

航空宇宙技術研究所報告

TECHNICAL REPORT OF NATIONAL AEROSPACE LABORATORY

TR-1301

Local 勾配法による動き測定精度

五味 広美

1996年7月

航空宇宙技術研究所

NATIONAL AEROSPACE LABORATORY

目 次

1. まえがき	1
1.1 構成	2
2. 動き検出アルゴリズム	2
2.1 理論	2
2.2 窓問題	3
2.3 移動距離の分解能と最大値	4
2.4 対象の変形と回転の影響	5
2.5 対象と背景との境界の検出	5
3. 実験	6
3.1 実験装置	6
(1) 対象	6
(2) 照明, ビデオカメラ, A/D変換器	6
(3) 撮影方式と計算機	7
3.2 処理プログラム	7
(1) ノイズ対策と空間フィルター	7
(2) 明度の空間的变化と時間的变化の計算	8
(3) 明度勾配の向きの計算	8
(4) オプティカルフローの方位の決め方	8
3.3 実験結果	8
(1) 画像の確認	8
(2) フィルターの効果	9
(3) 閾値 ΔE とフィルターの大きさ	9
(4) マルチチャンネルによる動き検出	12
(5) Local領域の大きさ	13
(6) Global法	14
(7) テンプレート・マッチング法	14
4. 考察	15
4.1 動き検出アルゴリズムの比較	15
(1) Local勾配法	15
(2) テンプレート・マッチング法	15
(3) 実験結果の比較	16
4.2 動き検出アルゴリズムの改良	17
4.3 動物の視覚から学ぶ	18
4.4 動き検出アルゴリズムの応用	19
5. まとめ	20
参考文献	21

Local 勾配法による動き測定精度*

五味 広美^{*1}

Accuracy and density of optical flows by gradient-based methods with local optimisation

Hiromi GOMI

ABSTRACT

The gradient-based method was evaluated as to its accuracy and space resolution in motion detection. The velocity was determined as the maximum of the optical flow components in the direction of the brightness gradient from the video image sequences in a miniature world. The standard deviations of measured velocity were 0.4 picture elements (p. e.) at a speed of 1.8 p. e., and 0.8 p. e. at a speed of 7.2 p. e. with four channels of space frequency. The method required about a thousand picture-element data to detect one velocity.

Keywords: motion detection, velocity measurement, gradient-based method, optical flow.

概 要

動き検出アルゴリズムとして勾配法とテンプレート・マッチング法に焦点を絞り調査を行なった。ヒトの視覚と同程度の空間解像度（視力）のビデオカメラと 1/500 の Boeing 747-200F の模型を使い、ヒトが約 2km の距離から実物を見た場合に相当する画像をコマ撮り撮影し、動き検出アルゴリズムの評価を行なった。オプティカルフローの明度勾配方向成分が最大値を取る方位をオプティカルフローの方位とする Local 勾配法を提案し、動きに応じて空間フィルターと閾値を調整することにより、1 画素程度の誤差で動きを検出し、航空機全体で約 15 点（視野角 0.3 度に 1 点の割合）の密度で動きを検出した。比較の為に、Global 勾配法とテンプレート・マッチング法による動き検出を行なった。

1. まえがき

動きの知覚は動物の視覚にとって最も基本的な機能である。動物が静止した外界の中を移動する時、動物の目には、近くのものほど大きく動き、遠くのものほとんど動かない様に見える。動物は自分の動きによって生じたこの微妙な網膜像の動きから対象までの距離を知り、周囲との衝突を避け安全に移動することが出来る。また、移動により網膜像は動いているにもかかわらず、動物に知覚されるのは安定した外界の中を自分が移動している

感覚である。この感覚を生じさせている処理により、外界で動くものがあつた場合、自分の動きによる外界の動きと切り離して、動くものを敏感に検出することが出来る。そして、動くものが危険なものかどうかを急いで判断して、危険な場合は逃避する行動に移らなければ生存が危うい。

動物が特別の努力もしないで視覚により外界を理解し行動しているので、人工視覚の実現が難しいこととは 60 年代の多くの研究者は認識していなかったが、計算機シミュレーションにより問題の難しさが明らかになった。3 次元の世界が眼にどの様に見えるのかは光学により厳密に計算可能で、コンピュータ・グラフィックスやバーチャル・リアリティの中心技術であるが、それに対して、視覚は、2 次元に圧縮された情報から 3 次元情報を復元する

* 平成 8 年 3 月 21 日受付 (received 21 March, 1996)

* 1 数理解析部 (Computational Sciences Division)

ので逆光学とも呼ばれ、数学的には逆問題を解くことに相当する。視覚には3次元の世界を復元させる為に多くのモジュールが並列的に存在するとして、陰影からの形状復元、距離からの形状復元、動きからの形状復元等の、Xからの形状復元（shape from X）の研究が行なわれている¹⁾。一般に、画像情報だけで外界の3次元情報を復元することは出来ず、外界に対する何らかの仮定、外界のモデルが必要である。

動き検出アルゴリズムの研究では、動きが別の方法で求まったとして、動き情報から対象の形状や運動を復元する研究が最初に行なわれ、その後、動きを画像処理により求める研究が行なわれる様になった²⁾。動きを画像処理により求める代表的なアルゴリズムは勾配法とテンプレート・マッチング法の2つで、70年代、80年代は計算量の少ない勾配法が優勢であった。計算機処理能力の向上に伴い、アルゴリズムが単純で並列化により高速化が可能なテンプレート・マッチング法が優勢になり、90年代に入り、動きをビデオレイトで計算可能な専用の計算機が開発されている。一見はなやかに見える人工視覚の研究だが、パターン認識の基本的問題（例えば、背景から対象を分離）が解決出来ず、研究初期に単純な問題を解決しただけで、人工視覚の研究は手詰まりの状態にある。

それにもかかわらず動き検出アルゴリズムの研究を推進すべきと考えるのは、第一に、初めに述べた様に動き検出は動物の視覚に共通する基本的な機能で、視覚モジュールの原型であると考えからである。工学的な観点からは、画像処理により得られる数個の動き情報がビデオカメラの手振れ防止に使われ役立っている様に、研究の各段階での成果に利用価値があるからである。また、動き情報は画像の明度情報ほど細かく求めることは出来ないが、情報量が少ないことは逆に明度情報より高度の後処理が可能であることを意味する。静止画の理解は主に1枚の画像を使うのでノイズの影響を受け易いが、視点移動や時系列の複数画像から得られる結果を統合することにより、誤差の影響を少なくし安定した動き情報を得ることが出来る。それらを可能にする条件は、今後とも計算機の処理能力が向上し続けることである。

この報告では動き検出アルゴリズムの見直しを行ない、アルゴリズムの動き検出の精度と密度を調べることを目的とする。この様な初歩的なレベルの研究からスタートするのは、ヒトの視覚とはいえ、網膜で検出されない情報を利用することは出来ないはずと考え、ヒトの視覚が検出している動き情報の精度と量とを確認する為である。

1.1 構成

第2章では、動き検出アルゴリズムの中で代表的な勾配法とテンプレート・マッチング法をレビューし、「正しい」動きを決定する為に解決しなければならない窓問題

について検討する。検出可能な動きの測定精度と範囲（最大値）を予測し、対象の変形、回転、背景の影響について述べる。

第3章では、オプティカルフローの明度勾配方向成分の方位分布から「正しい」動きを決定するLocal勾配法のアルゴリズムを提案し、航空機模型をコマ撮りした画像に対して適応し、実験により動き測定精度を求める。空間フィルターの大きさと閾値を最適化した4つの空間周波数チャンネルを使った動き検出を行ない、Local領域の大きさの影響を調べ、Global法とテンプレート・マッチング法による動き検出を行なう。

第4章では、動き検出アルゴリズムの実験結果の比較を行ない、今後の研究を進める方向と人工視覚を実現する為に動物の視覚研究から学ぶべきことを考察し、動き情報の応用として着陸時のパイロットの視覚支援を検討する。

2. 動き検出アルゴリズム

この章では、勾配法とテンプレート・マッチング法を5つの項目に関して比較検討することにより、動き検出アルゴリズムの違いを明確にする。

2.1 理論

視覚により検出可能な動きは目に見える動きで、実際の動きとは異なることがあり、この問題は視覚情報を使う限り原理的に避けることが出来ない。例えば、表面が一樣な球が回転していても静止している様に見えたり、移動中に対象の表面の傾きや照明条件が変化することにより明度が変化し実際の動きと異なる動きが見える場合がある。しかし、この様な場合はそれ程多くないので、一般に画像上の明度変化は対象の移動によってだけ生じる（対象と同じ運動をする視点から対象を見れば、対象の明度は変化しない）と仮定して、視覚情報から動きを検出している。この動きはオプティカルフロー、または、オプティックフローとも呼ばれている。

動き検出アルゴリズムは使用出来るハードウェア（動物の視覚では眼と脳、計算機視覚ではビデオカメラと計算機）により制約を受ける。動物の視覚では、網膜上に2次元的に配置された視細胞により明度分布が空間的に並列的に、時間的に連続的に測定されている。しかし、視細胞は一定の明るさの刺激が入力し続けると出力を出さなくなるので、網膜からの出力段階で既に明度の時間的な変化が明度情報として脳に伝えられている。それに対して、計算機視覚ではビデオカメラで画像を取り込む時、CCD画素の電荷を走査しながら読み出すので、明度は時間的に一定の間隔（ビデオレイト）で離散化されている

が、一定の明るさの刺激に対しては一定の出力をビデオカメラは出し続ける。したがって、動物の視覚と計算機視覚とでは適した動き検出アルゴリズムが異なることが考えられる。

ここでは動き検出アルゴリズムの中で代表的な勾配法（動物の視覚のアルゴリズムに近いと考えられる）とテンプレート・マッチング法（計算機視覚に適すると考えられる）について述べる。

① 勾配法

時刻 t 、座標 (x, y) に於ける明度 $E(x, y, t)$ が、 Δt 時間後に座標 $(x + \Delta x, y + \Delta y)$ に移動し、その間、明度の変化が無いものとする、

$$E(x + \Delta x, y + \Delta y, t + \Delta t) = E(x, y, t)$$

が成り立つ。明度が時間と位置に対してなめらかに変化するものとして、左辺をテイラー展開し、2次以上の項を無視すると、次式が得られる。

$$\frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} + \frac{\partial E}{\partial t} = 0$$

画像の明るさの空間的、時間的変化をそれぞれ、

$$\frac{\partial E}{\partial x} = E_x, \quad \frac{\partial E}{\partial y} = E_y, \quad \frac{\partial E}{\partial t} = E_t,$$

位置の時間変化、すなわち、速度（オプティカルフロー）の各成分を、

$$\frac{dx}{dt} = u, \quad \frac{dy}{dt} = v,$$

の短縮形に置き換えることにより、

$$E_x u + E_y v + E_t = 0 \quad (1)$$

のオプティカルフローの拘束方程式（Constraint equation）が導かれる²⁾。

明度の時間変化と空間変化は画像から測ることが出来るが、未知数は u, v の2つあるので（画像は一般的には2次元なので）、上の拘束方程式に加えて、もう1つの拘束条件が必要になる。このことは、上の拘束方程式から、オプティカルフローの明度勾配方向 $(E_x, E_y)^t$ 成分は

$$\frac{E_t}{\sqrt{E_x^2 + E_y^2}} \quad (2)$$

と決定出来るが、オプティカルフローの明度等高線に沿った方向成分に関しては何も情報が得られないことから解かる。これは窓問題（Aperture problem）として知られている。

② テンプレート・マッチング法

時刻 t における画像の一部を切り出したもの（これをテンプレート、鋳型と呼ぶ）を、時刻 $t + \Delta t$ の画像上を上下・左右に移動させ、重なっている画素間の明度の差が最小になる位置と、時刻 t の画像のテンプレートを切り出した位置とが対応しているとして、移動量を求める方法である。

対応の判定は、重なっている画素間の明度差の絶対値あるいは差の2乗をテンプレート全体で和（これを残差と呼んでいる）を取り、これが最小になる位置を対応点とするのが一般的である。2乗の方が明度差が大きな画素間の影響が大きく、対応の判定の感度が良いのではと予想したが、実際には両方とも同じ結果が得られたので、高速に計算出来る絶対値を選んだ。

テンプレート・マッチング法の理論はこの様に単純だが、画素間の明度差の計算量が大きいため、計算時間を短くする方法が主として研究されている³⁾。最近ではテンプレート・マッチングを高速に行なう専用のハードウェアの開発が行なわれている⁴⁾。

テンプレート・マッチング法は時刻 t と $t + \Delta t$ の2枚の画像が必要なので、動物の視覚が採用しているとは考えられない。現在の計算機視覚が動物の視覚に遠く及ばない現状を考えると、動物の視覚が採用していないアルゴリズムの将来性はないかもしれないが、テンプレート・マッチング法はノイズに強く、並列計算による高速化が容易で、計算機で扱う画像データは時間的にビデオレイトで離散化されているので、計算機視覚のアルゴリズムとして適している。

