第1節 視　覚

1. 視覚情報とその処理

1.1 はじめに

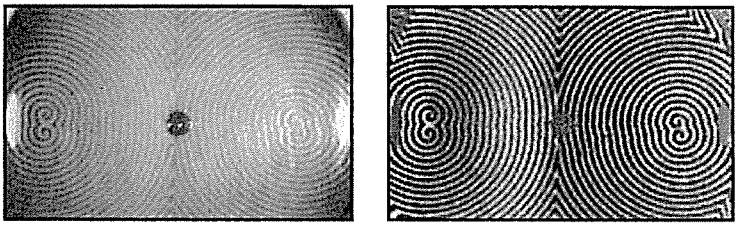
動物は視覚から多大な情報を得ている。進化の果ての存在としての人類の脳も，視覚情報の処理に膨大な資源を投資し，人工の仮想現実感では到底太刀打ちできない非常に精巧で強力な立体映像エンジンによるバーチャル・リアリティを獲得している。本稿では，人間の両眼立体視機能をはじめとする優れた三次元動画像処理能力をいかに人工的に実現するかという観点で，視覚情報処理に関する最近の研究や現在挑戦している問題や仮説(モデル)のいくつかを紹介する。

1.2 固視微動と画素時系列フィルタリング

われわれは眼球運動を積極的に行うことで，外界の多様な情報を入力している。この運動をアイマークレコーダで調べると，一点を見つめている固視の状態(250～350 ms)と急激な視線移動であるサッケード(十数ms)が繰り返し出現することが分かる。サッケード中は外界の情報は入力されず，固視期間の初期(100 ms以内)にのみ情報がサンプリングされる[1]。さらに，固視の期間においても眼は速く震えて(固視微動)いる。網膜に投影される像が完全に静止すると，外界は全く見えなくなることが実験的に確認されている[1]。これは，刺激の変化にのみ応答するという視細胞やニューロンの特性によるもので，生物特有のセンサの特徴と理解できる。フォトダイオードのような人工的なセンサは刺激の強さ(明るさ)に線形に応答する素子であり，網膜という非線型のセンサとは大きな特性の違いがある。

それでは，こうした人間の視覚センサの特性を積極的に模倣した場合，どのような利点が得られるであろうか。以下は，こうした観点から実現した各画素毎の時系列の局所時間フィルタリング処理[2]を紹介しよう。

ビデオやパソコンのムービープレイヤーが普及している今日，連続したシーン(動画像)を眺めているときと，その中の1シーンを静止画として見る場合とでは，画像の解像度やシーンから得られる情報量が甚だしく異なることを体験することが多い。1枚1枚の静止画処理の延長ではなく，連続的に入力される映像をある程度まとめて処理することの重要性が認識される。そこで，連続映像の中から局所的な時間スケールδtを考え，この時間中での画像内の各画素における輝度(濃淡)時系列の時間変化に注目した。画素(x, y)における輝度時系列$f(x, y, t)$と，δt時間内の平均輝度$<f(x, y, t)>_{\delta t}$との差を増幅し，適当なバイアスCを加えて新たな動画像$g(x, y, t, \delta t) = \alpha\{f(x, y, t) - <f(x, y, t)>_{\delta t}\} + C$を得ることができる。この考えは，各画素時系列のハイパスフィルタリングを行うことと等価であり，フレーム間の差分(微分)をとることとも通じ(δt＝1フレームとしたとき)，信号の変化がなければ出力が0であるというニューロンの特性と符合する。



