視差として物体エッジが広がっている。また、図 3 と次フレーム画像の2眼差分 Gray 画像との差分画像の例を、図 4 に示す。このフレーム間では2眼カメラはほとんど動いていない。従って、静止カメラでの背景差分と同じ効果で強度がほぼ消え、2眼による物体エッジもほぼ一定の幅である。

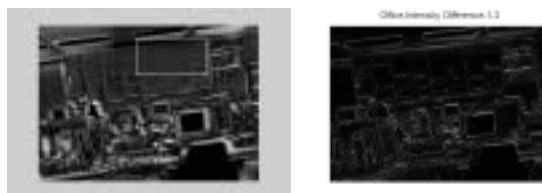


図 3: 左右カメラ差分画像

図 4: フレーム間差分

2眼カメラが大きく動いている場合、領域 A で強度マッチさせた画像を図 5 示し、図 5 と次のフレームの2眼差分 Gray 画像との差分画像の例を図 6 に示す。当然、図 6 はカメラ移動量に伴う物体エッジの広がりが存在する。これは、図 4 の物体エッジのオプティカルフローによる移動時間差拡がりとも見られることできる。



図 5: カメラ動時2眼差分

図 6: フレーム間差分

移動体が登場した場合の例を図 7, 図 8 に示す。2眼カメラが大きく動いていない場合には、画面分割の強度閾値処理により移動体の検出が可能である。



図 7: 移動体の 2 眼差分

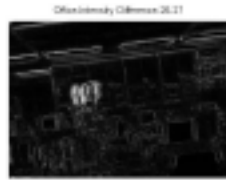


図 8: フレーム間差分

フレーム間差分画像において、x 方向 8 分割、y 方向 8 分割で 64 分割した場合の分割窓内の平均差分強度分布を、図 4、図 6、図 8 のフレーム例について、図 9、図 10、図 11、に示す。図 9-11 の縦軸は分割セルの平均差分強度値 (0 ~ 60) を示し、横軸は画像分割番号 (X 軸:1~8, Y 軸:1~8) を示す。図 9 では平均強度が全てのセルで 40 以下であるのに対して、図 10 は多くの窓に平均強度が 40 を超えるものが存在する。セル差分強度閾値 $I_{cell} = 40$ 、セル数閾値 $N_{cell} = 5$ は実験的に設定した。従って、図 10 の例の様に 5 個以上のセルに I_{cell} を超えるものがある場合は、そのフレームでの移動体検出処理をパスする。図 11 はセル座標 (4, 4) の位置にのみ平均強度が 40 を超えるものが存在する。このセル座標 (4, 4) を移動体候補として登録する。登録したセルはフレーム間で近傍座標相關の上、連続する 5 フレームで検定処理 (5out of 5) を実施して、移動体目標として登録する。