2.2 窓問題

小さな窓を通して画像の一部、例えば、直線の動きを見た場合、見える動きは直線に垂直方向の動きで、直線方向の動きは知覚されない。先に述べた様に、理論的に直線に垂直方向（明度の変化する方向）の動き成分を決定することは出来るが、直線方向（明度の変化しない方向）の動き成分は不定である。これは窓問題（Aperture problem）と呼ばれる。

画像から「正しい」（脚注）動きを決定出来る条件は画像が複数の方向の明度変化を含むことなので、広い視野を確保することにより窓問題解決の可能性が大きくなる。しかし、視野に比例して計算量が増加するので、外界に応じて視野を最適な大きさに変える必要がある。ヒトの視覚では、視力の優れた中心窩（視野角直径が1度以下）から、眼球運動により左右 ± 15 度、上8度、下12度、頭部運動

脚注）ここで言う「正しい」動きは全体的に見てより真の動きに近いという意味である。

により左右 ± 50 度, 上 20 度, 下 45 度, 身体の移動によりさらに広い範囲まで視野を変更することが可能である⁵⁾。

① 勾配法

・ Global 法

画像の明度にはノイズが含まれ, 背景と移動体との境界が解からないことに対応する為に, 画像全体で速度がなめらかに変化するというもう1つの拘束条件を導入し全体的に最適な速度場を決定する方法で, Local 法に対比して Global 法と呼ばれる。この方法は数学的には緩和法または弛緩法 (Relaxation method) (脚注1) である。Global 法は, 条件によっては Local 法より細かな空間分布で速度を求めることが出来るが, 画像全体で繰り返し計算をしなければならないので, 並列計算に向かず, 解が得られるまでの時間があらかじめ解からないので高速の動画像処理に適さない。

・ Local 法

ある領域 (数十画素以上, これを Local 領域と呼ぶことにする) が同じ速度で運動すると見なせるほど小さく (剛体の仮定), かつ, 複数の明度勾配方向を含むほど大きい条件を同時に満足出来ると仮定する。具体的な窓問題解決法には以下に示す幾つかのバリエーションが考えられる。

- i) Local 領域に含まれる画素ごとにオプティカルフローの拘束方程式を求め, それらを連立させて最小2乗解を求める方法⁷⁾ や, 拘束方程式の交点の最頻値を解とする方法⁸⁾。
- ii) Local 領域に含まれる画素ごとにオプティカルフローの明度勾配方向成分を求め, 横軸に明度勾配方向, 縦軸に明度勾配方向成分の大きさをプロットし, 最大値を取る方向を動きの方法 (この報告で提案, 詳しくは, 3.2 を参照)。
- iii) 入力画像に対して特定の方向の明度勾配方向だけを強調 (抽出) する空間フィルターをかけ, 独立な拘束方程式を2つ求め, それを連立させて解く方法⁹⁾。

② テンプレート・マッチング法

画像から切り出したテンプレートの明度に2次元的な空間分布があれば「正しい」動きを決定出来, 明度が1次元あるいは均一に空間分布していれば窓問題が生じる。対応点が安定した真の対応点である為には, 対応点での残差がセンサーとしてのビデオカメラの安定性によって予想される大きさ程度に小さいことに加えて, 対応点の周囲すべての向きで残差が急激に増加することが必要である。

対応点の周囲すべての向きで残差が急激に増加するのを厳密に確認することは困難なので, 対応点から適当な距離の幾つかの位置での残差が, 対応点の残差より大きいことを確認するだけで済ませることが多い。テンプレートの明度が均一になるにつれて, 方位ごとの明度の分散 (インタレスト値¹⁰⁾) が小さくなり, 窓問題が生じる可能性が高くなる。インタレスト値がある値以下の場合, テンプレート・マッチング探索をしないことにより, 窓問題を避けることが可能である。しかし, インタレスト値の計算に時間がかかる割に, 得られる動きの信頼性を高くする為にインタレスト値の閾値を高くすると得られる動きが少なくなる可能性が大きい。

2.3 移動距離の分解能と最大値

動きを計算する基準になる時間はビデオレイト (NTSC では, フィールドレイト約 $1/60$ 秒, フレームレイト約 $1/30$ 秒) 単位で任意に変えることが出来る。この方法は勾配法とテンプレート・マッチング法の両方に共通して使うことが出来るので, ここでは動きの計算に必要な移動距離測定の分解能と最大値について検討する (脚注2)。

① 勾配法

理論的前提として移動距離の範囲で明度の空間的变化率は一定でなければならないので, この条件を満足する範囲で移動距離の分解能を小さく, 最大値を大きくすることが出来る。しかし, 移動距離が大きくなるにつれて, 移動範囲で明度の空間的变化率が一定である領域が少なくなるので, 実用上の移動距離の最大値が存在する。デジタル画像を得る時点で, 明度は普通8ビット (0 ~ 255 ADU) 程度で離散化されるが, 明度差の最大値は100 ADU 程度で, 明度には最大10 ADU の誤差が含まれる。隣接するデジタル画素間で100 ADU 程度の明度差がある場合, 理論的には1 ~ 10 画素程度の移動距離の分解能が得られる。しかし, CCD 画素間の電荷の漏れは40 % 以上もあり, CCD 画素の感度は画素内で大きく変化するので, 1画素以下の分解能は信頼性が低い¹¹⁾。また, ビデオカメラの解像度をヒトの眼と同程度とした場合 (CCD 画素が見る角度は約1分), ビデオカメラが固定されていたとしても, ビデオカメラの視線を1 ~ 10 画素程度に安定させることは, 振動などの影響により難しい。10 画素間で100 ADU の明度差がある場合, 1画素程度の距離分解能で, 10対1程度のダイナミックレンジ (= 最大値分解能) が得られるが, 勾配法により達成出来る実用的な分解能とダイナミックレンジはこの程度と推定される。

脚注1) 画像の世界では弛緩法と言えば, Resenfeld らの Relaxation 法を応用した Labeling 法を意味するので, 注意が必要である⁶⁾。

脚注2) 動物の視覚では基準になる時間がビデオレイトより短いので, 移動距離を一定とし, 通過時間から速度を求める方法も考えられる¹²⁾。

測定可能な移動距離の最大値は画像の特性によるので、実験により決めることが出来る。

一般に、 n 個の測定値の平均を取ることににより、平均値に含まれるノイズは $1/\sqrt{n}$ になるので、100 画素程度の空間的なフィルターにより明度誤差は $1/10$ 程度になる。動き検出では時間平均を取ると動き検出の時定数が大きくなるので、空間的なフィルターを使わなければならない。また、明度の空間的変化率一定の勾配法的前提を成立させる為にも空間的なフィルターが必要である。しかし、フィルターを掛けることにより明度の空間的変化率も小さくなるので、ダイナミックレンジはそれ程拡大しないと考えられる。

移動距離の最大値を 10 画素とするには、直径 10 画素以上の空間フィルターをデジタル画像に掛けなければならないが、その計算を行なうには時間がかかるので、映像信号をローパスフィルターに通すことによりアナログ的に空間フィルターを掛け、AD 変換する方が現在の信号処理技術レベルでは実用的である。

② テンプレート・マッチング法

この方法の移動距離の最小分解能は 1 画素であり、最大値はテンプレート・マッチングの探索範囲で決まる。ただし、後で述べる様に、画素間の明度を内挿することにより移動距離の分解能を 1 画素以下にすることは可能である。

テンプレート・マッチングの探索範囲は検出可能な移動距離の大きさの上限を決定するので、探索範囲は大きい方が良いが、計算時間が増大する。テンプレートの大きさに較べて大きな探索範囲を取ると、誤対応に陥る危険性が増大するので、探索範囲はテンプレートの大きさ程度かそれ以下にするのが適当である。筆者が行なった大きなテンプレート（最大 256×256 画素）を使った実験では対応点からのテンプレートのずれがテンプレートの大きさの半分程度までは残差（不一致の程度を示す）がずれ量に応じて単調に増加したので、対応点を安定して探すことが出来た¹²⁾。

テンプレートを大きくすると、ノイズに強くなり、結果が安定し、窓が大きくなることにより窓問題が解決出来る場合が多くなるが、計算時間がかかる様になる。また、対象が変形した場合、対応点を見つけることが困難になると言われている。デジタル画像の大きさは一般に 512×512 画素程度なので、テンプレートの大きさは 3×3 画素から 15×15 画素程度に選ぶ場合が多い。

ホーム用ビデオカメラの手ぶれの影響を取り除く為に、森村ら¹³⁾は、画像信号をローパスフィルターに通してはかして飛び飛びにサンプリングしたデジタル画像に対して、テンプレート・マッチング法により、左右 ± 24 画素、上下 ± 8 ラインの範囲で、移動ベクトルを求めている。さらに大きな動きベクトルを検出する為に演算範囲を移

動させ、最大の動きベクトル検出範囲を左右 ± 47 画素、上下 ± 15 ラインとしている。対応点付近では残差は直線的に減少するものとし線形補間をして、元の画像の画素単位の正確さで対応点を求めている。これらの工夫により残差の計算量を $1/100$ 以下に削減している。

2.4 対象の変形と回転の影響

① 勾配法

移動距離の範囲で明度の空間的変化率が一定という理論的前提が満足される限り対象が変形したり回転しても正しい動きを検出出来る。狭い範囲では明度の空間分布は平面で近似出来るが、移動距離が大きくなるにつれて平面からのずれが大きくなり、変形や回転の影響が大きくなる。

② テンプレート・マッチング法

テンプレートが大きくなると変形や回転の影響を受け易くなるので、移動距離があまり大きくならない様に画像間の時間間隔を短くし、小さなテンプレートで動きを検出した方が良いが、時間間隔を短くすると検出される動きも小さくなり、相対的に測定精度が悪くなる。立体視の対応点問題で 256×256 画素の大きなテンプレートを使い、風景画を使用した実験では、10 % の倍率差（拡大・縮小）、8 度の傾き差（回転）があっても、テンプレートの並進運動だけで対応点探索は可能であった¹²⁾。

テンプレートを変形させて残差を小さくすることも可能だが、変形のモデルが必要になる。これは対象の運動を検出することを意味し、動き検出の上位の視覚機能と考えて、次の研究課題として計画している。

2.5 対象と背景との境界の検出

勾配法でもテンプレート・マッチング法でも対象と背景との境界により領域を分割して、それぞれの領域に対して動き検出アルゴリズムを適応する必要がある。領域分割は動きが連続か不連続かを評価する何らかの函数を考え、繰り返し計算により評価値を最小または最大にする方法が提案されている。この評価函数を決めることが難しいことと計算に時間がかかるので、境界の存在を無視して動きを求めると、一般に、対象と背景との境界付近の動きは対象と背景の平均的な動きが求まる可能性が高いが、しばしば周囲とは大きく異なる動きが計算されることがある。

① 勾配法

Global 法では、画像データに含まれるノイズの影響が少なくなる様に第 2 の拘束方程式に含まれる動きのなめらかさを決めるパラメータを調整すると、動きがなめらかに変化し境界の位置があいまいになる。境界の位置が明瞭になる様にパラメータを調整すると、動きがノイズ

に敏感に反応し必要以上に多数の境界が求まるようになる。Local 法では、境界の位置を正確に求める為に動き検出の空間分解能を上げるには Local 領域を小さくしなければならないが、そうするとノイズの影響を受け易くなる。

② テンプレート・マッチング法

Local 勾配法と同様な制約をテンプレート・マッチング法は受ける。テンプレート・マッチング法で移動対象の領域が解かれている場合、移動体のみを含むテンプレートを使うことにより、正確な動きを検出することが出来る。しかし、移動体のみを含むテンプレートを作る為には、境界の位置情報が必要になる。境界を検出する為にテンプレートを小さくし動き検出の空間分解能を上げると、ノイズの影響を受け易くなる。

3. 実験

動き検出アルゴリズムの中で動物の視覚に近いと考えられる Local 勾配法によりオブティカルフローの明度勾配方向成分を求め、その方位分布の最大値をオブティカルフロー（動き）とする方法を実験で検証する。実際の世界にアルゴリズムを適応した場合、どの程度の精度と密度とで動きが求められるかが実験の主な目的で、検証項目は以下の3点である。

- ① 画像データに含まれるノイズ対策（閾値と空間フィルターの大きさ）
- ② オブティカルフローの明度勾配方向成分の向きと大きさの精度と密度
- ③ ②の方位分布からオブティカルフローの向きと大きさの精度と密度（Local 領域の大きさに対する依存性）