(a) 原画像　　(b) フィルターを通した画像

図1　画素時系列フィルターによる動画像強調

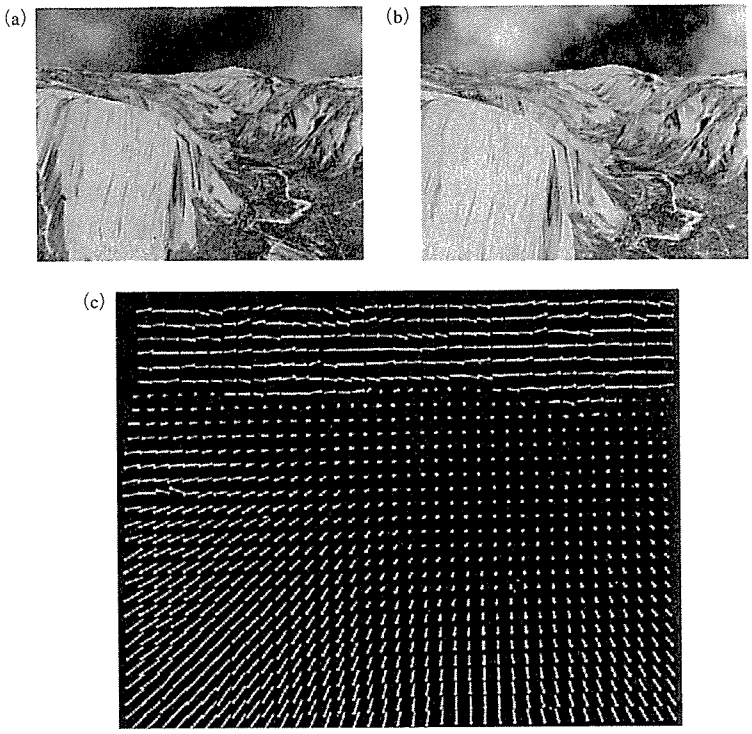


図2　照明の時間・空間変化する動画像(**a**), (**b**)からの運動ベクトル場(**c**)検出例

こうしたフィルタリングを行うことの利点を，実例を用いて以下に示す。**図1**(a), (b)は，濃淡値のダイナミックレンジが広く画面内の一部で微妙な変化が観測されるような動画像のコントラスト改善(画像強調)や運動強調の例である。**図2**は，照明の空間的不均一や時間的変動がある場合でも，運動ベクトル場の正確な検出が可能になることを示している[2]。理論やアルゴリズムの詳細は文献を参照いただきたいが，視覚情報の前処理として有効である。

1.3　両眼立体視を解く

鳥類の一部(フクロウやハヤブサなどの猛禽類)とほ乳動物の一部(霊長類など)に見られる両眼立体視の機能は，「カモフラージュ破り」という種の存続に重要な意味を持っていると言われる[3]。進化の中で偶然獲得されたと理解されている立体視機能は，人間に特有の豊かでスーパーリアルなイメージの世界をもたらしている。人間の立体視のメカニズムは複雑であり，単に両眼視差の情報だけでなく，陰影，相対運動，遮へい関係およびテクスチャー形状などの情報を総合して精巧な三次元世界を脳内に再現している[1]。右眼と左眼に与えられる画像の情報と，脳が左右の情報を統合して創生する三次元映像とは明らかに異なる。このことは，三次元ステレオグラムを平行視や交差視で解いたときにきわめて印象的な映像の出現として体験される。この映像は，左右の視覚がとらえた二次元画像ではなく，2枚の二次元画像から再構成された三次元世界そのものである。この意味で，脳は視覚という仮想現実機能を進化の中で獲得し，二つの二次元の眼から一つの三次元の“目”を実現したとも言える。

ところで，こうしたリアルタイムで三次元世界の動的映像を再構築する機能は，どのようにして人工的に実現できるのであろうか。単純には，一瞬，一瞬の両眼視差を解析し各画素毎の奥行き分布$z(X, Y)$と物体の絶対位置(x, y, z)や絶対スケールの情報が再構築できると考えられる。しかし，この方法では静止した三次元世界は再現できても，その中を運動する物体の運動情報(速度$\boldsymbol{v}$や重力の加速度$\boldsymbol{g}$などの物理定数)を解析するには不十分である。この解析には，左右それぞれの目(カメラ)がとらえる見かけの二次元速度ベクトル(オプティカルフロー)$\boldsymbol{v}=(V_X, V_Y)$の検出が必要となる[4]。すなわち，三次元世界を運動する物体の運動速度$\boldsymbol{v}=(u, v, w)$

