

ビジョンチップを用いたターゲットトラッキングに向けて

上山 英三^{1,3}, 八木 哲也^{2,3}, 細江 繁幸³

1 三重短期大学, 2 大阪大学大学院, 3 理化学研究所

Toward a Target Tracking with Vision Chips

Eizo UEYAMA^{1,3}, Tetsuya YAGI^{2,3}, and Shigeyuki HOSOE³

1 Tsu City College, 2 Graduate School of Osaka University, 3 RIKEN

1 はじめに

近年, 新たな産業の中核になることを期待してのことか, 国内外・産官学を揚げてのロボット工学の研究が盛んに進められている. 中でも人型の2足歩行ロボットの開発に力点が置かれている. その理由の一つ目としては, 見た目のインパクトやPR効果があげられるが, もっと重要なこととして, 人型二足歩行ロボットを構成するには, あらゆる学問分野にまたがる要素技術, さらにそれらを統合する技術が必要になるからである. つまり, あらゆる技術が, 底上げされることが期待されるからである.

例えば, 機械工学, 電気・電子工学, 情報工学, 材料工学などの工学のみならず, 生体にヒントを得るため, 解剖学, 大脳生理学, 神経科学などの医学分野の知見, さらに, 認知科学, 心理学にも密接に関わる横断的な学問分野だからである. また, 人間とは何かという問いに対して, 暫定的ではあるが, 人型ロボットを構成することは一つの解答になる. ただし, 哲学, 倫理学等の問題は依然不可欠な状況には変りない.

横道にそれだが, ロボット工学を構成する重要な要素技術として, 視覚, 聴覚, 触覚等のセンシング技術がある. 特に, 人間のようなしなやかな運動制御機構を実現するためには, 旧来の制御理論に加え, 視覚情報に基づく制御理論・機構の重要性が増している. というのも, 人間は外界から得る情報の約8割を視覚に依存しているため, 生物のように柔軟で適応性に富む視覚-運動機構を構成・実現するには, 生体の視覚機構の特性・長所を取り入れた人工の視覚機構の構成が必須になるからである.

その特性・長所とは, 生体の個々の視覚神経細胞の応答速度が遅いにも拘わらず, 並列処理機構により素早い視覚処理・認識が行われていることである. 特に, 視覚情報の持つ空間的 Markov 性¹⁾ (空間的に離れた点どうしの相関よりも近い点どうしの相関の方が高いという性質) と適合するように, その並列処理機構を実現する空間的なシナプス結合は, 空間的に近傍に位置するものに限定されていることが重要である.

計算機を用いて視覚処理を代替させることをコンピュータビジョンと言うが, 従来のコンピュータビジョン研究では, Von-Neumann 型アーキテクチャに基づく計算機を使用するため, 画像データを1次元的・逐次的にしか処理できない. つまり, 一度データをメモリ上に一列に配置し, CPU との間で随時処理と記憶を実行することしかできない. その際, 2次元データである画像の持つ空間的な局所相関の大きさを利用できず, 非効率な処理しかできない. そのため, 計算効率が上がらず, 特に処理速度に限界ができてしまうという欠点がある.

このような従来型のコンピュータビジョン研究の欠点を克服するために, 近年, 局所並列処理機構に基づく画像処理装置が様々に開発されている. 特に, 画像処理・視覚処理のフロントエンドとして, 人間の低位の視覚処理を実現する集積回路であるビジョンチップの開発・研究が盛んである. それらは, ICチップ上において受光素子部を抵抗ネットワークで局所結合した構造になっており, 画像の局所相関を有効利用して効率的で高速な処理が実現できる局所並列性を有している. 全ての素子間を結線することは, ノード数に対し結線数が二乗のオーダーで増加するという結線数の爆発を招き, ハードウェア上の実現も難しく, 当然非効率でもある.

