

TR - A - 0090

019

運動視と両眼立体視
~ 統合モデルの構築をめざして ~

緒形 昌美

1990. 8. 28

ATR 視聴覚機構研究所

〒619-02 京都府相楽郡精華町乾谷 ☎07749-5-1411

ATR Auditory and Visual Perception Research Laboratories

Inuidani, Sanpeidani, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1411
Facsimile: +81-7749-5-1408
Telex: 5452-516 ATR J

目次

1. まえがき	1
2. 運動検出モデル	2
2. 1 速度候補の検出法	2
2. 1. 1 スカラ・センサ	2
2. 1. 2 局所運動検出の原理	3
2. 1. 3 速度候補の検出	4
2. 1. 4 アルゴリズム	5
2. 2 速度情報の統合	5
2. 2. 1 運動の一様性	6
2. 2. 2 運動一様領域の抽出	6
2. 2. 3 速度平面の統合	7
2. 3 特徴点マッチング	7
2. 4 計算機実験	8
2. 5 考察	9
3. 時間的統合を考慮した運動検出モデルの検討	11
4. 位相検出による両眼立体視モデル	12
4. 1 両眼立体視の原理	13
4. 2 両眼視差検出法	13
4. 2. 1 運動検出と視差検出	13
4. 2. 2 位相情報を用いた視差検出	14
4. 3 空間周波数チャンネルの統合	15
4. 4 計算機実験	15
4. 5 Marrのモデルとの比較	17
4. 6 考察	17
5. 位相差符号による両眼立体視モデル	18
5. 1 従来法の改良	18
5. 1. 1 位相差符号の意味	18
5. 1. 2 位相差符号の検出	19
5. 2 空間周波数チャンネルの統合	19
5. 3 計算機実験	20
5. 4 考察	21
6. 運動検出モデルと両眼立体視モデルの統合	23
6. 1 運動速度と両眼視差の関係	23
6. 2 統合モデルの概略	23
6. 3 運動速度と両眼視差の整合法	24
7. むすび	26
謝辞	26
参考文献	27
付録	29
図表	

1. まえがき

3次元世界の投影としてテレビカメラ等によって得られた2次元画像には、外界に関するさまざまな手がかりが含まれているが、中でも運動や両眼視差は情景のセグメンテーションや3次元形状の復元など、より高次の処理を行うために用いることができ、それらを検出する方法はコンピュータ・ビジョンの中心的課題としてさかんに研究されてきた。その結果いくつかの優れた手法が提案されたが、それらはあくまでもある制限された環境においてのみ正しく動作し得るもので、自然環境の中であらゆる場面に対処できる柔軟な手法はいまだ確立されていない。一方、網膜像に含まれる運動や両眼視差は、我々人間にとっても外界を把握するための重要な手がかりである。実際、我々の視覚系はあらゆる場面においてこれらを実に巧妙に計算しており、その能力はコンピュータ・ビジョンのレベルを遙かに凌いでいる。従って、人間の視覚メカニズムを検討し計算モデルを構築することは、今日コンピュータ・ビジョン研究が抱えている問題を解決するための1つの有効なアプローチと考えられる。そこで本報告書では、人間の視知覚特性を考慮した運動検出、及び両眼視差検出の計算モデルを提案する。

また、人間の視覚系が優れた情報処理能力を実現している理由の1つとして、複数の手がかり間の相互作用をうまく利用していることが考えられる。通常、すべての手がかりは互いに関係をもっているため、単一の手がかりが曖昧であったり精度が低くても、複数の手がかりの整合をとることによって個々の信頼性を改善することが可能である。しかし、従来の研究は各手がかりを独立に検出することを目的としており、各手がかり間の相互作用はほとんど考慮されていない。ここでは運動視と両眼立体視の統合モデルの検討も行う。

2. では、心理実験によって得られた結果をもとに、フィルタリング処理とマッチング処理を組み合わせた2段階の運動検出モデルを提案する。

3. では、運動検出において、時間的積分による推定精度の向上と速度変化に対する追従性を両立させるために、検出モデルへのカルマン・フィルタの導入を検討する。

4. では、両眼視差検出と運動検出の類似性に着目し、従来行われてきた特徴点マッチングとは異なる信号処理的な視差検出法を提案する。

5. では生体モデルとしての可能性を考慮し、また計算速度の高速化を目的に、4. で述べた方法を改良する。

6. では、運動検出と両眼視差検出の相互作用について検討する。

2. 運動検出モデル¹

人間は運動情報のみから、運動物体の2次元的な形状を正確に知覚できることが知られている²。このような精密な形状切り出しを行うためには、その基となる運動情報をかなり精度よく計算する必要がある、なんらかのマッチング的な処理が行われていることが想像される。マッチング法は、コンピュータ・ビジョンで運動検出のためによく用いられているもので、時系列的に得られた画像間の特徴点を正しく対応付けることで、精度の高い速度を計算することができる³⁻⁶。しかし、膨大な数の特徴点を含むような画像を扱う場合、それらを正しく対応付けることが容易でないことは、これまでの研究からも明かである。複雑な情景を容易に処理する我々の視覚系が、入力データに対して初めからマッチング法を用いているとは考えにくい。

これに対し近年、人間の運動視知覚モデルとして、時空間フィルタリングを用いた方法がさかんに研究されている⁷⁻¹⁶。これは、輝度の時空間的な変化を用いて速度を計算するもので、対応問題を解く必要はない。しかし、本質的にあまり高い精度を期待できず、この方法だけでは正確な形状を抽出するには十分の速度を計算できないと思われる。

ランダム・ドット・キネマトグラムを用いたSatoの心理実験は、人間が運動情報から運動領域の形状を知覚するためにはフィルタリング的な処理とマッチング的な処理が両方働いていることを示唆している¹⁷。そこで我々は、フィルタリング法、及びマッチング法の利点を組合せた2段階処理によって、速度を検出するモデルを提案する。まずフィルタリング処理で大まかな運動情報を抽出する。実際には、可能性のあるいくつかの速度が速度候補として検出され、次のマッチング処理に渡される。これによって、マッチング処理における対応候補の数が絞られ、その負荷が軽減される。マッチング処理では、各速度候補に対してマッチングスコアが計算され、最高得点を得た速度を真の速度として採用する。

2. 1では、時空間フィルタリングによって局所的な運動候補を検出する方法について述べ、2. 3. では、より信頼性の高い速度候補を得るために、これらを空間的、及び空間周波数チャンネルに渡って統合する方法を示す。2. 4では、この速度候補を用いてマッチングを行い、速度を計算する方法について述べる。2. 5では自然画像を用いた計算機実験によって本モデルの有効性を示す。

2. 1 速度候補の検出法

筆者らは、これまでにA.B. Watsonらの考案したスカラ・センサ¹⁸からの出力を速度平面上で統合し、運動速度を計算するモデルを提案している¹⁴⁻¹⁶。このモデルでは、他の運動検出モデルと同様に、各空間位置ごとにただ1つの速度を算出して割り当てていた。ここでは、可能性のある速度すべてを、求めるべき速度の候補として検出する方法を提案する。速度の最終的な決定は後のマッチング処理によって行われる。

2. 1. 1 スカラ・センサ

視覚から得られる入力情報は、Fig. 2.1(a)に示すような2次元画像の時系列と

考えられる。このような時空間データに対し、帯域通過形の時空間フィルタを施す。これをスカラ・センサと呼ぶ。スカラ・センサは2次元の実空間平面上及び空間周波数平面上において局所性を持っている。従って、空間位置 (x, y) 及び空間周波数 (f_x, f_y) に選択性を持つスカラ・センサは、Fig. 2.1(b)に示すように、自分の受持ちの局所領域において、 θ_s 方向に周波数 u_s で変化する正弦波状の格子縞パターンを見ていることになる。ここで、

$$u_s = (f_x^2 + f_y^2)^{1/2} \quad (2.1)$$

$$\theta_s = \tan^{-1}(f_y / f_x) \quad (2.2)$$

である。このことは、スカラ・センサが本質的にaperture problem (小さな穴から方向性の強いパターンの運動を見た場合、速度に曖昧性が生じる)¹⁸を抱えており、2次元の速度を一意に決定できないことを示している。しかし、このセンサは速度に関して拘束条件を与えることができる。一般に、Fig. 2.1(b)のような格子縞パターンが運動する場合、パターンの空間周波数 (u_s, θ_s) と速度 $(v$: 速さ、 θ : 向き)、及び運動によって生じる時間的輝度変化の周波数 f_t の間には、次の関係がある。

$$f_t = u_s \cdot v \cdot \cos(\theta - \theta_s) \quad (2.3)$$

u_s, θ_s は、スカラ・センサのパラメータで既知であるから、センサ出力に於ける時間周波数 f_t を検出し、(2.3)式に代入することで、速度の拘束条件が得られる。

ここで、 θ_s はもともとパターンの方向に関するパラメータであるから、その定義域は

$$-\pi/2 < \theta_s < \pi/2 \quad (2.4)$$

と考えられる。このとき、速度の向き θ が

$$\theta_s - \pi/2 < \theta < \theta_s + \pi/2 \quad (2.5)$$

を満たしていれば、(2.3)式の f_t は正の値をとり、それ以外の場合には負の値となる。しかし、実際に時間周波数として検出できるのはその絶対値であり、それを(2.3)式に代入しても正しい拘束条件が得られない。Watsonらは θ_s の定義域を

$$-\pi < \theta_s < \pi \quad (2.6)$$

とし、各スカラ・センサを、自分が選択性を示す θ_s に対して(2.3)式の f_t を負にするような向きの運動には時間変化の出力を出さないように構成している。つまり、スカラ・センサに運動の向きに対する選択性を持たせている。従って、出力が得られない場合には拘束条件をつくることができないが、 f_t が検出された場合にはそれを用いて正しい拘束条件を作ることができる。

2. 1. 2 局所運動検出の原理

選択性の異なる複数のスカラ・センサから得られた拘束条件を、速度平面上で統合し、2次元速度を求める方法について説明する。

いま、Fig. 2.2(a)に示すような平面を考える。2つの軸 v_x, v_y は、それぞれ2次元速度の水平方向及び垂直方向の成分を表わす。任意の速度はこの平面上の1点で表現できる。各点の原点からの距離は速さにあたり、 v_x 軸からの角度が運動の向きを示す。

いま、単一のスカラ・センサが与える速度の拘束条件(2.3)式を、速度平面上で

考えてみる。この式の両辺を空間周波数 u_s で割ると、

$$f_t / u_s = v \cdot \cos(\theta - \theta_s) \quad (2.7)$$

となり、その右辺は速度の θ_s の向きへの射影成分を表している。つまり、求めるべき速度は、Fig.2.2(a)に示すように、速度平面上において長さ f_t / u_s 、向き θ_s のベクトルの終点を通り、それと直交するような直線上のいずれかの点であることになる^{16,19}。このような直線の式は

$$v_y = - (1 / \tan \theta_s) \cdot v_x + (f_t / u_s) / \sin \theta_s \quad (2.8)$$

となるが、これは(2.7)式を変形することによっても得られる。

単一のスカラ・センサは、2.1.1で述べたようにaperture problemを抱えており、(2.8)式の直線上のどの点が真の速度であるかを決定することはできない。2次元の速度を求めるためには、向き θ_s の異なる複数のスカラ・センサからの情報を統合する必要がある。向きに対する選択性が異なれば、そこから得られる速度候補の直線は、速度平面上で異なった傾きを持ち、互いに交差する。各直線は、皆同じ速度の候補として得られたものであるからそれらは理論的には1点で交わるはずである。この交点が求めるべき速度となる(Fig.2.2(b))。

Adelsonらの心理実験は、このような向きの異なる複数のセンサの出力統合が、人間の運動検出機構においてもなされていることを示唆している²⁰。

Fig.2.3に本モデルの構成を示す。左に並んだ円がスカラ・センサを表し、その中の矢印は各々が選択性を示す運動の向きを表している。各スカラ・センサ出力に於ける時間周波数を検出し、(2.8)式の拘束条件式によって速度平面上に直線を描く。それらの交点として2次元速度を求める。

図中のスカラ・センサはすべて同じ空間周波数に選択性を持っており、全体として1つの空間周波数チャンネルを構成している。つまり、これによってある空間周波数成分の運動を検出することができる。空間フィルタの中心周波数 u_s を変えることで、異なるチャンネルを構成することができ、種々の周波数成分の運動が求められる。

2.1.3 速度候補の検出

空間周波数の成分に対する時間周波数 f_t を正確に検出できれば、原理的には2本の直線の交点として2次元速度を求めることができる。すなわち、 θ_s の異なる2つのスカラ・センサを用いれば十分である。しかし、本質的にフィルタリングによって、このような空間周波数と時間周波数を正確に対応することは不可能である。なぜならば、空間フィルタが帯域幅を持つために、センサ出力において複数の時間周波数が検出されるが、その中から u_s と対応する真の時間周波数を選ぶ手段がなく、どれを用いても誤差を生じる可能性があるからである^{15,16}。この問題の1つの解決法として、Fig.2.3に示すように複数のスカラ・センサを用い、各センサからの情報の整合をとることが考えられる。実際、従来モデルの多くはこの方法を用いている。しかし、各センサから得られる情報が誤差を含んでいる以上、どのような整合法を用いても、結果として正しい速度が得られる保証はない。そこで我々は、フィルタリング処理だけで速度を一意に決定することを放棄し、ここでは可能性のある速度候補を検出するにとどめる。最終的な決定は、より精

度の高いマッチング処理によって行う。以下に、速度候補の検出法を示す。

いま、あるスカラ・センサの出力における時間周波数分布が、ある広がりを持っていたとする。このとき、どの時間周波数も u_s と対応する可能性がある。従って、速度に関する拘束条件は、Fig. 2.2(a) に示すような直線ではなく、Fig. 2.4 のような広がりを持った帯状の領域となる。方向性の異なる複数のセンサによって得られた帯状領域の共通部分として速度候補を得ることができる。

2. 1. 4 アルゴリズム

ある空間周波数チャンネルにおいて、各空間位置 (i, j) での速度候補を求める方法を示す。速度平面のための配列 $VEL(v)$ を用意し、各要素を 0 で初期化しておく。ここで v は 2 次元速度 (v_x, v_y) を表す。あるスカラ・センサの出力に於ける時間周波数分布を調べ、しきい値以上のパワーを持つ周波数の最小値 f_{tmin} と最大値 f_{tmax} を検出する。この 2 つの周波数に対して速度平面上に (2.8) 式の直線を描く。これは、(2.8) 式を満たす速度 v に対して (2.9) 式のように得点を与えることで実現する。

$$VEL(v) = VEL(v) + 1 \quad (2.9)$$

さらに、これら 2 本の直線で囲まれた領域に存在する速度 v に対しても (2.9) 式によって得点を与える。同じチャンネル内の他のセンサについても、同様に速度平面に得点を加算していき、最も高い得点を得た速度の集合 $V_{i,j} = \{v_{ij0}, v_{ij1}, v_{ij2}, \dots\}$ を (i, j) における速度候補とする。

2. 2. 運動情報の統合

2. 1 では局所的な情報から各位置における速度候補を検出する方法について述べた。本章では、こうして得られた速度に関する情報を大局的に統合し、より信頼性の高い速度候補を求める方法を提案する。

統合法としては、次の 2 種類が考えられる。

1) 空間的統合

2) 空間周波数チャンネル間の統合

1) は、空間的に隣接する点における情報を統合するものである。画像上で隣接する点は同じ速度で運動している可能性が高く、それらを統合することで、より信頼性の高い運動情報が得られると考えられる。また、2) は空間的に同じ位置におけるいくつかの空間周波数チャンネルからの情報を統合するものである。一般に、ある空間位置に着目した場合、そこにはいろいろな空間周波数成分が混在しているが、通常それらの集合が 1 つの運動物体を表現しており、各成分は同じ速度を持っていると考えられる。従って、精度よく運動検出をおこなうためには対応する空間周波数チャンネルからの情報を統合することが望ましい。

このような統合を行う場合、どの範囲まで統合すべきかが問題となる。例えば、空間的統合の場合、できるだけ広い範囲の情報を統合したほうが信頼性は向上するが、異なる速度の領域までも統合してしまうと結果に悪い影響を及ぼす。チャンネル間の統合についても同様なことが言える。そこで、ここでは統合を行う前に、運動が一様である領域を抽出し、その領域内で情報の統合を行う。

なお、猫や猿の視覚皮質には、上に述べたような運動の一様領域や不連続点の

抽出を行なう機構が存在することが示されている²¹。

2. 2. 1 運動の一様性

我々は運動の一様領域を、“おおまかな運動の向きが一致してる領域”と定義する。これは実際問題としては、この時点で我々が利用できる運動情報が2. 1. 3で得られたおおまかなものであり、速さまで含めてもあまり意味を持たないと思われるからである。さらに、運動知覚の初期段階では、運動の向きのみが重要な役割を果たすと考えられており^{18, 22}、この段階では、運動の向きのみを考慮することで十分と思われる。ここでは、領域抽出に用いる、運動の向きに関する情報について説明する。

2. 1. 2で述べたように、各空間位置には、空間周波数、及び、運動の向きに対する選択性が異なる複数のスカラ・センサが用意されている。周波数選択性が等しいセンサの集まりは、その位置において1つの空間周波数チャンネルを構成する。ここでは、N個のチャンネルを用意し、各チャンネルは周波数 $u_s(n)$ に選択的であるとする($n=0, 1, 2, \dots, N-1$)。また、各チャンネルは、L個のスカラ・センサを持っており、各々 $\theta_s(l)$ の向きの運動に選択的であるとする($l=0, 1, 2, \dots, L-1$)。いま、各スカラ・センサを $S_{i,j,n}(l)$ で表し、それが反応しているか否かのみに着目する。すなわち、

$$S_{i,j,n}(l) = \begin{cases} 0 \cdots \text{無反応} \\ 1 \cdots \text{反応} \end{cases} \quad (2.10)$$

とする。ここで、 (i, j) は空間的な位置を示す。 (i, j) において、チャンネル n のスカラ・センサ l が反応するかどうかは、その $\theta_s(l)$ に対して、そこに含まれる運動の向き θ が(2.5)式を満たしているかどうか依存する。センサ l が反応していれば、運動の向きが(2.5)式の範囲であることがわかる。従って、向きの異なるセンサの反応を組み合わせることによって、Fig. 2.5に示すように運動の向きをある範囲に限定できる。つまり、これらのセンサ群の反応パターン

$$S(i, j, n) = \{ S_{i,j,n}(0), S_{i,j,n}(1), \dots, S_{i,j,n}(L-1) \} \quad (2.11)$$

は運動の向きを反映していることになる。この反応パターンを表すベクトル $S(i, j, n)$ に基づいて運動の一様性を調べる。

2. 2. 2 運動一様領域の抽出

$S(i, j, n)$ は、空間位置と空間周波数チャンネルの関数となっているが、この3次元空間の中で $S(i, j, n)$ による領域分割を行う(Fig. 2.6)。すなわち、同じ反応パターン $S(i, j, n)$ を持ち、この3次元空間において連結している点は同じ運動領域に含まれているとする。 $S(i, j, n)$ が同じでも連結していない点は別の領域とみなす。また、連結していても $S(i, j, n)$ が異なれば、やはり別の領域とする。

実際には、まず各点 (i, j, n) についてそれが、なんらかの運動領域に含まれているかどうかを判定する。これは、反応パターン $S(i, j, n)$ と、その点を囲む近

傍内の各点の反応パターン $S(i+\Delta i, j+\Delta j, n+\Delta n)$ を比較することによって行う。近傍内の反応パターンがすべて同じであれば、点 (i, j, n) はある運動領域に含まれているとみなして抽出する。近傍内に異なる反応パターンを持つ点がある場合には、 (i, j, n) は運動領域に含まれないとする。ここで、運動領域に含まれていない点とは、異なる運動領域の境界の点である。つぎに、抽出された点の連結性を調べる。連結している点の集合を、1つの運動領域 R_m ($m=0, 1, 2, \dots, M$) とする。

2. 2. 3 速度平面の統合

2. 1. 3では、各 (i, j, n) ごとに速度平面を作り、速度候補を検出した。しかし、2. 2. 2で得られた各領域 R_m 中の点はすべて同じ速度で運動していると考えられ、これらの情報を統合することによってより信頼性の高い運動が得られる。そこで、領域 R_k に含まれる (i, j, n) の速度平面上の情報をすべて1つの速度平面上にプロットしなおし、これによって得られた新しい速度平面から速度候補を検出する。すなわち、 (i, j, n) の速度平面を $VEL_{i,j,n}(v)$ とすると、(2.12)式によって新しい速度平面 $VEL'_m(v)$ をつくる。

$$VEL'_m(v) = \sum_{(i,j,n) \in R_m} VEL_{i,j,n}(v) \quad (2.12)$$

この $VEL'_m(v)$ で最も高い得点を得た速度を領域 R_m の速度候補の集合 $V_m = \{v_{m0}, v_{m1}, v_{m2}, \dots\}$ とする。

2. 3 特徴点マッチング

連続する時刻の2枚の特徴画像の間でマッチングを行うことによって、2. 2. 2で得られた各運動領域 R_m の各時刻における2次元形状を抽出するとともに、その領域を構成している各位置 (i, j) の速度を速度候補 V_m の中から決定する。

初めに、マッチング処理を施す特徴画像について説明する。いま、画像のある1つの特徴に着目する。これは、例えば、画素値の極大点や、極小点、または、エッジに相当するゼロ交差点¹⁸などが考えられる。特徴画像 $F_{n1}(i, j, k)$ は、入力データが空間周波数チャンネル n を通過した後の時刻 k における位置 (i, j) に、この特徴が存在するか否かを表すもので、(2.13)式のような2値画像の時系列である。

$$F_{n1}(i, j, k) = \begin{cases} 1 & \dots \text{特徴点} \\ 0 & \dots \text{その他} \end{cases} \quad (2.13)$$

ここで、1は着目する特徴の種類を表す。異なるいくつかの特徴に対してこのような特徴画像 F_0, F_1, F_2, \dots を生成する。

この特徴画像の上で次のようにマッチングを行う。いま、速度候補 V_m の中の1つの速度を $v = (v_x, v_y)$ とする。ある時刻 k においてチャンネル n の位置 (i, j) に存在する点 A が、この速度 v で運動していたとすると、この点は次の時刻 $k+1$ には Fig. 2.7 に示すように $(i+v_x \cdot \Delta t, j+v_y \cdot \Delta t)$ に存在するはずである (Δt は時刻 k と $k+1$ の時間間隔)。そこで、特徴画像 F_{n1} において、2つの時空間位置 (i, j, k) と $(i+v_x \cdot \Delta t, j+v_y \cdot \Delta t, k+1)$ における特徴点の有無が一致するかどうかを調べる。もし一致していれば、 (i, j, k) は速度 v で運動していた可能

性がある。用意した各特徴画像について (i, j, k) における速度 v の可能性を調べ、(2.14)式によって速度 v の (i, j, k) における得点を計算する。ただし、領域 R_m がどの空間周波数チャンネルにまたがっているかは既知であるから、その特徴画像に対してのみ加算を行う。

$$MS_{ijk}(v) = (\sum \text{XOR}(F_{ni}(i, j, k), F_{ni}(i+v_x \cdot \Delta t, j+v_y \cdot \Delta t, k+1))) / \text{num}(m) \quad (2.14)$$

ここで、 Σ は領域 R_m がまたがっているチャンネルでの加算を表し、 $\text{num}(m)$ はそのチャンネル数を示す。また、 $\text{XOR}(a, b)$ は a と b の排他論理積を示す。 V_m 中のすべての速度に対して $MS_{ijk}(v)$ を計算しそれらの中で最高得点求める。その得点があるしきい値をこえる場合には、時刻 k の位置 (i, j) は運動領域 R_k に含まれているとして抽出し、最高得点を与えた速度 v_{max} をその点の速度とする。抽出した点の集合は、運動領域 R_m の 2次元形状を構成する。種々の V_m に対して同様にマッチングを行うことにより、各時刻の各位置に対して速度を割り付けることができる。

2. 4. 計算機実験

本モデルの有効性を確かめるために、テレビカメラから取り入れた実データを用いて計算機実験を行った。ここで用いる画像の大きさは、 $256\text{pixel} \times 256\text{pixel}$ であり、時間方向には 16frame を用いた。モデルは、3つの空間周波数チャンネルから成っており、各チャンネルの中心周波数は、それぞれ 0.25cycle/pixel , 0.125cycle/pixel , 0.0625cycle/pixel とした。また、周波数に関する通過帯域幅(半値幅)は、心理物理や電気生理の実験によって得られている知見を考慮し、1オクターブとした。さらに、各空間周波数チャンネルにおけるスカラ・センサの数を10とし、それぞれ $0^\circ, 36^\circ, 72^\circ, \dots, 324^\circ$ の向きの運動に選択性を持たせた。これは、周波数に関する帯域幅が1オクターブである場合、各センサの運動の向きに関する帯域幅がほぼ 38° となるため、全ての向きをカバーするためには10個のセンサが必要となるからである。速度平面の軸の単位は、 0.125pixel/frame とした。マッチングに用いる特徴としては、方向性、空間周波数、位相が異なる30種類のゼロ交差点を用いた。