動き検出アルゴリズムには、動きが正しく求まる前提条件が幾つかあるので、検出しようとする動きの大きさに合わせて、装置と処理プログラムを最適化して、実験をしなければならない。

3.1 実験装置

(1) 対象

管制塔から見た滑走路端の飛行機を対象とする。飛行機までの距離が増加するにつれて機体表面の特徴が見えにくくなり、動き検出が難しくなるが、地上にいる飛行機の監視距離としては2 km 程度が最長である。通常の天候では管制塔から飛行機が明瞭に見える距離である。

実験としては実際の飛行場で撮影した画像に対して動き検出アルゴリズムを検証することも考えられるが、飛行機の正確な動きが解からないこと、撮影のやり直しに手間がかかるので、動き検出アルゴリズムの研究の最初の段階としては、模型を使い実験室で撮影することにした。

実験室で取れる最大の距離が5 m 程度なので、縮尺1/400 程度の模型が最適だが、模型を特注した場合に製作にかかる費用と時間と出来ばえと、既製のミニチュアモデルのそれらを比較して、今回は既製の Herpa Miniaturmdelle 社（ドイツ）の Boeing 747-200 F の1/500 の模型（脚注）を使うことにした。模型は金属製で、車輪は回転するが他の部分に較べて大きめで、模型の表面は鏡面的に光を反射し、一部に接合線が見えるが、機体の塗装は均一に仕上がっているため、今回の実験には充分使えるものと判断した。このシリーズの模型には普通の旅客機もあるが、今回は客室窓の無い荷物機を選び、特徴（模様）の少ない機体に対してどの程度の動きが求まるかを調べた。

模型の大きさに合わせて1/500 の縮尺で実験を行なった。模型の移動は糸で引っ張ることにより連続的に動かすことも可能であるが、(3)で述べる様にコマ撮り撮影することにしたので、模型を移動台に載せて、1/100 mm の正確さで動かした。1 軸の移動台の上に幅120 mm の無地の厚紙を置き滑走路とし、その上に模型を載せた。したがって、模型と一緒に紙（滑走路）も移動という不自然さはある。背景は木目模様の見える合板を使った。

ビデオカメラと模型との間に、管制塔の高さを80 m として、その1/500 の高さの160 mm の差を付けた。滑走路端あるいはその付近の誘導路上の飛行機は、管制塔から見て、ほぼ視線方向に動く場合が多い。画像処理は原理的に視線方向の動き検出の感度が悪く、また、正確な実験の設定も難しいので、飛行機は視線に対して直角に動く様に移動台を配置した。

(2) 照明、ビデオカメラ、AD 変換器

模型の表面は鏡面的に光を反射するので、照明は実験室天井の蛍光灯を使った。照明光が反射してビデオ出力が飽和しない様に、模型の位置を調整し、蛍光灯と模型の間に蛍光灯の光を拡散させる包装用プラスチックを置いた。その結果、模型とビデオカメラ間の距離は3.55 m となった。

ビデオカメラの解像度をヒトの眼と同程度にする為に、1/3 インチ CCD に焦点距離50 mm のレンズを着け、CCD 画素の視野角を約0.5 分とした。ビデオカメラの CCD 画素とデジタル画像の画素とが1対1の対応が取れたデジタル画像を得る為に、ビデオカメラとAD 変換ボードのクロック周波数が同じもの（14.31818 MHz）を使い、AD 変換ボードからビデオカメラへ水平と垂直の同期信号を入力し、水平同期信号の同軸ケーブルの長さを変えタイミングを調整した¹¹⁾。

勾配法の理論的前提を満足させる為には、 γ 特性は1で

脚注) その後、1995 年には日本製の1/400 のプラスチック模型が発売された。車輪が格納された状態で、表面の仕上げにむらがある。

なければならない。

(3) 撮影方式と計算機

入力画像はコマ撮り撮影をし、AD 変換した画像データを計算機のハードディスクに記録したものを使い、処理結果にビット単位の再現性を確保した。

動き検出アルゴリズムが使われる条件では、通常の連続撮影で画像データを取得し、実時間処理をしなければならないが、使用した計算機 MC 68040 (25 MHz) では実時間処理は出来ないで、計算機の処理能力の点から、入力画像を 1 度記録して処理することが必要である。アナログ方式の光ディスクは手元にあるが、旧式で SN 比が良くないので使わなかった。処理結果にビット単位の再現性を確保する為に、AD 変換した画像データをハードディスクやメモリーの記録媒体に実時間で記録する手段はまだ高価なので、コマ撮り撮影をすることにした。

撮影対象が大きく動く場合、飛び越し走査の影響により連続撮影とコマ撮り撮影の画像データは異なることが考えられる。コマ撮り撮影では対象を固定しているので飛び越し走査の影響が解からないので、最終的には連続撮影のデータを使った実験が必要になる。ビデオカメラのノイズや照明の変動がランダムであれば、連続とコマ撮り撮影とではそれらの影響の差は出ない。

将来、動き情報を使いビデオカメラで対象をトラッキングする実験を行なう場合、画像取り込み時間と計算機の処理時間の合計がカメラ制御遅れ時間になるので、高速処理はトラッキングが上手く行くかどうかを決定する重要なファクターとなる。

3.2 処理プログラム

(1) ノイズ対策と空間フィルタ

実験に使用したビデオカメラにより得られる画像の画素明度には、CCD 画素感度のばらつき、ビデオ信号処理のドリフト等による標準偏差 2.1 ADU の変動と、蛍光灯照明のちらつきによる 1.8 ADU の変動が含まれる¹¹⁾。オプティカルフローの計算では空間的、時間的な明度差を使うが、それらの標準偏差は再生性の定理により約 4 ADU となる。明度差が小さいと、(2)式で計算するオプティカルフローに含まれる誤差が大きくなるので、明度差が小さい場合はオプティカルフローを計算しない方が良い。この閾値を ΔE とすると、閾値 ΔE を大きくすればノイズの影響が少なくなり、信頼性の高い動きだけが求まるようになるが、求まる動きの数は減少する。明度差の理論的最大値は 8 ビットで AD 交換した場合 255 ADU だが、実験で得られる画像では最大 100 ADU 程度であるので、 ΔE が 10 ADU 以下の場合はオプティカルフローの計算をしないことにした。時間的な明度差は(2)式の分子に表われるので、空間的な明度差程の影響はないが、信頼

性の低いオプティカルフローを計算すると処理量が増え、かつ、全体として精度が低下すると考え、空間的な明度差と同様に計算しないことにした。

$$E_x^2 + E_y^2 \geq \Delta E^2 \quad \text{かつ} \quad |E_t| \geq \Delta E$$

このことは、ある閾値以上の刺激がなければ動物の神経系が反応しないことに対応して自然である。

また、勾配法では動き検出範囲で明度の空間的な変化率が一定でなければならないが、ビデオカメラから得られる生の画像がこの条件を満足することはほとんど無いので、ローパスの空間フィルタで有害な高周波成分を取り除く必要がある。しかし、空間的な明度差は最大で 255 ADU なので、広い範囲で変化率を一定にすると、大きな動きの検出には良いが、変化率が小さいのでノイズの影響を受け易く、小さな動きを正確に検出出来ない。したがって、ある大きさの動きを検出する為には動きの大きさに応じてフィルタを掛ける必要があり、動きとフィルタの大きさの関係は実験により決めなければならない。

フィルタとしてはガウス函数の方が良いが、計算の簡単なトップハット型の円形フィルタを使用した。画像にフィルタを掛ける操作は、フィルタ内に含まれる画素の明度の平均をフィルタ中心の画素の明度とすることである。

フィルタは Fig. 1., Table 1. に示す f 3, f 7, f 14, f 27 (f の後の数字はフィルタの直径 (画素) を表わす) の 4 種類とした。最大のフィルタを画像に掛ける処理時間は約 15 分かかるので (脚注), 入力画像にフィルタ

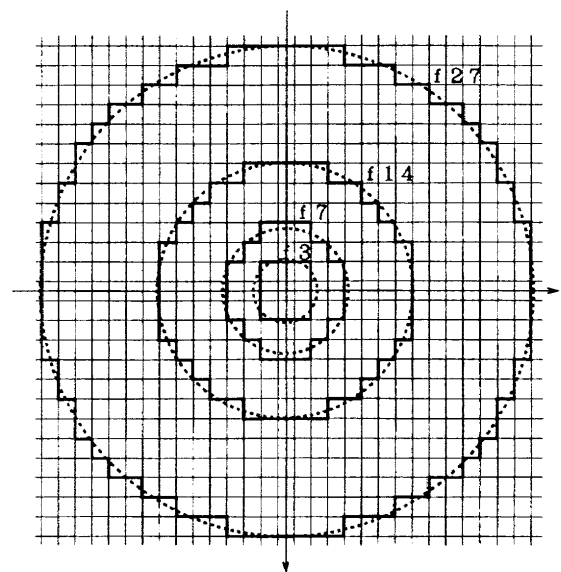


Fig. 1. Top hat filters.

脚注) プログラム上のミス为了避免る為に、画像データにポインターを使わず、配列を使った。64 k バイト以上の大きな配列は、OS-9 の c コンパイラは remote 変数扱いにするので処理に時間がかかる。

Table 1. Parameters of a gradient method.
(Unit: p. e. < abbreviation of picture element)

Filter symbolic name		f 3	f 7	f 14	f 27
Filter size	hor. x ver.	3x3	7x7	15x13	29x25
	p. e. numbers	9	37	151	573
Support $\partial x, \partial y$		1	2	4	8

ーを掛けた画像データをハードディスクに保存して、それを使って動き検出アルゴリズムの研究を行なった。さらに大きなフィルターも考えたが、計算機の処理能力とハードディスクの記憶容量を考えて止めた。

今回はコマ撮り画像を使ったので、時間的フィルターを掛けることは出来ない。

(2) 明度の空間的变化と時間的变化の計算

明度の変化は

$$E_x = \frac{E_{x1} + E_{x2}}{2}$$

$$E_y = \frac{E_{y1} + E_{y2}}{2} \times \frac{6.35}{7.4}$$

$$E_t = E_2 - E_1$$

で計算した。添字 1, 2 は時間差がある画像データを表わす。CCD の画素は縦 $7.4 \mu\text{m}$ 、横 $6.35 \mu\text{m}$ なので、y 方向の変化率は画素の縦横比の補正を行なった。

勾配法で速度を求める場合の前提（移動範囲で明度変化が一定）を満足させる為に、2つの画像の明度の空間的变化の差をその位置での明度で割ったものが 0.01 以下とした。

$$\frac{|E_{x2} - E_{x1}|}{(E_2 + E_1) \cdot 2} < 0.01$$

$$\frac{|E_{y2} - E_{y1}|}{(E_2 + E_1) \cdot 2} < 0.01$$

さらに、変化率を安定させる為に、使用するフィルターの大きさに比例して明度変化率を計算する ∂x と ∂y を大きくした (Table1.)。

(3) 明度勾配の向き計算

明度勾配の向きは $(E_x, E_y)^t$ で与えられるが、x 軸に対する角度 θ の 1 変数で表わした方が、実験結果の整理に便利である。角度 θ は、

$$\tan \theta = \frac{E_y}{E_x}$$

で与えられる。

角度 θ を求める時、ノイズ対策により $E_x^2 + E_y^2 \geq 10^2$ ではあるが、 θ が ± 90 度に近づくと E_x の値が小さくなり、

ノイズの影響を受け易くなるので、例えば、 $0 \leq \theta \leq 90$ 度の範囲では、次の様な工夫をした。

$$\frac{E_y}{E_x} \leq 1 \text{ の時, } \theta = \tan^{-1} \frac{E_y}{E_x}$$

$$\frac{E_x}{E_y} < 1 \text{ の時, } \theta = 90 - \tan^{-1} \frac{E_x}{E_y}$$

初めは、向きを求めるのに逆三角関数を使うと計算時間がかかると考え、 $\tan \theta$ の値で向きを分類したが、向きを分類する間隔を変えるごとに $\tan \theta$ の値を修正しなければならないので、プログラムの単純な逆三角関数を使うことにした。

(4) オプティカルフローの方位の決め方

(2)の方法で計算した明度の変化率を(2)式に代入してオプティカルフローの明度勾配方向成分の大きさを、(3)の方法で明度勾配の方向を、画素ごとに求める。横軸に明度勾配の方向、縦軸に明度勾配方向成分の大きさをプロットし、大きさが最大値を取る方位をオプティカルフローの方位とした。

オプティカルフローの大きさを 1、オプティカルフローの方位からのずれを θ とすると、明度勾配方向の成分は $\cos \theta$ で分布し、 ± 15 度と ± 30 度ずれた向きの成分はそれぞれ 0.966, 0.866 なので、オプティカルフローの方位の分解能はあまり良くないことが予想される。

3.3 実験結果

コマ撮りにより i0 から i8 までの 9 枚の画像をハードディスクに取り込んだ。フィルターと閾値 ΔE の大きさを変え、i0 を基準として、i1 から i6 の画像データとの間で、動き検出を行なった。画像 i7 は移動台を基準の位置まで戻した画像データで、バックラッシュのチェックに使用した。画像 i8 は照明を追加し明るくし、模型と背景の境界が明瞭になる様にした画像である。

(1) 画像の確認

模型が写っている画像の部分を 15 のサブ領域に分割し (Fig. 2., この図の左上の円は使用した 4 つのフィルターの大きさを示す)、各サブ領域の明度 (横 $28 \times$ 縦 60 画素) を A4 版の紙に印刷し、模型の明度が飽和していない (255 ADU を越えていない) ことを確認した。画像 i8 の明度を印刷し、飛行機模型の画像を構成する画素数を数え、約 9200 画素を得た。

焦点距離 50 mm、カメラ・模型間の距離 3.55 m の場合、CCD 画素の横の長さ $6.35 \mu\text{m}$ は模型の位置で 0.45 mm に対応する。これは、模型全長 139 mm が画像上に 314 画素で写っていることと一致している。

明度変化の大きいところ (背景と機体との境界) を選

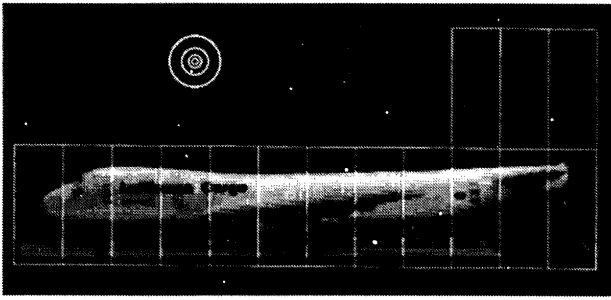


Fig. 2. Input image data of the miniature.