そのビジョンチップの例として生体の網膜での視覚処理を模倣した, いわゆる人工網膜と呼ばれるものがある. ただし, 人工網膜とは言え, 実際の生体の眼球に埋め込んで網膜の役割を果すような視覚デバイスの開発および, サイボーグ技術の人体への臨床応用は, かなり先になると思われる. ここでいう人工網膜とは, コンピュータに接続され, 比較的低レベルの画像処理を局所並列処理で行う視覚処理デバイスのことである. 低レベルの画像処理とは, 生体の初期視覚²⁾における空間フィルタリング等の処理である. 人間の網膜レベルでは, ラプラシアンガウシアン型の空間フィルタリングが実現されていると考えられており, それは平滑化と輪郭強調を同時に行うものである. それがノイズ除去と注視対象を背景から抽出する処理の最初のステップだと考えられている.

上記の性質をもつビジョンチップは二種類ある。ひとつはデジタル型のものであり、もうひとつはアナログ型のものである。前者の例としては、東京大学の石川研究室で開発されたものがある。それは、サンプリング時間が1msと非常に短いという長所を持つものの、照明の変化に対して敏感という短所をもつ。ただし、高サンプリングレートという特性を生かして、ビジュアルフィードバック制御も高速で行うことにより、全体の制御を高速・高精度で行うことができる。その実現例として、パッキングロボット³⁾がある。

一方、アナログ型の例として、大阪大学の八木研究室で開発されたものが挙げられる⁴⁾。それは、サンプリング時間こそ数10msと遅いものであるが、照明条件の変化にも対応できる。生体を模倣した視覚処理システムを構築することを考えると、このような柔軟性を備えたデバイスが望ましいと考えられる。

さて、本稿で考察することは、上記のアナログ型のビジョンチップを内蔵したカメラを用いて、動く対象物体を追跡すること、即ち、ターゲットトラッキングと呼ばれる処理機構をどのように構築するのかということである。そのために、まず、アナログ型のビジョンチップ内蔵カメラを概説し、次にその性質に基づいた運動検出機構のシミュレーションを行う。そして、どのような制御方法を用いて、ターゲットトラッキングを実現していくのかということを考えていく。

2 アナログビジョンチップ内蔵カメラ

まず、大阪大学八木研究室で開発されたアナログ型のビジョンチップを内蔵したカメラシステムを紹介する。それは NeuralimageTM 社により、レンズ系と制御ボード一体型のカメラとして商品化されており、Intelligent Vision Sensor Unit "ASPECTUS" という名で販売されている。その特徴は、まず、Fig.1にあるように、USB 端子でコンピュータに接続可能で、電源の供給を受けながら、コンピュータに並列処理結果の画像を実時間で転送する。

出力としては、空間的平滑化・2次微分画像や時間差分などの処理結果がある。また、GUIを含むインターフェイスソフトウェアの充実も図られ、様々なプログラミング環境にも適応するよう工夫されている。

そして、輪郭や移動物体の重心等が照明条件の変化によらずに、安定的な出力画像が得られるというアナログビジョンチップの特徴を忠実に実現している。

なお、同社の前バージョンのカメラのチップにおいて

は、受光素子間の空間相互作用が六角格子に基づいていたが、"ASPECTUS" では、正方格子に基づいており、プログラミングが容易になっている。

本稿では、実際に上記の"ASPECTUS"を用いて、動画像処理を実現することを最初のステップとしてめざす。そこで、空間的に平滑化と2次微分された画像と、エッジ(ゼロ交差)の情報を用いて、ランダムドットパターンで合成された動画(ランダムドットキネマトグラム: RDKと略記, Fig.2)から動き(オブティカルフロー: OFと略記)を検出・推定する方法について考察することを目標とする。なお、Fig.2(左図)のように、ランダムパターンの上を同様のパターンが平行移動するような動画を観察すると、Fig.2(右図)のように、中心部の正方領域が手前に浮き出て知覚される⁵⁾。動きの検出方法には、マッチング法、勾配法、エネルギー法、相関法などがある⁶⁾。また、上記と同様のアナログビジョンチップを用いて、運動するエッジに基づくロバストなオブティカルフローの検出を行った研究⁷⁾もある。本稿では、平滑化と2次微分された画像から、勾配法に基づく方法で運動を検出・推定する方法を提案する。