Fig. 2.8に用いた時系列画像の第1frameを示す。画像中の手は時間とともに、ほぼ真下に移動している。速さは、約 0.9frame/pixel である。周囲のテクスチャは静止している。

Fig. 2.9にフィルタリング処理のみによって得られた速度場を示す。Fig. 2.9(a)-(c)は、単一の空間周波数チャンネルの結果を示す。これらは2. で述べた方法を次のように修正して計算した。すなわち、速度は、各位置で得られた速度平面において、最大得点を得た速度集合の重心として求めた。また、Fig. 2.9(d)は、3つのチャンネルの出力を統合して得られた結果を示す。このチャンネルの統合は、同じ位置における異なるチャンネルによって得られた速度平面を重ね合わせることによって実現した。手首のあたりで速度が与えられていないのは、そこでのコントラストが低いためスカラ・センサが応答できず、静止しているとみなされたためである。

Fig. 2.10に2段階モデルによって得られた速度場を示す。フィルタリングのみによって得られた結果に比べ、位置によるばらつきが殆どなく、高い精度の推定値（下向き、速さ0.75pixels/frame）が得られている。また、手首のあたりで速度が得られているのは、このあたりでもゼロ交差点が検出され、正しくマッチングが行われたためと考えられる。Fig. 2.11に本モデルで抽出された運動一様領域を示す。各画像は、Fig. 2.6に示した3次元空間の各空間周波数におけるスライスを示している。各チャンネルで抽出された領域（白点で示した位置）は、周波数方向に連結しており、全体として1つの（手に対応した）運動領域を構成している。この運動領域における統合によって生成された速度平面をFig. 2.12に示す。

ここで、Fig. 2.11は、フィルタリング処理のみによる運動領域の切り出し（セグメンテーション）の結果とみなすことができる。各画像は、単一チャンネルによって得られたもので、高い空間周波数チャンネルではノイズが多く、低い周波数チャンネルでは分解能が悪い。いずれのチャンネルも手の指などは識別不可能である。これに対し、2段階モデルによるセグメンテーションの結果をFig. 2.13に示す。Fig. 2.13(a)は、速度候補集合の中の速度が割り付けられた位置を示す。Fig. 2.13(b)は、(a)に平滑化（孤立点除去）を施した結果である。Fig. 2.11に比べてはっきりとした手の形状が得られている。

次に、複数の運動物体を含む、より一般的な自然の情景に対して本モデルを適当した。1フレームをFig. 2.14に示す。2本の紅葉の枝がそれぞれ右（上の枝）または左（下の枝）向きに運動している（背景の葉は静止している）。16フレーム間の各枝の運動は、平行運動だけでなく振動成分や回転成分など、時間フィルタリングを用いる手法にとってはかなり雑音要因を含んでいる。また、Fig. 2.8に示した例では、運動情報を用いなくても、輝度エッジだけで手を分離できるが、この例では、各枝の動きだけがそれらを分離する手がかりとなっている。

このデータに対して、本モデルは2つの運動一様領域を抽出した。Fig. 2.15(a)、及び(b)に各領域から得られた速度場を示す。それぞれ、上の枝、下の枝の運動をよく表している。Fig. 2.16にセグメンテーションの結果を示す。上下の枝がよく分離され、それぞれに属する葉の形状も識別可能な程度に抽出されている（下方の枝や葉に隠されている上方の枝または葉の一部は抽出されていない）。

2. 5 考察

従来、運動検出法として別々に用いられてきた特徴マッチング法とフィルタリング法を組合せ、正確な運動速度を検出する方法を提案した。本来マッチング法が抱える対応問題の負荷を、フィルタリング法によって軽減する。これによって、正しい対応付けが可能となり、フィルタリング法の持つ精度の問題を克服できる。人間の運動視知覚において、このような2つの方法が併用されていることは、心理実験によって示唆されているが、計算機実験によって、このアプローチが実際に有効であることが示された。

最後に、本モデルの限界について述べておく。本モデルでは、速さだけが異なる運動領域が隣接している場合には、これらを分割することができない。これは、運動の向きのみで求めた運動領域 R_m で各画素をラベル付けするためである。また、異なる場所における領域 R_m と R_n が同じ速度 v で運動している場合、各領域の速

度候補の集合 V_m 、 V_n はどちらも速度 v を含んでおり、この v でマッチングを行った場合には、どちらの領域の中の画素も高いマッチングスコアを得て抽出されてしまう。すなわち、領域 R_m と R_n を区別することができない。第一の問題を解決するためには、運動一様領域を抽出する際に、おおまかな速さの情報も考慮する必要がある。また、第二の問題は、得られた全ての速度候補集合を、全ての位置に対して適用することによって生じるものであり、これを解決する1つの方法としては、運動一様領域が持つ位置の情報を用いて、ある速度候補集合を適用する範囲を制限することが考えられる。

3. 時間的統合を考慮した 運動検出モデルの検討

前章で述べた我々のモデルも含め、従来の運動検出モデルのいくつかは、局所的に得られた推定速度の精度を向上させるために、実空間上や空間周波数領域上で情報の統合（空間的統合）を行っているが、時間的統合についてはあまり研究されていない。しかし、運動は時空間において生じるものであり、空間だけでなく時間的にも情報を統合することで、より精度の高い推定値が得られることが期待される。ここでは、運動情報の時間的統合について検討する。

フィルタリング処理による運動検出法では、用いた時間フィルタの持続時間の間にサンプルされたフレームの情報が統合され、その間の平均的な速度が計算される。従って、速度が一定であれば積分効果によってより精度の高い推定値を得ることができるが、持続時間の中で速度に変化が生じた場合には、精度は低下してしまう。速度の変化に追従するためには、持続時間の短い時間フィルタを用いればよいが、この場合積分効果は減少してしまう。一方、マッチング処理による運動検出法では、基本的に2フレームから速度を計算するので、速度変化に対する追従性は優れているが、時間的積分効果は期待できない。2つの処理を組み合わせた我々のモデルでは、積分効果と速度変化に対する追従性をある程度両立していると思われる。すなわち、ある程度速度変化が小さければ、時間フィルタの持続時間中に存在した速度全てがフィルタリング段階において検出され、マッチング処理によって、その中から正しい速度を各フレームに割り付けることができる。しかし、速度変化が大きい場合にはフィルタリングによる速度候補の計算自体がうまく行われなくなってしまう。

我々は、時間的統合における積分効果と速度変化に対する追従性のトレード・オフを克服するために、運動検出モデルにカルマン・フィルタ²³を導入することを検討している。カルマン・フィルタは、ノイズに汚された観測値から2乗誤差を最小にする推定値を逐次的に求める最適フィルタである。時々刻々得られる観測値に対して、ある時刻の推定値は、その時刻において観測された値と前の時刻における推定値から決定される。

Matties²⁴らやHeel²⁵らは、このカルマン・フィルタを時系列画像から奥行き復元を行うために用いているが、我々はこれを運動検出モデルに導入することを検討している。Fig.3.1にその構成図を示す。各円はスカラ・センサのような運動方向に選択性を持つ時空間フィルタであり、それぞれ、2次元速度のある向きへの射影成分を出力する。このとき速度変化に対する追従性をよくするために、できるだけ少ないフレームによって計算を行う。時刻 k におけるセンサ n による観測値 v_n^k と、時刻 $k-1$ において得られた対応する射影成分の推定値 v_n^{k-1} からカルマン・フィルタによって時刻 k の推定値 v_n^k を求める。各センサから得られた v_n^k ($n=0, 1, 2, \dots$)の整合をとることによって、2次元速度の推定値 v^k を計算する。 v^k は遅延されたあと、各センサが選択性を示す方向に分解され、次の時刻における射影成分の推定に用いられる。この方法によれば、速度変化に対して高い追従性を保ちながら、再帰的な処理による積分効果が期待できる。

4. 位相検出による両眼立体視モデル²⁶

外界に存在する物体を見た場合、その像は左右の網膜上において、若干異なる位置に投影される。この位置の差（両眼視差）は、観測者と対象物体との距離によって生ずるもので、我々が奥行きを知覚するための重要な手がかりとなっている。

生体が視差を検出するメカニズムについては、視覚心理学のテーマとして古くから研究されてきた。また、このような両眼による立体視は、ロボット等への応用が期待され、工学の分野でも多くの研究がなされてきた。こうして、今日迄に様々な視差検出アルゴリズムが提案されてきたが、視差検出を特徴点の対応問題（特徴点マッチング）に帰着させているという点では、どれも共通している。すなわち、あらかじめ左右の画像からなんらかの特徴点（局所的な空間パターンの場合もある）を抽出しておき、一方の画像の特徴点が、他方のどの特徴点に対応するかを見つけ出すことで視差を検出している。従って、これらの研究の中心課題は、用いる特徴点の選択と、それらの対応付けの方法を開発することにあった。

このようなマッチング的な手法には、計算量の問題がつきまとう。なぜならば、ある特徴点に対応する点を決定するためには、複数の候補点に対してなんらかの評価関数を計算し、さらにそれらを比較することによって最適値を与えるものを捜さなければならないからである。また、視差が、特徴点の存在する位置でしか得られないことも大きな欠点となっている。

ところで、複数の画像から位置のずれを求めるという意味では、時系列画像に含まれる運動速度を求める処理は、両眼視差の検出と非常に類似している。しかし、これまでに提案された視差検出の方法が全て、特徴点マッチングに基づいているのに対し、運動検出法としては、その他に時空間フィルタリングを用いた信号処理的な方法が提案されている。この方法によれば、画像上の全ての位置において速度を求めることができ、またフィルタリングを前処理とみなせば、各位置での速度はその位置の画素値だけから計算できる。このようなフィルタリングによる方法は、主に生体の運動検出モデルとして提案されているが、もし運動速度が生体においてこのようなメカニズムで検出されているならば、類似した処理である視差検出においても同様な方法が取られていることは十分に考えられる。

そこで、ここでは特徴マッチングによらない信号処理的な両眼視差検出法を提案する。本方法では、視差は左右の画像の位相差として検出される。各画像の位相は、位相特性の異なる2種類の空間帯域通過型フィルタからの出力の比によって計算される。フィルタの中心周波数は、検出可能な最大視差を制限するが、これは空間分解能とトレードオフの関係にある。この問題は複数の空間周波数チャンネルを統合することによって解決する。

4. 1では、両眼立体視の原理について簡単に触れ、4. 2で視差を検出する方法について述べる。また、4. 3では、複数の空間周波数チャンネルの統合法を説明する。4. 4では、ランダム・ドット・ステレオグラムを用いたシミュレーションの結果を示す。4. 5では、本方法と、従来法の中で最も有力視されているMarrのモデルとの比較を行う。

4. 1 両眼立体視の原理

Fig.4.1に両眼立体視の原理を示す。 f_l 、 f_r は左右の網膜の中心窩で、 p_1 が注視点となる。角度 θ_l 、 θ_r はそれぞれ、各網膜上における、点 p_2 の像の中心窩からのずれを表し、その差 ($\theta_l - \theta_r$)が p_2 の両眼視差となる。両眼間隔を a 、注視点までの距離を D 、注視点から p_2 までの奥行き方向の距離を ΔD とすると、両眼視差は幾何学的な関係によって次のように表される。

$$\begin{aligned}\delta &= \theta_l - \theta_r = \alpha_1 - \alpha_2 \\ &\approx (a/D) - (a/(D+\Delta D)) \\ &\approx a\Delta D/D^2\end{aligned}\quad (4. 1)$$

このように、視差は注視点に対する相対的な奥行き情報を表している。なんらかの方法（例えば、生体では輻輳など）によって注視点までの距離 D を知ることができれば、対象までの絶対的な距離を計算することができる。

4. 2 両眼視差検出法

運動検出と視差検出の類似性については既に述べたが、これらは全く同じではなく、従って運動検出のために提案されたモデルをそのまま視差検出に用いることはできない。ここでは、両者を比較することにより、視差検出が抱える固有の問題を指摘し、それらを克服する新しい手法を提案する。

4. 2. 1 運動検出と視差検出

運動検出と視差検出の大きな違いとしては次のようなものが考えられる。

- 1) 求めるべき”ずれ”の次元
- 2) 計算に用いることのできる画像の数
- 3) 観測点の数

ここでは、1)についてはあまり考慮する必要がない。なぜならば、運動検出が2次元のずれ（速度）を対象とするのに対し、視差検出では水平方向の1次元のずれを求めればよいからである。この意味では、従来の運動検出法は全て、視差の検出に適用可能である。

一方、2)に関しては、視差検出固有の問題が生じる。すなわち、運動検出では時系列的に得られる3枚以上の画像を利用できるのに対して、視差は左右2枚の画像から求めなければならない。実際、時間波形の相関^{9, 10}や、時間周波数¹¹を利用した運動検出モデルは、10~20枚程度の画像を用い、時間的な積分によってその計算精度を上げているが、2枚の画像では正しい速度を計算することはできず、それらを視差検出に直接適用することは困難である。

3)に関しては、工学的に応用する際に特に問題となる。例えば、両眼による視差検出では2つの光学系を必要とするが、それらの特性が異なっていると、左右の画像の平均輝度やコントラストが変化してしまう。また、2つのカメラの位置を正確に設定しないと、各画像の間に奥行きとは関係のない上下方向のずれなどが生じてしまう。視差検出にとっては、左右画像の差は奥行きによってのみ生じていることが望ましく、上記のようなものはすべてノイズと考えられる。勿論、これらのノイズは、各光学系の特性をそろえ、それらを正確に設置することで回避できるが、視差検出のアルゴリズムとしても、これらのノイズに対して柔軟性

を持つべきである。人間の視覚系は、左右の網膜像にこれらの変化がある程度存在しても、正確に視差を検出できることが知られている^{18, 27, 28}。運動検出は、1つの眼だけから取り込まれた画像を用いるために、このようなことはあまり問題にならず、モデル構築の際にも考慮されていない。特にMarr-Ullmanモデル¹⁸をはじめとする勾配法に基づいた運動検出法は、コントラストの変化に弱く、視差検出には向かない。

以上のことより、視差検出のためには2枚の画像からずれを検出でき、画像間の平均輝度やコントラストの変化、さらにカメラの回転などによる上下方向のずれにあまり影響されないような方法を考える必要がある。この意味で、輝度変化の位相差を用いた吉田らの運動検出法²⁹は有望である。位相差は2枚の画像から計算でき、また平均輝度やコントラストに影響されない。そこで我々は、両眼画像の位相差から視差を求める方法を考える。

4. 2. 2 位相情報を用いた視差検出

左右画像における輝度変化の位相差を求めるために、まず各画像の位相を求める。ただし、これは局所的な輝度変化から求められなければならない。なぜならば、視差は位置によって変化するからである。このために、各画像にそれぞれ次のような2種類の空間フィルタを施す。

$$g_c(x, y) = a_c(x)b(y), \quad (4. 2)$$

$$g_s(x, y) = a_s(x)b(y). \quad (4. 3)$$

ここで、

$$a_c(x) = \exp[-(x/\sigma)^2] \cos(2\pi u_s x), \quad (4. 4)$$

$$a_s(x) = \exp[-(x/\sigma)^2] \sin(2\pi u_s x), \quad (4. 5)$$

$$b(y) = \exp[-(y/\sigma)^2]. \quad (4. 6)$$

(4.2)、(4.3)式は2次元のガボアフィルタと呼ばれ、生体の視覚皮質における単純細胞の受容野関数の近似としてよく用いられており、Fig. 4.2に示すような形をしている。x方向（水平方向）はガウス関数とコサインまたはサイン関数との積であり、y方向はガウス関数である。これらは、x、y両方向のガウス関数によって、その広がり σ に応じた空間的な局所性を与えられている。また、周波数面上でも局所性を持っており、周波数 $(u_s, 0)$ を中心とした帯域通過型フィルタになっている。従って、これらのフィルタは、局所領域において、水平方向に周波数 u_s で変化する成分を抽出する。

ところで、ある位置 (x, y) におけるフィルタ g_c 、 g_s の出力値 $o_c(x, y)$ 、 $o_s(x, y)$ を考えると、それらは (x, y) を中心とした局所的な輝度変化の、関数 $\cos(2\pi u_s x)$ と $\sin(2\pi u_s x)$ への射影成分と考えることができる。従って、それらの比から、次のように周波数 u_s 成分の位相を計算することができる。

$$p(x, y) = \tan^{-1}(o_s(x, y)/o_c(x, y)) \quad (4. 7)$$

左右の画像に対してこのように位相 $p_l(x, y)$ 、 $p_r(x, y)$ を求め、その差

$$dp(x, y) = p_l(x, y) - p_r(x, y) \quad (4. 8)$$

から視差を検出する。すなわち、

$$d(x, y) = (dp(x, y)/2\pi) \lambda \quad (4. 9)$$

ここで、 λ はフィルタの中心周波数に対応する波長 ($\lambda = 1/u_s$)、 $d(x, y)$ は位置(

x, y)における視差である。本手法の構成をFig. 4.3に示す。

本方法が、平均輝度やコントラストの変化に影響されないことは既に述べたが、 y 方向に広がりを持ったフィルタを用いることによって、ある程度上下方向のずれがある場合にも対処できると思われる。

4. 3 空間周波数チャンネルの統合

4. 2. 2では1組のガボア・フィルタを用いて視差を検出する方法を示したが、このとき検出できる視差の範囲はその中心周波数 u_s によって制限される。これは、位相差として検出できる範囲が

$$-\pi < p_l - p_r < \pi \quad (4. 10)$$

であるため、これに対応する視差範囲は、

$$-\lambda/2 < d(x, y) < \lambda/2 \quad (4. 11)$$

となる。大きな視差を検出するためには、中心周波数の小さいフィルタを用いる必要があるが、このようなフィルタは空間分解能が悪く、奥行きが不連続な部分では、その近傍の広い範囲で正しい視差を計算することができなくなる。

ここでは、中心周波数が異なる複数のフィルタ対（空間周波数チャンネル）を用い、各チャンネルから得られる視差情報を統合することによって、広い視差範囲と高い空間分解能を両立する方法について説明する。

Fig. 4.4に処理の流れを示す。なお、各チャンネルは、中心周波数が小さいものから順にCHANNEL 1, CHANNEL 2, ...と表すことにする。また、視差マップの初期値を $d_0(x, y) = 0$ として与えておく。CHANNEL k では、前のチャンネルで得られた視差マップ $d_{k-1}(x, y)$ と、このチャンネルでフィルタリングされた左右画像から新しい視差マップ $d_k(x, y)$ を作成する。このために、まずフィルタリングされた各画像の位相を、(4.7)式によって求め、それらの位相差を次式によって計算する。

$$dp' = p_l(x, y) - p_r(x + d_{k-1}(x, y), y) \quad (4. 12)$$

この位相差は、前のチャンネルで得られた視差の誤差によって生じるものであり、これを(4.9)式に代入して求まる値 $d'(x, y)$ によって、 $d_{k-1}(x, y)$ を補正することができる。すなわち、

$$d_k(x, y) = d_{k-1}(x, y) + d'(x, y) \quad (4. 13)$$

こうして得られた新しい視差マップを次のチャンネルに渡す。この処理を、低空間周波数チャンネルから順に行っていく。

各チャンネルは、前チャンネルまでに得られた視差値の誤差を求めればよいので、その誤差の大きさが自分の受持ち範囲に入っていれば、正しく働くことができる。これによって、大きな視差に対しても、中心周波数の高いチャンネルが有効になり、高い分解能を維持できる。

このように、誤差分をより高いチャンネルで求める方法は、多くの運動検出法で採用されている²⁹。また視差検出のためのMarrの第2アルゴリズムでは¹⁸、眼球運動と対応させて同様な方法が用いられている。

4. 4 計算機実験

本手法の有効性を確かめるために、ランダム・ドット・ステレオグラムを用い

てシミュレーションを行った。Fig. 4.5に入力データの例を示す。各画像の大きさは256pixel×256pixelである。左画像は50%のランダム・ドット・パターンであり、右画像は左画像の中央128pixel×128pixelの正方領域（ターゲット）を水平方向にd pixelずらして作られている。つまり、このずれの量がターゲットの視差となり、その他の部分（ターゲットの周辺）は視差0である。

視差検出には、3つの空間周波数チャンネルを用いた。各チャンネルの中心周波数 u_s はそれぞれ、0.0625, 0.125, 0.25cycle/pixelである。よって、検出可能な視差範囲はそれぞれ、8, 4, 2pixel未満になる。また各チャンネルの広がり σ は0.795/ u_s とした。これによって、それらの周波数領域上における広がり1オクターブになっている。以下では、各チャンネルを中心周波数の低い順にch1, ch2, ch3と呼ぶことにする。

Fig. 4.6に、d=1のときの結果を示す。これらは、各空間周波数チャンネルが独立に働いた時の結果であり、上は視差0として検出された画素を、下は視差1として検出された画素を黒点で示してある。また、Fig. 4.7に3チャンネルを統合した場合の結果を示す。最終的な結果は一番右に示されたもので、左の2つは低周波数チャンネルにおける途中経過である。視差がいずれのチャンネルでも検出可能なので、最も高い周波数チャンネルch3で良好な結果が得られているが、チャンネルの統合によって、より正確に視差が求まっている。これは低周波数チャンネルによる空間的な積分効果が有効に高周波数チャンネルに伝搬されたためと考えられる。

Fig. 4.8に、d=3に対する各単一チャンネルの結果を示す。ただし、視差0の周辺領域では殆ど正しい値が得られているので、ここでは、視差3として検出された画素だけを示した。d=3は、ch3の検出可能範囲を超えており、正しい視差が得られない。従って、1つのチャンネルで最も分解能よく視差を求めるためには、ch2を用いなければならない。しかし、3チャンネルの統合によって得られたFig. 4.9の結果は明らかにFig. 4.8(b)よりも優れている。これは、ch1やch2によって得られた視差の誤差がch3の検出範囲内にあり、このチャンネルも有効に働いたことを示している。ところで、ch3にとって、d=3は位相差にして $3\pi/2$ にあたるが、これは $-\pi/2$ として検出されるので、結果として-1の視差が得られるはずである。Fig. 4.10に、ch1によって視差が-1と見なされた画素を示す。

次に、コントラスト変化が本手法に与える影響を調べてみた。なお、ここではd=3とした。また平均輝度は一定にした。Fig. 4.11に結果を示す。横軸は、左画像に対する右画像のコントラストを表す。縦軸は正答率である。正答率は、結果として得られた視差と、はじめに与えたdが一致している画素数を全画素数で割ることによって求めた。コントラストの広い範囲でほぼ一定の正答率が得られている。

本手法は、原理的にはコントラストの影響を受けないはずであるが、結果は0.2あたりから正答率が低下し始める。これは、ガボア・フィルタの周波数応答がガウス形であることによる。すなわち、各ガボア・フィルタは、完全に直流分を遮断できないために、平均輝度に対してコントラストが小さくなると、その平均輝度の影響が大きくなり、正しい位相を求めることができなくなる。この影響は、中心周波数の低いチャンネルほど大きくなる。

この問題は、完全に直流成分を遮断するようなフィルタを用いることによって

解決できるが、このようなコントラストによる視差検出の性能低下が生体においても生じるかどうかは興味深い。

4. 5 Marrのモデルとの関係

従来の視差検出モデルとして、最も有力なもの1つにMarrの第2アルゴリズム¹⁸がある。これは、左右の画像にラプラシアンGフィルタをかけ、その出力において検出されたゼロ交差点を用いて特徴点マッチングを行う。対応点探索を、用いたフィルタの広がりに応じた範囲だけを行うことによって、誤対応を回避している。これは、我々の方法と同様、視差の検出範囲と空間分解能の間のトレード・オフを引き起こす。なぜならば、大きなフィルタは大きな視差を検出できるが、低い空間周波数に選択性を持つために分解能が悪くなるからである。彼は、大きなフィルタ（低周波数チャンネル）で得られた対応点をより小さなフィルタ（高周波数チャンネル）における探索開始点として用いることでこの問題を解決しているが、これは低周波数チャンネルで得られた視差だけ離れた位置の位相差を計算する我々のチャンネル統合法と同様の考え方による。

しかしながら、Marrのモデルが我々のものと本質的に異なるのは、各チャンネルにおけるずれの検出法である。すなわち、特徴点マッチングに基づく彼の方法は特徴点が存在する場所でのみ、ずれを検出できる。これに対して、輝度変化の位相差を用いた我々の方法では、原理的にすべての位置でずれが検出可能である。