び、画素間の明度を補間する方法で画像上での移動量を確認した。模型の移動量と画像から求めた移動量との差は、機体の移動（左右）方向が ± 0.1 画素程度（Table 2.），上下方向が 0.01 画素程度であった。この移動量の精度の差は、入力画像の特徴と CCD カメラの空間分解能の違いが原因と考えられる。左右方向の移動を測定した機首と機尾との境界は曲率が大きいため正確な左右方向の位置を決めることが難しいのに対し，上下方向の移動の測定に使用した機体と背景の境界はほぼ水平の直線で正確な上下方向の位置を決めることは簡単であった。さらに，CCD カメラは電荷の呼び出し機構の関係から水平（左右）解像度が，垂直解像度に較べて悪いので，画素間の明度の補間精度も悪くなったと考えられる。

Table 2. Image displacement against i_0 in x-direction.
(Unit: p. e.)

Image number	i 1	i 2	i 3	i 4	i 5	i 6	i 7
True value	0.2	0.4 _s	0.9	1.8	3.6	7.2	0.0
Meas. at nose	0.1	0.3	0.7	1.7	3.4	7.1	-0.1
Meas. at tail	0.3	0.6	0.9	1.8 ₄	3.6 ₄	7.1 ₄	-0.0 ₃

(2) フィルターの効果

模型の機首・機尾と背景の境界の明度変化がフィルターを掛けることによりなだらかになる様子を Fig. 3. に示す。フィルターの直径程度の範囲で明度変化がほぼ一定になっている。

(3) 閾値 ΔE とフィルターの大きさ

閾値 ΔE を大きくすればノイズの影響が少なくなり，信頼性の高い動きだけが求まるようになるが，動きが求まる点の数は減少する。CCD 画素感度のばらつきの影響は空間的平均を取ることでよりデータの数 n （円形のフィルターに含まれる画素数）の平方根に逆比例して小さくなる（ n の指数 $p = -0.5$ ）ので，最適な大きさの ΔE はこの付近にあると考え，実験を行なった。平均を取らない画像データ（ $n = 1$ ）に対しては $\Delta E = 10$ ADU とし，これを基準として， $p = -0.7, -0.5, -0.3$ に対する ΔE （Table 3.）に対して航空機全体でどの程度の動きが求まるかを調べた。

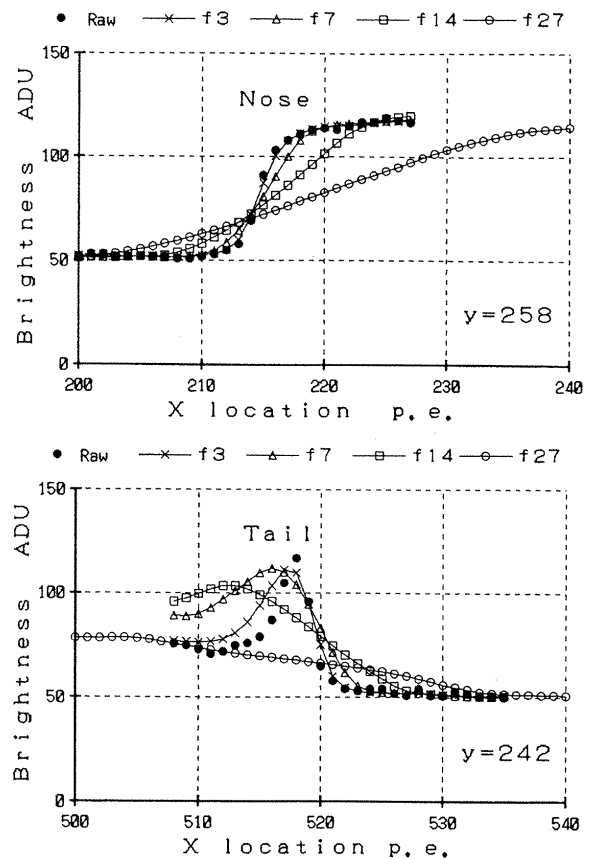


Fig. 3. Brightness variations versus filter sizes at the miniature nose and tail.

Table 3. Threshold ΔE . (unit: ADU)

Power	f 3	f 7	f 14	f 27
-0.7	2.15	0.80	0.30	0.12
-0.5	3.33	1.64	0.81	0.42
-0.3	5.17	3.38	1.34	1.49

得られた結果の評価基準として次の基準を採用した。

- 基準 1：オプティカルフローの明度勾配方向成分の方位分布から，そのピーク（オプティカルフローの向き）が解かること。
- 基準 2： i_0 と i_7 の間ではオプティカルフローが検出されないこと。
- 基準 3：オプティカルフローが求まる画素（位置）が広く分布し数が多いこと。
- 基準 4：オプティカルフローのばらつきが小さいこと。

基準 1 による評価： i_1 から i_6 までの画像データのオプティカルフロー（脚注）を 6 度間隔のヒストグラムにし

脚注) 正確には「オプティカルフローの明度勾配方向成分」と書くべきだが，混乱の無い限り，「オプティカルフロー」と書く。

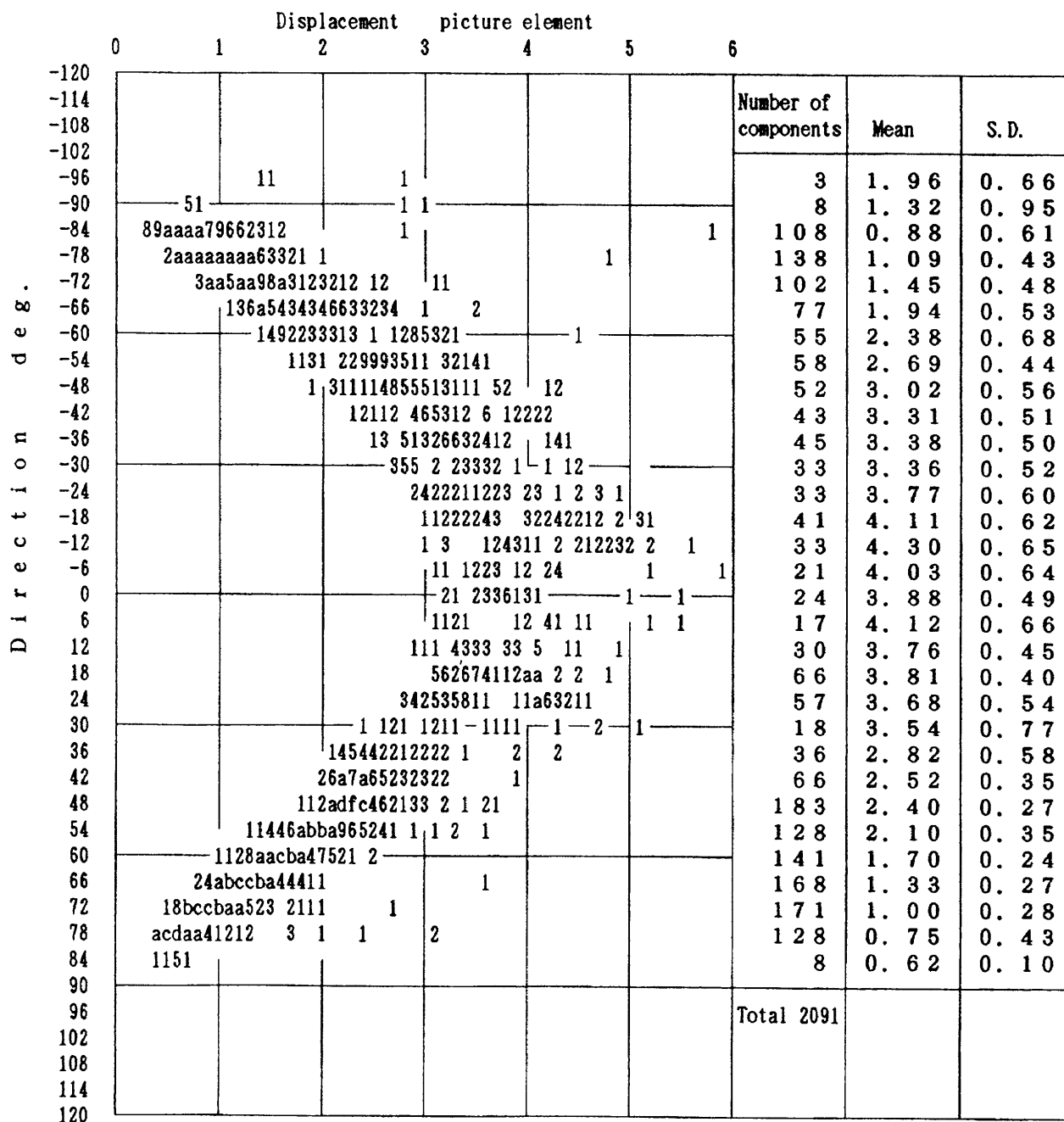


Fig. 4. Histogramme of optical flow components in the direction of the brightness reduced between i0f14 and i5f14.
(Note: a = 10 ~ 19, b = 20 ~ 29, c = 30 ~ 39,...)

(その一例を Fig. 4. に示す), ヒトが見て判定した結果を Table 4. に示す。×印はデータがばらついてピークが求まらないことを示し, 数字 0 はオプティカルフローが全然求まらない (オプティカルフローが求まった画素数が全体で 0 画素) ことを示す。ピークが求まった場合は, 機体の移動方向 ± 15 度に含まれるオプティカルフローの平均値 (脚注), 標準偏差, オプティカルフローが求まった

画素数が示してある。入力画像の再現性 0.1 画素に較べて, 検出された標準偏差はかなり大きかった。

基準 2 による評価: 基準 2 は基準 1 と同じ内容 (SN 比が良いこと) である。画像 i0 と i7 との間ではオプティカルフローが求まらないのが理想的である。あるいは, 移動量 0 として求まれば良いが, 画像データにノイズがあるのでばらついたオプティカルフローが求まると考えられる。Table 4. の右端列には i0 と i7 との間で検出された全ての方向のオプティカルフローの数, 平均値 (m), 標準偏差 (σ) が示してある。Table 4. では, 検出されたオプ

脚注) オプティカルフローの明度勾配方向成分が cosine 関数で変化するものとする, 移動方向 ± 15 度の平均値は移動方向の値 (最大値) の 0.99 倍である。

Table 4. Means and standard deviations and numbers of optical flows.

