

移動物体像の抽出技術

A New Image Extraction Method for Moving Objects

川端 敦 谷藤 真也 諸岡 泰男
Atsushi KAWABATA Shinya TANIFUJI Yasuo MOROOKA

(株)日立製作所日立研究所
Hitachi Research Laboratory, Hitachi Ltd.

Abstract

This paper deals with a new method to extract a moving object from time varying image, based on a human vision recognition model. At the first, all edges of an input image are enhanced by Hitachi's Image Signal Processor. A background image which does not move or moves very slowly is gradually constructed into a background memory. The difference between the input image (edge enhanced image) and the background image (edge enhanced image) is calculated in each pixels. After that, the image of a moving objects is extracted according to the value of this difference.

[1]. はじめに

マニピレータ、ロボット等の発達により、近年特にFA (Factory Automation) に関する研究が、多くなされている。これは高度に知能化された生産設備に関する研究、と言い換える事ができるが、高性能な手足、あるいは触覚を得て、初めて実現が可能である。

現在、小型高出力アクチュエータの開発、マイクロプロセッサを使用した制御法の確立等が、原動力となって、マニピレータ、ロボットの駆動部、あるいは駆動方法に関しては、かなりの成果が、得られている。それに比べると、各種センサー、特に視覚センサーから入力された情報の利用に関する研究は、大きく遅れており、今後FA化を進めるにあたり、大きな障害になるであろう。

視覚センサーは、光が遮られた事を判断する、フォト・トランジスタを使用した簡単な物から、撮像素子を使用した物まで各種有り、最近ではCCD撮像素子、あるいはMOS撮像素子に、直接処理回路を付加したスマート・センサと呼ばれる物も、考案されている(1)。これらのセンサーから入力された信号は、多くの場合デジタル変換され、計算機によって処理される。

センサとして撮像素子を選んだ場合、処理装置に入力される情報は、膨大な量となるが、必要としている情報は、例えばある物が見えているか、いない

か、あるいは、あらかじめ記憶してある物のうちの、どれが見えたか、といったわずかな情報である事が多い。処理装置は画像情報から、上で述べた様な情報を抽出しなければならず、そのため極めて大きな処理能力が、必要とされている。以前は、このような処理能力を備えた装置と言えば、大型計算機、ミニ・コンピュータ、あるいは高価な専用ハードウェアしかなかったが、ME (Micro Electronics) の発達により、現在、極めて高性能で比較的安価な、デバイスが開発されつつある。このようなデバイスは安価であるがためにその用いられ方が、大型計算機等とは根本的に異なっている。従って、このようなデバイスを用いることを前提とした、画像処理技術の研究は、今後盛んになるものと、予想される。

本報告は、新しい画像処理方法に関するヒントを生体に求め、その持っている高度な視覚情報処理機構を模倣する装置に関するものである。

生体は脳に限らず、いろいろな場面で巧みな情報処理を行なっているが、視覚系も、その中の一つであると考えられている。この視覚系の情報処理機構を生理学的、心理学的に調べ、得られた知見に基づいた情報処理装置の研究が、現在進められている。視覚系に関しては、背景像と、その背景に比べて相対的に小さな物体像を呈示し、物体像を固定したまま背景像を動かした場合、人間は、物体像の移動を知覚することが、心理実験によって確かめられてい

る。この現象に着目し、人間の視覚情報処理機構に関して、あるモデルを想定し、計算機でシミュレーションを行ない、あるいは、電子回路を用いて実現することによって、背景に対して運動している移動物体像を抽出することのできる画像処理装置に関して成果が得られたので、報告する。

〔2〕、移動物体を抽出する視覚系の生体モデル

人間は、外界の様子を網膜上に投影し、それを神経の興奮の度合いで示される信号に変換し、神経繊維を通してその信号を脳に伝えることによって、外界の情景を理解する。本章では、生物の運動知覚 (Motion Perception) に関する知見に基づき、運動知覚、特に移動物体に対する、生体の情報処理機構のモデルを提案する。

網膜には多くの光に感ずる細胞 (視細胞) が密集して存在しており、そこに投影された像は、外側膝状体を通して、大脳の視覚野に伝えられる。大脳の視覚野では、数々の処理が、段階的に行なわれている。