4. 6 考察

両眼視差検出のために、特徴点マッチングによらない信号処理的な方法を提案した。視差は、両眼画像の輝度変化の位相差として検出し、また、“検出可能な視差範囲”と“空間分解能”という相反する問題を、複数の空間周波数チャンネルを統合することによって解決した。ランダム・ドット・パターンを用いたシミュレーションでは、本方法がコントラストの変化に強いこと、複数チャンネルの統合が有効に働くこと確かめた。

上下方向のずれに対する効果を調べることは今後の課題である。また、ここでは水平方向の視差を検出するために、垂直方向に選択性を持つ空間フィルタのみを用いたが、他の方向性を持つフィルタの利用についても、検討する必要がある。

このような信号処理的な視差検出が、生体によって行われているかどうかはさだかでないが、心理実験などを通してその可能性を調べることは興味深いと思われる。

5. 位相差符号による 両眼立体視モデル³¹

前章では、画像の輝度変化の局所的な位相情報を用いた視差検出法を提案した。また同様の視差検出法はJopson and Jenkin³²やSanger³³によっても提案されている。これらは、特徴点マッチング法が抱える対応問題を回避でき、画像のいたるところで視差を求めることができる。さらに、平均輝度やコントラストの変化に影響されにくいという利点を持っている。しかし、この方法は生体のモデルとしては妥当性に欠ける。なぜならば、各画像の位相を求めるためには \tan^{-1} の計算が必要となるがこのような関数が生体中で計算されているとは考えにくいからである。また、実際に計算機上にインプリメントする場合、関数 \tan^{-1} の計算には四則演算等と比べて多くの時間を必要とし、これも従来法の欠点といえる。ここでは従来法の視差検出法を改良し、左右各画像の位相を計算することなく、両画像の位相差の符号(-、0、+)のみから視差を計算する方法を提案する。本方法では、各空間周波数チャンネルは、求めるべき視差の取り得る範囲を限定するだけであるが、複数のチャンネルで得られた視差範囲を考慮することによって、最終的に視差を求める。

5. 1では、これを改良して、位相差の符号を検出する方法について説明する。また5. 2では、複数の空間周波数チャンネルで得られた位相差符号を、低周波数チャンネルから高周波数チャンネルへ伝搬することによって視差を求める方法を示す。5. 3章ではランダム・ドット・ステレオグラムを用いたシミュレーションによって、本方法と我々の従来法との比較を行う。

5. 1 従来法の改良

4. で述べた視差検出法は、特徴マッチング法に較べて処理が簡単であり、また平均輝度やコントラスト変化等に強いという利点を有するが、位相を求めるために \tan^{-1} の計算が必要となり、生体モデルとしての妥当性や計算コストの面で問題を残している。ここでは、位相差から正しい視差を求めるのではなく、位相差の符号をもとに視差に関するおおまかな情報、すなわち、視差の解が取り得る範囲を求めることを考える。

5. 1. 1 位相差符号の意味

いま、中心周波数 $u_s (=1/\lambda)$ のガボア・フィルタ対を考えると、このフィルタで検出可能な視差範囲は(4.11)式で与えられ、この範囲の視差に対しては、左右の画像の位相差の符号は視差の符号(注視点に対して奥か手前かに対応する)と一致する。従って、位相差符号が正であれば、求めるべき視差は

$$0 < d < \lambda / 2 \quad (5. 1)$$

の範囲に存在すると考えられる、またそれが負であれば視差が取り得る範囲は

$$-\lambda / 2 < d < 0 \quad (5. 2)$$

に限定される。このように、位相差符号を計算することによって、求めるべき視差が取り得る範囲を、検出可能な全範囲の半分に絞ることができる。

5. 1. 2 位相差符号の検出

Fig.5.1に示すように、左右の画像に対するガボア・フィルタ g_c , g_s の出力をそれぞれ

$$P_L(x, y) = (O_{Lc}(x, y), O_{Ls}(x, y)) \quad (5. 3)$$

$$P_R(x, y) = (O_{Rc}(x, y), O_{Rs}(x, y)) \quad (5. 4)$$

なるベクトルとして考える。このとき、各画像の位相は、それぞれのベクトルが水平軸となす角度 θ_L , θ_R で表され、両画像の位相差の符号とは、文字どおり $\theta_L - \theta_R$ の符号である。しかし、この符号は2つのベクトルの相対的な位相関係、すなわち、 P_L に対して P_R が反時計回りの方向に存在するか、時計回りの方向に存在しているかによって決まるもので、両出力ベクトルの外積

$$\begin{aligned} P_L(x, y) \times P_R(x, y) &= |P_L| |P_R| \sin(\theta_L - \theta_R) \\ &= O_{Lc}(x, y) \cdot O_{Rs}(x, y) - O_{Ls}(x, y) \cdot O_{Rc}(x, y) \end{aligned} \quad (5. 5)$$

の符号として検出できる。よって、位相差符号を求めるだけであれば各位相を計算する必要はなく、 \tan^{-1} の計算を回避できる。

ここで、(5.5)式の値が0となる場合、これは必ずしも位相差0を意味しない。位相差180°の場合にも外積は0となる。また、たとえ実際の位相差が0であっても、ノイズなどの影響で外積が正確に0になるとは限らない。そこで、位相差0を検出するために、出力ベクトルの内積を計算する。

$$\begin{aligned} P_L(x, y) \cdot P_R(x, y) &= |P_L| |P_R| \cos(\theta_L - \theta_R) \\ &= O_{Lc}(x, y) \cdot O_{Rc}(x, y) + O_{Ls}(x, y) \cdot O_{Rs}(x, y) \end{aligned} \quad (5. 6)$$

この値が大きいほど位相差0の可能性が高い。我々は、(5.5)式、(5.6)式を比較し、次のように位相差符号を決定する。

$$\begin{aligned} \text{内積} > |\text{外積}| &\cdots \text{位相差符号 } 0 \\ \text{内積} \leq |\text{外積}| & \\ \text{外積} > 0 &\cdots \text{位相差符号 } + \\ \text{外積} < 0 &\cdots \text{位相差符号 } - \end{aligned} \quad (5. 7)$$

この関係をFig.5.2に示す。検出された位相差符号が+、または-の場合には、求めるべき視差は(5.1)式または(5.2)式の範囲に存在するものとし、位相差符号0の場合にはその範囲を、

$$-\lambda/4 < d < \lambda/4 \quad (5. 8)$$

とみなすことにする。位相差符号の検出法の構成をFig.5.3に示す。

5. 2 空間周波数チャンネルの統合

5. 1で述べた方法によって求めた位相差符号が、正しく視差符号に対応するためには、用いたガボア・フィルタの中心周波数が、求める視差に対して(4.11)式を満たしていなければならない。すなわち、大きい視差を扱うためには、低い中心周波数を持つフィルタを用いる必要がある。一方、検出された位相差符号が示す視差候補の範囲の大きさは、(5.1)式または(5.2)式からも明かなように、フ

フィルタの中心周波数に反比例する。従って、高周波数のフィルタは求めるべき視差を狭い範囲に絞ることができるが、低周波数フィルタでは解の曖昧性が大きくなる。これらのことから、求まる視差の曖昧性と、扱うことのできる視差の上限はトレード・オフの関係にある。ここでは、中心周波数が異なる複数のフィルタ対（空間周波数チャンネル）を用い、各チャンネルから得られる位相差符号を統合することによって、大きな視差に対しても曖昧性の少ない解を求める方法を提案する。

Fig. 5.4に本方法の概略を示す。なお、各チャンネルは、中心周波数が小さいものから順にCH1, CH2, . . . と表すことにする。また、CH1の中心周波数 u_{s1} （波長 $\lambda_1=1/u_{s1}$ ）は十分に小さいものとし、入力画像に含まれるすべての視差は、(4.11)式の範囲に存在すると仮定する。いま、このチャンネルで検出された位相差符号を+とすると、求めるべき視差は、図中の黒で示された範囲に限定される。そこで、次のチャンネルCH2では、この範囲の細分化を行う。これは、位相差符号を検出するとき、左画像の位置 (x, y) に対して右画像の位置 $(x+d_1, y)$ におけるフィルタ出力を用いて外積 $P_L(x, y) \times P_R(x+d_1, y)$ 、内積 $P_L(x, y) \cdot P_R(x+d_1, y)$ を計算することによって実現できる。ここで、 d_1 は、チャンネルCH1で得られた視差範囲の中心の値であり、Fig. 5.4の場合 $\lambda/4$ である。CH2は、CH1に比べて小さい視差しか扱うことができないが、その中心周波数 u_{s2} または波長 λ_2 が

$$u_{s2} < 2u_{s1} \quad \text{または} \quad \lambda_2 > \lambda_1/2 \quad (5.9)$$

を満たしていれば、CH1によって与えられた範囲のすべての視差に対して正しい位相差符号を与えることができる。しかも、このチャンネルCH2は、より狭い範囲に解を絞ることができる。

一般にチャンネルCH n では、前のチャンネルCH $n-1$ で得られた視差範囲の中心値 d_{n-1} を用いて、外積 $P_L(x, y) \times P_R(x+d_{n-1}, y)$ 、内積 $P_L(x, y) \cdot P_R(x+d_{n-1}, y)$ を計算し、位相差符号を求める。これによって、得られる視差範囲の中心値 d_n を次のチャンネルに伝搬する。このように、低い周波数チャンネルで得られた視差範囲を高い周波数チャンネルで細分化していくことによって、大きな視差に対しても曖昧性を少なくすることができる。十分高い空間周波数チャンネルを用意することにより、精度の高い視差を求めることが可能となる。

ここで述べた複数チャンネルの統合方法は、Marrが彼の第2アルゴリズムの中で、眼球運動と対応づけて提案した方法¹⁸と等価である。

5.3 計算機実験

ランダム・ドット・ステレオグラムを用いて、本方法と4. で述べた従来手法の計算時間と計算精度を比較した。なお、以下では本手法をPDS (Phase Difference Sign) 法、従来法をPD (Phase Difference) 法と呼ぶことにする。計算には、UNIXマシンのMASSCOMP5600を用い、計算精度は、計算された視差と、入力画像に与えた視差が一致している画素数を全画素数で割ることによって求めた値（以後、正答率と呼ぶ）で表した。

入力データとしては、4. と同じランダム・ドット・ステレオグラムを用いた。

視差検出には、中心周波数が1オクターブづつ離れた3つの空間周波数チャンネルを用いた。各チャンネルの中心周波数 u_s はそれぞれ、0.0625, 0.125, 0.25cyc

les/pixelであり、検出可能な視差範囲はそれぞれ、8, 4, 2pixel未満になる。従って、最も高い周波数チャンネルは、視差の解を、-1, 0または+1に限定でき、位相差符号のみを用いる本方法でも、 ± 7 pixelsまでの視差ならば一意に求めることができる。また各チャンネルの広がり σ は $0.795/\mu s$ とした。これによって、それらの周波数領域上における広がりが1オクターブになっている。

Fig.5.5、Fig.5.6に、ターゲットの視差が1、及び3の場合の視差検出結果を示す。図中の各画像は、各手法によって推定された視差が d であった位置を示している。また、Table 5.1に、それぞれの計算時間と正当率を示す。どちらの視差に対しても、計算時間に関してはPDS法が2倍以上速いが、正答率はPD法の方が2%程度優れている。

この正答率の違いは、次のような理由によるものと考えられる。すなわち、PD法は、各チャンネルで視差を計算するので、低い周波数チャンネルでの計算誤差をより高いチャンネルが修復することが可能であるが、PDS法では、1オクターブ毎にチャンネルを用意した場合、低い周波数チャンネルが一旦視差符号を間違えると、それをより高い周波数チャンネルで修復することは不可能となる。このようなPDS法の欠点を克服する1つの方法としては、隣接するチャンネルの中心周波数の間隔を小さくすることが考えられる。そこで、上記の3チャンネルの他に、中心周波数が、 0.0883883 、 0.1767767 cycles/pixelである2つのチャンネルを用意し、5つのチャンネルで視差検出を行ってみた。この場合、チャンネルの中心周波数間隔は約0.5オクターブになっている。正答率、及び計算時間をTable 5.2に示す。PD法による正答率はほとんど変化していないが、PDS法では、PD法とほぼ同程度まで正答率の改善がなされている。しかも、5チャンネルによるPDS法は3チャンネルのPD法よりも計算時間が短く、総合的に評価した場合、PDS法はPD法と同程度、あるいはそれ以上の性能を持つと考えられる。5チャンネルを用いたときの両手法による視差検出結果をFig.5.7、Fig.5.8に示す。

5. 4 考察

両眼立体視検出法として、左右画像における輝度変化の位相差符号(-, 0, +)から視差を求める方法を提案した。本方法では、単一の空間周波数チャンネルは、求めるべき視差が取り得る範囲を限定するだけであるが、これを複数の空間周波数チャンネル間で伝搬することによって最終的に視差を求めることができる。本方法は、我々の従来法と同様、輝度変化の位相情報を用いているために、特徴マッチング法が抱える対応問題を解く必要がない。また、平均輝度やコントラストの変化に影響されにくいという利点をもつ。一方、従来法とは異なり \tan^{-1} の計算を必要としないため、計算速度が速く、しかも生体モデルとしての可能性がより高いと思われる。また、適当な視差に対しては、位相差の符号は、視差の符号(注視点を中心に奥にあるか手前にあるか)と等しく、それを検出することは、生理学の研究において両眼立体視に関与する細胞として発見されているnear cell、far cellの機能を実現しているようにも考えられる。

ランダム・ドット・パターンを用いたシミュレーションでは、位相差から視差を計算する我々の従来法との比較を行い、計算時間と計算精度を考慮した総合的な性能にはあまり差がみられないことが確かめられた。しかし、Fig.5.5、Fig.5

.6、Fig.5.7、及びFig.5.8の視差検出の結果からわかるように、両手法では誤差の現れ方が異なっている。すなわち、正答率が同じでも、位相差符号のみを用いた本方法の結果には局所的な（孤立点的な）誤差が目だつ。これは、両手法の視差0を検出する能力が、他の視差を検出する能力よりも優れていると仮定することによって次のように説明できる。従来法では、低い周波数チャンネルでも視差を計算できるので、そのチャンネルで正しい視差が求まる可能性がある。しかも、視差一定の領域がある程度広がっている場所では、低周波数チャンネルのほうが積分効果によって精度がよくなる。もし、正しい視差が低周波数チャンネルで求まってしまえば、より高い周波数チャンネルでは視差0を求めればよい。これに対し本方法では、一般的には最も高い周波数チャンネルまでは正しい視差を求めることはできない。従って、このチャンネルは常に0以外の視差を算出することが要求され、低周波数チャンネルの積分効果の恩恵（視差0を求めること）を受けることができない。上記の仮定（0視差検出の優位性）の真偽については今後検討する必要があるが、シミュレーション結果として現れた孤立点ノイズを除去するためには、各周波数チャンネルにおいてなんらかの空間的相互作用を導入する必要がある。

6. 運動検出モデルと 両眼立体視モデルの統合

従来、運動検出は1本の時系列画像に対して行われてきた。また両眼視差の検出は、ある時刻に左右の眼から得られた2枚の画像だけを入力データと見なしてきた。しかし、外界における物体の3次元運動を観測する場合には、その運動の奥行き方向の成分によって両眼視差が時間的に変化する。そしてこの視差の時間的変化は両眼に生じた2次元運動と独立ではない。ここでは、両眼で得られた2つの速度と両眼視差の関係をもとに、運動視、両眼立体視の統合モデルについて検討する。

6. 1 運動速度と両眼視差の関係

Fig. 6.1に、外界で運動している物体を両眼で観測した場合の、左右の眼に生じる2次元運動速度と両眼視差の時間的変化の関係を示す。 F_L 、及び F_R は、左右の画像上の原点(中心窩)を示す。いま、3次元世界に存在する点 P が時刻 k において3次元速度 V^k で運動していたとする。この点は、左右の画像においてそれぞれ P_L^k 、 P_R^k に投影されるが、これらの位置における画像上での速度 v_L^k 、 v_R^k が検出できれば、次の時刻 $k+1$ における点 P の像の位置を知ることができる。よって、ある時刻における各画像上での運動を検出することで、次の時刻における視差を予測することができる。しかし、運動から得られたこの視差の予測値は、 $k+1$ での左右画像から検出された視差と一致しているはずである。従って、このような状況では、入力データとして両眼から得られる2本の時系列を用い、これらから得られる運動と視差の整合をとることによって、推定値の精度を改善することができると思われる。

6. 2 統合モデルの概略

Fig. 6.2に運動視・両眼立体視統合モデルの構想の概略を示す。ここでは、時間的な統合を実現するために、3. で述べたカルマン・フィルタを用いて逐次的に処理を行うことを考えている。

左右の眼に入力された各時系列画像に対して、2.、及び3. で述べたような運動検出モデルを適用し、時刻 k の運動速度 v_L^k 、 v_R^k を求める。このとき、3. でも述べたように、速度変化に対する追従性を考慮して、できるだけ少ないフレームから速度を計算する。また、この時刻における両眼画像から視差 d^k を4.、5. で提案したようなモデルで検出する。これらの観測値と、前の時刻 $k-1$ で得られた推定値 v_L^{k-1} 、 v_R^{k-1} 、 d^{k-1} を用いて、カルマン・フィルタによって時刻 k における推定値を v_L^k 、 v_R^k 、 d^k 計算する。さらに、時刻 $k+1$ における視差 d^{k+1} を観測し、これと先に求めた時刻 k の推定値の整合をとることによって新しい推定値 v_L^k 、 v_R^k 、 d^k 、 d^{k+1} を計算する。この中で、 v_L^k 、 v_R^k 、 d^k は時刻 k の速度、及び視差として出力される。また、 v_L^k 、 v_R^k 、 d^{k+1} は次の時刻 $k+1$ における推定値を求めるために、遅延後フィード・バックされる。 d^k をフィード・バックしないのは、視差は運動によって時々刻々変

化するためである。ここで、Fig.6.2において、視差 d^k が時刻 k における左画像上の点 P_L^k に対して得られたものであるとすると、 d^{k+1} は時刻 $k+1$ における左画像上の点 $P_L^k + v_L^k \cdot \Delta t$ に対する視差を表している (Δt はフレーム間の時間間隔)。

このモデルによれば、運動、両眼視差の統合と時間的統合によって、より精度の高い推定値が得られることが期待できる。

6. 3 運動速度と両眼視差の整合法

Fig.6.2のモデルでは、速度と視差の推定値 $v'_L{}^k$ 、 $v'_R{}^k$ 、 d'^k と観測値 d^{k+1} の整合をいかにとるかが問題となる。ここではその1つの解決法について述べる。

解くべき問題をFig.6.3に示す。L、Rはそれぞれ右眼、左眼の画像を表す(簡単のために1次元の画像で示す)。また、 v_L^k 、 v_R^k は速度そのものではなく、速度にフレーム間の時間間隔(定数)を掛けた移動距離を表すものとする。時刻 k における移動距離(速度)及び視差としては、本来カルマン・フィルタによって得られた推定値 $v'_L{}^k$ 、 $v'_R{}^k$ 、 d'^k を用いるが、ここでは説明の都合上、これらも"観測値"と呼び、 v_L^k 、 v_R^k 、 d^k で表すことにする。

いま、時刻 k における左眼画像上の位置 P_L^k に存在する点に対して、速度 v_L^k が求められれば、この点の時刻 $k+1$ における位置 P_L^{k+1} を予測することができる。また、位置 P_L^{k+1} に対する視差 d^{k+1} が検出できれば、これに対する右眼画像上の位置 P_R^{k+1} が予測できる。一方、 P_L^k に対応する右画像上の位置 P_R^k は視差 d^k によって得られ、この位置における点の速度 v_R^k がわかれば、その点の時刻 $k+1$ における位置 P_R^{k+1} が予測できる。各位置および時刻における速度と視差はこれまでに述べたようなモデルで検出できるので、我々は P_L^k に対する P_R^{k+1} の相対的な位置を、

$$v d_1 = (v_L^k + d^{k+1}) \quad (6. 1)$$

$$v d_2 = (v_R^k + d^k) \quad (6. 2)$$

という異なる2つの式で予測することができ、原理的にはそれらの結果は一致しているはずである(Fig.6.3(a))。すなわち、

$$\begin{aligned} v d_1 - v d_2 &= (v_L^k + d^{k+1}) - (v_R^k + d^k) \\ &= 0 \end{aligned} \quad (6. 3)$$

しかし、観測された速度や視差は誤差を含んでおり、必ずしも(6.3)式が満たされるとはかぎらない。いま各観測値の誤差を Δv_L^k 、 Δv_R^k 、 Δd^k 、 Δd^{k+1} とすれば、通常は

$$\begin{aligned} & [(v_L^k + \Delta v_L^k) + (d^{k+1} + \Delta d^{k+1})] \\ & - [(v_R^k + \Delta v_R^k) + (d^k + \Delta d^k)] = 0 \end{aligned} \quad (6. 4)$$

となる(Fig.6.3(b))。ここでは、(6.4)式における各誤差を求め、それによって観測値を補正することによって、運動速度と両眼視差の情報の整合をとる。

各誤差成分を求めるために、我々は(6.4)式の条件のもとで、次式を最小にすることを考える。

$$\begin{aligned} & (\Delta v_R^k)^2 + (\Delta v_L^k)^2 \\ & + (\Delta d^k)^2 + (\Delta d^{k+1})^2 \rightarrow \text{最小} \end{aligned} \quad (6. 5)$$

これは、(6.3)式を満たす誤差(補正值)の組合せの中から、これら観測誤差の合

計が最小となるものを求めることを意味する。しかし、最小自乗法を用いてこれを解くと、

$$\begin{aligned}\Delta v_R^k &= e - 3d^{k+1} \\ \Delta v_L^k &= -d^{k+1} \\ \Delta d^k &= -d^{k+1}\end{aligned}\quad (6.6)$$

但し、

$$\begin{aligned}e &= v d_2 - v d_1 \\ &= (v_R^k + d^k) - (v_L^k + d^{k+1})\end{aligned}\quad (6.7)$$

となり、一意に解を求めるためには、さらに拘束条件が必要となる。このような拘束条件としては、例えば

$$\Delta v_L^k + \Delta d^{k+1} = e / 2 \quad (6.8)$$

または、

$$\Delta v_R^k + \Delta d^k = -e / 2 \quad (6.9)$$

が考えられる。ここで、(6.7)式からもわかるように、 e は P_R^{k+1} の相対位置を表す2つの予測値 $v d_1 (= v_L^k + d^{k+1})$ 、 $v d_2 (= v_R^k + d^k)$ の差である。従って、(6.8)式、(6.9)式は、 P_R^{k+1} を2つの予測位置 $P_L^k + v d_1$ と $P_L^k + v d_2$ の中央

$$P_R^{k+1} = P_L^k + (v d_1 + v d_2) / 2 \quad (6.10)$$

に拘束することを意味している。

また、各観測値 (v_L^k 、 v_R^k 、 d^k 、 d^{k+1}) の信頼度を考慮した拘束条件を考えることもできる。すなわち、なんらかの方法で各観測値の信頼度 (分散等) を計算し、さらにそれらを用いて $v d_1$ と $v d_2$ の信頼度 p_1 、 p_2 を計算する。各信頼度の逆数を、予測位置誤差 e に対する $v d_1$ 、及び $v d_2$ の寄与度として、 e を次のように各観測値誤差に分配する。

$$\Delta v_L^k + \Delta d^{k+1} = e \cdot p_2 / (p_1 + p_2) \quad (6.11)$$

$$\Delta v_R^k + \Delta d^k = -e \cdot p_1 / (p_1 + p_2) \quad (6.12)$$

単に $v d_1$ と $v d_2$ の平均値を用いる (6.8) 式 (または (6.9) 式) の拘束条件に比べ、(6.11) 式 (または (6.12) 式) では、各観測値の信頼度を考慮しており、より精度の高い推定値が得られると思われる。

なお、2. ~ 5. で提案したモデルでは、観測値の信頼度を反映している量として次のようなものが考えられる。

運動 . . . 速度平面上の得点分布の広がり。得点が高い範囲に分布している場合には信頼度が低く、1つの速度のまわりに集中している場合には信頼度が高いと考えられる。

両眼視差 . . . フィルタ出力から計算した内積。これは、得られた視差が0である可能性を示すが、5. 4で述べたように視差検出器が視差0に対して優位性を持っているとすれば、内積の値が大きいほど得られた視差の信頼度が高いと思われる。

これらを用いて、実際にどのように信頼度を計算するかは今後検討を要する。

7. むすび

視覚系の重要なモジュールである運動視、及び両眼立体視について、心理実験などで得られている人間の視知覚機構に関する知見を考慮してそれらのモデル化をおこない、またそれらの統合について検討した。