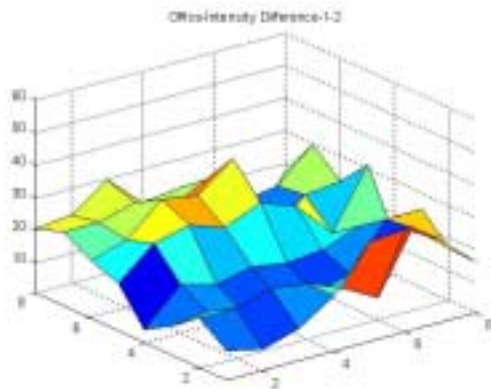


図 9: 図 4 の強度分布

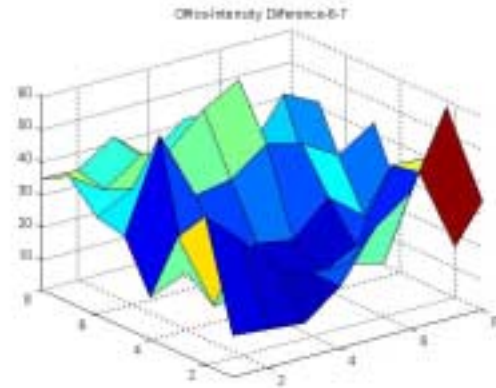


図 10: 図 6 の強度分布

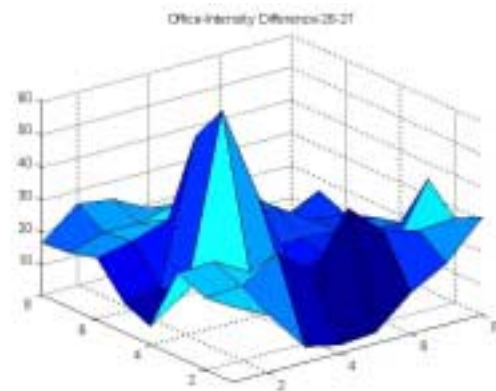


図 11: 図 8 の強度分布

3.2 動く 2 眼カメラでの移動体の追尾

大域的処理において検出した移動体に対して、追尾処理した右カメラ追尾画像の例を図 12、図 13、図 14、図 15、に示す。



図 12: No.27 フレーム

図 13: No.41 フレーム

図 12 の例では、大きな白枠は追尾処理の Blob 処理対象範囲を示し、小さい白窓が検出された Blob 重心位置を示す。この小さい白窓に対して色比率ヒ

ストグラムを得る。追尾確立と共に基準の色比率ヒストグラムとして登録する。

図 13 の例では、大きな白枠と小さい白窓は、図 12 と同じ意味であるが、白窓の色比率ヒストグラムと基準色比率ヒストグラムとの色類似性検定に失敗した為、大きな枠内で色類似性サーチを実施して色類似性検定に成功した色類似性重心位置を小さい緑窓で示している。



図 14: No.110 フレーム 図 15: No.200 フレーム

図 14 の例では、大域的処理では移動体検出がないが、既存の追尾確立した目標が存在する場合に、大きな白枠内で色類似性サーチを実施して色類似性検定に成功した色類似性重心位置を小さい緑枠で示している。ここでは、局所的な特徴点对応により前フレーム画像と現フレーム画像との間の座標対応量を求め、大きな白枠を現フレーム画像でシフトさせた後に色類似性検定を行っている。

図 15 の例では、目標が遮蔽されている為、色類似性だけでは目標捕捉継続が困難になり初めている状況を示している。

長いフレーム期間での追尾目標の軌跡の例を、図 16, 図 17 に示すが、図 16 は色類似性識別を CHI で行った場合であり、図 17 は色類似性識別を BC で行った場合である。図 16, 図 17 において、横軸は画像フレーム番号を示し、縦軸は画像 X 座標 (ピクセル値) を示している。

オフィス内歩行者がフレーム番号 20 から画面内左側から登場する。フレーム番号 21 から大域的検出処理が始まり、フレーム番号 26 でオフィス歩行者の追尾に入り、フレーム番号 60 付近で歩行者は座席に座りほぼ静止状態になる。しかし、2 眼カメラ本体が動いている為に X 座標値が変化している

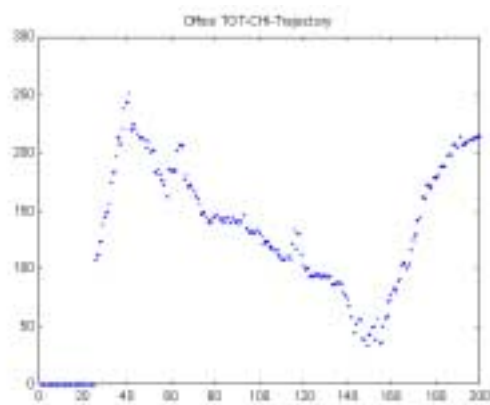


図 16: 追尾軌跡 CHI ケース

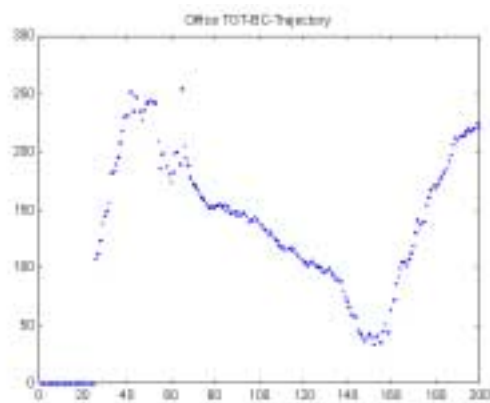


図 17: 追尾軌跡 BC ケース

ことが見てとれる。カメラの動きのままフレーム番号 200 までその目標を追尾している状態の軌跡を図 16, 図 17 は示している。

色類似性相関尺度としては CHI 方式、BC 方式共にほぼ同等の識別性能であるが、いずれの方式も、類似色の局所的なピークの存在に引かれる場合があり、また、目標が遮蔽された場合の色類似相関処理の終了判定に難しい面を持っている。特徴点对応と共存させるとの併用が必要である。