Filter	Threshold ΔE ADU	Input image (reference image : i 0)						
		i 1	i 2	i 3	i 4	i 5	i 6	i 7
N o	1 0	×	×	×	×	×	×	0
	(10)	×	0.6/0.4 55	1.0/0.8 116	1.8/1.2 105	×	×	0.4/0.2 7
f 3	2.1 5	×	0.5/0.3 215	0.9/0.4 162	1.9/0.5 87	×	×	0.4/0.3 2813
	3.3 3	×	0.5/0.2 125	1.0/0.2 122	1.9/0.4 64	×	×	0.4/0.3 474
	5.1 7	0.3/0.1 9	0.5/0.2 67	0.9/0.2 78	1.9/0.2 43	×	×	0.2/0.0 6
f 7	0.8 0	×	0.5/0.3 309	1.1/0.4 304	2.0/0.6 200	4.2/0.8 100	×	0.5/0.4 4042
	1.6 4	×	0.6/0.2 81	1.0/0.2 193	1.9/0.4 138	4.2/0.8 50	×	0.4/0.2 24
	3.3 8	0	0	0.9/0.1 48	1.9/0.2 77	4.2/0.2 20	×	0
f 1 4	0.3 0	0.8/0.4 98	0.8/0.4 254	1.4/0.7 419	2.3/0.9 358	3.9/1.5 195	×	0.7/0.7 5857
	0.8 1	0	0.7/0.0 5	1.3/0.4 71	2.2/0.4 215	4.0/0.6 125	7.8/1.7 41	0
	1.3 4	0	0	0	2.1/0.1 15	4.0/0.4 60	8.1/0.5 16	0
f 2 7	0.1 2	1.2/1.2 120	1.1/0.8 197	1.4/1.2 453	2.2/0.9 598	3.4/1.3 581	6.0/2.8 384	0.8/0.8 8905
	0.4 2	0	0	0	2.2/0.3 115	4.0/0.5 256	7.2/0.8 235	0
	1.4 9	0	0	0	0	0	7.6/0.3 55	0

Note: display position

mean / s.d. number

unit: p.e. for mean and s.d..

0 : number is zero.

× : no peak is detectable.

ティカルフローが $m + 3 \times \sigma$ 以下では信頼性が低いと考え、その範囲が網掛けしてある。

基準3による評価：Table 4.に示した移動方向 ± 15 度の範囲に含まれるオブティカルフローは、フィルターと ΔE が大きくなるにつれて、模型の一部でしか求まらなくなった。フィルター **f27** を使った場合、**i4** で $\Delta E = 0.42$ と **i6** で $\Delta E = 1.49$ の時は機首の部分だけでしかオブティカルフローが求まらなかった。また、小さなフィルターを使った場合、動きが大きくなるにつれ、機首と機尾でオブティカルフローが求まりにくくなり、**f14** を使った場合でも、**i6** に対しては機首と機尾の部分ではほとんどオブティカルフローが求まらなかった。

オブティカルフローが求まる画素数はフィルターの面積（含まれる画素数）に比例して大きくなると考えられるので、オブティカルフローが求まる画素数の比較はフィルターの大きさを考慮しなければならない。

基準4による評価：基準4の判定としてはオブティカルフローの標準偏差を使った。

その他の評価：

・生画像データを使った場合

閾値 $\Delta E = 10$ とした場合、オブティカルフローが計算される画素数は航空機全体で5画素以下と大変少なく、オブティカルフローの方位分布から移動方向を求めることが出来なかった。

生データの画像は1画素以下の移動検出に適すると考えられるが、1画素以下の移動では E_{x1} と E_{x2} 、 E_{y1} と E_{y2} との差が小さい条件を満足するのが難しいので、この制限を外し、オブティカルフローを求めた。この条件を外すことによりオブティカルフローの求まる画素数は増えたが、データのばらつきがかなり大きくなった。画像 **i0** と **i7** からオブティカルフローが求まる画素数は7画素と少ないが、 $m = 0.35$ 画素、 $\sigma = 0.18$ 画素なので、オブティカルフローの大きさが $m + 3 \times \sigma (= 0.89$ 画素) 以上の範囲で高い信頼性でオブティカルフローが求まるとすると、1画素以下のオブティカルフローを高い信頼性で求めることが出来ない。したがって、生画像データに対して勾配法を適用すべきではない。

・フィルター **f3** を掛けた画像を使った場合

画像 **i7** に対して、 $\Delta E = 3.33$ の時、動きが求まる画素数は474画素、 $m = 0.43$ 画素、 $\sigma = 0.31$ 画素なので、動きの大きさが1.4画素以下のオブティカルフローを高い信頼性で求めることが出来ない。

閾値 $\Delta E = 5.17$ の時は、**i7** に対して動きが求まる画素数は6画素と少なく、平均値も標準偏差も小さくなるので、**i1** に対する動き (0.2 画素) も求めことが出来たが、オブティカルフローの求まる画素数が少なく、実用的では無い。画像 **i2**、**i3** に対してはかなり良い結果が得られたが、**f3**

以外のフィルターでは1画素以下の動き検出は難しい。

・フィルター **f14**、**f27** を掛けた画像を使った場合

指数 $p = -0.7$ に対して、かなりのオブティカルフローが求まるが、得られる値は真の動きとは異なり、データのばらつきも大きかった。フィルター **f27** の場合、移動台の動きが含まれているものがあつた。

以上をまとめて、総合的に評価して、Table 4.では、最適と思われる結果を太い線で囲んである。ほぼ最適なフィルターは模型の動きに比例した大きさが良く、最適な閾値 ΔE は $p = -0.5$ 付近にある。 $p = -0.5$ の場合、空間的变化率を計算する ∂x と ∂y をフィルターの大きさに比例して取ったので、 ΔE と ∂x または ∂y の積は、ほぼ3.3と一定になる。これは、データのばらつき (標準偏差、ほぼ ΔE に比例) はデータ数の平方根 (フィルタの直径) に逆比例するに対して、フィルターの直径に比例する ∂x または ∂y を掛けているからである。

(4) マルチチャンネルによる動き検出

入力画像を、(3)で求めた最適なフィルターと ΔE を組み合わせた4つの空間周波数チャンネル (Table. 4 で \square で囲んだ) で並列処理を行ない、それらの処理結果を統合し動きを検出する方法を検討した。画像 **i6** に対する4つのチャンネルの出力 (6度間隔ごとの平均値) は、予想される様に、そのチャンネルに適した大きさの動きに対しては良い結果を出力するが、それ以外では大きくばらついた (Fig. 5.)。これら4つのチャンネルを統合する為に、次の3つの論理を採用した。

i) 各チャンネルの検出範囲をそのチャンネルが検出し易い動きの $1/\sqrt{2}$ 倍から $\sqrt{2}$ 倍の範囲とした。ただ

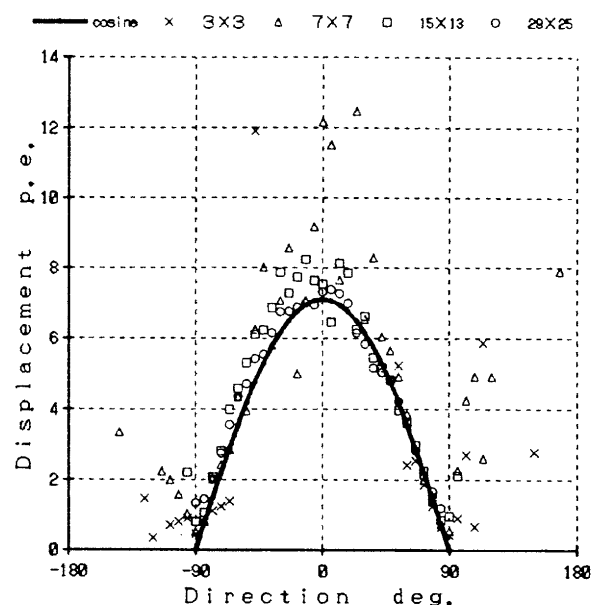


Fig. 5. Angular distribution of optical flow components in the direction of the brightness with different filter sizes.

し、最も小さい動きを検出するチャンネルの下限は0とした。

- ii) 明度勾配方向の方位分布の6度間隔当たりのデータ数が10以下の場合には出力を出さない。
- iii) 180度方位が異なる動きが検出された場合は出力を打ち消し合わせる。

統合した出力の明度勾配方向の方位分布を Fig. 6. に示す。論理 i) の適応の仕方は、画素ごとに求めた動きの内、 $1/\sqrt{2}$ 倍から $\sqrt{2}$ 倍の動きの平均値を出力する方法と、ヒストグラムの間隔ごとの動きの平均値を求めてから、平均値の値が $1/\sqrt{2}$ 倍から $\sqrt{2}$ 倍の範囲だけを出力する方法の2つが考えられる。前者の方法では、正しい動きの大きさが、そのチャンネルが検出し易い動きの $1/\sqrt{2}$ 倍から $\sqrt{2}$ 倍の範囲に無い場合、正しい動きと大きく異なる結果が得られたので、後者の方法を採用した。また、最も小さい動きを検出するチャンネルの下限を0としたのは、どの程度まで小さな動きを正確に検出可能かを確認する為である。論理 ii) と iii) は無くても方位分布の最大値を決めることは可能だが、正しい動きの方位からのずれが ± 90 度以上の結果を打ち消すのに有効であった。

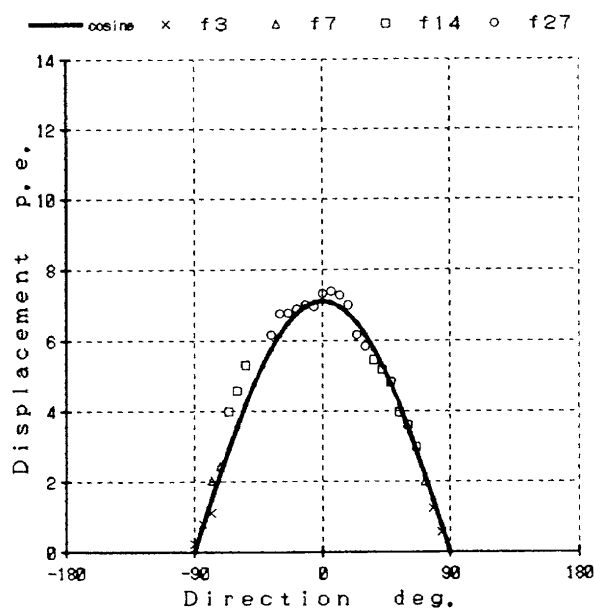


Fig. 6. Angular distribution of optical flow components in the direction of the brightness with optimal filter sizes.

(5) Local 領域の大きさ

今までの実験では Local 領域の大きさを対象の模型全体をカバーする様にして模型全体として1つの動き情報を抽出し、Local 領域が大きければ正確に動きを検出出来ることを確認した。次の課題として、どの程度まで Local 領域を小さくしても、正しい動きを検出することが可能かを実験した。

Local 領域の大きさとして、ヒトの第1次視覚野のハイパーコラムの大きさがカバーする網膜の大きさに近い、

15のサブ領域を採用し、画像 i4 を使って実験を行なった。各 Local 領域の6度間隔のヒストグラムの度数(データ数)が0, 1のところはかなりあり (Table 5.), その結果、平均値がかなりばらつくようになった。それでも ± 15 度のデータの平均を取るによりかなりなめらかに変化する様になり (Fig. 7.), その最大値を動きの向きとすると、15の Local 領域のうち、1つは窓問題により尾翼前縁に垂直な方向の動きを出力したが、その他の14の Local 領域はほぼ正しい向き (± 24 度) と大きさ (1.67 ~ 2.67 画素) の動きを出力した (Fig. 8.). 15の Local 領域のオプティカルフローの平均は2.0画素、標準偏差は0.34画素、向きの平均は+2度、標準偏差は17度、窓問題を解決出来なかった領域を除いた14の Local 領域のオプティカルフローの平均は2.1画素、標準偏差は0.25画素、向きの平均は-1度、標準偏差は13度であった。

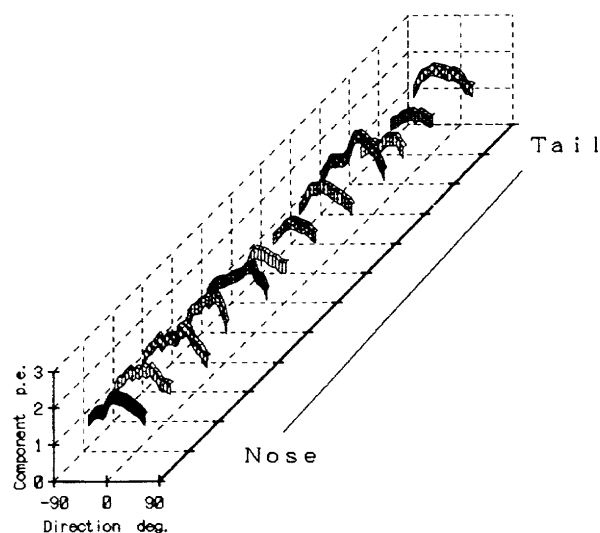


Fig. 7. Local angular distributions of optical flow components.

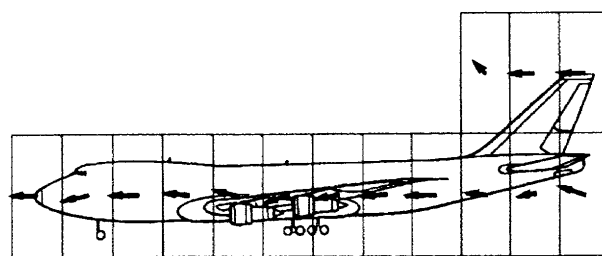


Fig. 8. Optical flows by gradient-based method with local optimisation.