運動視では、時空間フィルタリングによる信号処理的なアプローチと特徴点マッチング法を組み合わせ、時系列画像に含まれている速度を精度よく検出する方法を提案した。また、時間的統合による推定精度の向上と速度変化に対する追従性を両立させるために、カルマン・フィルタによる逐次的な運動検出の方法を検討した。

両眼立体視では、その処理の運動検出との類似性に着目し、特徴マッチングによらない信号処理的な視差検出法を提案した。また、生体モデルとしての妥当性と計算速度の高速化を目的として、このモデルの改良を行った。第一のモデルでは、視差を両眼画像における輝度変化の局所的な位相差として検出するが、改良したモデルでは、位相差の符号のみから視差検出を行う。

運動視と両眼立体視の統合については検討段階であり、ここではそのおおまかな構想について述べた。外界において運動している物体を観測する場合、両眼に投影される運動と視差の時間的変化が互いに関係していることから、運動検出および両眼立体視の2つの処理の統合によって、それぞれの推定値の精度を向上させることができると思われる。また、両眼における2次元速度を両眼視差によって対応づけることができるので、外界における3次元速度を計算することもできる。

本研究では、運動視と両眼立体視という視覚にとって重要な2つの処理を対象としたが、我々人間の優れた視覚情報処理は、これらの他にもテクスチャや陰影などさまざまな情報を検出し、それらの間の整合をとることによって実現されていると考えられる。従って、よりロバストな視覚システムを構築するためには、これらすべての手がかりを検出するモデルと各モデルによって検出された情報の相互作用を考えていく必要がある。

謝辞

本報告書は、著者がATR視聴覚機構研究所において進めてきた運動視、および両眼立体視の研究成果をまとめたものである。このような研究の機会を与えていただいたATR視聴覚機構研究所 淀川英司社長、上野圭一視覚室長、梅田三千雄元視覚室長（現、大阪電気通信大学教授）に感謝いたします。また、本研究を進めるにあたり終始熱心にご指導いただいた佐藤隆夫主幹研究員には心から感謝の意を表します。さらに、日頃御討論頂いた視聴覚機構研究所の諸氏にも感謝いたします。

文献

1. 緒形, 佐藤, "2段階処理による運動物体の形状知覚モデル", 信学技報, PRU89-44 (1989).
2. B. Julesz, Foundation of Cyclopean Perception, Univ. of Chicago Press(1971).
3. S. Ullman, Interpretation of visual motion (MIT Press, 1979).
4. J. K. Tsotsos, J. Mylopoulos, H. D. Covvey and S. Zucker, "A framework for visual motion understanding", IEEE Trans., PAMI-2, 563-573 (1980).
5. D. T. Lawton, "Processing translational motion sequences", Comput. Vision, Graphics, and Image Process., 22, 116-144 (1983).
6. J. Little, H. Bulthoff and T. Poggio, "Parallel optical flow computation", in Proceedings of Image Understanding Workshop, 915-920(Morgan Kaufmann, Los Altos, CA, 1987).
7. W. Reichardt, "Autocorrelation, a principle for the evaluation of sensory information by the central neurons system", in Sensory Communication, W. A. Rosenblith, ed. (Wiley, New York, 1961).
8. D. Marr and S. Ullman, "Directional selectivity and its use in early visual processing", Proc. R. Soc. London Ser. B211, 151-180 (1981).
9. J.P.H. van Santem and G. Sperling, "Elaborated Reichardt detectors", J. Opt. Soc. Am. A, 2, 300-320(1985).
10. E.H. Adelson and J.R. Bergen, "Spatiotemporal energy models for the perception of motion", J. Opt. Soc. Am. A, 2, 284-299(1985).
11. A.B. Watson and A.J. Ahumada, "Model of human visual-motion sensing", J. Opt. Soc. Am. A, 2, 322-341(1985).
12. D. H. Heeger, "Model of extraction of image flow", J. Opt. Soc. Am. A 4, 8, 1455-1471(1987).
13. 佐藤, 緒形, "動きを見るメカニズムとそのモデル化", テレビ誌, 42, 9, 904-910(1988).
14. 緒形, 佐藤, "時空間フィルタを用いた運動視知覚のモデル", 信学技報, IE87-107(1987).
15. 緒形, 佐藤, "運動視知覚モデルにおける推定速度の精度の改善", 昭63信学春季全大, D-12 (1988).
16. 緒形, 佐藤, "空間周波数チャンネル間の相互作用を考慮した運動視知覚モデル", 信学論, D-II, J73-D-II, 2, 207-214 (1990).
17. T. Sato, "Reversed apparent motion with random dot patterns", Vision Res., 29, 1749-1758 (1989).
18. D. Marr, VISION, W. H. Freeman and Company, New York(1982); 乾, 安藤 (訳) "ビジョン", 産業図書(1987).
19. E. C. Hildreth, The measurement of visual motion

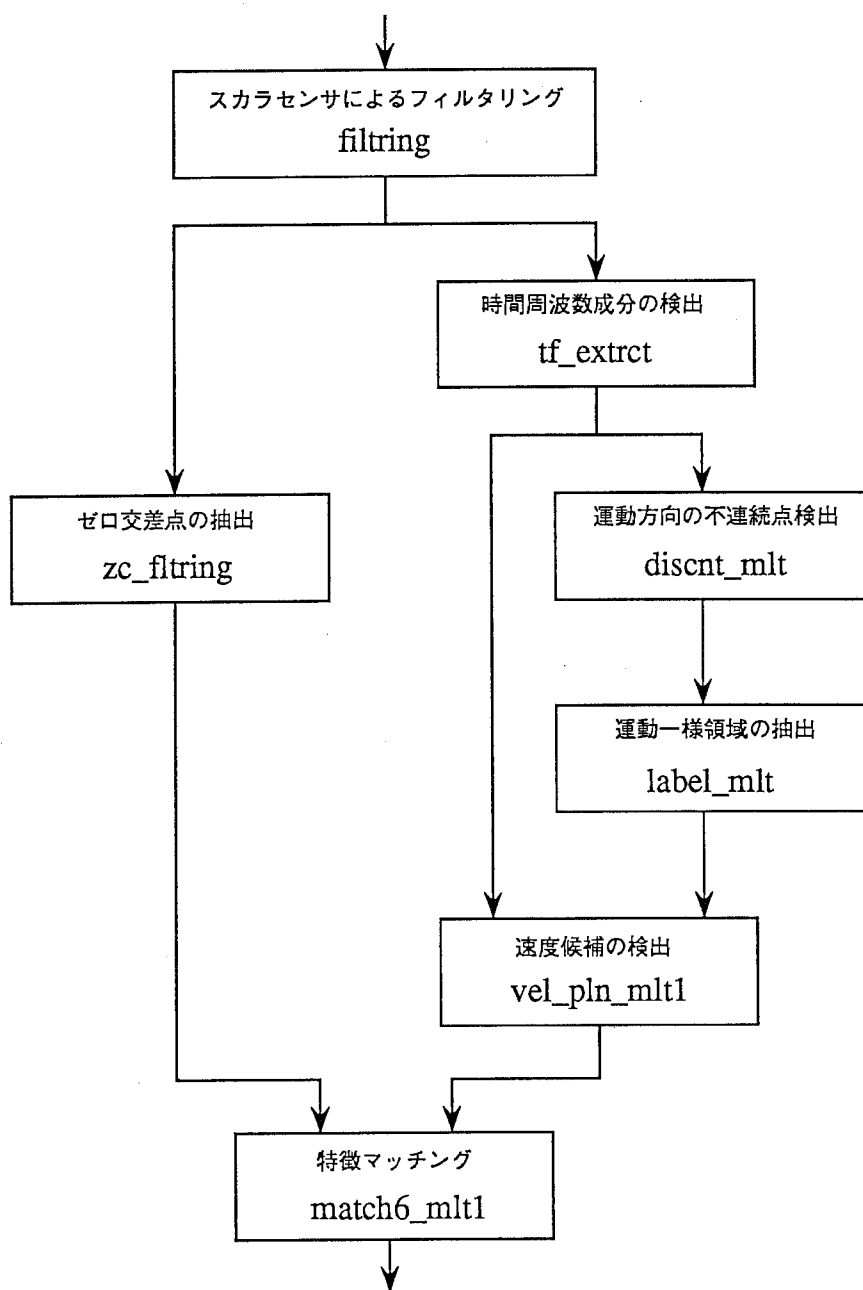
- (MIT Press, 1983).
20. E. H. Adelson and J. A. Movshon, "Phenomenal coherence of moving visual patterns", *Nature* 300, 523-525 (1982).
 21. G. A. Orban and B. Gulyas, "IMAGE SEGMENTATION BY MOTION:CONTROL MECHANISMS AND IMPLEMENTATION IN NEURAL NETWORKS", NATO ASI Series, F41, 149-157(1988).
 22. K. Nakayama and G. H. Silverman: "Serial and parallel processing of visual feature conjunctions", *Nature*, 320, 6059, 264-265 (1986).
 23. 有本卓, カルマン・フィルター, 産業図書 (1977)
 24. L. Mattheis and Takeo Kanade, "Kalman filter-based algorithms for estimating depth from image sequences", *International Journal of Computer Vision*, 3, 209-236 (1989).
 25. J. Heel, "Direct dynamic motion vision"
 26. 緒形、佐藤, "位相検出による両眼立体視モデル", 信学技報, MBE89-129, 29-36 (1990).
 27. 畑田, "生理光学14-奥行き知覚特性", O plus E, 71, 108-118 (1985).
 28. 畑田, "生理光学15-3次元ディスプレイ", O plus E, 73, 98-109 (1985).
 29. 吉田, 渡部, "視覚系における運動知覚機構のモデル", TV学会技術報告 VVI89-30, 11, 16, 31-36 (1987).
 30. 佐藤, 佐々木, "動画像における動きベクトルの階層的推定方", 信学論 (D), J69-D, pp.771-776(1986).
 31. 緒形, "両眼立体視のモデル", VISION, 2, 2, 56-68 (1990).
 32. A. D. Jepson and M. R. Jenkin, "The fast computation of disparity from phase differences", in *Proceeding of CVPR*, 398-403(1989)
 33. T. D. Sanger, "Stereo disparity computation using Gabor filters", *Biol. Cybern.* 59, 405-418(1988).

付録

本報告書で述べた運動検出モデル、および両眼立体視モデルはUNIXマシン MASSCOMP5600 (hm07) 上にC言語によってインプリメントされているが、ここでは、各モデルを構成する関数を処理の流れに沿って示す。なお各関数のソースは、関数名に.cをつけたファイルに存在する。

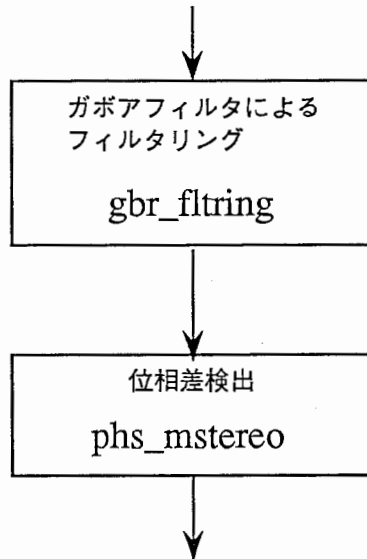
付録1 2段階処理による運動検出モデル

本モデルのための関数はディレクトリ /usr/ogata/MOTION/ の下に存在する。また、スカラ・センサは関数 s_snsrs (/usr/ogata/MOTION/MK_SNSR/) によって生成することができる。



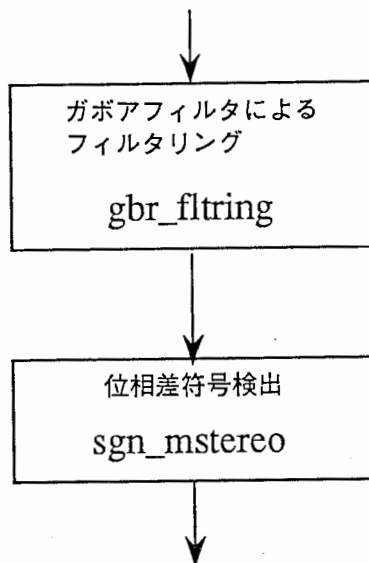
付録2 位相検出による両眼立体視モデル

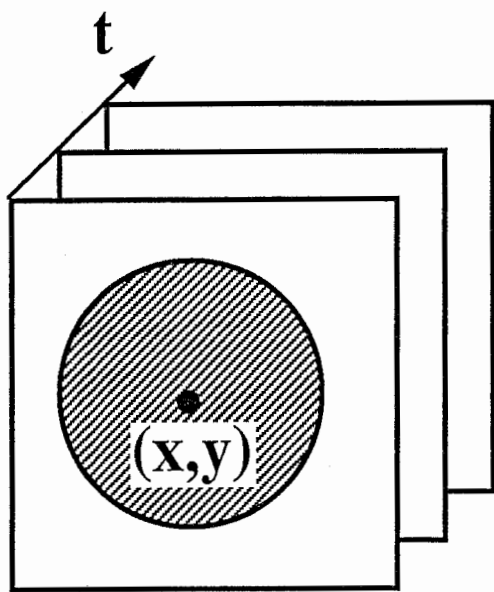
本モデルのための関数は、ディレクトリ /usr/ogata/STEREO/ の下に存在する。ガボア・フィルタは同じディレクトリにある関数 gbr_filter によって生成することができる。



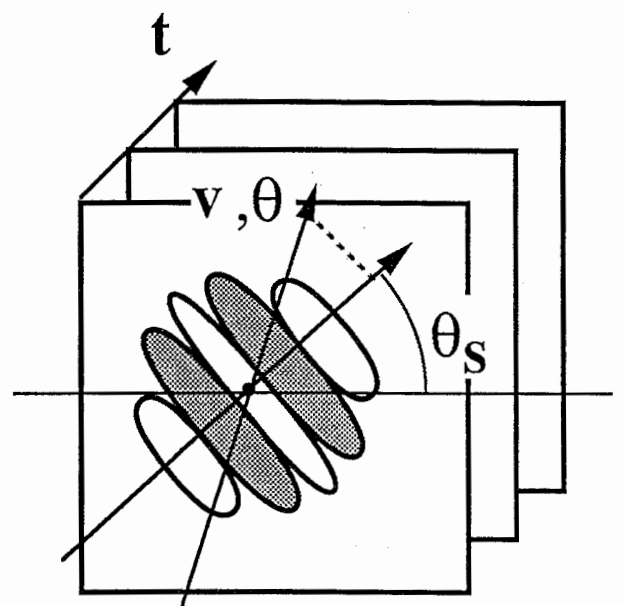
付録3 位相差符号による両眼立体視モデル

本モデルのための関数は、ディレクトリ /usr/ogata/STEREO/ の下に存在する。ガボア・フィルタは同じディレクトリにある関数 gbr_filter によって生成することができる。