3 運動ベクトルの推定方法⁸⁾

Neuralimage 社製のアナログビジョンチップは、入力画像に対して、ラプラシアンフィルタの差分(DoG)の畳み込み込んだもの、およびその輪郭を出力できる。本稿で示すシミュレーションでも、その二つの情報に基づいてRDKからオブティカルフローの推定を試みる。ただし、ここでのRDKはFig.2とは異なり、全体が一様に、ある方向に動くものとする。

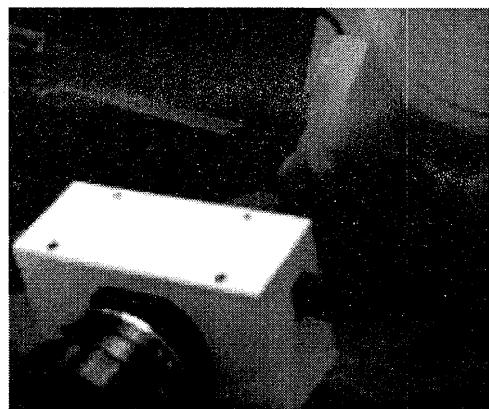


Fig. 1 Camera system plugged into PC

3.1 D o Gフィルタとゼロ交差

離散化された2次元平面上における k フレーム目の RDK の画像を $I_{i,j}^k$ とする。また、標準偏差 σ で特徴付けられるガウシアンフィルタを $G(\sigma)$ と表記する。その際、ラプラシアンガウシアンフィルタを最もよく近似する D o G フィルタは、 $G(\sigma) - G(1.6\sigma)$ である²⁾。RDK に D o G フィルタを畳み込んだものを $f_{i,j}^k$ と表記すると、

$$f_{i,j}^k = I_{i,j}^k * \{G(\sigma) - G(1.6\sigma)\} \quad (1)$$

となる。この畳み込みは空間的な平滑化を行いつつ、輪郭を抽出するための特徴を強調する作用を持つ。

そして、 $f_{i,j}^k$ のゼロ交差 $z_{i,j}^k$ を、微小な閾値 ϵ を用いて

$$z_{i,j}^k = \begin{cases} 1, & f_{i-1,j}^k f_{i+1,j}^k < -\epsilon \text{ or } f_{i,j}^k f_{i,j+1}^k < -\epsilon \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

のように検出することで、画像の輪郭が得られる。

3.2 オプティカルフローの検出

$f_{i,j}^k$ からオプティカルフロー (OF) を検出する方法として、空間的局所最適化 (SLO) 法⁹⁾を用いる。

まず、時間差分が下添え字を表すように、

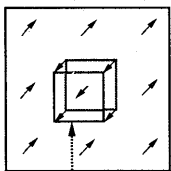
$$f_t = f_{i,j}^k - f_{i,j}^{k-1}$$

とし、そして、フレーム数を表す上添え字も省略する。 x 方向 y 方向の空間差分 (Sobel オペレータを使用) も、位置を表す下添え字 i,j を省略して、それぞれ、 f_x, f_y と表記する。OF (u, v) を求める際の勾配法の基本制約式は

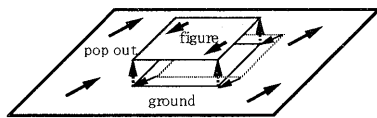
$$f_t + u f_x + v f_y = 0 \quad (3)$$

であるが、一意な解を求めるには何らかの制約条件が必要となる。そこで、 (u, v) は局所空間的に一定であると仮定し、最小二乗法によってこれを求める。つまり、和をとる範囲を 5×5 の正方領域として、

$$E = \sum \{f_t + u f_x + v f_y\}^2 \quad (4)$$



(a) Shift based RDK



(b) Result of perception

Fig. 2 Shift based RDK and results of perception

を最小にする (u, v) を、

$$u = \{(\sum f_y^2)(\sum f_x f_t) - (\sum f_x f_y)(\sum f_y f_t)\} / d \quad (5)$$

$$v = \{(\sum f_x^2)(\sum f_y f_t) - (\sum f_x f_y)(\sum f_x f_t)\} / d \quad (6)$$