模型の動きは1.8画素なので、最大で1画素程度の誤差があることになる。この例では、全体で約25000画素からなる領域に含まれる約9200画素の対象に対して1525画素でオプティカルフローが求まり、15の動きを決めることが出来た。この処理には約11秒かかったが、この時間には方位分布からその最大値を決める時間とフィルターを掛ける時間は含まれていない。

Table 5. Number of picture elements where velocities are reduced between i0f7 and i4f7 image data.

Direction deg.	Local block number															Dir. total
	L01	L02	L03	L04	L05	L06	L07	L08	L09	L10	L11	L12	L13	L14	L15	
-90					2	5	1 3	3			2		1			26
-84					1 3	2 0	5 8	1 2	6		1 1	2	4			126
-78				6	1 0	1 5	1 5	1 1		2	3	6	7			75
-72				9	1	4	4	6	1	4	4	5	1 5	1	2	56
-66		1		1 2	1		1	6	2	8	2	4	1 1		1	49
-60		1		1 1	3	2	3	3	7	5		5	1 8		1	59
-54		1	1	1 4	4	2	3	3	3	6		1	1 9			57
-48		1		3	4	1	3	1	1	5	1	1	5	4	1	31
-42			1	2	5	3	1	6	3	4		1	4	4	2	36
-36				1	4	2		6		1		4	4	3	1	26
-30		1	2		5	3	1	2	1	1	1	1	1	1	1	21
-24			1		4	1	3	2	1		2	1	3	6	2	26
-18		2	3		1	1	2	1	2	1			2	4	2	21
-12		1	2	2	1	1	1	4	5	4			2	2	3	28
-6		1	1	2	5	2		2	2	2		3	9	2	2	33
0		1		1	3	3	1		3	1	1	2	2	1	2	21
6		2	2	4	7		1	1		3		2	2	2	2	26
12		3	2	1	6	1	2	1		2	3		4	1	4	30
18		2	3 6	2	2				1	2	2	1	4	1	7	60
24			1	2	9	1		2			1	1	3	3		23
30		3	1	2	3	5	1	4	1		3	2	6	7	2	40
36			1	3	8	1		1	1	6	6	1	6	6	4	44
42	3	1 4	1	4	4	4	2	3		1	3	1	1 8	2	7	67
48	2 7	7 6	2	7	3	5	2	1	2	1	4	4	3 0	3	7	174
54		1	5	8	3	5	1		1	4	5	5	1 3	4	4	59
60		1	4	1 1	1 2	3		6		4	2	1	7	1 1		62
66			1	5	2 5	6	1	1 6		1	8	2	6	1 5	7	93
72					9	7	1	7			6	3	1 7	2 2	7	79
78					7	5	5	4			6		1 0	1 8	2	57
84					2		1					8	5		2	18
90															1	1
96															1	1
L. total	30	112	67	112	166	108	126	114	43	68	76	65	238	123	77	1525

(6) Global 法

Local 法を評価する為に Horn & Schunck² の Global 法との比較を行なった。フィルター f27 を掛けた i0 と i6 との間で、横方向に 7 画素おきに 50 点、縦方向に 6 画素おきに 20 点の 1000 点の動きを計算した。

Global 法では動きを計算する領域の動き境界条件を与えなければならないが、これを自由境界条件とするのが自然である。今回使用した画像では模型の背景の明度は比較的均一なので、自由境界条件を採用すると、背景も模型と一緒に移動する結果が得られた。計算領域の境界での動きを 0 とすると、背景の動きは模型境界から計算領域の境界に向けて単調に減少した。また、機首と機尾の動きは正確に求まるが、その間の比較的明度の均一な

機体部分の動きは小さくなった。

Global 法では繰り返し計算をし、解の収束判定が必要になるが、横方向が 50 点なので、繰り返し計算の回数は 100 回とした。処理時間は 1000 点の動きを計算するのに 132 秒かかったが、1000 点の動きの相互依存性が大きいので、実質的な動き点数はかなり少ない。

(7) テンプレート・マッチング法

15 × 15 画素の 61 個のテンプレートで航空機をカバーし、i0 と i4 との間で動きを求めた結果、31 個のテンプレートが 2 画素、4 個のテンプレートが 1 画素の動きを出力した。垂直尾翼の部分で窓問題の解決に失敗し、また、右下には移動台の動きが検出されている (Fig. 9.)。

明度分布が均一な場合、テンプレート・マッチング法

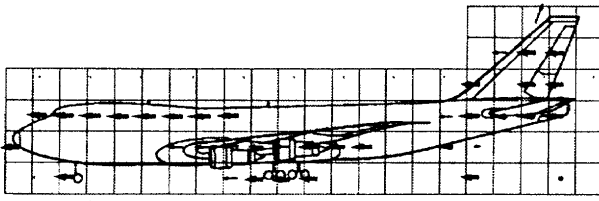


Fig. 9. Template matching of 15 × 15 p. e. size.

でも結果の信頼性を確認することにより信頼性の低い結果を取り除くことが可能だが、信頼性の確認は以外に手間取る。この信頼性の確認をしない場合の処理時間は37秒であった。Figure 9. では、対応点からの2画素の距離の残差が対応点の残差の2倍以上の場合は太い矢印で、対応点の残差の1.5倍以上の場合は細い矢印でオプティカルフローが示してある。入力画像間の動き1.8画素に対して、対応点からの2画素の距離の残差が対応点の残差の2倍以上のオプティカルフローの平均は約1.9画素であった。

4. 考察

4.1 動き検出アルゴリズムの比較

局所的な動き検出アルゴリズムの実験により解かったことを参考にして、勾配法とテンプレート・マッチング法とを検討し、Local 勾配法の実験結果の比較を行なう。動き検出アルゴリズムは局所的に限定された画像情報から動き情報を計算出来ること（独立性または並列性）が望ましいので、Global 勾配法は検討の対象から外す。

(1) Local 勾配法

一般に、ノイズ（高周波成分）の影響を取り除く為にローパスの空間フィルタを使うので、その時、ノイズと一緒に細かな画像の特徴も失われてしまい、小さな動きを正確に求めるのが難しい。勾配法では1画素以下の動きを検出することも可能であるが、フィルタをかけてもノイズが残るので、対象の動きが0の場合でも1画素近い動きが検出される。勾配法は1～2画素程度の動きが高い密度で求まる。大きな空間フィルタを使えば大きな動きの検出は可能であるが、画像の解像度が低下するので実質的な検出密度は減少する。

明度勾配方向の動きを検出するには空間と時間に対する濃度勾配 E_x , E_y , E_t を計算する最少4画素の明度データがあれば良いが、必要な画素数が少ないことは計算が簡単であることを意味するが、同時にノイズに弱いことも意味する。濃度勾配の計算で異なる画素データを使うことにより、6画素のデータが必要になる。この報告では2枚の画面で空間的な濃度勾配を計算し、その平均を取ったので10画素のデータを使った。

さらに、個々の微係数を2点から計算する代わりにより多くの点を使いノイズの影響を少なくする方法も考えられる（4点を使った例が文献14に紹介）。同様な効果は、濃度勾配の計算に使う明度を1個の画素データの代わりにその画素を中心とする画素明度の平均を使うことにより、多数の画素データを使うことが可能になり、ノイズに対して強くなる。

しかし、時間平均を計算するには多数の画像データが必要になり、画素明度の時間平均を取るのには難しい。照明の時間変動の影響を除けば、ノイズに対して時間平均と空間平均とは同じ程度の効果がある。また、画像の時間平均を取ることで動きの検出が遅れるので、時間平均は避けるべきである。時間平均が必要な場合には、動きを計算するのに必要な最小限の2枚の画面から局所的な動きを求めてから動きの時間平均を取ることで、計算量とメモリー空間容量を節約することが出来る。節約により時定数の異なるチャンネルで並列的に動きの検出が可能になる。

今回の実験では模型表面に特徴の少ない飛行機を使った為に、動きが検出された場所は主に模型と背景との境界であった。背景の明度はほぼ均一であり、模型の明度とは異なるので、3.3の(2)に示した様に空間フィルタを掛けることにより境界の明度はなだらかに変化する様になる。模型の移動に連れて境界のなだらかな明度勾配も移動し、境界に垂直な方向の動き成分が求まる。今回は背景と模型の明度がそれぞれ比較的均一であったので、境界付近の動きが正確に求まったが、勾配法で動きが求まる位置は模型上でなければならない。

今回の実験では、オプティカルフローの明度勾配方向成分の方位分布で最大値を取る方向を移動方向としたが、一般に、すべての方位成分が求まらないし、また、成分には誤差が含まれるので、移動方向を決めることは意外に難しい。単純に最大値を取る方法では最大値を取るデータにより移動方向が決まり、それ以外のデータが使われていない。得られたすべてのオプティカルフローの平均を取ることで、正しい動きの方向成分は残り、それに垂直な成分は打ち消し合うので、移動方向を決定することが出来る。3.3の(5)にこの方法を適応すると、14のLocal領域での向きは-13～+8度とかなり良くなった。ただし、この方法でも特定の方位だけしか動きが求まっていないヒストグラムでは正しい方向を出力しない。

(2) テンプレート・マッチング法

関心のある領域を前画像の中から参照テンプレートとして選び、現画像との一致度を評価しながら動きの検出範囲を全領域探索し、一致度の最も高い位置を現在の位置とする方法なので、デジタル計算機に適したアルゴリズムである。動き検出範囲が探索範囲に限定され、探索

範囲に比例して処理時間が増加し、動きの精度は画素単位である。フィルターを掛けない解像度の高い画像に対して、 3×3 画素の小さなテンプレートでも2画面 \times 9画素のデータを使い、 15×15 画素のテンプレートでは2画面 \times 225画素のデータを使うので、勾配法に較べて高い信頼性の動き検出が可能である。動きの探索範囲をテンプレートの大きさより小さくすることにより、信頼性を高めることが可能である。しかし、さらに信頼性の高い動きを求めるには、勾配法で明度勾配方向成分の方位分布から動きを求める場合と同様に、パターン認識の面倒な問題を解決しなければならない。

勾配法に較べてテンプレート・マッチング法の優れている点は、動き0が検出可能であることである。対象を追跡する場合、勾配法では対象が止まってしまうと対象が何処に居るのが解からなくなるのに対し、テンプレート・マッチング法では何処に居るのが解かる。しかし、テンプレート・マッチング法では、初めに、画像のどの部分を参照テンプレートとするのかは外部から指示を与える必要がある。

(3) 実験結果の比較

この報告ではLocal勾配法に重点を置いて実験をしたので、ここでは他の研究者により行なわれたLocal勾配法の実験結果との比較を行なう。

① Barron, Fleet & Beauchemin の実験¹⁴⁾

彼らは動き検出アルゴリズムの比較を最も広範囲に行ない、4グループ(勾配法、テンプレート・マッチング法、エネルギー法、位相法)の9種類のアルゴリズムを、合成画像(4種類)とビデオ実画像(4種類)に適応し、動き検出の精度と信頼性(標準偏差)と密度とを計算した。その中で、他の実験との比較に適するのは、大きな木が画面全体を占めている写真に対してビデオカメラを平行移動させながら取り込んだ画像を使った実験である。この実験ではLucas & KanadeのLocal勾配法の方がテンプレート・マッチング法よりも良い精度と信頼性を得ている。実画像を使った彼らの実験では対象の正しい動きを求めず、その結果、精度と信頼性を計算することが出来ず、結果を定性的に議論しているだけなので、他の実験との比較が出来ない。

画像にガウスフィルターを掛けて得られる実数画像に較べて明度を8ビットで量子化した整数画像では、動きの誤差が、Local勾配法で50%増、Global勾配法では数倍になり、動き検出の精度と信頼性に対する画像に含まれるノイズの影響が大きいことを彼らは報告している。誤差の測度として角度を彼らは採用しているので、実験結果の解釈が難しいが、2画素程度の動きに対して、Local勾配法の誤差と標準偏差は数%と報告している。

② Perrone の実験¹⁵⁾

中心に1個、その円周に12個のゼロ交差オペレータから成る2次元動き検出器を使い、中心と周囲とのオペレータ間の通過時刻の差により30度間隔の動きの方位分布を求め、方位分布からエッジに垂直方向(明度勾配方向)のオプティカルフローを求め、複数のエッジの動きの活動度(重み)を計算し動きを決定している。空間フィルターの大きさは 7×7 画素で、エッジ方向を決める13個のオペレータを含む領域は 15×15 画素である。彼の方法は動物の視覚が採用している方法に近いが、オプティカルフローの明度勾配方向の向きとその大きさはこの報告の2.1に示す様に直接計算出来るので、計算機視覚として彼の計算アルゴリズムは効率が悪い。