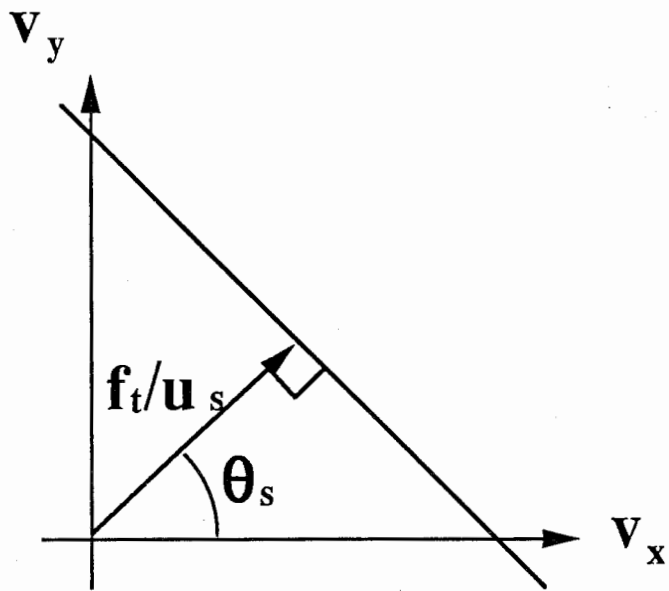


(a) 入力パターン

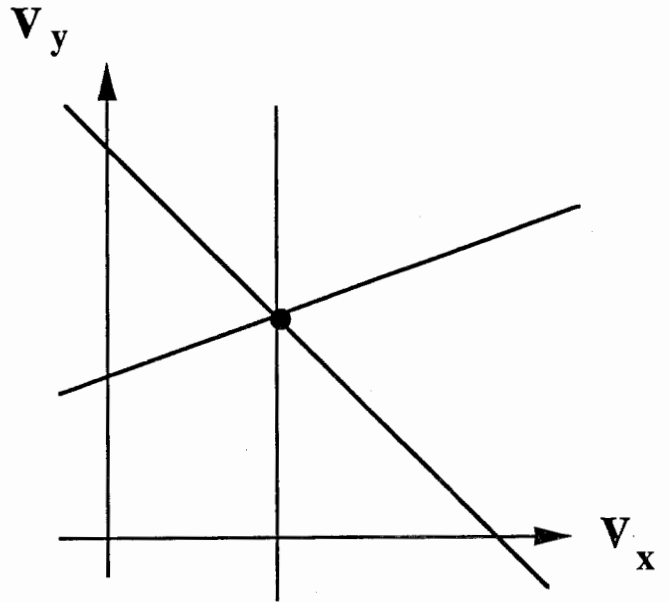


(b) 出力パターン

Fig.2.1 スカラセンサの原理



(a) 速度に関する制約条件



(b) 2次元速度の決定

Fig.2.2 速度平面

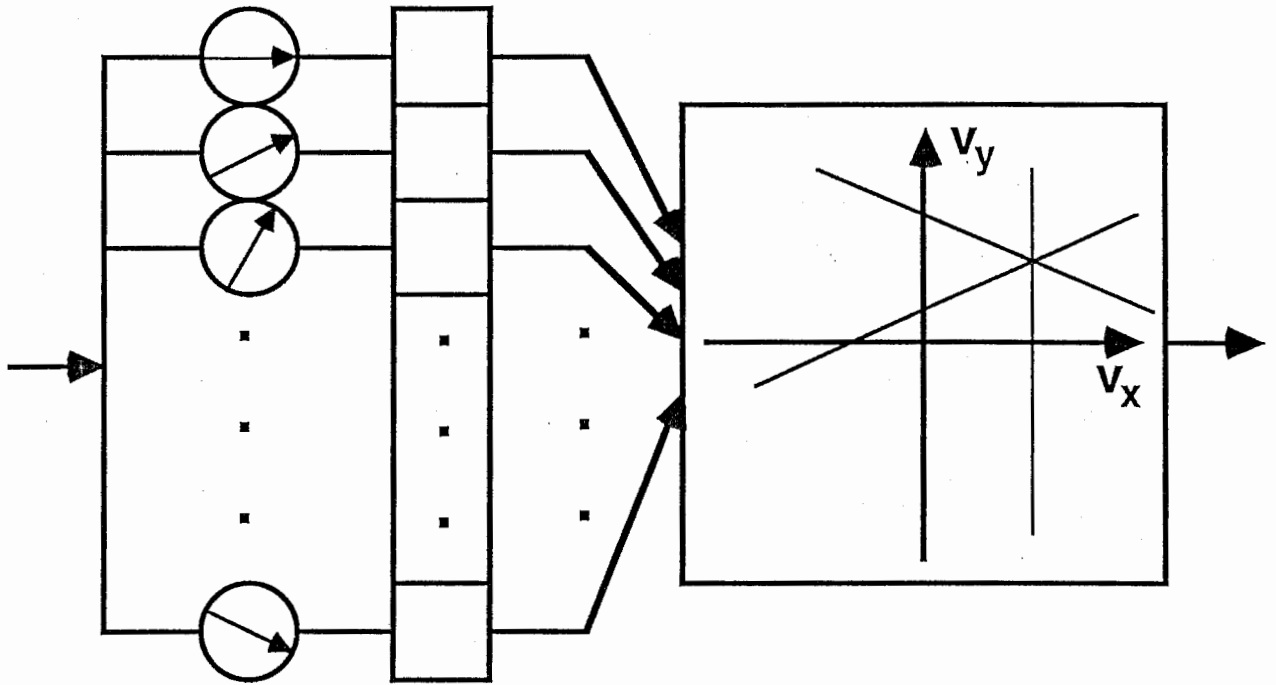


Fig.2.3 単一空間周波数チャンネルの構成

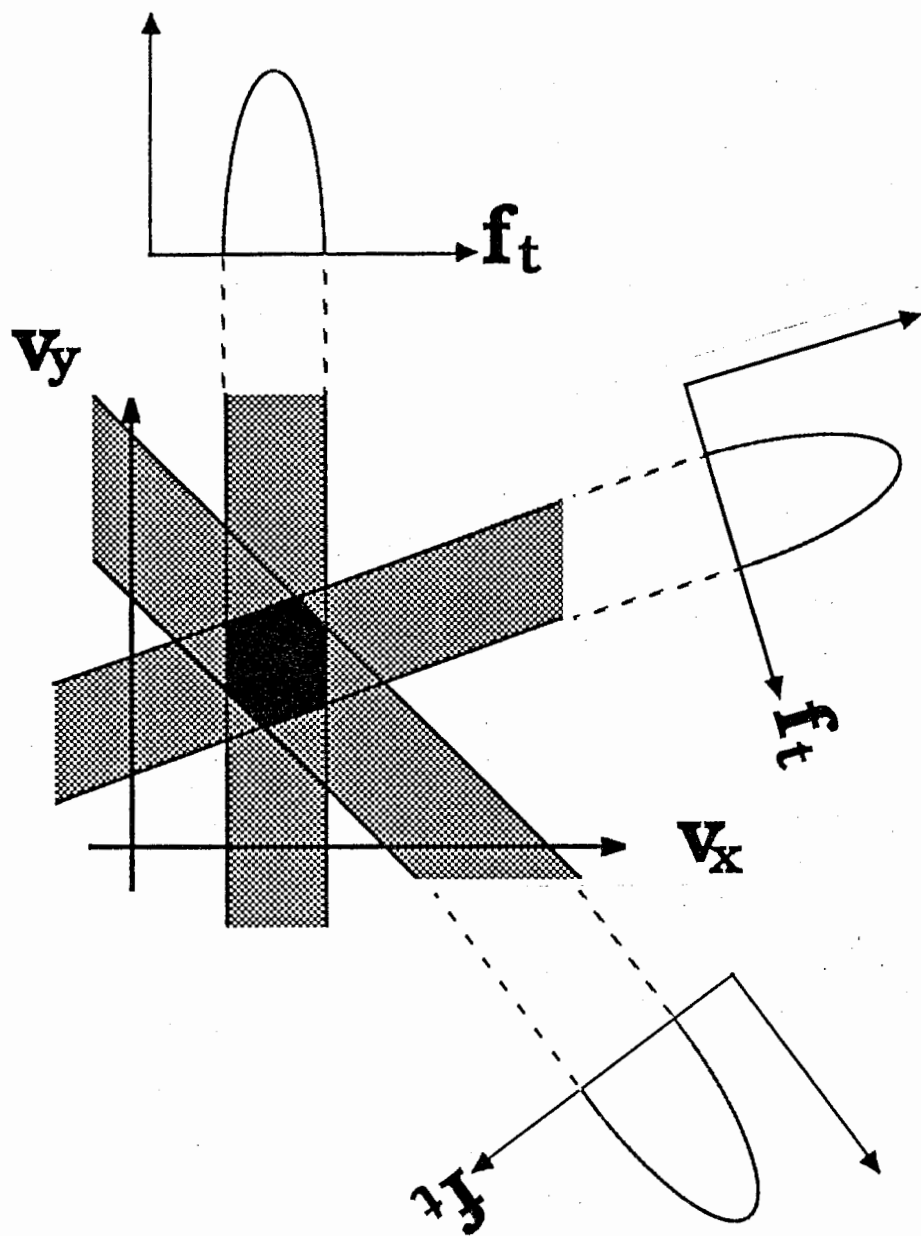


Fig.2.4 速度候補

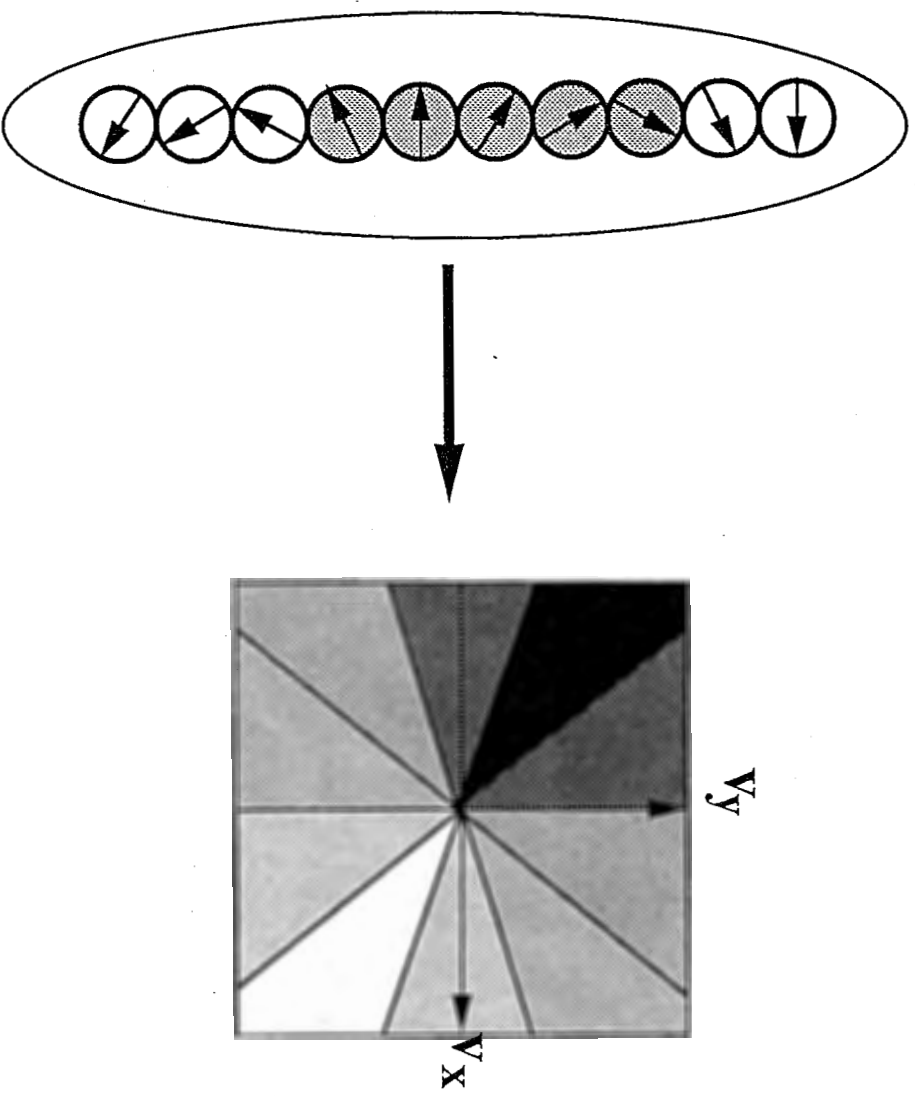


Fig.2.5 運動の向きに対する反応パターン

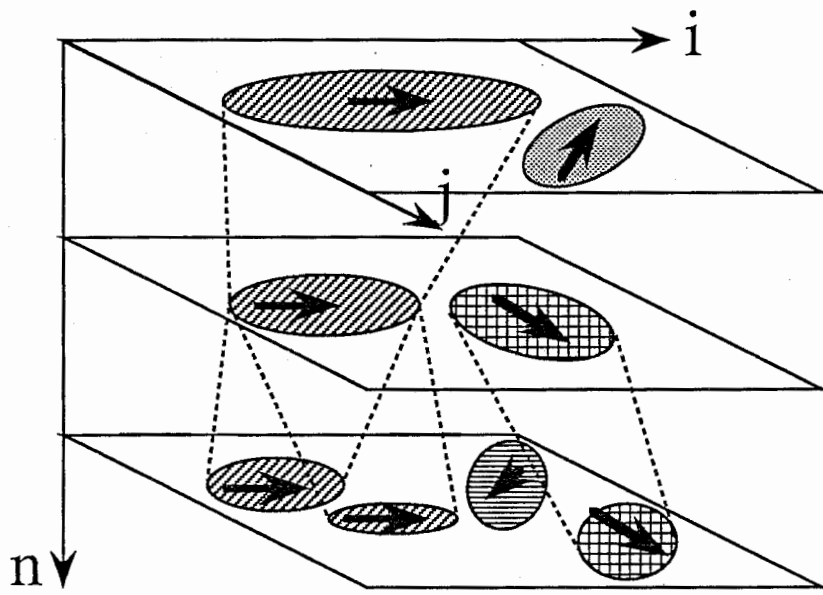


Fig.2.6 運動一様領域の抽出

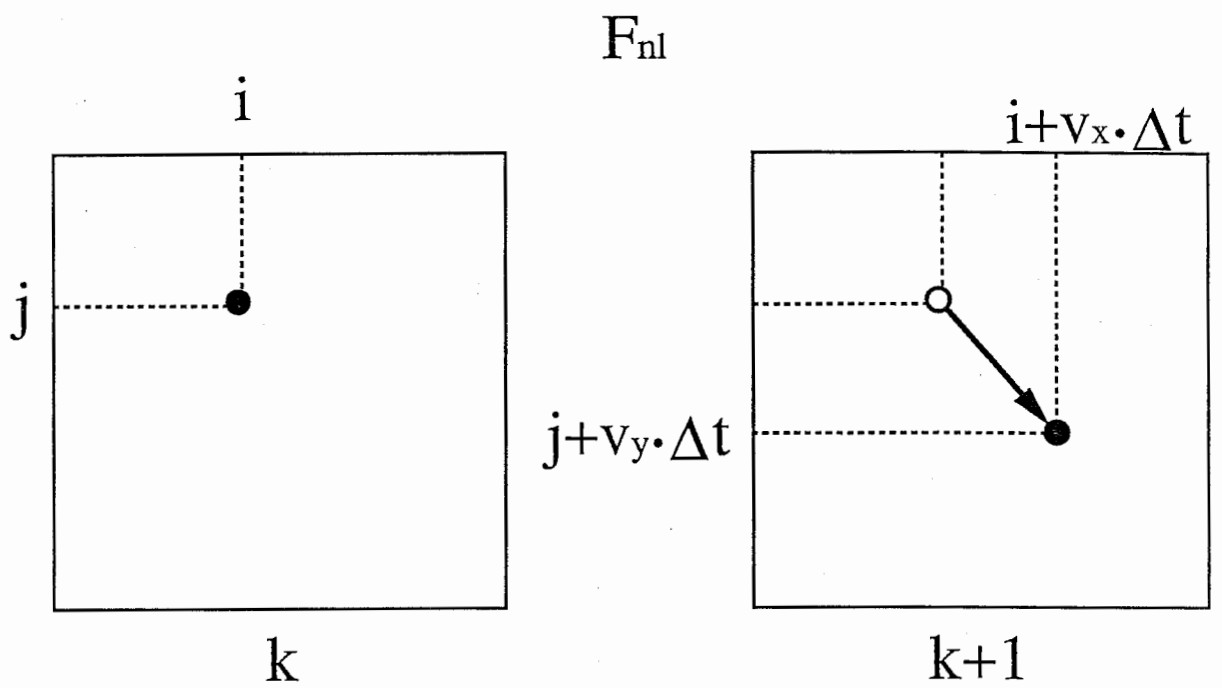


Fig.2.7 マッチング処理

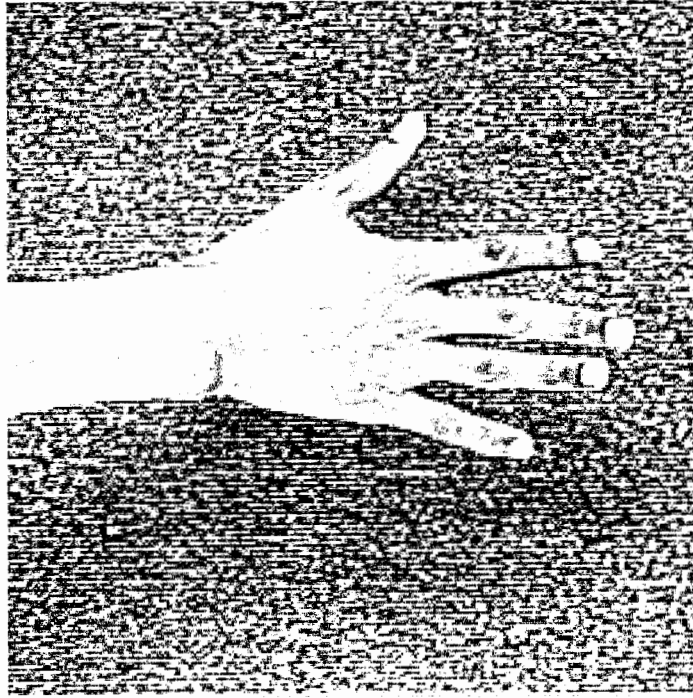
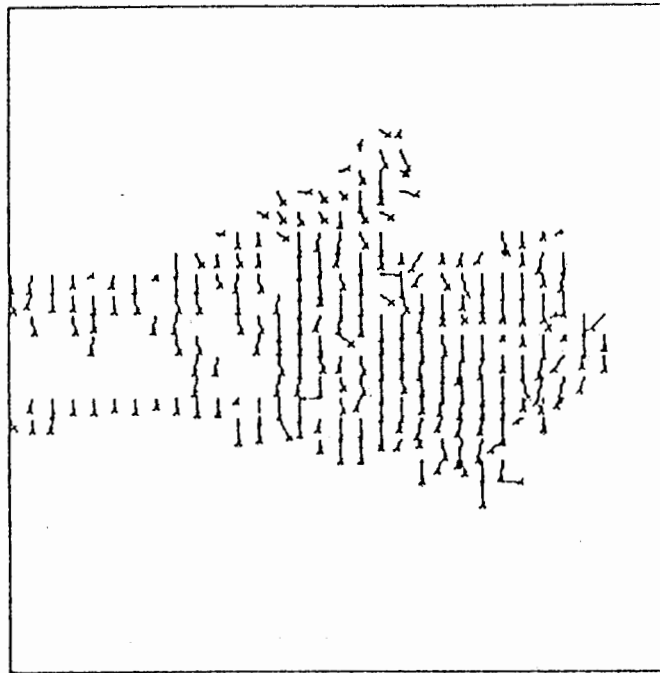
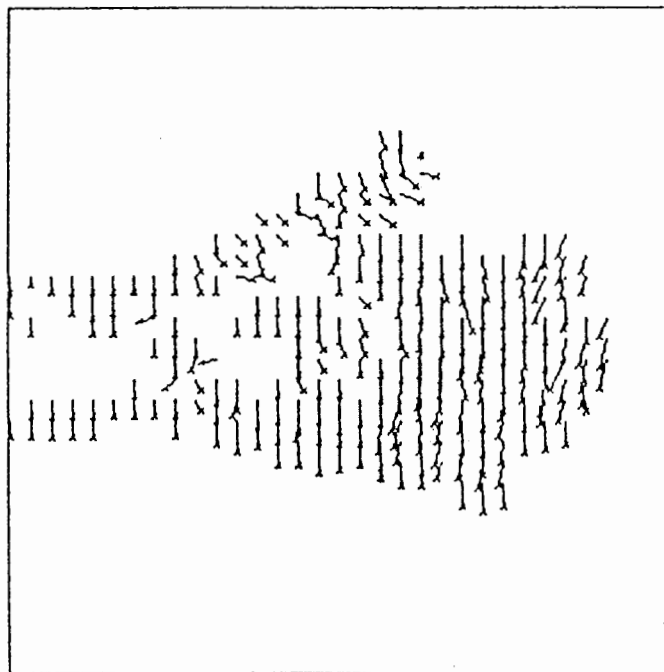


Fig.2.8 入力画像

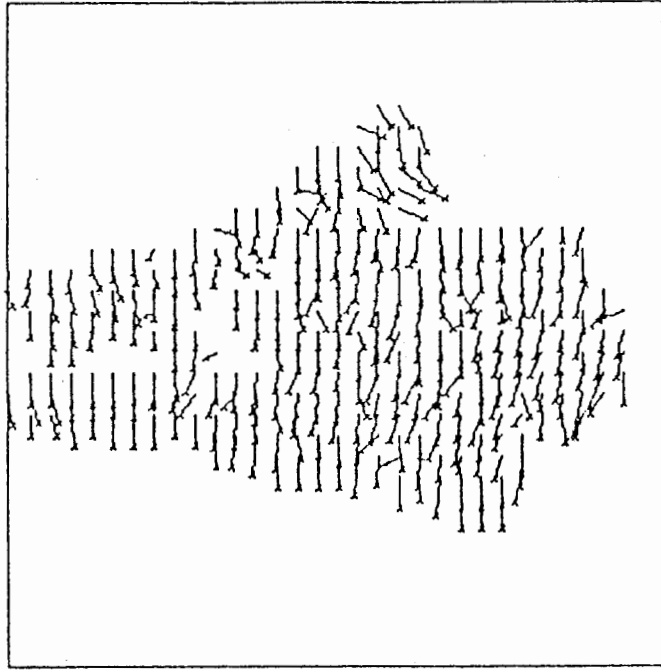


(a) $u_s=0.25$

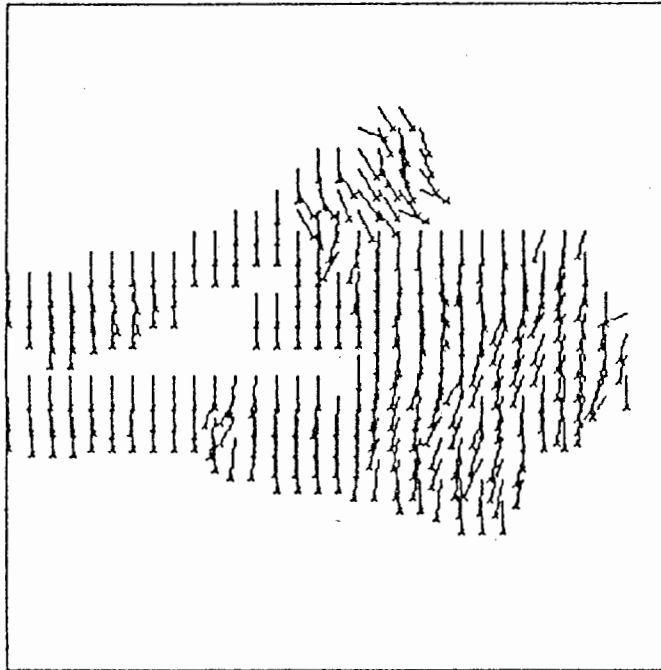


(b) $u_s=0.125$

Fig.2.9 フィルタリング処理のみによる速度場



(c) $u_s=0.0625$



(d) 3チャンネル統合

Fig.2.9 フィルタリング処理のみによる速度場(続き)

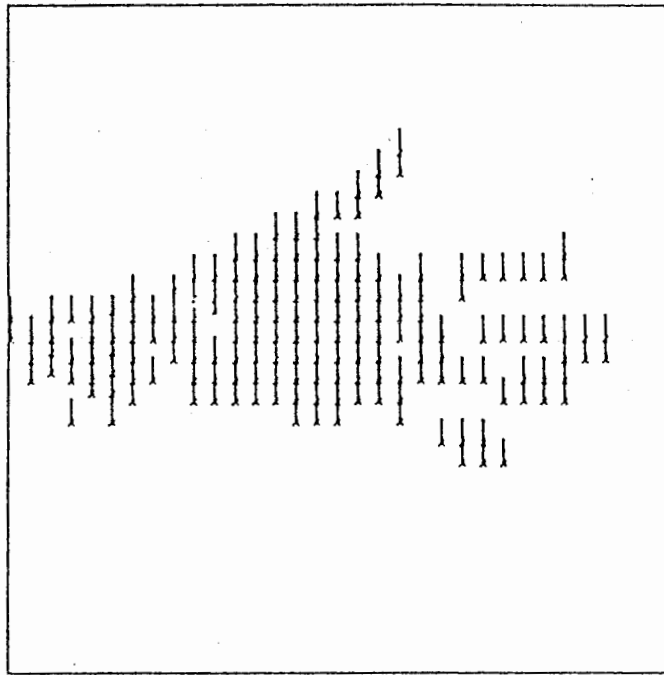
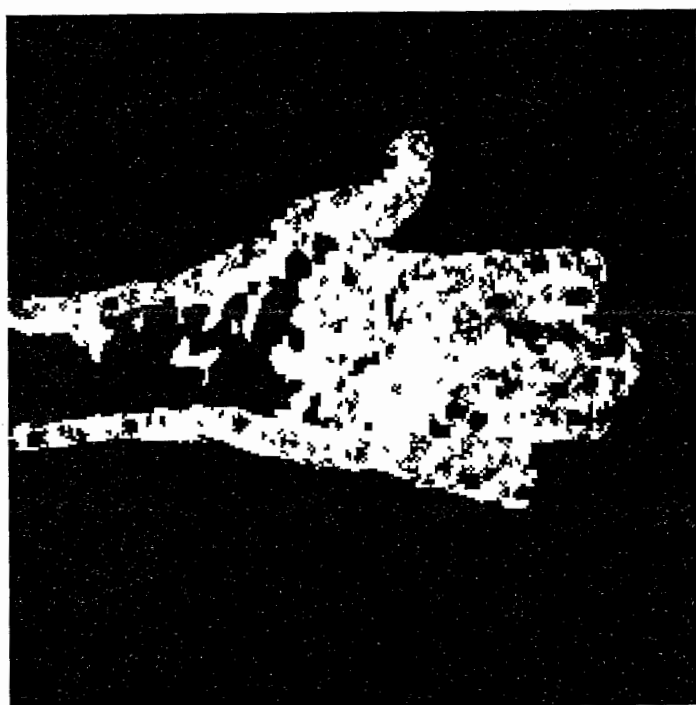
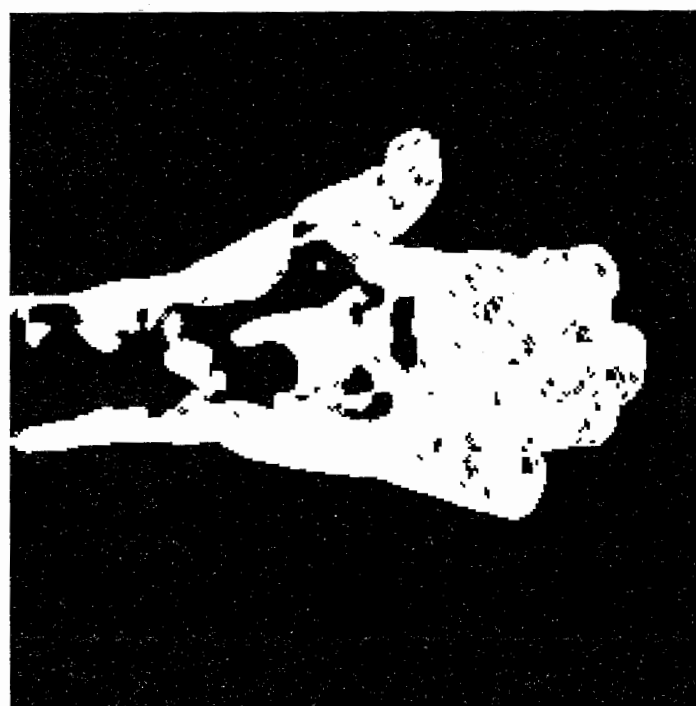


Fig.2.10 2段階モデルによる速度場

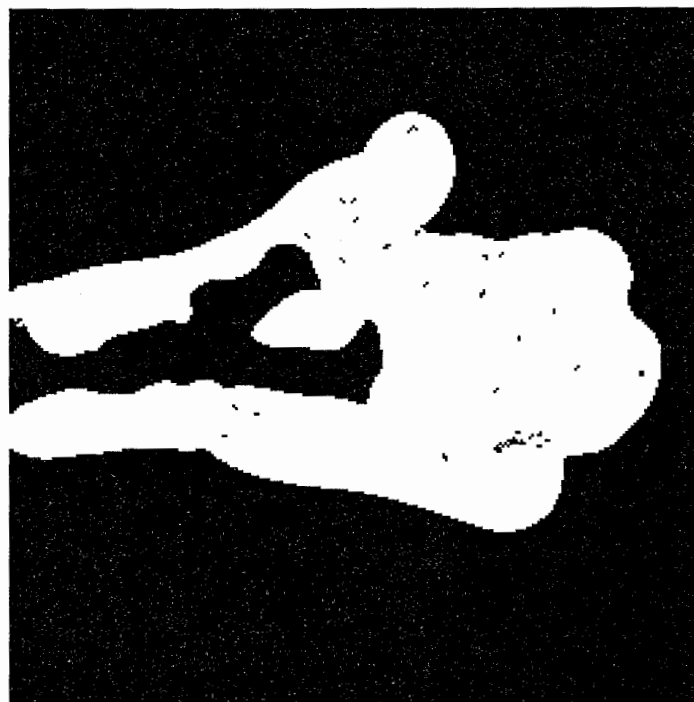


$U_s=0.25$



$U_s=0.125$

Fig.2.11 運動一様領域



$U_s=0.0625$

Fig.2.11 運動一様領域(続き)

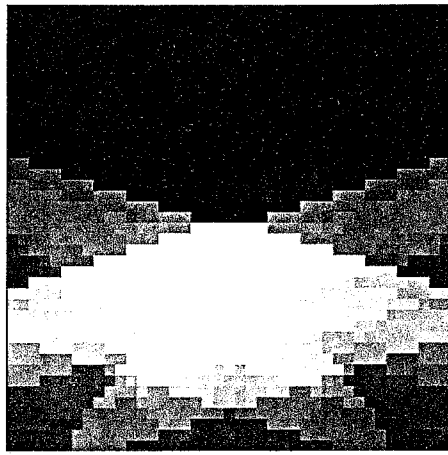
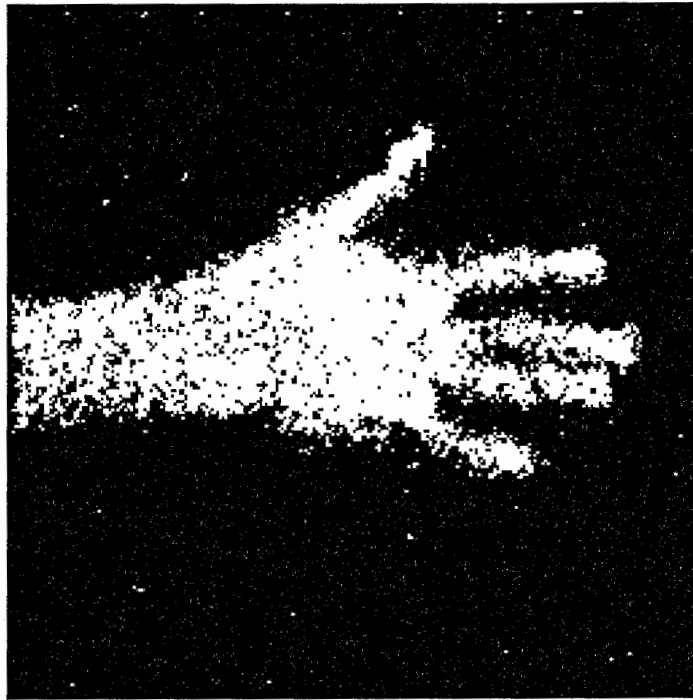