4 結びと今後の課題

本論文では視差を積極的に利用した大域的検出処理と追尾処理を新たに提案し、Affine カメラで近似

できる近接2眼カメラの運動下で、移動体を検出し追尾することが可能であることを示した。

ここでは大域的検出処理でのセル内差分強度閾値 I_{cell} 、検出セル数の閾値 N_{cell} 、検出累積回数の閾値 D_{th} 、は実験的に選択したが、移動体の視差による検出条件範囲に影響を与える。従って、初期設定した背景マッチング領域の奥行き距離とそれより手前にある物体の距離との間の視差の違いの問題であり、移動体検出条件と初期整合距離との関連性も今後の課題である。また従来、視差や奥行きマップに関して多くの研究がなされてきた [9, 10, 11, 12]。視差や奥行きマップと移動体検出閾値との関連性も今後の課題である。

従来、カメラ運動と物体との分離は、多くの研究がなされてきた。[3, 4, 5, 6, 7, 8] ここでは、画像の特徴点が抽出され画像間での対応付けがなされ特徴点対応行列が特異値分解等によって計算されるものであり、剛体エッジや模様などの特徴が明確である物体に対しては分離が可能であった。また、大域的整合性、部分空間分離法や幾何学的モデル選択 [7, 8] も示されているが、ここでの方法との関連性も今後の課題である。

更に、本文中に考察した残像や眼球運動(衝動性、滑動性) saccade 抑制、[16, 17] との関連性も今後の課題である。

参考文献

[1] O.D.Faugeras, Three Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, Boston, 1993.

[2] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK, 2000.

[3] C.Tomasi and T.Kanade, "Detection and Tracking of Point Features," CMU Tech.Rep.CMU-CS-91-132,1991

[4] J.P.Costeira and T.Kanade, "A multibody factorization method for independently moving objects", Int. J. Computer Vision, 29-3,(1998), pp159-179

[5] Z.Zhang, "Parameter Estimation Technique; A Tutorial with Application to Conic Fitting", Research Report No.2676, INRIA Sophia-Antipolis, October 1995

[6] K.Kanatani, "Statistical Optimization for Geometric Computation: Theory and Practice, Elsevier Science,Amsterdam, the Netherlands, 1996

[7] 黒澤典義、金谷健一,"アフィン空間分離法による運動物体の分離", 2001-CVIM-125,p25-32

[8] 金沢靖、金谷健一,"大域的な整合性を保証するロバストな画像の対応づけ", 2003-CVIM-136,p171-178

[9] D.Marr and T.Poggio,"A Computational Theory of Human Stereo Vision," Proc. Royal Soc. London B, vol.204, pp301-328,1979

[10] M.Okutomi and T.Kanade, "A Multiple-baseline stereo," In Proc.IEEE conf. on CVPR,pp63-69,1991

[11] C.L.Zetnik and T.Kanade,"A Cooperative Algorithm for Stereo Matching and Occlusion Detection", IEEE Trans. PAMI,vol22,No.7, pp675-684, 2000

[12] M.Agrawal and L.Davis,"Window-based, discontinuity preserving stereo," Proc.IEEE conference on Computer Vision and Pattern Recognition,pages 1-8, 2003

[13] M.J.Swain, L.D.Ballard, "Color Indexing", Intern.J.Comp.Vis. 7:11-32, 1975

[14] T.Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection", IEEE Trans. Commun. Thech. COM-15: pp52-60, 1967

[15] 金吉雅人,"歩行者グループの追尾",FIT2002 No.3 I-20,p39-40,2002

[16] 斉藤秀昭、森晃徳編、"視覚認知と聴覚認知," オーム出版局、1999, ISBN4-274-07876-0

[17] 日本視覚学界編、"視覚情報処理ハンドブック," 2000, ISBN4-259-10157