ハンマーを振り下ろす画像に対してハンマーと手とは均一に運動すると仮定して動きを求めている。エッジの動きを求める方法として方位分布をcosineカーブで近似して最大値を求める方法では、動きの平均値は2.54画素/フレーム、標準偏差は0.37画素、向きは267度(ほぼ下方)であった。同じ画像から直接ヒトが求めた動きは 2.0 ± 0.5 画素/フレームであった。活動度(重み)から計算した全体としての動きは1.9画素/フレームであった。

エッジの動きを求める方法として単純に方位分布の最大値を取る方法は、エッジの動きの方向が ± 15 度の誤差を含むにもかかわらず、動きの平均値は2.52画素/フレーム、標準偏差は0.053画素、向きは263度であった。活動度(重み)から計算した全体としての動きは1.8画素/フレームであった。彼はcosineカーブで近似する方法より誤差が大きいと報告しているにもかかわらず、この方が標準偏差が小さくなっているため、正しい標準偏差は0.53画素の可能性が考えられる。

③ 千葉&小沢の実験⁸⁾

明度50 ADUの背景に対して明度120 ADUの長方形を左に3画素、上に2画素移動(大きさ3.61画素、方位角146度)する画像に、平均0 ADU、標準偏差2.6 ADUのホワイトノイズを加えた合成画像に対して勾配法を適応した。空間フィルターは 7×7 画素で、Local領域の大きさは 5×5 画素である。予想される様に、動きは長方形と背景との境界付近でのみ求まり、最小2乗法によるLocal法を適応した場合、動きの大きさの平均は3.64画素、標準偏差0.47画素、方位角の平均は140度、標準偏差は6度であった。

④ 今回の実験

実験に使用したビデオカメラの画素感度のばらつきの標準偏差は2.1 ADUであり、模型の動きが1.8画素で、 7×7 画素の空間フィルターを使った場合、個々の画素のオプティカルフローは平均2.0画素、標準偏差0.4画素であった。模型全体を15個の領域に分割して求めたLocal

ごとの動きは、大きさの平均2.1画素、標準偏差0.3画素、方位の平均+2度、標準偏差17度であった。模型全体から求めた動きは2.0画素、方位はほぼ0度であった。

以上、4つの実験を比較すると、①の誤差の標準偏差は小さ過ぎる。今回の実験では誤差の標準偏差は動きの大きさに依存し、また、空間フィルターの大きさやLocal領域の大きさにも依存したが、②、③、④の実験では、誤差の標準偏差は動きの13～20%であった (Table 6.)。信頼係数を99%とすると、検出された動きの大きさには1画素程度の誤差が含まれていると考えなければならない。

Table 6. Accuracy of optical flows by gradient-based methods with local optimisation. (Unit: p. e.)

	Mean	S.D.
Perrone(1990)	2.54	0.37
Chiba & Ozawa(1991)	3.6	0.47
Present study(1996)	2.1	0.3

③の方位角の標準偏差は小さい様に思われるが、直線に対して垂直な動きを求めているので精度が良いのかもしれない。オプティカルフローの明度勾配方向成分の方位分布から動きの方位を決める方法は、原理的に方位に対する感度が悪く、オプティカルフローの見かけの大きさ (明度勾配方向成分の大きさ) が標準偏差の大きさ、例えば20%小さくなる方位は±37度である。

画像処理による動き検出は、画素ごと、あるいは、Local領域ごとの精度はこの様に悪いが、画像全体として動きを求めると、画像に含まれる誤差が対象の大きさに較べて小さくなり無視出来るので、全体としての動きはかなり良い精度で求まる様になると考えられる。

動き検出アルゴリズムの実験に使用する機器はほぼ共通していて、ビデオカメラ、8ビットのAD変換器、256画素×256画素から512画素×512画素程度のデジタル画像を使い、検出する動きの大きさに合わせて最良の結果が得られる様にアルゴリズムのパラメータを調整しているので、同程度の結果を得ることが出来ると考えられる。

一般のテンプレート・マッチング法では対応点の位置が画素単位で量子化されマッチングされるので、誤差と標準偏差は1/2画素程度と推定される。

4.2 動き検出アルゴリズムの改良

勾配法では模型飛行機を背景から分離出来そうな位に明度勾配方向の動きが検出可能な場合もあったが、模型

の動きに応じてフィルターを大きくすると動きの検出される部分は一部に偏り、模型全体の動きを推定するのは困難になった。テンプレート・マッチング法では動きをLocal勾配法より高い密度で求めることが出来たが、機体の中央部分の動きが求まらず、模型の前部と後部との2つのブロックの動きとしてそれぞれが連結して求まった。したがって、動き情報だけでは2つの対象が同じ方向に移動しているのか、1つのものが移動しているのかの区別がつかない。この問題を解決する方法としては次の方法が考えられる。

① 局所領域の拡大

局所的に明度が均一で動きが決められない場合、局所領域の大きさを大きくすることにより、極端な場合は対象全体を1つの局所領域でおおうことにより、対象の動きを連続して求めることが可能になる。

Local勾配法の実験では4つの大きさのフィルターを使っているのので、既にこの方法を採用していることになるが、フィルターの大きさにより検出に適する動きの大きさが異なるので、この方法は向かない。

テンプレート・マッチング法ではテンプレートの大きさによらず画素単位の精度で動きを検出することが可能なので、この方法は有効と考えられる。しかし、テンプレートの大きさが15×15画素と30×30画素の実験では、前部と後部との間で動きが求まらない距離は同じ2区間であった (Fig. 10.)。距離が小さくならなかった理由としては、テンプレートが大きくなると、残差が対応点でそれ程小さくならないので、対応点から2画素の距離の残差が対応点の残差の1.5倍以上という条件が厳しくなり、動きが得られなくなったことが原因の1つと考えられる。テンプレートを大きくすると計算時間がかかる様になるので、飛び飛びの画素からなるテンプレートを採用し、テンプレートを構成する画素数を一定とする等の対策が必要になる。さらに領域が大きくなると対象の回転などによる像の歪みが問題になり、テンプレートの単純な平行移動だけでは不十分になり、テンプレートの回転、拡大、縮小等のアフィン変換が必要になる。

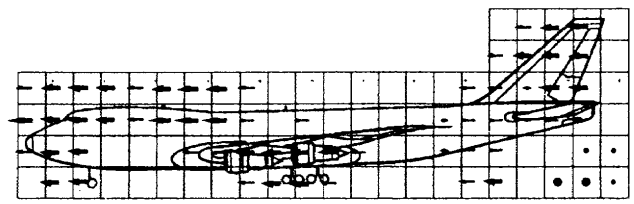


Fig. 10. Template matching of 30 × 30 p. e. size.

② 動き情報以外の情報との統合

明度が時間変化する領域や色・明度の一致度による領域や距離情報と、動きが検出された領域との論理和を取ることにより対象の背景からの切り出しの可能性が高まる。明度が時間変化する領域は動きが検出される領域と重なる可能性が大きい、明度の時間変化の方が単純な現象なので動きが検出される領域より広範囲をカバーする可能性が高い。同じ色や明度の部分は同じ対象に属する可能性が高く、色や明度が均一な部分は動きが検出出来ない、動きが検出された領域との統合は重要である。

③ 運動検出アルゴリズム

その他に知識やモデルを使うトップダウン的な方法も考えられる。例えば、2つの領域が1つの剛体運動として近似出来る場合は、2つの領域は1つの物であるという運動モデルを仮定し、運動パラメータを計算することが可能である。局所的な動きから運動パラメータを求めることはパターン認識の上位レベルの難しい問題である。現在、私の関心は知識やモデルに依存しないボトムアップの動き検出アルゴリズムにあるが、運動検出アルゴリズムはごく近い将来の課題と考えている。

4.3 動物の視覚から学ぶ

現在の計算機視覚のレベルは画像処理のレベルで、動物の視覚に遠く及ばないので、視覚機能を工学的に実現する為には、視覚機能を既に実現している動物の視覚から多くのことを学ぶ必要がある。視覚機能を実現する人工視覚システムはかなり複雑で、それを構成する要素が多く、論理的に工学的に積み上げることは不可能である。さらに、地球上の風景や物は地球上固有のもので、それらを理解する為には視覚処理系のパラメータを調整することが必要であるが、動物の視覚では進化と学習によりパラメータ調整をし、生存に必要な情報の獲得には成功していると考えられるからである。

① 眼

外界から画像情報を取り入れる視覚センサーとしては、特に議論することなく、手近なビデオカメラが使われている。ビデオカメラはヒトが見て自然に見える画像が得られる様に作られている。ビデオカメラとヒトの眼とでは視野、空間解像度、時間分解能が異なり、ビデオカメラで得られた画像データにはヒトの眼で得られた視覚情報が全て含まれてはいない。特に、ビデオレイト（NTSCでは、1秒間30画面）により離散化され、転送により明度情報が読み出されるので、動き検出アルゴリズムは大きな制約を受ける。

魚眼レンズを使い、ヒトの眼の様に視野中心から周辺にかけて解像度が低下する画像を使う研究も行なわれているが、現在のビデオカメラを使った場合、視野中心の

解像度がヒトの中心窩に較べて悪い。今回の実験の関心は、動き検出アルゴリズムに使う画像の解像度をヒトの網膜中心窩のそれと同程度にした場合、局所的な動き情報がどの程度の密度で求まるかにある。視野の広さを犠牲にし、ヒトの中心窩と同程度の解像度が得られる様にレンズの焦点距離を選び、中心窩の視覚アルゴリズムの研究を行なっている。視覚アルゴリズムの評価をする時、ヒトの視覚との比較を行ない、人工視覚の目的の1つがヒトの視覚の代行であるので、可能な限りヒトの眼と同程度の画像情報を外界から取り込む様に努力している。ヒトに見えないものは人工視覚にも見えない可能性が大きく、外界から取り入れた画像情報に含まれていない情報は後の計算機処理をどの様にしても、抽出することは出来ないと考えるからである。

高品位テレビカメラ、いわゆるハイビジョンカメラを使うことにより、ヒトの眼に一步近づくことが出来ると考えられるが、実験設備費用と処理時間を含めて、論理的にそうする必然性がなければならない。

② 神経生理学

動物の視覚では、例えば、網膜像の局所的な明度情報の変化から局所的な動きを検出するモジュールの下位レベルと、多数のモジュールからの出力とその他の情報を統合し対象の全体としての動き、すなわち運動を検出する上位レベルのいくつかのレベルに分けて考えられている¹⁶⁾。第1次視覚野の傾きコラムとMT野の存在が確認されている。大脳皮質上のコラムの大きさは0.5 mm程度の大きさだが、網膜中心窩に相当する位置で、傾きコラム1周期の視角は約10分、MT野の神経細胞の受容野は数度と報告されている。傾きコラムは下位レベルの動き検出能力を決め、MT野の神経細胞は上位レベルの運動を検出していると考えれば、ヒトの視覚では下位レベルの動き検出に数百画素の情報を使っている。

第1次視覚野の傾きコラムは刺激の運動方向ではなく、刺激の傾きに対して反応する。勾配法では明度勾配方向の動き成分だけを定めることが出来る。この2つの事実から、1周期の傾きコラムをLocal領域の大きさとして動きを検出するアルゴリズムを今回提案した。

③ 心理学

明度勾配方向の動き分布からオブティカルフローを求める方法、いわゆる「窓問題」を解決する方法には2.2に示した様に幾つかある。Adelson & Movshon¹⁷⁾は異なる方向に移動する2つ格子縞を同時に呈示した場合どの方向にオブティカルフローが知覚されるかを調べる心理物理学実験を行ない、ヒトの視覚の高次の運動解析がどの様に行なわれているか解明しようとしている。

また、動き検出が可能な範囲を予想し、動き検出アルゴリズムを作り、実験をしないと、アルゴリズムの正し

い評価が出来ない。動き検出可能な範囲は実験のし易さから決めることも考えられるが、応用の視点からはヒトの視覚を基準に決めるべきである。心理物理学の実験によればヒトの視覚では対象の速度と移動距離が適当でないと運動は知覚されない。刺激条件により異なるが、運動距離閾（運動が知覚出来る最小の距離）は視角で8～22秒、運動速度閾は1～2分/秒、運動速度頂（運動を知覚出来る最大の速度）は50度/秒である¹⁸⁾。森は、「運動距離閾が視力より小さいことは、静的な位置を先に知覚し、その2点の位置の差から動きを知覚するのではない」ことを主張している。しかし、ヒトの視力は視角で1分程度だが、副尺視力（2線分間のずれを見分ける能力）は2～4秒程度なので、副尺視力のアルゴリズムが解明されれば運動距離閾のアルゴリズムを説明することが可能になるかもしれない。また、森は速度に関しては、1度/秒以下の低速と、1～50度/秒の中速と、50度/秒以上の高速とでは異なる処理系が関与している可能性を指摘している。