Fig.2.12 速度平面



(a) セグメンテーション

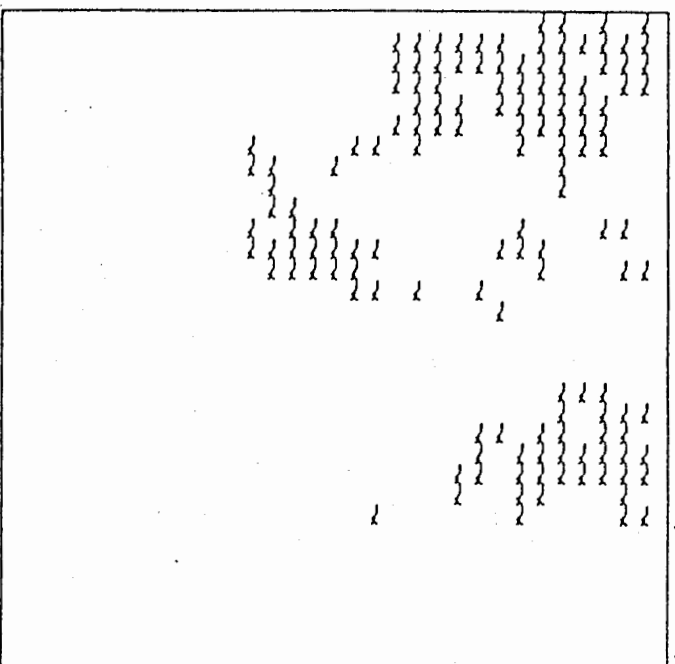


(b) 孤立点除去

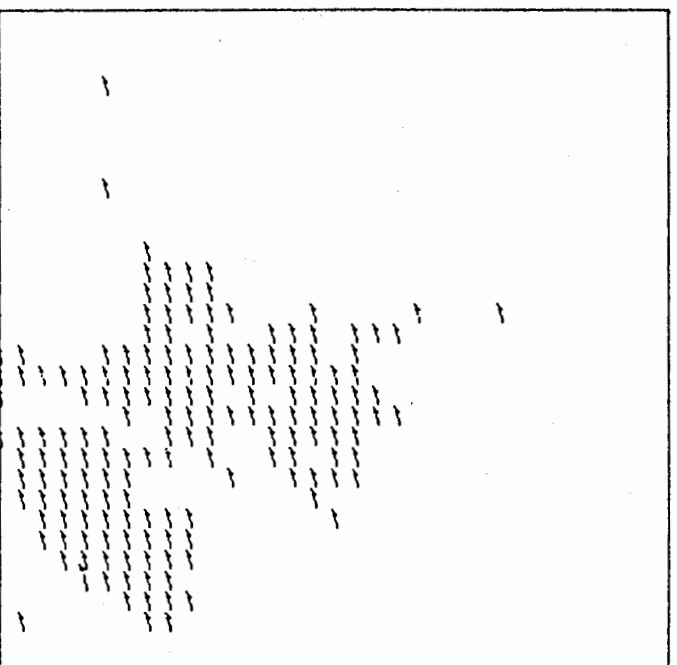
Fig.2.13 運動領域の形状抽出結果



Fig.2.14 入力画像

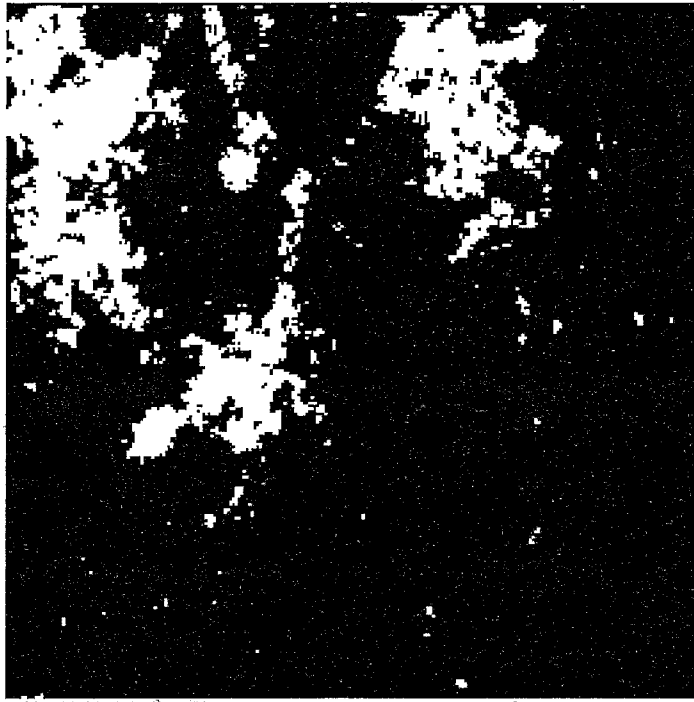


(a) 上の枝

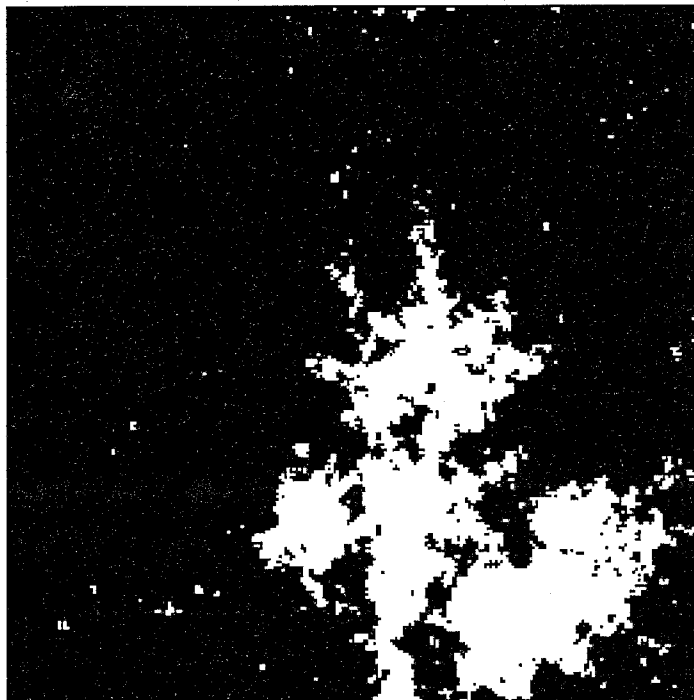


(b) 下の枝

Fig.2.15 2段階モデルによる速度場



(a) 上の枝



(b) 下の枝

Fig.2.16 運動領域の形状抽出結果

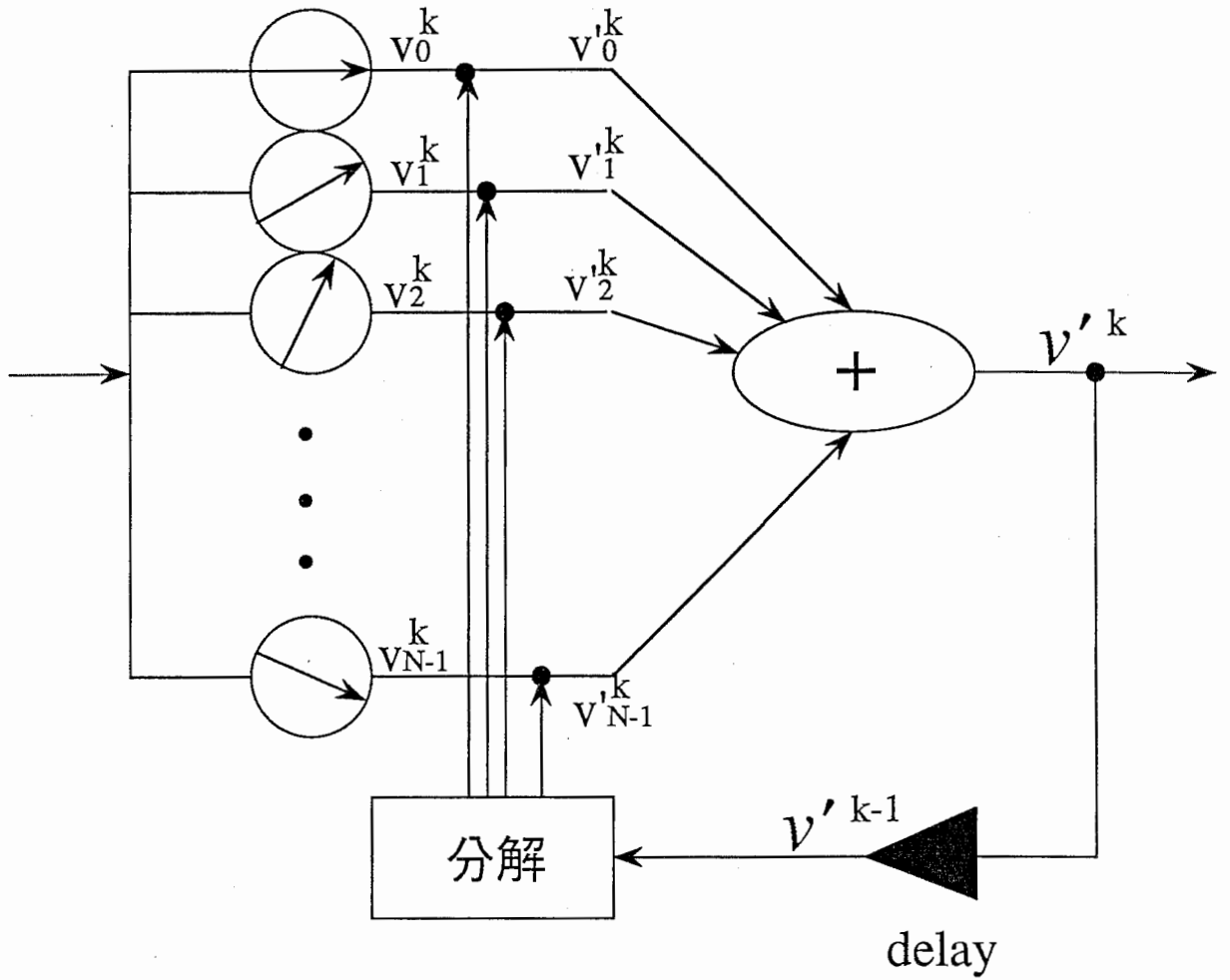


Fig.3.1 カルマンフィルタを用いた
運動検出モデル

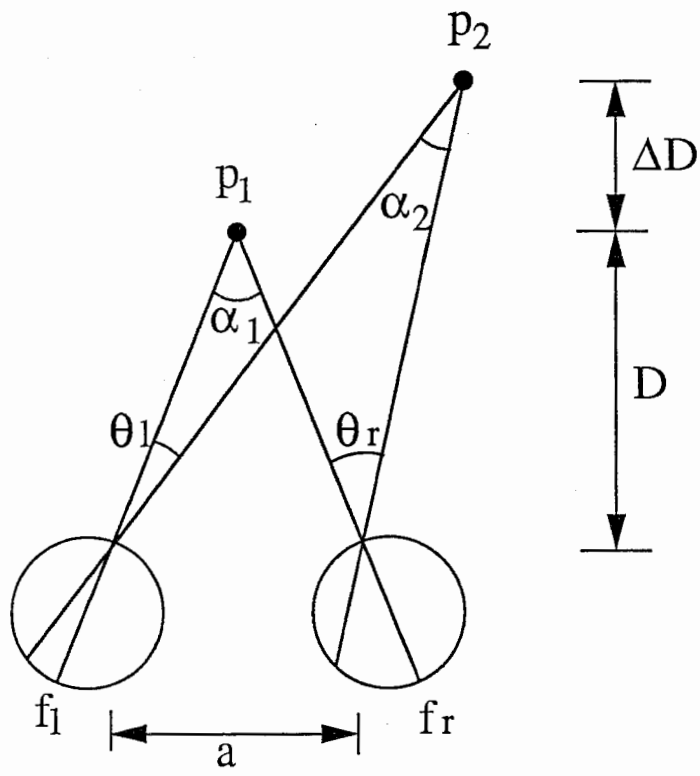


Fig.4.1 両眼立体視の原理

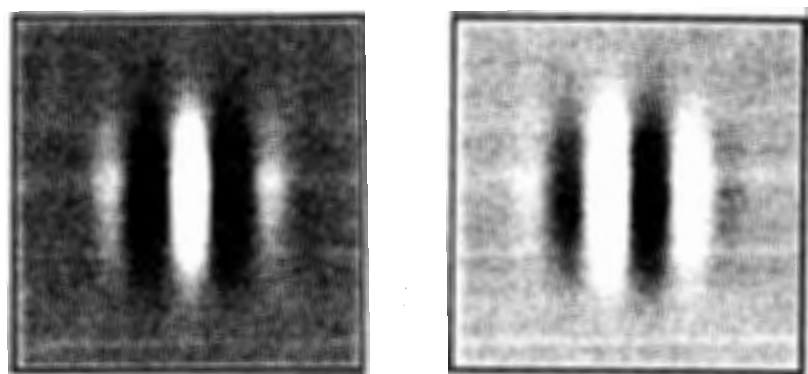


Fig.4.2 2次元ガボアフィルタ

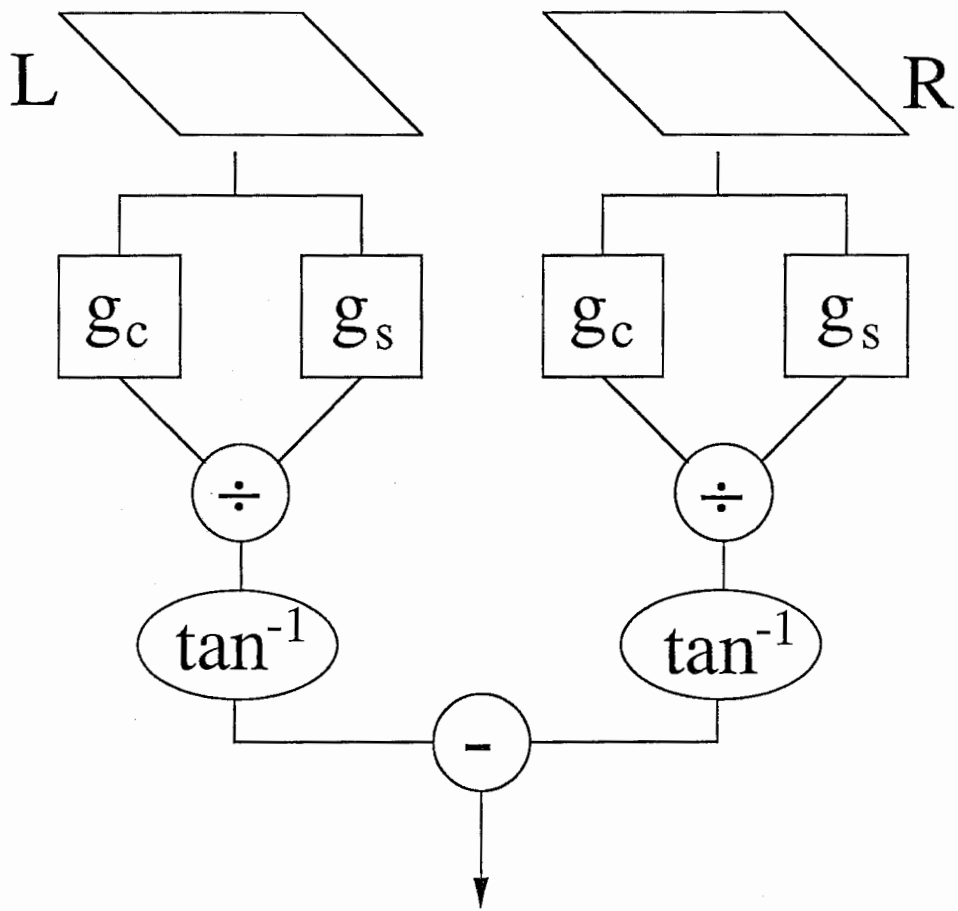


Fig.4.3 両眼視検出法

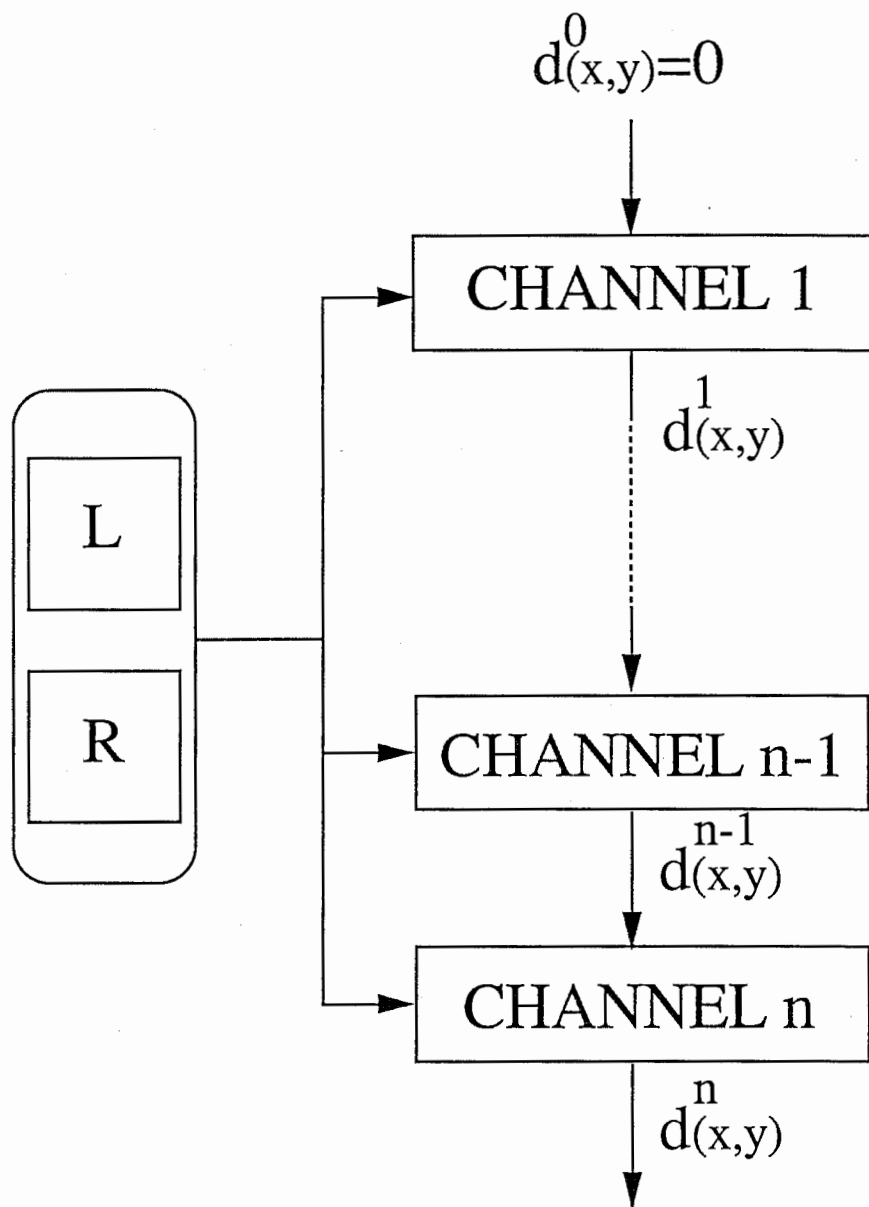


Fig.4.4 空間周波数チャンネルの統合

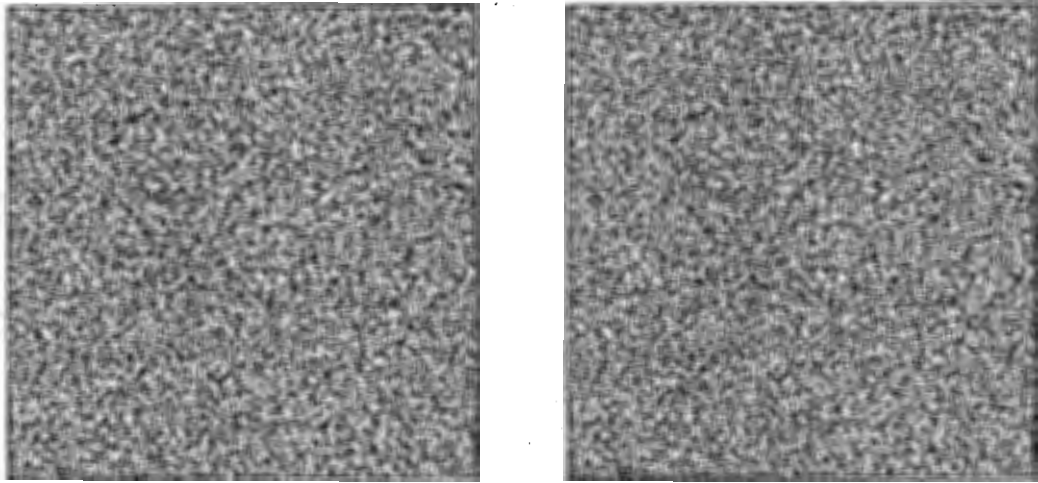
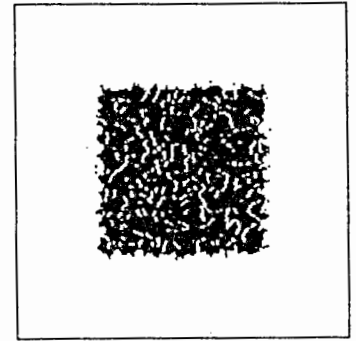
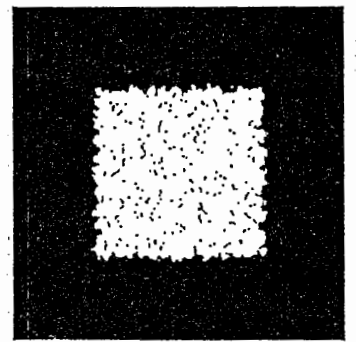
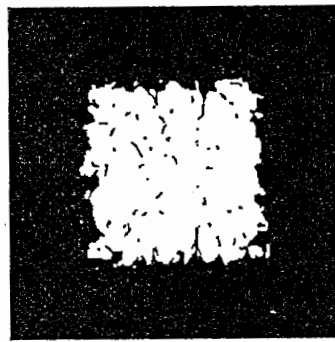
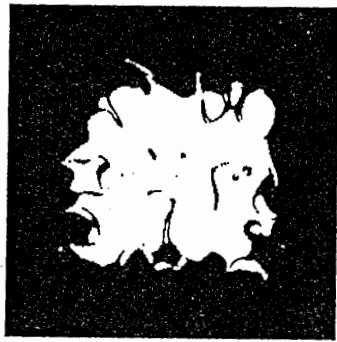


Fig.4.5 ランダムドットステレオグラムの例



(a) ch1

(b) ch2

(c) ch3

Fig.4.6 単一チャンネルによる視差検出結果(d=1)

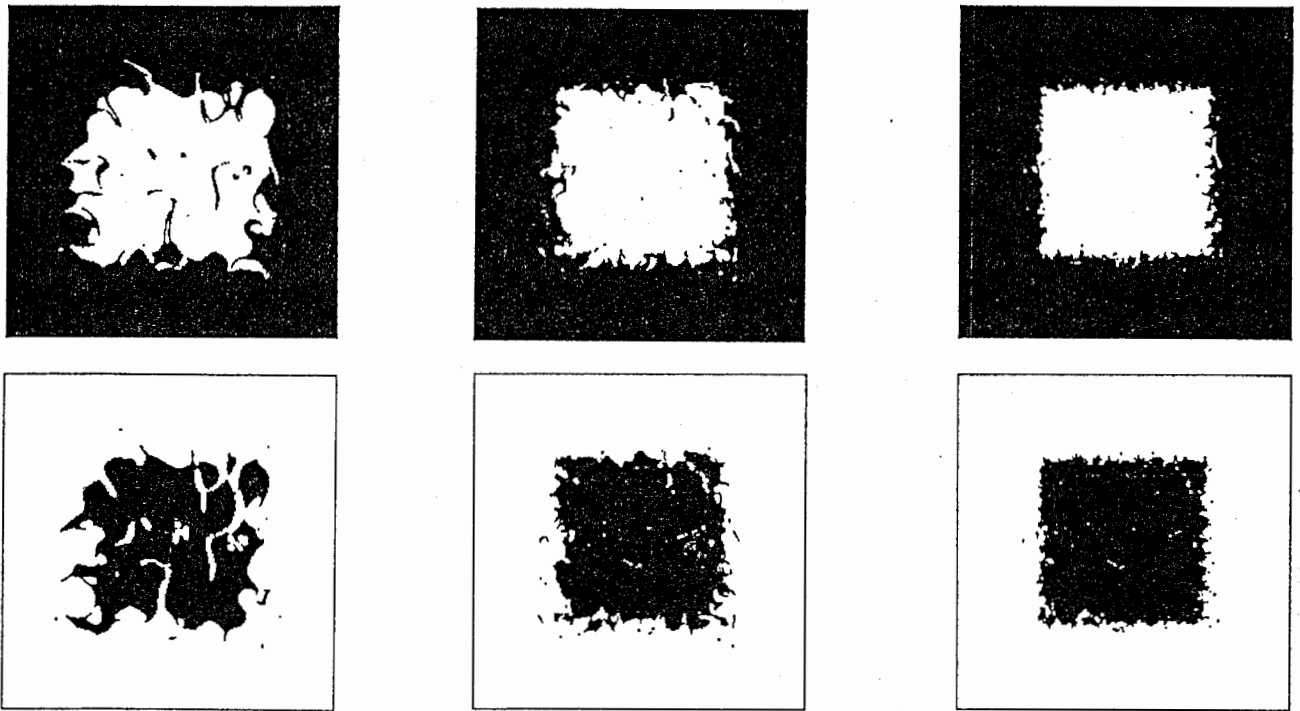
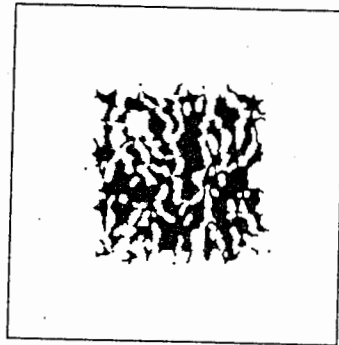


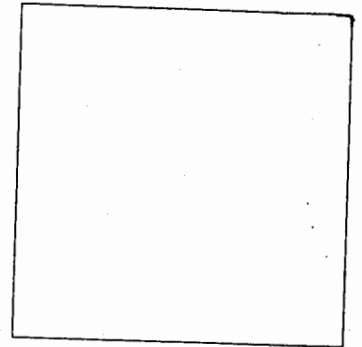
Fig.4.7チャンネル統合による視差検出結果(d=1)