視細胞は、X型細胞、Y型細胞という、2種類のタイプに分類できる。X型細胞は、光の強さに比例した反応を示し、Y型細胞は、光の強さの微分値に比例した反応を示す。

Y型細胞は、受光光度の時間微分を出力するのでこの信号を使用して、移動物体に対して反応する視覚系の生体モデルを作ることができる。網膜を固定したと仮定する場合、背景は静止した画像を、移動物体は、移動する画像を、網膜上に形成する。ある細胞が背景を捉えている場合には、その細胞の捉えた光量は変化しないが、移動物体を捉えた場合には大きく変化する。Y型細胞は背景を捉えている時には、その捉えている光量が変化しないため、小さな値を出力するのみであるが、移動物体を捉えた場合には光量が変化するために、大きな値を出力する。

このY型細胞の出力に基づく移動物体認識過程のモデルは、数多く提案されている(2)。これらのモデルは全て、網膜上の像の移動によって移動物体を認識するものであり、言わば、処理のための座標系が網膜に固定されていると、考えることができる。従って、網膜が動いた場合には静止しているはずの像が、網膜上で移動し、その結果、背景も含めた全体の像が移動物体として認識されてしまうということになる。

一方、我々の経験から「移動物体認識」という問題を考えてみた場合、上記のY型細胞のみで、認識を行なっているとは考えにくい。我々の目は常に振動しており、かつ視点を変えており、時には非常に速い速度で視点を飛ばしている。にもかかわらず、我々は正しく移動物体を認識する事ができる。つまり我々の視覚情報処理系は、移動物体を認識する際に、網膜に固定されているY型細胞の出力による情報処理だけではなく、他の情報処理機構を利用している可能性がある。

心理学の分野に於いても、運動知覚に関する実験が、数多くなされておられ、次の様な結論が、得られている。「知覚される心理的運動は、必ずしも、一定の物理的座標系上の位置移動に一致するわけではない。すなわち、ある対象物の運動は、その現象系の内部で決定された心理的座標系 (基準系と呼ぶ) 上の動きとして知覚される(3)。」ここで言う、心理的座標系とは、(a) 全体の重心が静止して見える座標系、(b) 直線運動をしている点が静止して見える座標系、(c) 取り囲んでいる図形が静止して見える座標系、(d) 背景が静止して見える座標系、等である。

上で述べた様な心理学的実験から得られた結果を支持する様な、生理学的実験結果も得られている。Bridgemanによれば、サルは視覚前野に有る神経細胞は、その1/3が背景に相対的な運動がある場合にのみ反応する神経細胞であることが、確認されている。又、酒田はサルの頭頂葉を調べ、取り囲んでいる図形が静止して見える座標系となる様な心理実験において、誘導される運動 (実際には存在しない運動) に反応する神経細胞について報告している。

以上の事実を考慮して、特に「背景に固定した座標系」に注目し、この座標系を実現している人間の視覚情報処理系に対して、次に示すようなモデルを仮定した。図1にモデルの概念図を示す。目が、図に示した様な風景を捉えると、網膜あるいは外側膝状体にある、適当な大きさの受容野を持った細胞がエッジの強調を行なう。その信号は、大脳の視覚野に入力され、数々の特徴抽出が行なわれる。この特徴抽出機構は網膜像をかなりのレベルまで抽象化するものと考えられる。しかしここでは便宜上、扱われる情報は全て絵画的情報で示してある。本モデルでは、特徴抽出され抽象化された網膜像は連合野に入力され、その結果として、入力された網膜像から

連想される像が、連合野内に形成される。連合野には網膜像が入力されるだけでなく、眼球運動中枢からの信号、自分はどこにいて、どちらの方向を見ているか、という情報等が、入力されるとする。連合野ではそれらの情報が総合的に処理され、網膜が捉えるはずの像が、(像といっても、前述の様に絵画的な情報ではなく、高度に抽象化されたものである。) 連想的に形成される。図1の場合では、「家」

「鳥」の像が入力された場合、過去に見慣れた像、あるいは長時間視野内に存在する像、ここでは「家」の像が、網膜が捉えるはずであった像として、連想される。この連想された像と網膜が捉えた像とが連合野内で比較され、その結果網膜が捉えた像に於いて、今回新たに加わった像が抽出される。図1では、連想された「家」の像と、網膜が捉えた「鳥」「家」の像が比較され、その差異成分つまり「鳥」の像が抽出されている。