は，オプティカルフローと以下のような関係にあり，

$$V_X=\left(\frac{hI}{zF}\right)\left(u-\frac{wx}{z}\right)$$
$$V_Y=\left(\frac{hI}{zF}\right)\left(v-\frac{wy}{z}\right) \quad (1)$$

オプティカルフローは，

$$f_X V_X+f_Y V_Y+f_t=0 \quad (2)$$

を満たす(基礎拘束式)。式(1)，(2)および$X=\frac{hIx}{z}$，$Y=\frac{hIy}{z}$より

$$\left\{(uf_X+vf_Y)-w\frac{(Xf_X+Yf_Y)}{hI}\right\}\times\frac{hI}{zF}+f_t=0 \quad (3)$$

が導かれる[5]。ここで，$f_X=\partial f/\partial X$，$f_Y=\partial f/\partial Y$，$f_t=\partial f/\partial t$であり，$h$はレンズの焦点距離，$I$はカメラの装置定数(光電素子上の1画素の逆数)，$F$は動画像のサンプリング周波数を表す。

ところで，オプティカルフローの検出法はHorn & Schunck[4]以来，おびただしい数のアルゴリズムが提案されている[6]。この中で，人間の視覚機能のように，照明の空間的不均一や時間的変動がある場合に対応できる検出法の提案は比較的少ない。われわれは，輝度の保存則を前提とした一般化した基礎拘束式を提案することで，照明の不均一の問題や物体の変形等に対処できるモデルを提案した。すなわち，輝度の保存則式(2)を

$$f_X V_X+f_Y V_Y+f_t=-f\mathrm{div}\boldsymbol{V}+\phi \quad (4)$$

と拡張し，光学モデルを検討することで，

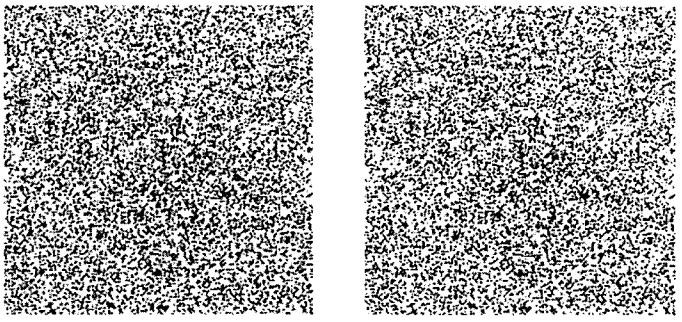
$$\phi=fW \quad (5)$$

の関係を得た[7]。ϕは輝度の生成消滅速度を表し，Wが照明の時間・空間的不均一を表現する。物体の変形，遮へいあるいは奥行き方向の物体運動を表現する$\mathrm{div}\boldsymbol{V}$の項の取扱いは未解決のままである。当面$\mathrm{div}\boldsymbol{V}=0$と考えたとしても，式(4)は見かけの速度場であるオプティカルフローではなく，照明等に左右されない真の運動ベクトル場を検出する理論として有用である。

ϕや$\mathrm{div}\boldsymbol{V}$を含まない場合でも，式(3)の拘束式を解くのは容易ではない。両眼視差よりzが既知としても未知数は三つあり，他の拘束条件(知識，常識あるいはモデル)が必要である。通常は，速度場の滑らかさの拘束を加えた正則化手法(変分原理と繰返し計算により解く)の導入が常とう手段であるが，決定すべき未知数が多いと推定精度が悪くなる。また，どのモデルを選べば十分なのかという問題も残されている。なお，1.2項で述べたように，画素時系列のフィルタリングという前処理を行えば，式(2)の単純なモデルを用いても照明の不均一に左右されない真の運動ベクトル場を求めることが可能である。各画素毎に時間微分特性を持つ素子が，運動をとらえるのに適することを示唆している。