(a) ch1



(b) ch2



(c) ch3

Fig.4.8 単一チャンネルによる視差検出結果($d=3$)

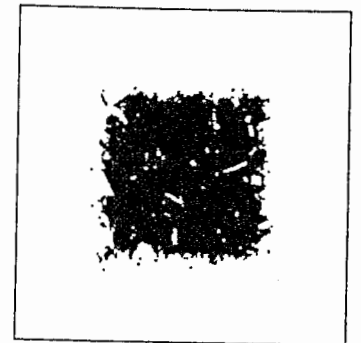
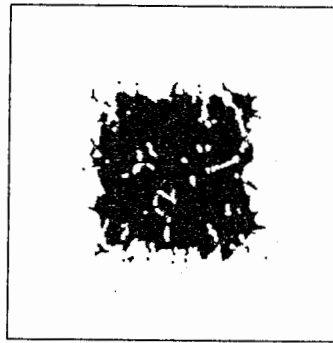


Fig.4.9 チャンネル統合による視差検出結果($d=3$)

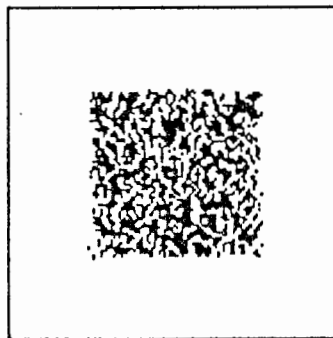


Fig.4.10 $d=3$ に対するch3の結果

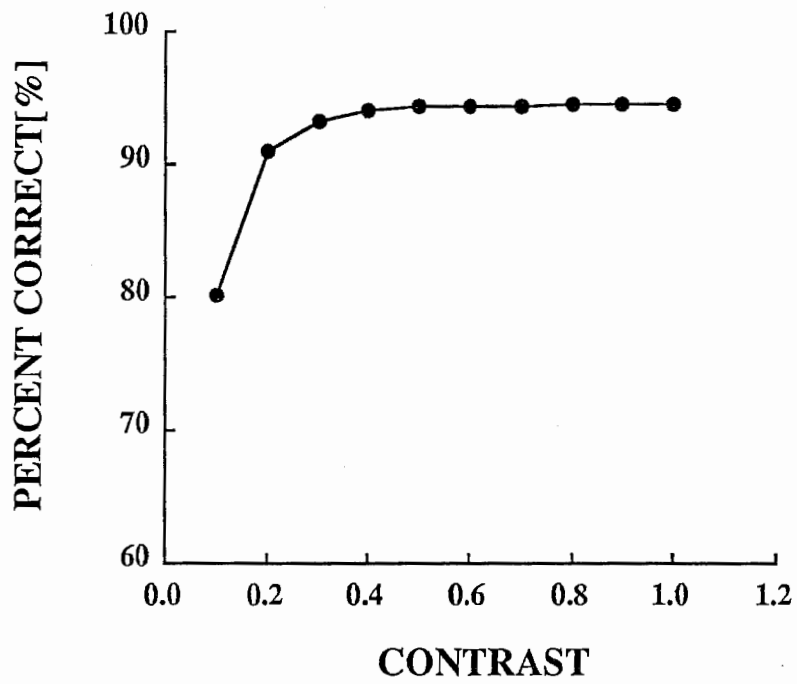


Fig.4.11 コントラストによる
視差検出能力の変化

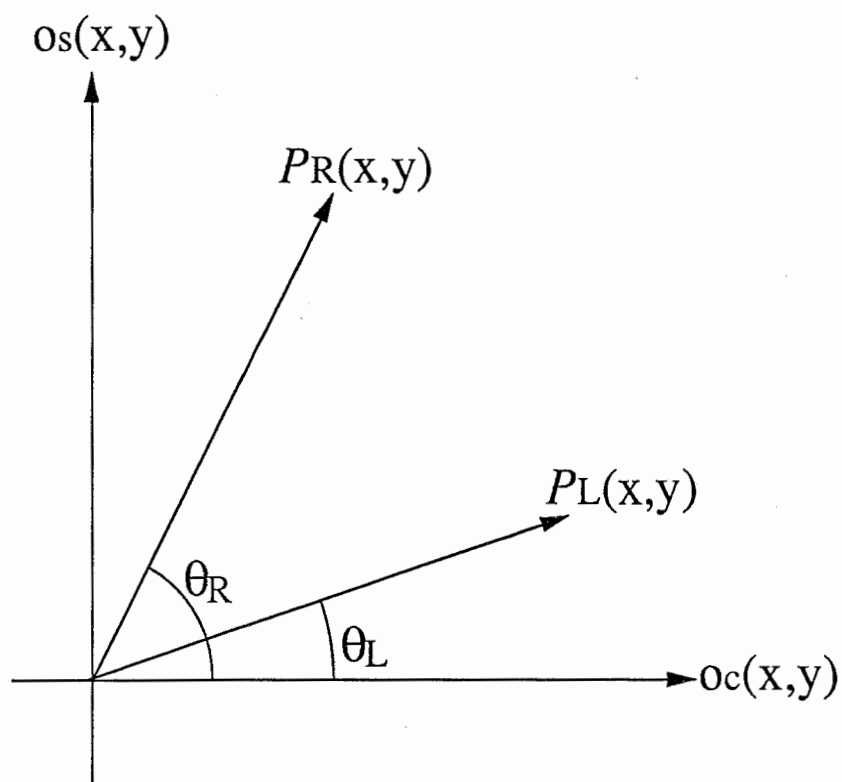


Fig.5.1 ガボアフィルタ対からの
出力ベクトル

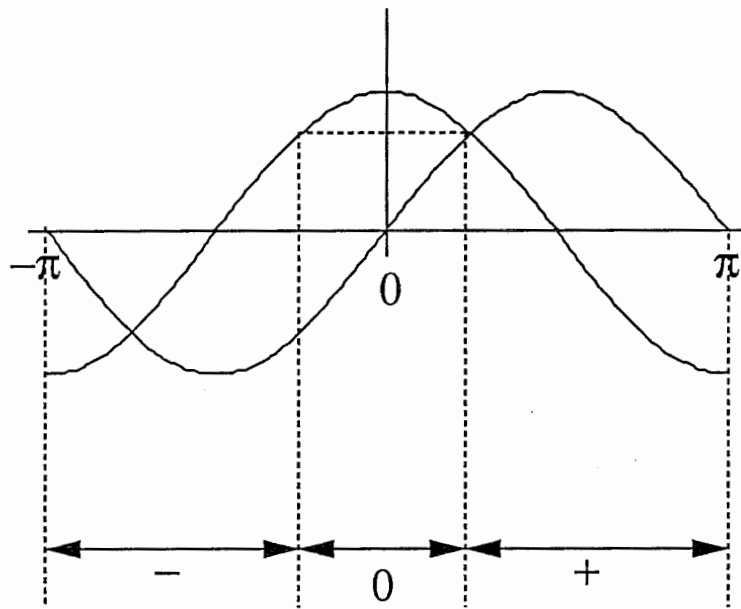


Fig.5.2 位相差符号

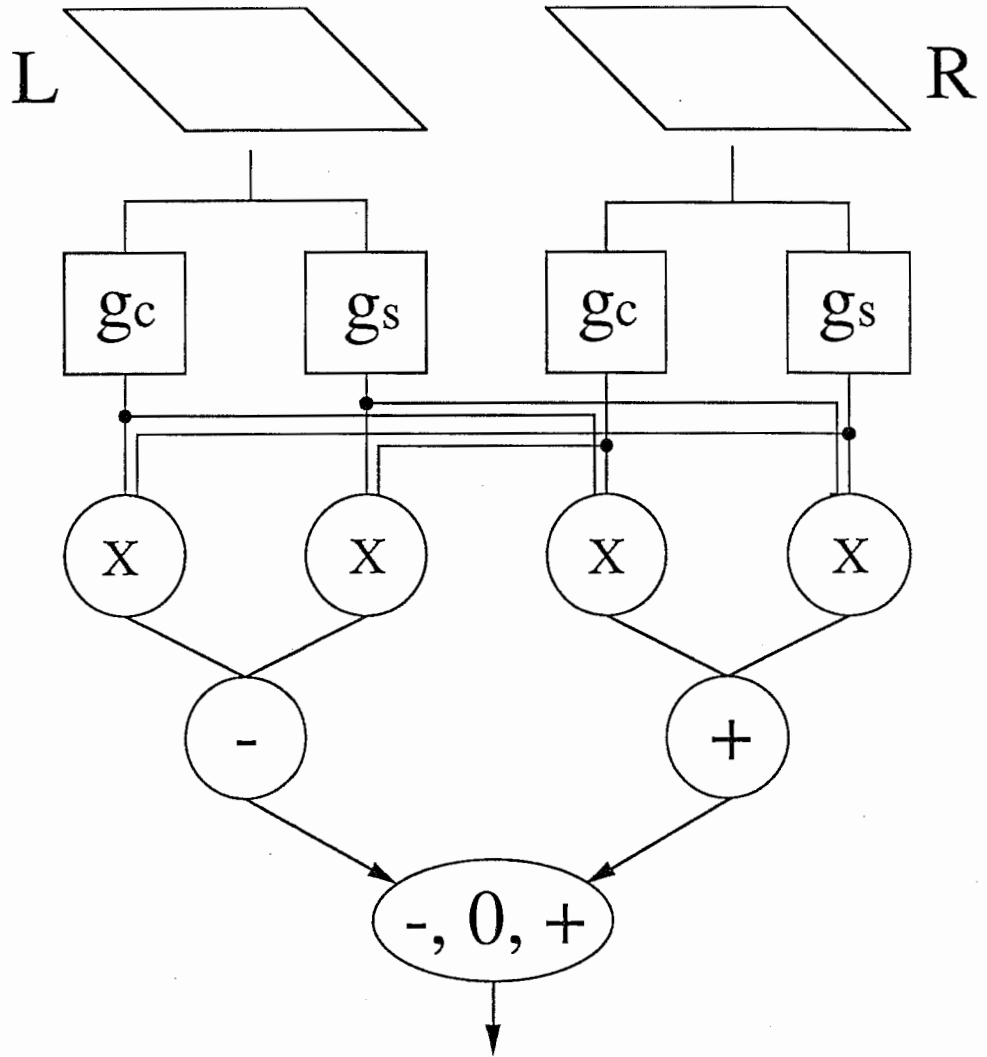


Fig.5.3 位相差符号検出法

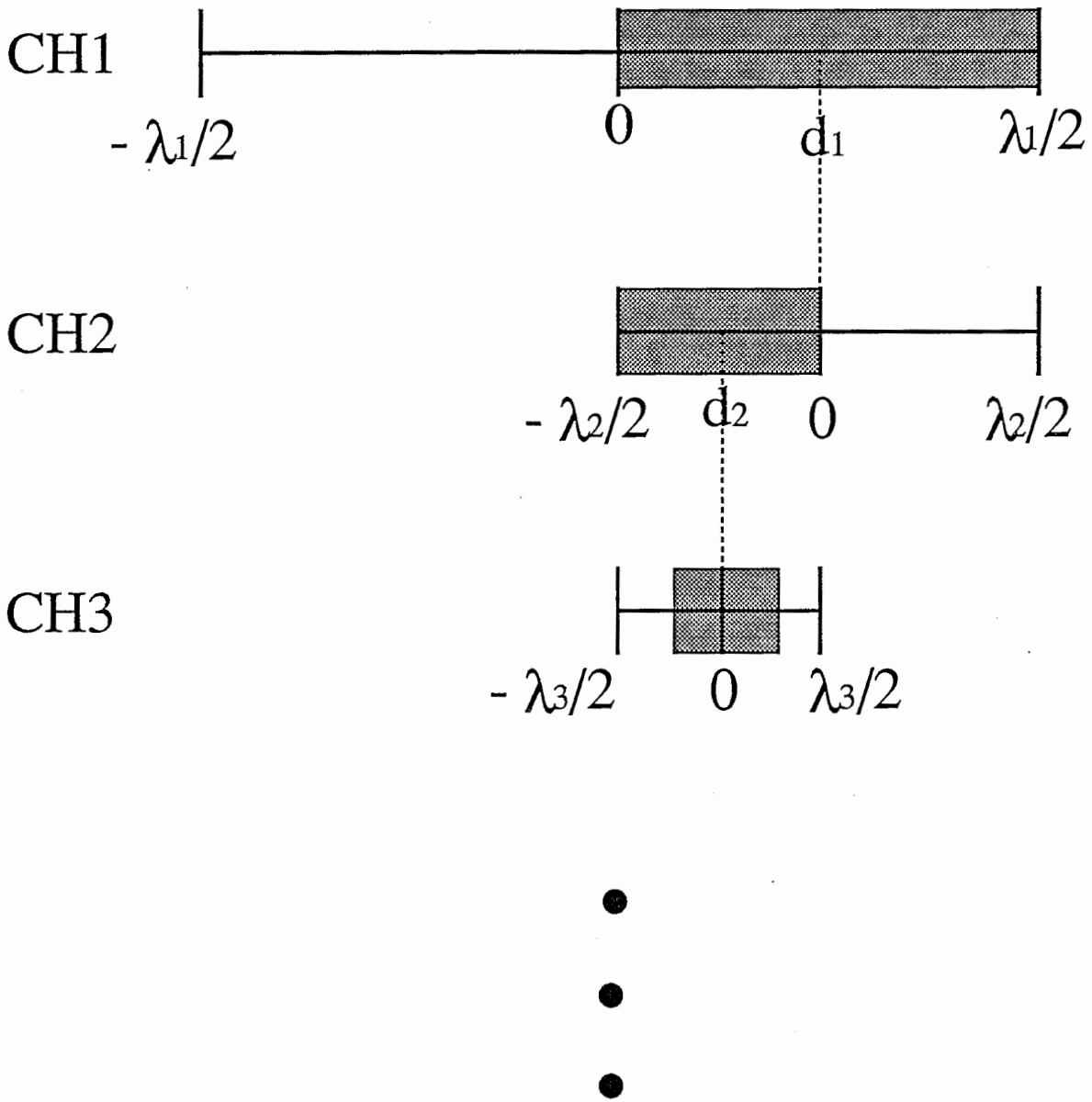


Fig.5.4 周波数チャンネル間の位相差符号の伝搬

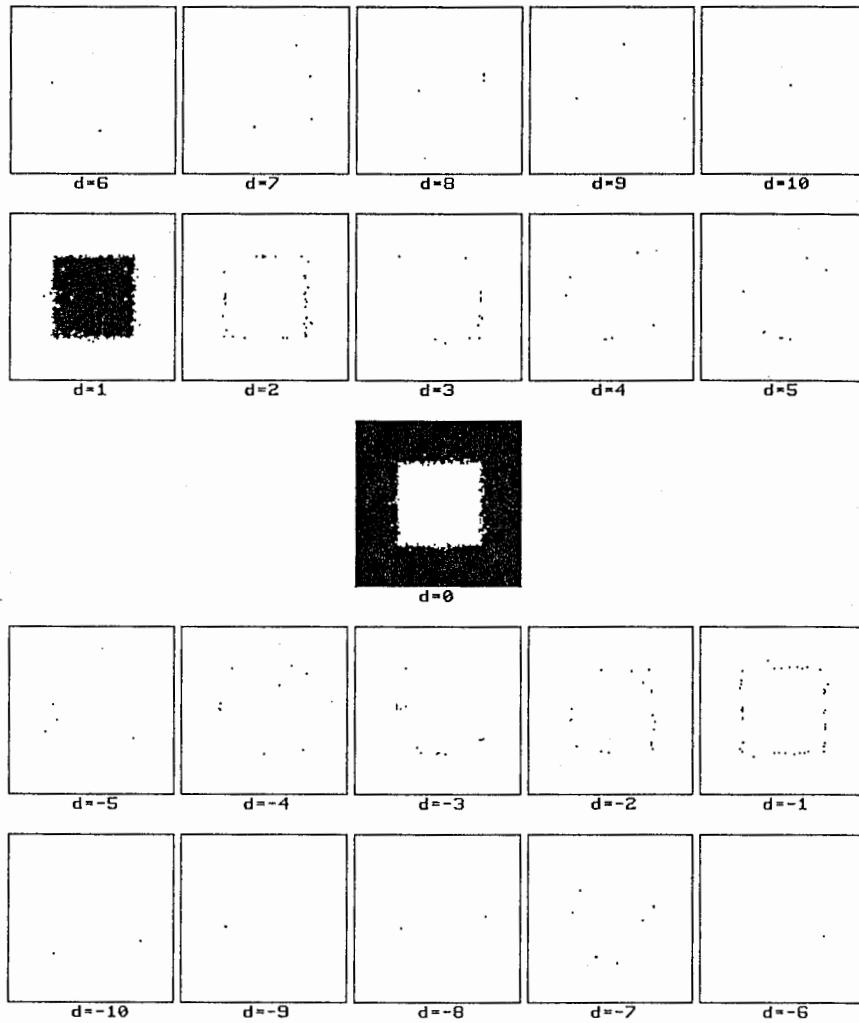


Fig.5.5(a) PD法による視差検出結果
(視差1、3チャンネル)

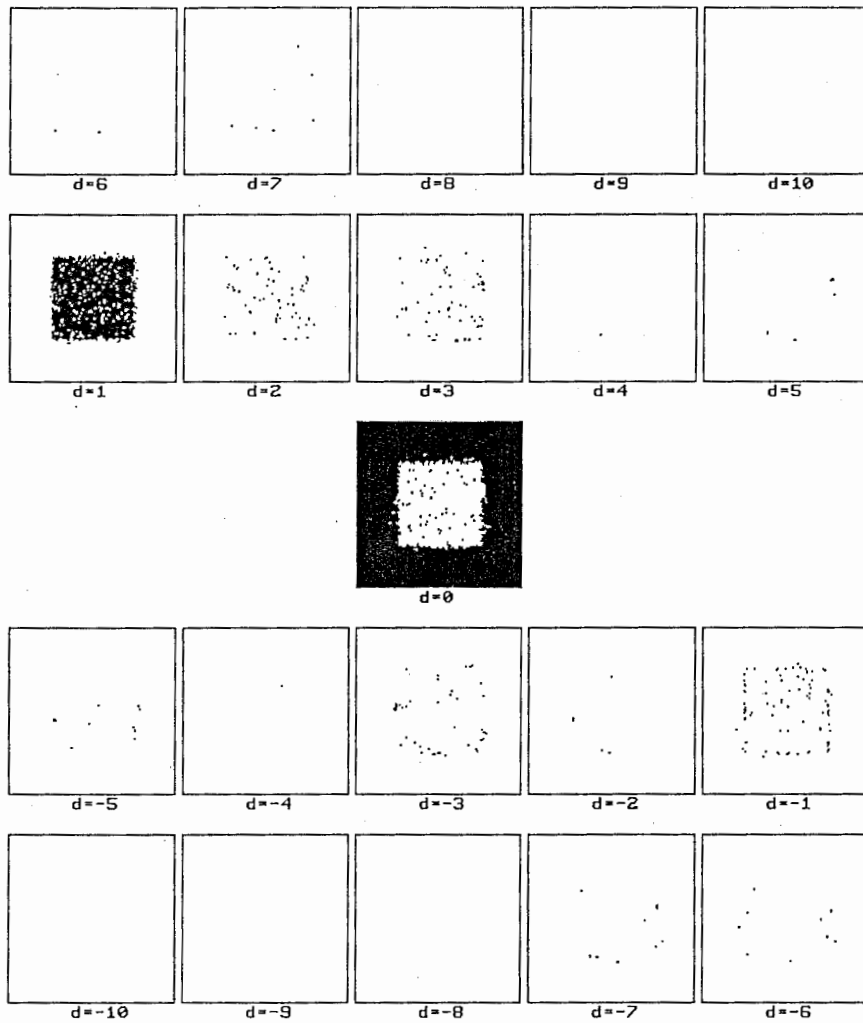


Fig.5.5(b) PDS法による視差検出結果
(視差1、3チャンネル)

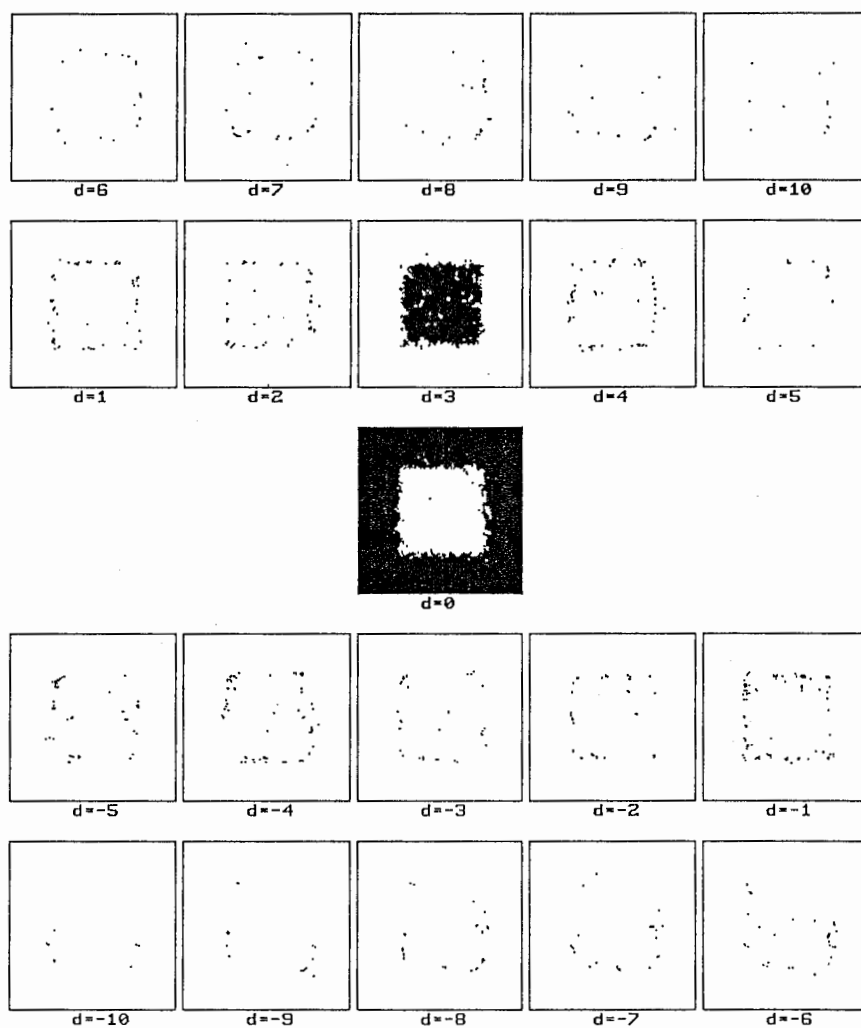


Fig.5.6(a) PD法による視差検出結果
(視差3、3チャンネル)

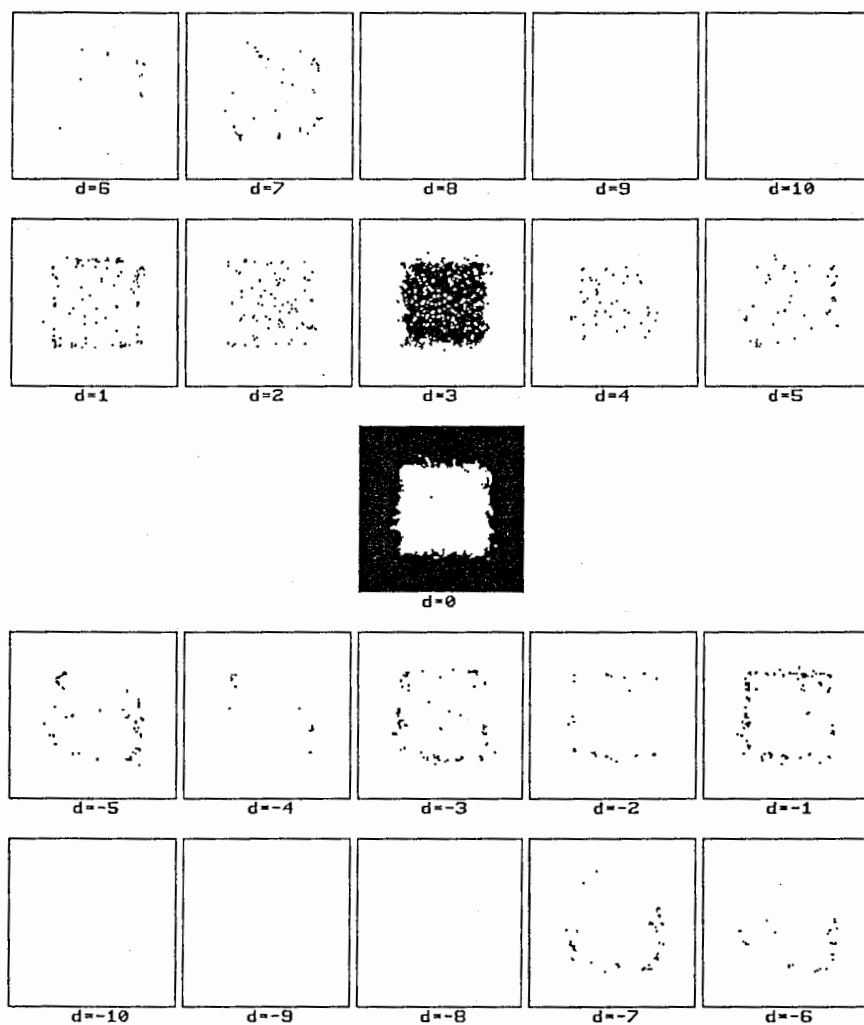


Fig.5.6(b) PDS法による視差検出結果
(視差3、3チャンネル)