のように求める⁹⁾。ここで、 d は以下である

$$d = (\sum f_x f_y)^2 - (\sum f_x^2)(\sum f_y^2) \quad (7)$$

3.3 OFの選別抽出

上記のOFは、ノイズによる影響を受けやすい一方、 f の傾きの小さなところでは正しい値が得られにくい。そこで、ゼロ交差の部分からOFを検出することにする。即ち、 $z_{i,j} = 1$ となる位置 i,j から $(u_{i,j}, v_{i,j})$ を抽出する。

さらに、ゼロ交差の情報を生かして、OFのはずれ値を除外することを考える。 $z_{i,j} = 1$ となる位置 i,j 上で求められている $(u_{i,j}^k, v_{i,j}^k)$ を逆に遡り、その位置がゼロ交差であったかどうかを確認する。もしそうなら、そのOFの信頼性は高く、そうでないなら信頼性は低い。そこで、

$$w_{i,j}^k = \begin{cases} 1, & z_{i,j}^k = 1 \text{ and } z_{i-u_{i,j}^k, j-v_{i,j}^k}^{k-1} = 1 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

を求め、 $w_{i,j} = 1$ となる位置から $(u_{i,j}, v_{i,j})$ を抽出する。

3.4 反応拡散方程式によるOFの伝播

前節で抽出したOFの観測値 $(\tilde{u}_{i,j}, \tilde{v}_{i,j})$ を画面全体に行き渡らせるために反応拡散方程式を用いる。その時間反復によって得られる値をOFの推定値 $(\hat{u}_{i,j}, \hat{v}_{i,j})$ とする。

前節で説明したように、OFの観測値には違いがある。まず一つ目として $z_{i,j}^k$ を用いるものである。即ち、

$$\dot{\hat{u}}_{i,j} = -(\hat{u}_{i,j} - \tilde{u}_{i,j}^k) z_{i,j}^k + D \Delta \hat{u}_{i,j} \quad (9)$$

$$\dot{\hat{v}}_{i,j} = -(\hat{v}_{i,j} - \tilde{v}_{i,j}^k) z_{i,j}^k + D \Delta \hat{v}_{i,j} \quad (10)$$

の時間発展によってOFの推定値を求める方法である。反応項にゼロ交差 $z_{i,j}^k$ の位置のみからの情報が観測値として与えられることが反映されている。

一方で、ゼロ交差のフレーム間移動が検出されたOFの観測値に一致するかどうかを検証し、一致しないものを除外した $w_{i,j}^k$ を用いる方法について前節で説明した。これについても同様に、 $w_{i,j}^k$ を用いた反応拡散方程式

$$\dot{\hat{u}}_{i,j} = -(\hat{u}_{i,j} - \tilde{u}_{i,j}^k) w_{i,j}^k + D \Delta \hat{u}_{i,j} \quad (11)$$

$$\dot{\hat{v}}_{i,j} = -(\hat{v}_{i,j} - \tilde{v}_{i,j}^k) w_{i,j}^k + D \Delta \hat{v}_{i,j} \quad (12)$$

の時間発展によってOFの推定値を求めていく。

4 シミュレーション

本稿で用いたRDKは 32×32 画素で定義された、白黒2値のランダムパターンからなる動画 $I_{i,j}^k$ である (Fig.4(a)). 画面内全体に $(\bar{u}, \bar{v}) = (-1, -1)$ のベクトルを与え、右上から左下にパターンを平行移動させた。