1.4 チューリング不安定性によりRDSを解く

両眼立体視において両眼視差から奥行き分布$z(X, Y)$を求めるには，通常，テンプレートマッチングや特徴マッチングなどの手法が用いられる。しかし，人間の視覚システムがこうしたマッチングの手法を採用しているか否かは定かではない。視覚システムは，視細胞や神経回路網で構成されている。



(a)　(b)

図3　ランダムドットステレオグラムの例(口絵⑩参照)

神経細胞の可塑性を基礎とするニューラルネットワークの理論は，視覚や聴覚の機能理解にも適用されいくつかの成果を挙げてきた。それらは，階層ネットワークにおけるバックプロパゲーション(シナプスの結合係数の調整)による教師付き学習アルゴリズムや，ヘッブルール(結合しているニューロンが同時発火すれば結合が強まる)による教師なし学習などである[1)]。

ところで，**図3**(a)，(b)のような一対のランダムドットステレオグラム(RDS)は，両眼視差以外の奥行きの手がかりを持たない。片目で眺めている限りは，ランダムな点の集合にすぎない。しかし，左右の視野を平行法か交差法で融合するとやがて立体形状が出現する。この映像は非常にリアルで，人間の視覚システムが三次元世界を再構築できる仮想現実映像の創生機械であることを認識させるのに十分である。RDSはベラ・ユレッシュの提案によるものであるが，通信技術の分野で知られていたステレオ視の常識を認知科学の領域に導入し，この分野の進展に大きく寄与している。

一方，乾敏郎氏[1)]が指摘しているように，RDSの謎は

1) 両眼の対応問題を解くこと
2) 両眼視差を正確に測定すること
3) 面を補間すること
4) 不連続面を明確に知覚すること

の四つの側面がある。RDSの解法に関しては多くの提案[8)]があるが，ここでは人間の視覚の機能を模倣した最近のわれわれの提案[9)]を紹介しよう。

まず，RDSに隠されている立体形状を自分の目で再現する場合を想起してみよう。まず，両眼の映像を融合させ，その後自然と立体形状が浮かび上がってくるのを待っている。この間，意識的あるいは無意識的に，融合する左右の像を水平方向に微妙に動かし調整している。そのうち，背景のような視差がほぼ一定(～0)で，画像中に大きな面積を占める平面的領域が最初に三次元形状として認識される。時間の経過につれ，三次元形状の詳細部分が見えてくる。以下は，こうした手順を考慮し，両眼の対応問題を領域分割問題に置き換え，二次元の神経回路網のモデルと共通した特徴を示す反応拡散系の自己組織化を利用したモデルで解いた例[9)]を示す。すなわち，各画素の視差の最適値を一つずつ計算するのではなく，ある視差dに相当する一定の水平方向の相対並進を2枚の画像間で行った後，XOR(排他的論理和)演算で融合し，視差Dの対応領域を可視化する。視差Dの領域とD以外の領域が異なるテクスチャーを持つ画像として浮かび上がってくる(**図4**参照)。この場合，視差は水平方向のみと仮定している(エピポーラ拘束)。こうして，ステレオ対応問題はテクスチャーの異なる領域分割問題へと変換される。この領域分割問題を，2変数の反応拡散モデル(化学反応と拡散による物質の生成消滅を示す発展方程式)とチューリング不安定性(生物の形態形成を説明する促進・抑制物質による反応拡散モデル。抑制物質の拡散が速いとき安定なパターンが形成される)で解くことが可能である(**図5**，**図6**参照)。神経回路の特性を持つ非線形の反応拡散モデル(反応項が二次以上の非線形関数モデル)により，視覚情報処理が実行できることは脳の視覚アルゴリズムの理解への本質的なアプローチと考えることができよう。最近，こうしたパターンの自己組織化と画像(視覚)情報処理が融合する新しい学問領域に関する研究報告が見られるようになった[10)]。

1.5　おわりに

人間の視覚機能の特徴を模倣したいくつかのアルゴリズムを紹介してきた。人間の視覚機能の中で最も特徴的なのは，立体視と運動の認識と思われる。動態視力として知られている運動物体の検出と運動