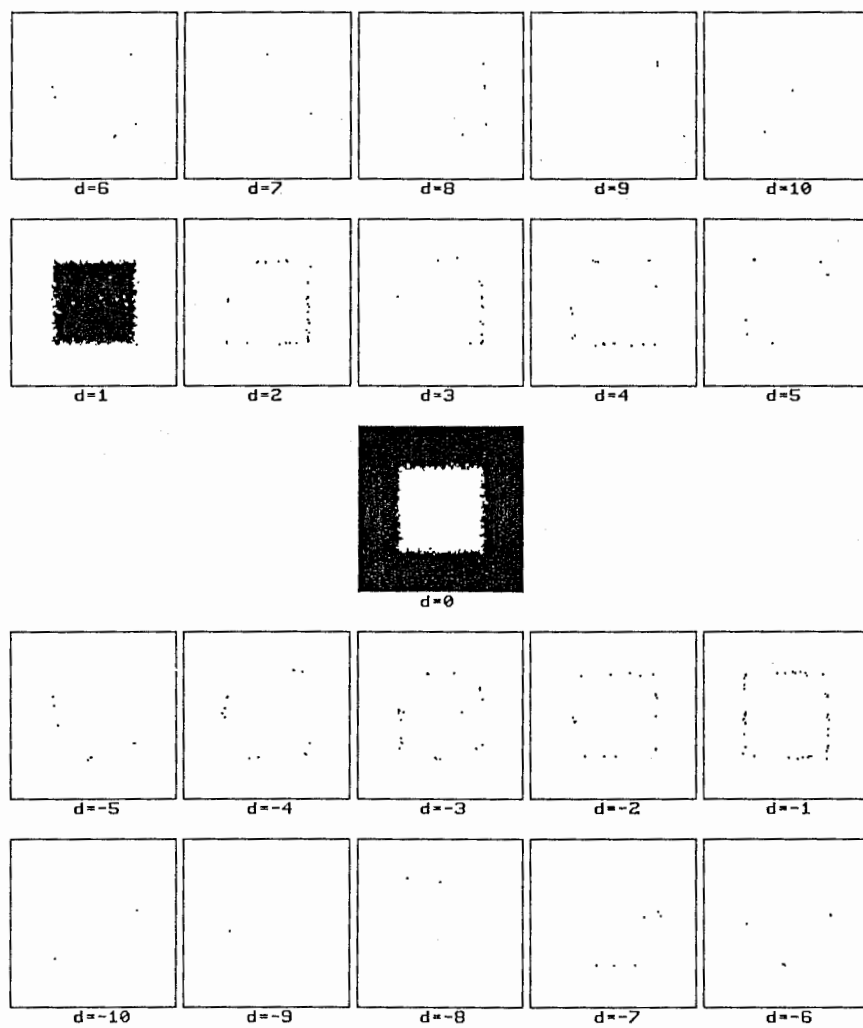


Fig.5.7(a) PD法による視差検出結果
(視差1、5チャンネル)

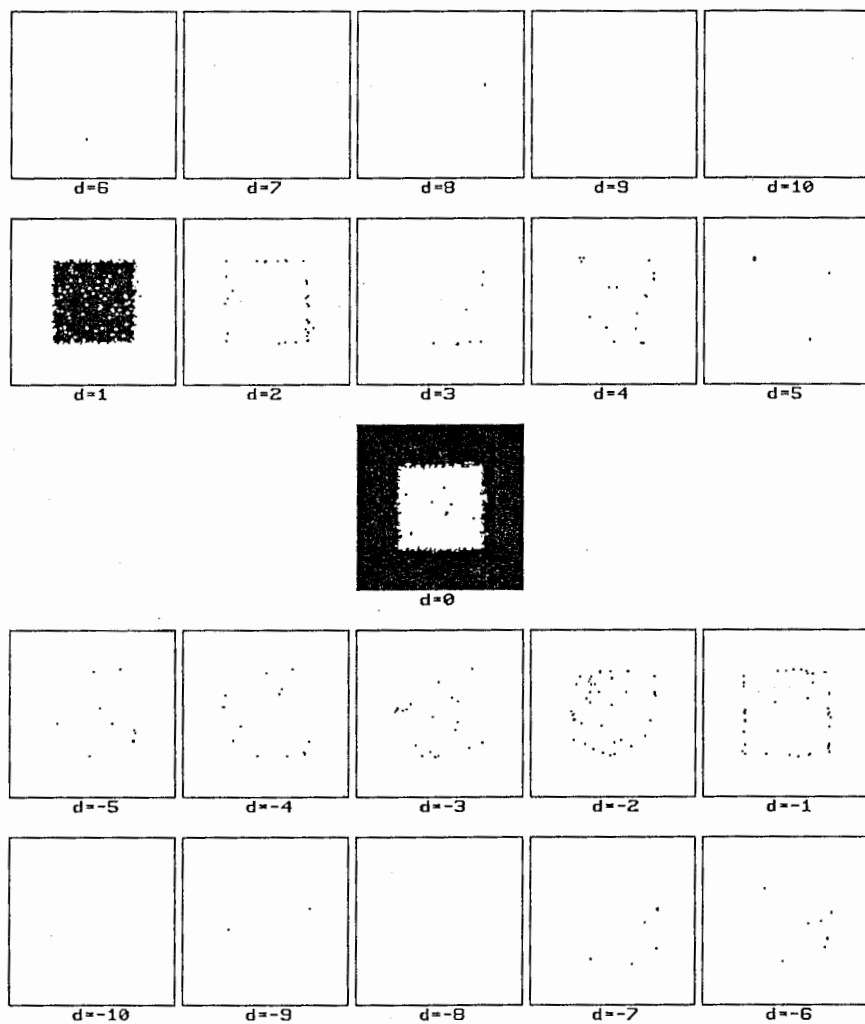


Fig.5.7(b) PDS法による視差検出結果
(視差1、5チャンネル)

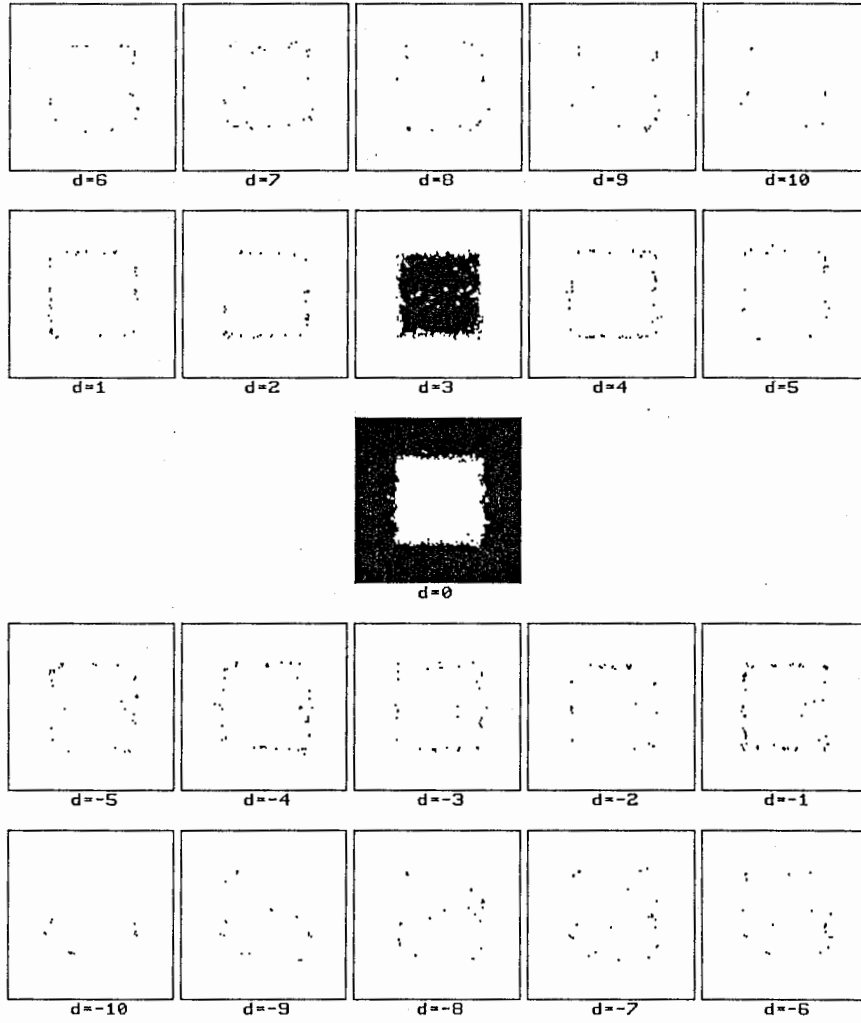


Fig.5.8(a) PD法による視差検出結果
(視差3、5チャンネル)

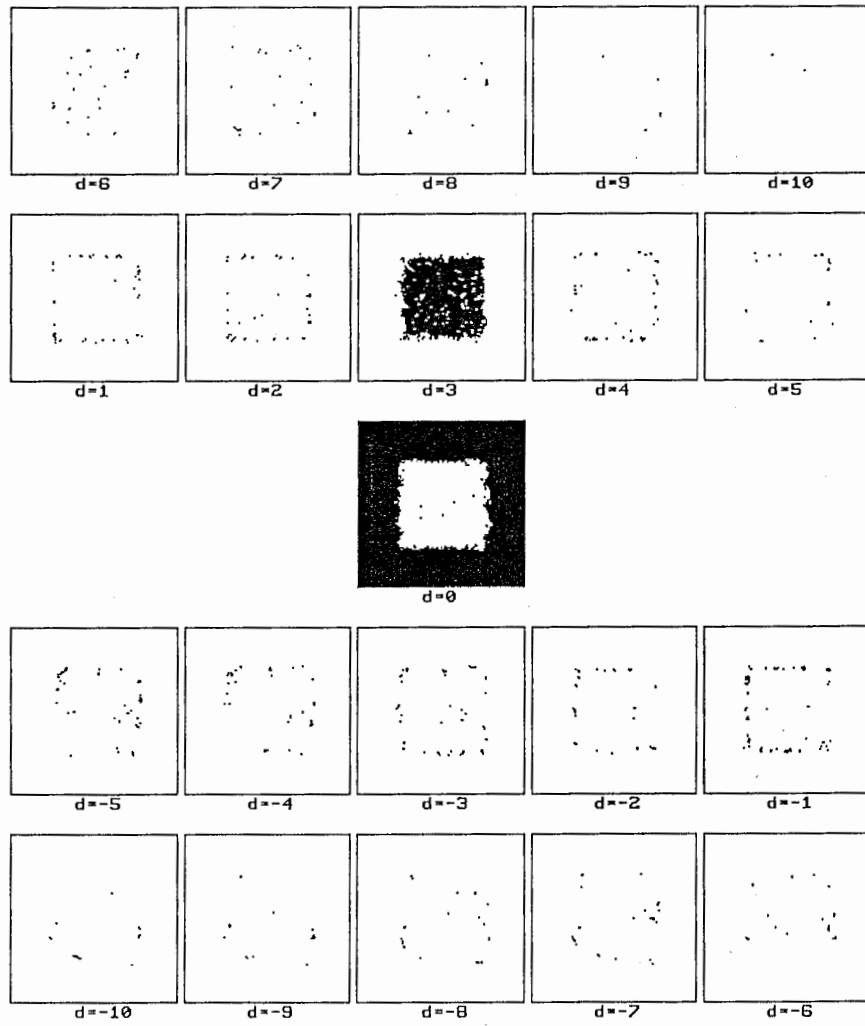


Fig.5.8(b) PDS法による視差検出結果
(視差3、5チャンネル)

Table 5.1 P D 法と P D S 法の性能比較
(3チャンネル)

		視差 1	視差 3
P D 法	正答率	98.0%	94.5%
	計算時間	40sec	
P D S 法	正答率	96.6%	92.6%
	計算時間	17sec	

Table 5.2 P D 法と P D S 法の性能比較
(5チャンネル)

		視差 1	視差 3
P D 法	正答率	97.9%	95.0%
	計算時間	70sec	
P D S 法	正答率	97.8%	94.8%
	計算時間	26sec	

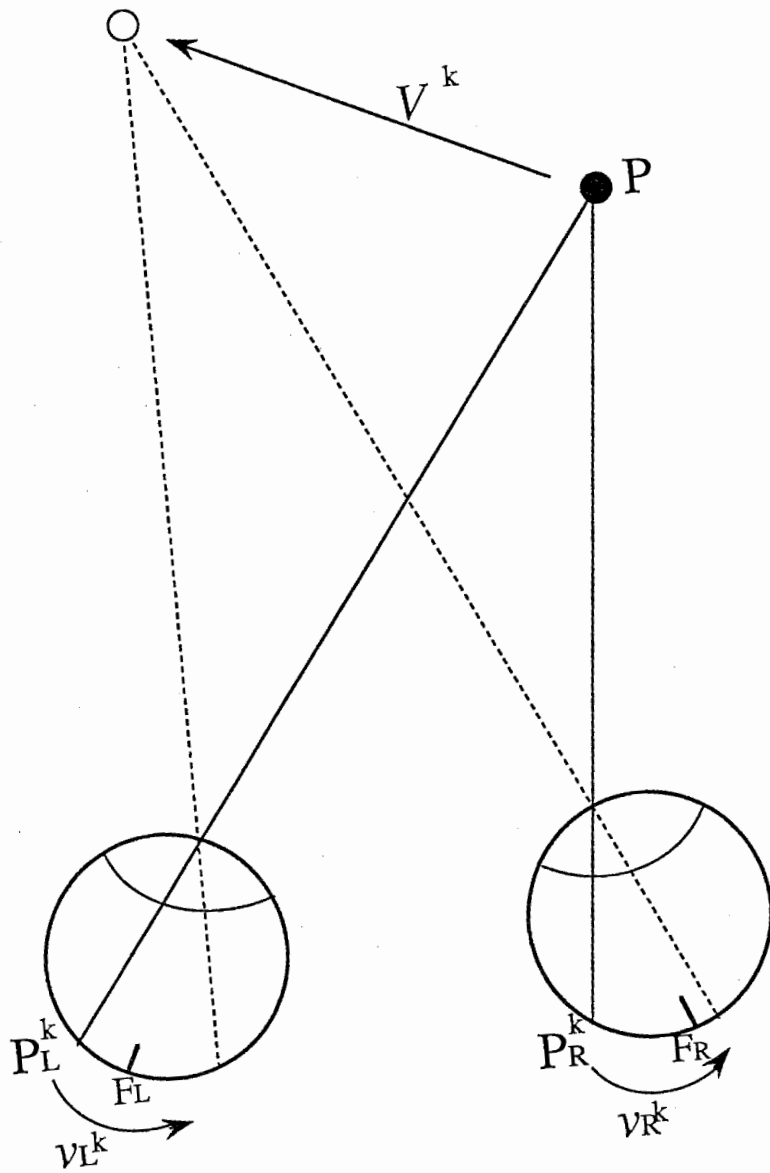


Fig.6.1 両眼における運動速度と
視差の時間的変化の関係

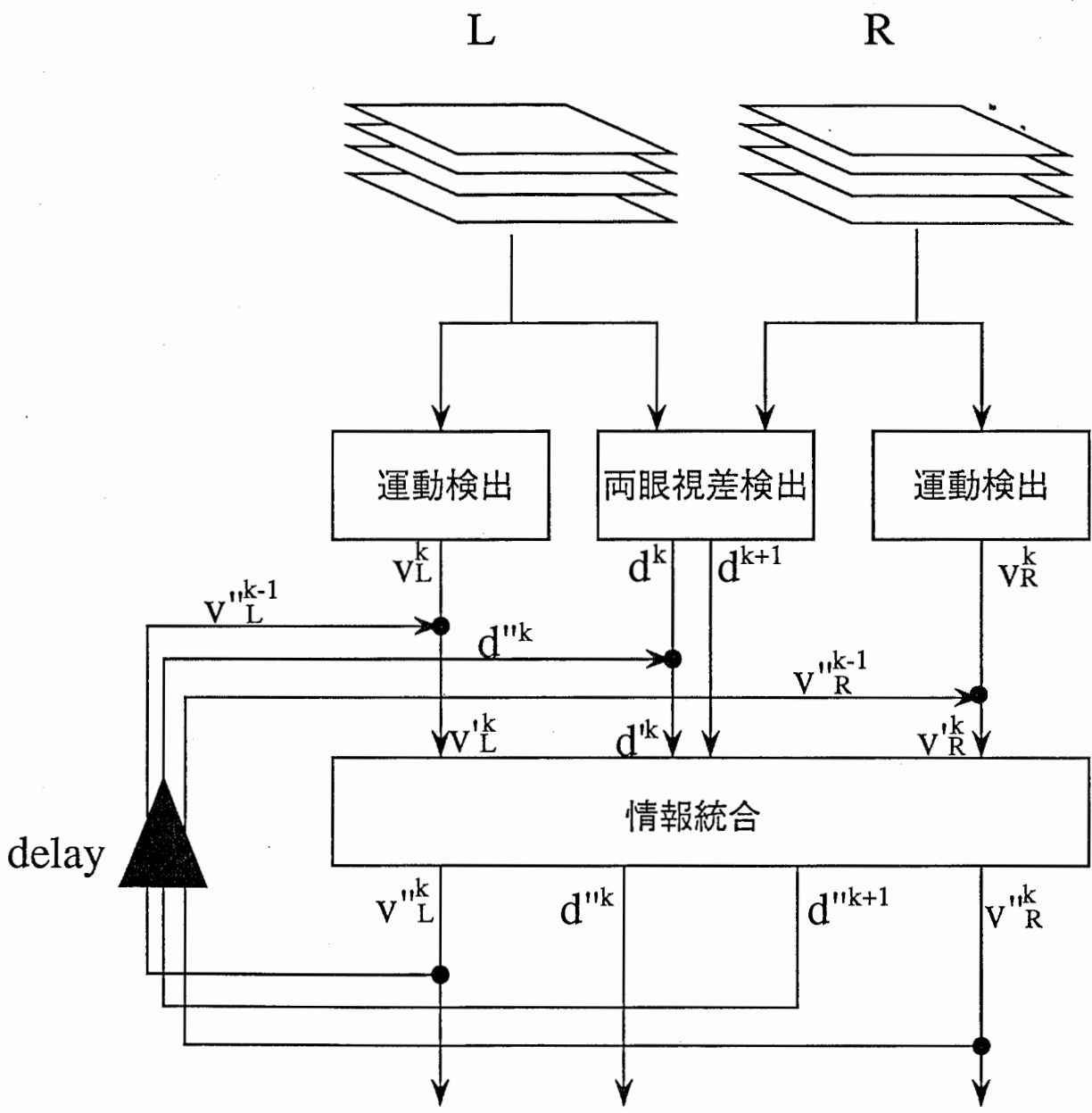
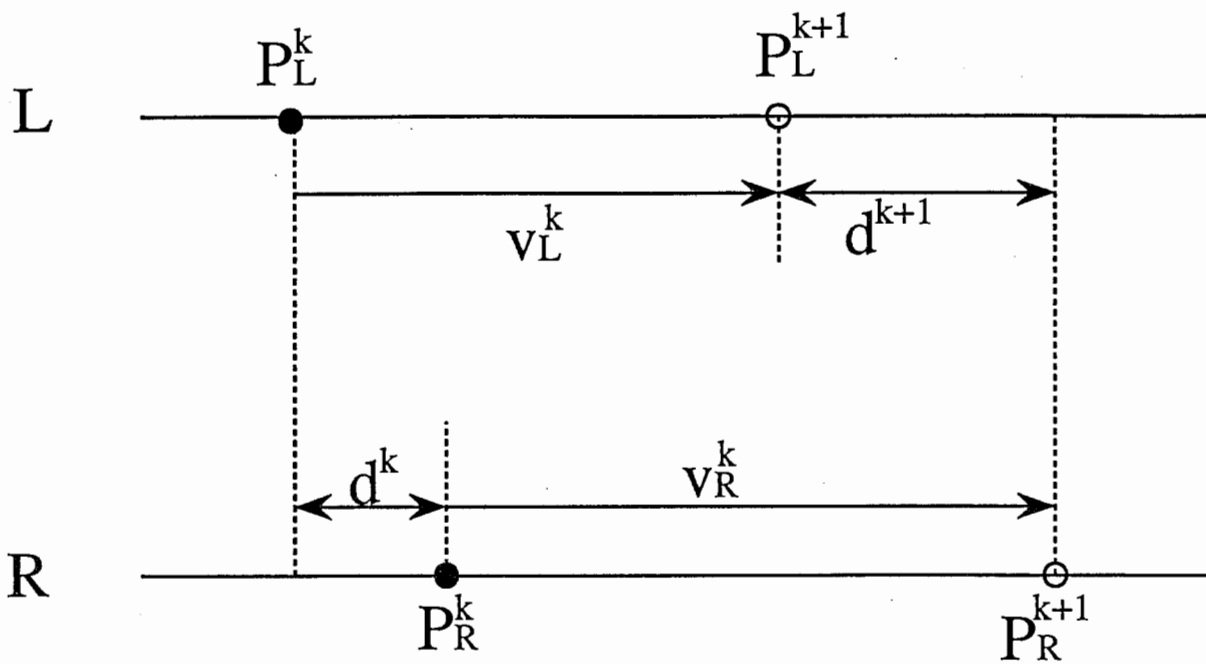
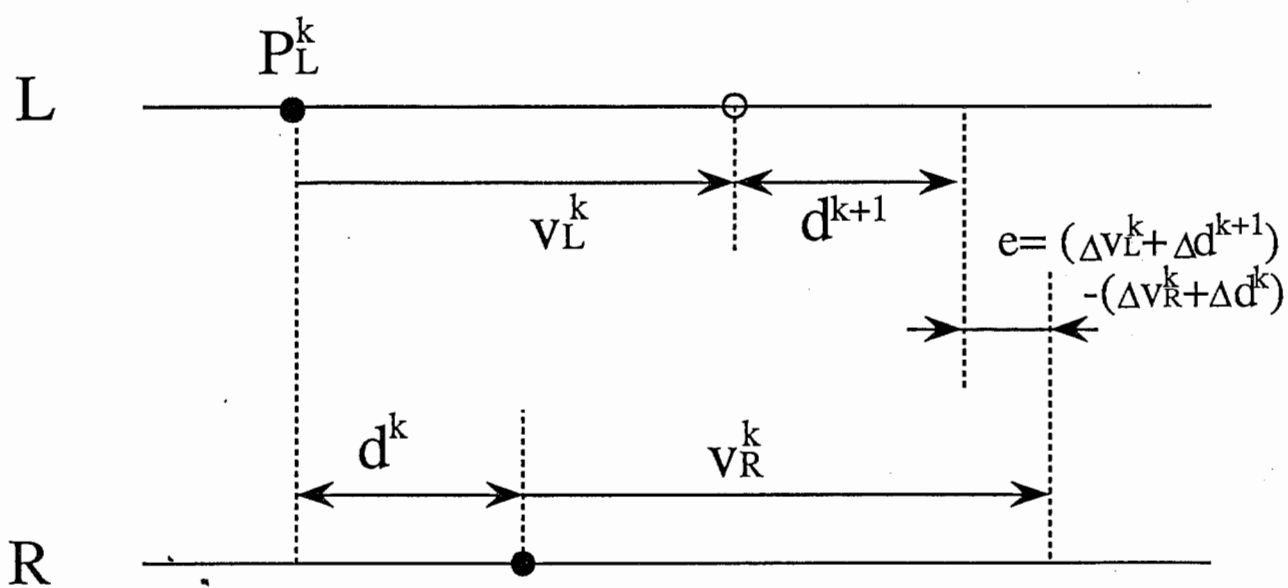


Fig.6.2 運動視モデルと両眼立体視モデルの統合



(a) 原理



(b) 観測誤差を含む場合

Fig.6.3 運動速度と両眼視差の整合法