人間が行なっている連想記憶のモデルと考えられている相関行列を、使用した連想記憶において、Novelty-Filter (新奇性フィルタ) の名で知られているフィルタが、ある(4)。このフィルタは、入力された情報の中から過去に記憶した情報には含まれていない今回新たに加わった情報を抽出するものである。このフィルタの動作が、上で述べた、新たに加わった像を抽出する動作と類似している、と考えられる。移動物体を抽出する細胞の層は、多くの細胞で構成されているが、その1つ1つの細胞が、前に述べた酒田の実験に用いられた細胞と同じような特性を持っていると考えられる。つまり、1つ1つの細胞は網膜に対して、適当な大きさの受容野を持っており、対象物がその受容野内で背景に対して相対的に移動した場合に限ってその細胞が発火する、と考えられる。背景像を記憶する機構は、多くの細胞の複雑な結合によって実現されているものと考えられるが、細胞間の結合の強さは下式の様に変化すると仮定する。

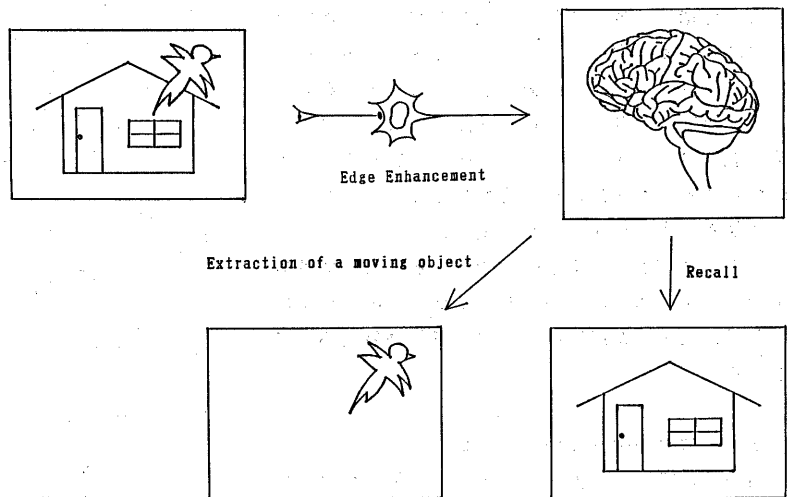


Fig.1 A model of a human visual processing

$$W_{ij}(t) = a \cdot W_{ij}(t-1) + (1-a) \cdot X_i(t) \cdot X_j(t) \quad (1)$$

$X_i(t), X_j(t)$: それぞれ t 時点での i 番目、 j 番目の細胞の発火レベル

$$(-1 \leq X_i, X_j \leq 1)$$

a : 記憶定数 ($0 < a < 1$)

$W_{ij}(t)$: t 時点での i 番目の細胞から j 番目の細胞に対する結合係数

(1) 式右辺第1項は、記憶動作を示す項であり、第2項は記憶を更新する項である。 a は、記憶の程度を示す係数である。本式は、学習方程式の名で知られている式であり、しばしば登場するパターンは、強く記憶され、めったに登場しないパターンは、あまり記憶されず、しかも時間の経過と共に、忘れる傾向にあるということを示す。

この原理により背景像が正確に構成される様子を、模式的に図2に示す。(A1), (A2), (A3) は、入力される画像の系列であり、(B1), (B2), (B3), (B4) は、背景に関する記憶が、次第に形成される様子を示している。最初、背景としては、何も記憶されていない状態であると仮定する。「鳥」と「家」を含む画像(A1)が入力されると、その像に関する記憶が(B1)の様に少しだけ背景像として記憶される。次に「猫」と「家」を含む入力画像(A2)が入力されると、背景像に於いては(B2)のように、引き続き存在している

「家」の像は、強調され、「猫」の像が少しだけ記憶され、今回は存在しない「鳥」の像が、弱まっている。同様に「人」と「家」を含む(A3)が入力された場合、(B3)の様に「家」の像は、さらに強く記憶され、他の像は「家」の像に比べると弱く記憶される。最終的には、背景像として(B4)に示す様な「家」の像のみが構成される。