これにDOGフィルタを施したもの $f_{i,j}^k$ が、Fig.4(e) である。画素の幅は1としてあり、これに対して画素一つ半の幅、即ち $\sigma = 1.5$ という値で平滑化されている。

この $f_{i,j}^k$ からゼロ交差 $z_{i,j}^k$ を求めたものが、Fig.4(b) である。このとき、 $\epsilon = 0.0001$ とした。

Fig.4(c) はゼロ交差 $z_{i,j}^k$ 上で検出されたOFである。SLO法では 5×5 の範囲でOFが一定と仮定する。そのため、画面の境界から2画素分にある周辺部分で検出されるものは信頼性が低いものとして除外した。

Fig.4(f) は式 (8) に従って、抽出された $w_{i,j}^k$ であり、Fig.4(g) はその上で検出されたOFである。Fig.4(c) と比較すると、極端な誤りが除外されていることが分かる。

Fig.4(d)(h) は、それぞれ、方程式 (9)(10) と同 (11)(12) の時間発展によって得られたOFの推定値である。これらの差異を定量化するため、Fig.3に、与えた動きベクトル (\bar{u}, \bar{v}) と得られた推定値 (\hat{u}, \hat{v}) との間の平均二乗誤差

$$E = \frac{1}{2N^2} \sum_{i=1}^N \sum_{j=1}^N \{(\hat{u}_{ij} - \bar{u}_{ij})^2 + (\hat{v}_{ij} - \bar{v}_{ij})^2\} \quad (13)$$

の時間プロットを示す。実線は方程式 (9)(10)、破線は方程式 (9)(10) を用いた場合の時間発展で得られたものである。後者の方が計算量が多いにも拘わらず、第6フレームまでは誤差が大きい。誤りの割合が多くても、OFの検出箇所が多いからだと考えられる。しかし、第6フレーム以降は大小関係が逆転して、誤差が小さくなっている。

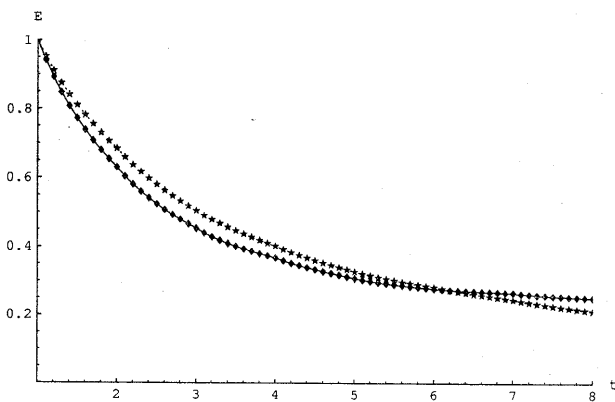


Fig. 3 Time plots of mean squared error E

5 運動ベクトル検出・推定のまとめ

本稿では、Neuralimage社製アナログビジョンチップ内臓カメラの使用を前提として、運動視知覚の問題について考察した。特にそれがDOGフィルタと輪郭検出等を並列的に高速に実現することに着目し、RDKをDOGフィルタにかけて空間的に微分可能な画像にした上で、オプティカルフローの検出を試みた。その際、ゼロ交差の情報を利用して信頼性の高いOFを得る方法として、得られた検出値を、拡散方程式により、画面全体に行き渡らせることで、画面全体の推定値を得ることを試みた。

ゼロ交差とその履歴及び現時点でのOFとの関係に基づいて、OFのはずれ値を削除することを試みたが、計算量や手間の割には、効果が出始めるのに時間がかかり、その意味は現時点ではあまり見出せなかった。しかし、今後、Fig.2のように、背景の上を別の小領域が移動するようなオクルージョンが生じるRDKへ問題を拡張する際には、この情報が図と地の分離の知覚に対する有用な手がかりになる可能性が考えられる⁵⁾。

このような図地分離の問題やや主観的輪郭の抽出等のさらに高次の認知過程を実現することが今後の課題である。一方で、上記ビジョンチップを用いて実際の問題に取り組むことも課題である。それについて、ターゲットトラッキングを例に挙げて考察してみたい。