4.4 動き検出アルゴリズムの応用

時系列画像から動きを検出するアルゴリズムは、ビデオカメラの手振れ防止、ビデオ動画の画像圧縮等に既に利用されているが、ここでは飛行機の着陸に関連して述べる。

全日空の機長として20年以上の経験を持つ前根が1989年11月に行なった講演「易しい着陸テクニック」¹⁹⁾で、次の様に述べている。少し長くなるが引用する。

飛行教官をやるにあたって、着陸を分析したことがあります。私自身、セスナと同時に、目の高さが5メートル以上も異なるトライスターに乗務しており、小型機からワイドボディまでの共通の着陸方式を必要としていたからです。そして、人間の目の特性を活かすことで、案外、簡単かつシンプルに着陸できることを知りました。

ファイナル・アプローチに入るとパイロットは一直線に目標地点を目指します。一般にAiming Point (AP) とよぶ地点です。実際にはなかなか一直線というわけにはいきませんから、パイロットにしる車輪にしる、向かっていく地点はつねにそれぞれのAP近くを動きまわってでしょう。こうした地点をそれぞれの接近点 Ground Proximity Point 略して、GPPと呼ぶことにします。

ファイナル・アプローチで、パイロットは自分がどの地点へ向かっているかをほぼ正確につかめます。二つの目があるお陰と、いわゆる『遠近法の原理』で、GPP以外の景色が外側へ移動していくからです。

アプローチさえ正しければ、パイロットのGPPは、ぴったりとAPに一致しているはずで、フレア開始高度でパイロットが操縦桿を引いていくと、GPPはAPか

ら離れていき、タッチダウンを迎えるのです。

ただ、幾つかの勘どころはあります。ひとつはGPPを見つける練習です。いちばんの勘どころはフレア開始高度です。

Gibsonの教科書²⁰⁾には、第2次世界大戦中の航空心理学の研究成果として、次の様な記述がある。

Figure 11. は滑走路に着陸しようとしているパイロットの視界である。幾つかの代表点での動きベクトルにより全体的な動きの変化する様子がほぼ解かる。全ての速度は水平線上で零になる。拡大焦点はその時の進入路と地上との交点である。進入角が急になれば拡大焦点は下がり、進入角が浅くなれば拡大焦点は水平線に向かって上昇する。したがって、拡大焦点はパイロットにとって車輪が滑走路に接地する位置を示すものであり、オーバー接地かアンダー接地かのキューである。適正でない接地はしばしば致命的なので、パイロットはこの様なキューに極めて重大な関心を持っている。

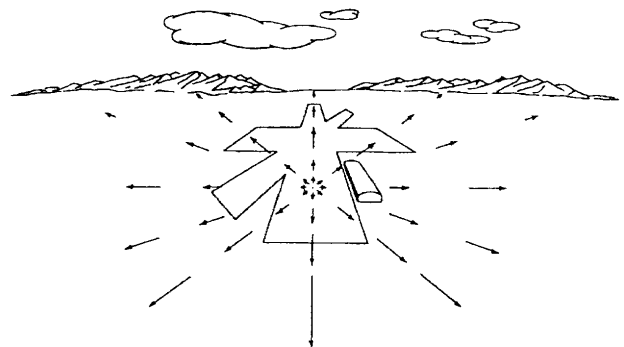


Fig. 11. The optical flow at landing (J. J. Gibson 1950).

Marrの教科書¹⁾には、次の様な記述がある。

生徒に飛行機の着陸を教えるとき、飛行教官は時々刻々評価される着陸地点が各時点での拡大焦点であるということを説明するのに時間をかけるのである。これには、注意の集中と学習を必要とする。というのはそれが生得的な反射ではないからである。したがって、拡大焦点が歩行の制御に重要な役割を果たしているというGibsonの仮説はおそらく人間に関しては間違っているものと思われる。ただし、鳥に対してはより適切なものかもしれない。

この報告で検討した動き検出アルゴリズムを、操縦席に設置したビデオカメラにより得られる画像に適応することにより、動きを検出することは出来る。拡大焦点は動きベクトルの交点なので、向きの異なる複数の動きが求めれば原理的に拡大焦点を求めることが出来る。一般に動きに誤差が含まれるので、動きベクトルの交点は複

数求まるが、拡大焦点は交点の重心とすれば良い。

Verri & Poggio²¹ は検出される動きに含まれる誤差が大きいので、誤差に強い拡大焦点の様な特異点を利用することを提案している。この方法は対象に関する情報が最小限ですむ。上の場合は飛行場が平面的であるという情報または仮定があれば良い。対象が立体的である場合も、対象が動かないのであれば、計算はかなり複雑になるが、原理的には計算が可能である。この方法を着陸に応用する場合、特異点の位置情報だけでは滑走路に対する方位が解からないので、ほぼ滑走路を正しく見る最終進入経路に飛行機を誘導する方法と、GPP の目印として滑走路の中央に視覚的な特徴がなければならない。

現在使われている ILS や現在開発試験が行なわれている MLS, GPS 等は地上施設や衛星からのサポートが必要なのに対して、この方法は自律的である。しかし、霧や雪等に弱い。

着陸の様な飛行のクリティカルフェイズでは機体の状態の変化を強く速く検出出来る必要がある。着陸時に突風を受けた場合、最初に、かつ、強く感じられるのは加速度である。機体の姿勢が変化し、突風の影響が視覚的に感じられるまでには時間がかかり、視覚だけを頼りに操縦しては事故になる場合がある。画像情報から直接得られるものは位置情報である。動き情報は位置情報の時間変化として計算されるので、ノイズに弱い。加速度情報は位置情報の2階微分なので、さらにノイズに弱い。したがって、画像情報を使った航法には加速度計の併用が必要である。

現在の計算機視覚は、工場の生産現場の様に照明条件を制御出来、対象が限定されている場合には機能するが、日常の世界は複雑過ぎて適用が難しい。汎用性のある計算機視覚の前段階として限られた視覚機能の計算機視覚を中間目標とし、それが活躍出来る分野を開拓し、その中で計算機視覚の研究と評価と開発を進めて行くことが必要である。その様な分野として道路や駐車場の自動車交通の監視が考えられるが、計算機視覚の対象として飛行場は自動車の駐車場に較べて次の点で有利である。

- ①飛行場に入出入りする飛行機と車両は種類と移動が制限されている。
- ②飛行場は施設規模が大きく、安全対策に使う金額も大きい。

5. まとめ

画像情報から動きを検出するアルゴリズムとして、勾配法とテンプレート・マッチング法の調査を行ない、2 km 先の大型旅客機の動きをどの程度検出することが出

来るかを実験により調べた。航空機模型とヒトの眼と同程度の解像度を持ったビデオカメラを使い、1/500 のスケールで実験を行ない、Local 勾配法を中心にアルゴリズムの評価を行なった。

Local 勾配法に関して、

- ①模型の動きが小さい場合、背景から模型を分離出来る位に高い密度でオプティカルフロー（見かけの動き）の明度勾配方向の成分が境界の位置で求まった。
- ②オプティカルフローの検出精度を良くするには空間フィルターと閾値を大きくする必要があるが、あまり大きくすると、オプティカルフローの数が減少し、局所的にしか検出されなくなった。最適な空間フィルターはオプティカルフローの大きさ程度であった。
- ③ビデオ画像に含まれるノイズの大きさから、生データを使いオプティカルフローを求める場合、閾値は10 ADU とし、明度差が10 ADU 以下の場合は動き検出の処理をしなかった。空間フィルターを掛けたデータを使う場合、ノイズの大きさがほぼ空間フィルターに含まれる画素数の平方根に逆比例するので、それに応じて閾値を小さくした。
- ④最適なフィルターと閾値を使った場合、動きが1～7画素の実験範囲で、画素ごとのオプティカルフローの標準偏差は動きの20～10%程度であった。
- ⑤窓問題を解決する為に、オプティカルフローの明度勾配方向成分が最大値を取る方位をオプティカルフローの方位とする方法を提案した。
- ⑥Local 領域の大きさはヒトの第1視覚野のハイパーコラムの視野程度まで小さくすることが出来、飛行機全体で15個の動きを決定することが出来た。Local 領域ごとのオプティカルフローの大きさの標準偏差は動きの約20%，方位の標準偏差は約20度であった。
- ⑦今回の実験結果は他の実験結果とほぼ同程度の精度であり、信頼区間を3倍の標準偏差とすると、勾配法の測定値には1画素程度の誤差が含まれる。
- ⑧動きが主に求まった所は対象と背景の境界であったが、今回の実験では背景が比較的均一なので、正しい動きが求まった。
- ⑨画素ごと、Local 領域ごとの精度に較べ、画像全体のオプティカルフローの平均と模型の動きとは、大きさで0.2画素、方位で数度の誤差で一致した。

その他、

- ⑩Global 勾配法では、背景が均一な画像なので、対象の動きに引きずられ、背景も移動した。
- ⑪テンプレート・マッチング法ではLocal 勾配法より高い密度でオプティカルフローが求まった。

参考文献

- 1) Marr, D.; Vision(1982), Freeman, New York.
乾敏郎, 安藤広志訳; ビジョン(1987), 産業図書, 東京.
- 2) Horn, B. K. P. and Schunck, B. G.; Determining optical flow, Artificial Intelligence 17(1981), 185-203.
- 3) 尾上守夫, 前田紀彦, 斎藤優; 残差逐次検定法による画像の重ね合わせ, 情報処理 17(1976), 634-640.
- 4) 内山隆, 沢崎直之, 青木孝, 森田俊彦, 稲本康, 佐藤雅彦, 稲葉雅幸, 井上博允; ビデオレートトラッキングビジョンの実用化, 第12回日本ロボット学会学術講演会予稿集(1994), 345-346.
- 5) 畑中豊彦, 福田忠彦; テレビジョン・画像情報工学ハンドブック 1編3章 視覚(1990), オーム社, 東京.
- 6) 田村秀行, 坂根茂幸, 富田文明, 横矢直和, 金子正秀, 坂上勝彦; SPIDER user's manual(1980), 電子技術総合研究所.
- 7) Kearney, J. K., Thompson, W. B. and Boley, D. L.; Optical flow estimation: An error analysis of gradient-based methods with local optimaization, IEEE trans. PAMI 9(1987), 229-244.
- 8) 千葉昌孝, 小沢慎治; 明度拘束式のモードによるオプティカルフローの検出, テレビジョン学会誌 45(1991), 1199-1206.
- 9) Srinivasan, M. V.; Generalized gradient schemes for the measurement of two-dimensional image motion, Biol. Cybern. 63(1990), 421-431.
- 10) Barnard, S. T. and Thompson, W. B.; Disparity analysis of images, IEEE trans. PAMI 2(1980), 333-340.
- 11) 五味広美; 視覚センサーとしてのビデオカメラ (その2), 航空宇宙技術研究所報告 TR-1231(1994).
- 12) 五味広美; ヒトの視覚に学んだビデオカメラの視線制御による対応点探索, 航空宇宙技術研究所報告 TR-1185(1992).
- 13) 森村淳, 魚森謙也, 石井浩史, 赤堀裕志, 北村好徳, 坂口隆; 純電子式画像揺れ補正技術, National Technical Report 37(1991), 308-314.
- 14) Barron, J. L., Fleet, D. J. and Beauchemin, S. S.; Performance of optical flow techniques, Int. J. Comp. Vision 12(1994), 43-77.
- 15) Perrone, J. A.; Simple technique for optical flow estimation, JOSA A7(1990), 264-278.
- 16) 斎藤秀昭; 視覚神経系の構造とその情報処理, 情報処理 30(1989), 114-128.
- 17) Adelson, E. H. and Movshon, J. A.; Phenomenal coherence of moving visual patterns, Nature 300(1982), 523-525.
- 18) 永野俊, 梶真寿, 森晃徳; 視覚系の情報処理(1993), 啓学出版, 東京.
- 19) 前根; 易しい着陸テクニック, 安全飛行 143(1990), 23-25, 全日空.
- 20) Gibson, J. J.; The perception of the visual world(1950), Houghton Mifflin, Boston.
Reprinted in 1974 by Greenwood Press.
- 21) Verri, A. and Poggio, T.; Motion field and optical flow: Qualitative properties, IEEE Trans. PAMI 11(1989), 490-498.

航空宇宙技術研究所報告1301号

平成8年7月発行

発行所 航空宇宙技術研究所
東京都調布市深大寺東町7丁目44番地1
電話三鷹(0422)47-5911(大代表)〒182
印刷所 株式会社 東京プレス
東京都板橋区桜川2-27-12

Printed in Japan