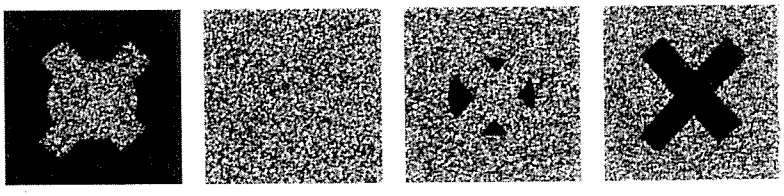
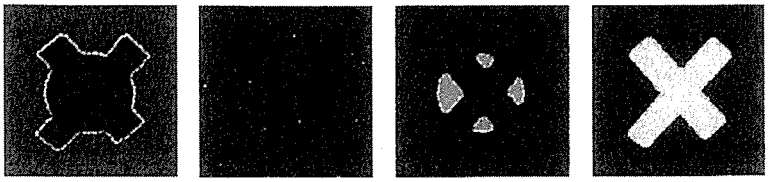
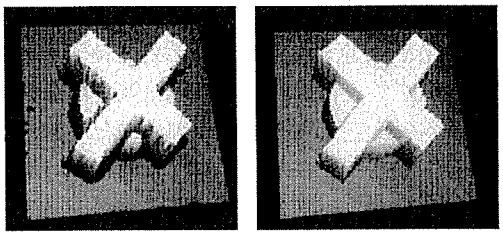


図4　視差**D**を変えながら排他的論理和(XOR)演算で左右の**RDS**画像を融合する(口絵⑩参照)



D=0　　D=1　　D=2　　D=3

図 5　反応拡散モデルとチューリング不安定性を用いて領域分割を自己組織化する
（口絵カラー⑩参照）



(a)　　(b)

図 6　復元した三次元形状(**a**)と理論値(**b**)
（口絵カラー⑩参照）

の予測は，動物である人間にとって生存に深くかかわる能力であろう。実際，車の運転は自他の生命の危険を含む重要なタスクであり，動態視の能力が問われるタスクである。人間とロボットの視覚能力の決定的な違いを示すポイントでもある。にもかかわらず，動態視力や運動錯視の問題を情報処理の観点からアルゴリズムとして実現するという視点は，従来ほとんど見られない。スポーツロボットのように素早い反射運動を実現するには不可欠の機能であり，次世紀の重要な研究テーマの一つと考える。動的な映像情報を，まとまった形で把握し解析する能力が，人工の視覚システムに求められる。静止画処理の延長ではない，新しいアプローチでの動画像処理アルゴリズム開発が必要である。

【参考・引用文献】

1）乾敏郎：Q＆Aでわかる脳と視覚，サイエンス社（1993）.

2）H.Miike, L.Zhang, T.Sakurai and H.Yamada：*Pattern Recog. Letters*, **20**, 451-461（1999）.

3）下条信輔：視覚の冒険，産業図書（1995）.

4）B.K.P.Horn and B.Schunck：*Artificial Intelligence*, **17**, 185-203（1981）.

5）近藤拓也，山際貴志，山中光司，山本正信：電子情報通信学会論文誌，**J-80-D2**，247-255（1997）.

6）J.L.Barron, D.J.Fleet and S.S.Beauchemin：*Int. J. Computer Vision*, **12**, 43-77（1994）.

7）L.Zhang, T.Sakurai and H.Miik：*Image and Vision computing*, **17**, 309-320（1999）.

8）平井有三：視覚と記憶の情報処理，培風館（1995）.

9）A.Nomura, M.Ichikawa and H.Miike：in Proceedings of 10th DAAAM International conference, pp.385-386（1999）.

10）伊藤聡，湯浅秀男，伊藤正美：反応拡散方程式を用いた自己想起型連想記憶による画像認識，計測自動制御学会論文誌，**30**，97-103（1994）.

〈三池　秀敏〉

生物とその機能モデルキーワード

ヒトの視覚(とくにステレオグラム)；両眼立体視；オプティカルフロー；チューリング不安定性