6 ターゲットトラッキング

ターゲットトラッキングとは、Fig.5のように、カメラ画面内にある動く対象物体を、カメラの中心で捕らえるように、カメラをのせる回転台のパン角・チルト角を適切に制御することである。人間の場合、二つの眼球運動、即ち Saccade と Smooth Pursuit により、物体を追跡しており、それぞれ、位置推定と速度推定に基づいているとされている¹⁰⁾。以下、これらについて概説する。

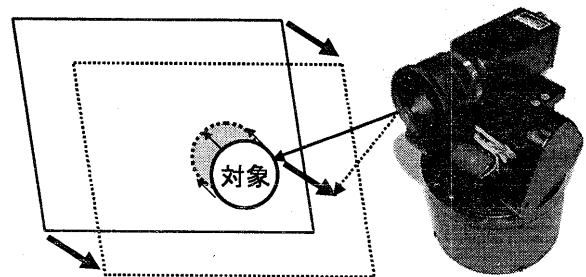
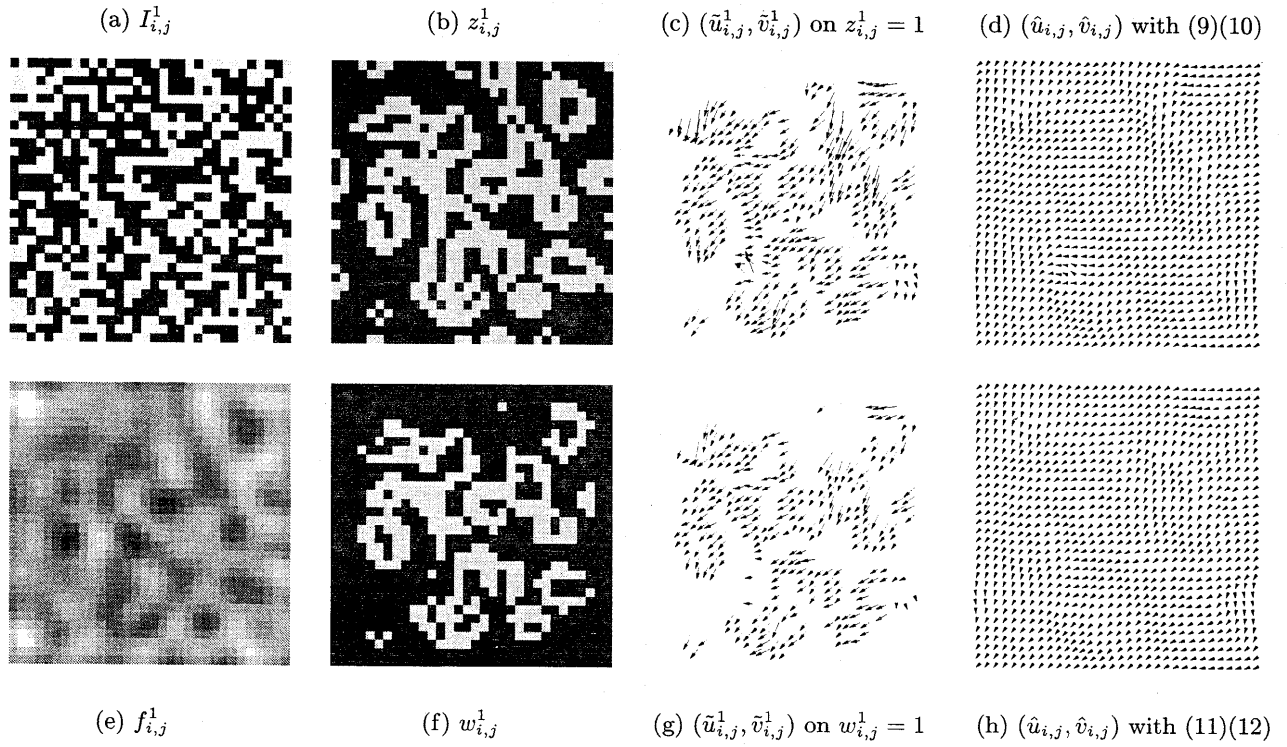
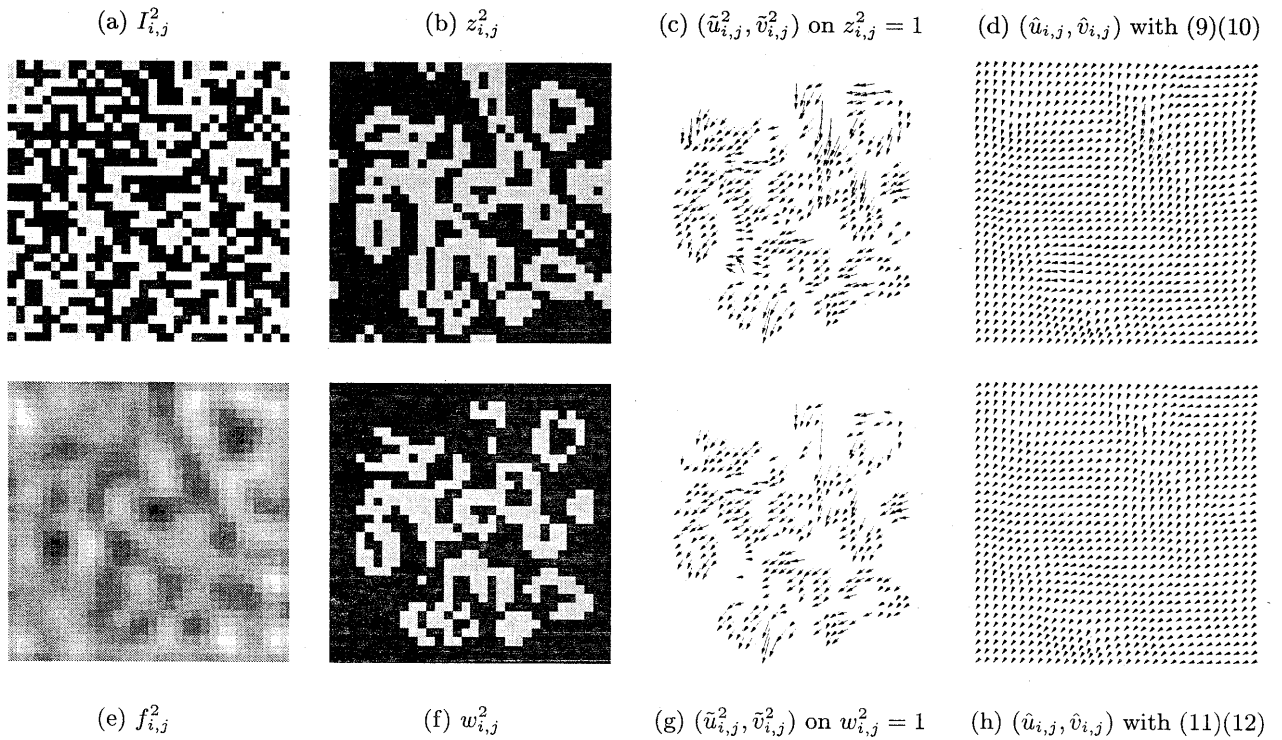


Fig. 5 Conceptual chart of target tracking



[1] A set of results of the first frame



[2] A set of results of the second frame

Fig. 4 Simulation result

6.1 移動物体の位置推定

移動物体の検出の第一歩は、フレーム間の時間差分をとることである。この場合、回転台が静止している間は物体の検出が可能だが（初期状態）、回転台が動き出すと、背景と物体共に時間差分が生じる（過渡状態）。回転台が物体を完全に捉えることができれば、物体部分には時間差分が生じなくなる（定常状態）（Fig.6）。

そこで、フレーム間差分だけでなく、エッジの勾配と合わせた統合画像から、その重心を求めて、物体位置を推定する方法が考えられる。両者の情報とも ASPECTUS から直接あるいは間接的だが効率的に抽出できる。

6.2 移動物体の速度推定

移動物体の速度は、式 (6) に示したように、時間差分と、 x, y 両方向の空間勾配を用いて、SLO法によって求めることができる。この情報の抽出にも ASPECTUS を利用することで、CPUの負荷を軽減できる。

6.3 位置と速度の統合

位置と速度を利用してカメラ回転台の角度・角速度を決定することは、眼球運動における Saccade と Smooth pursuit に対応する。また、人間の視覚系には、大細胞系と小細胞系とがあり、それぞれ空間差分と、時間差分に関係している。ASPECTUS の二つの出力チャンネルはこれらに対応しており、人間の眼球運動のような、滑らかで柔軟な物体追跡の実現に適していると考えられる。

7 おわりに

本稿では、アナログビジョンチップ内蔵カメラと、それを前提とした運動ベクトル場推定のシミュレーションを紹介した。また、そこから得られる情報を如何に利用して、ターゲットトラッキングを行うかについても考察した。現在、ハーモニックドライブ社でカメラ回転台が製作中である。回転角の制御法を考えつつ、実際の回転台を用いた実験で、従来法との比較を行うことが望まれる。



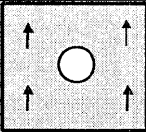
	初期状態	過渡状態	定常状態
対象	動	動	静
背景	静	動	動
			

Fig. 6 Initial, transient and stable state

謝辞

本研究は、平成 19 年度日本学術振興会科学研究補助金（若手研究 (B)）(# 18700190) の補助を受けて行われました。感謝致します。また、大阪大学臨床医工学融合研究センターの下ノ村和弘氏、株式会社 Neuralimage の高見涼太郎氏および同社の鶴殿直嗣氏から有益なご意見を頂きました。ここに感謝の意を表します。

参考文献

- [1] 土屋和雄, 中西健二: 不連続性を含む確率場のベイズ推定, 計測自動制御学会論文集, 28-4, pp.514-518 (1992).
- [2] デビッド・マー: ビジョン-視覚の計算理論と脳内表現-, 産業図書 (1987).
- [3] A. Namiki, Y. Nakabo, I. Ishii, and M. Ishikawa: Ims Sensory-Motor Fusion System, IEEE Trans. Mechatronics, 5, 3, 244/252 (2000)
- [4] S.Kameda and T.Yagi: An Analog VLSI Chip Emulating Sustained and Transient Response Channels of the Vertebrate Retina, IEEE Trans. Neural Networks, 14, 5, 1405/1412 (2003)
- [5] 上山英三, 湯浅秀男, 細江繁幸, 伊藤正美: 反応拡散方程式を用いた動きによる図地分離-形成されたパターンの界面と主観的輪郭-, 電子情報通信学会論文誌, J-81-D-II-12, pp.2767-2778 (1998).
- [6] 蘆田: 視覚運動検出モデルと最近の動向, VISION, 15, 1, 11/16 (2003)
- [7] 秋山, 羅, 大西, 八木, 細江: シリコン網膜を用いたオプティカルフローの局所計算方式, 電子情報通信学会論文誌, J87-D-II, 7, 1462/1472 (2004)
- [8] 上山英三, 八木哲也, 細江繁幸: アナログビジョンチップに基づく運動視知覚の一考察, 第 19 回自律分散システム・シンポジウム資料, pp.103-106, (2007).
- [9] 三池, 古賀編: パソコンによる動画像処理, 森北出版, (1993)
- [10] 竹内義則, 汪 増福, 大西 昇, 杉江 昇: 眼球の Saccade 運動と Pursuit 運動に示唆を得た移動物体の実時間追跡, 日本ロボット学会誌, 15, 3, 474/480 (1997)