以上が今回提案する生体に関する移動物体認識機構のモデルである。容易にわかる様にこのモデルは「背景に固定した座標系」による移動物体認識を説明することができる。

[3] . 移動物体を抽出する装置

前節で提案した視覚系のモデルを使用して実際に移動物体を抽出する場合、背景を連想的に想起する機構の実現が問題となる。電子回路、あるいは計算機プログラムを用いて、この連想的想起を行なう方法は数々提案されているが、現時点で満足に利用できる物はない。従って連想記憶機構に代わって通常の記憶機構を用い、連想記憶機構に相当する動作をさせる方法を考慮する必要がある。本報では、背景を想起する機構として、次に示す2種類の方法に関して考慮し、それぞれ装置として組み上げた場合の有効性に関して検討した。

(1). 背景を想起する機構として、相関行列を使用した連想記憶機構を用いる方法が考えられる。以下に相関行列を使用した移動物体抽出法を述べる。図3に示すように移動物体像と背景像を含む入力画像は、2次差分演算が施された後に1次元ベクトル表現に変換され正規化されて、連想記憶部に入力される。連想記憶部では、以下に示す方法により相関行列を更新する。

$$M(t) = a \cdot M(t-1) + (1 - a) \cdot X(t) \cdot X(t) \quad (2)$$

M(t) : t 時点での相関行列
 a : 記憶定数 (0 < a < 1)
 X(t) : t 時点での正規化された1次元画像ベクトル

また、以下により想起を行なう。

$$r(t) = M(t) \cdot X(t) \quad (3)$$

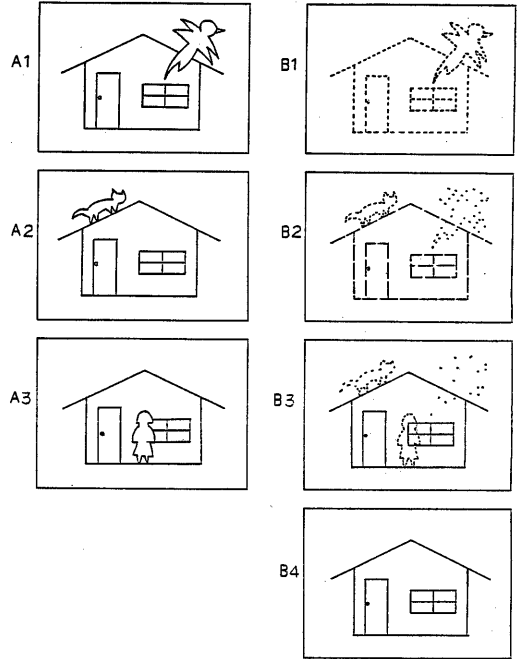


Fig.2 The process of constructing the background

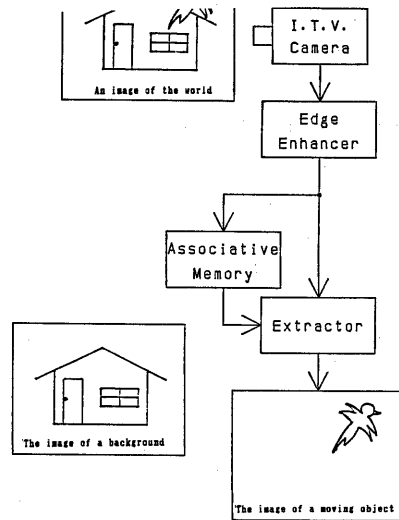


Fig.3 The blockdiagram of extraction

r(t) : t 時点での想起ベクトル

(3) 式によって、想起されたベクトル r(t) と X(t) とを比較することによって、移動物体のエッジに関する情報を抽出することができる。ところで、

ここで用いた相関行列を構成するために、必要とされる記憶量は、膨大なものである。例えば、128画素×128画素の画像を取り扱う場合に、必要となる記憶量は、256Mワードとなり、現時点では、これだけの記憶量を持つことは、とうてい、不可能である。従って、シミュレーションを行なう場合には、(2)式、(3)式を変形して、(4)式とし、使用メモリー数を、減じている。

$$r(t) = (1-a) \cdot \sum_{i=0}^t a \cdot X(i) \cdot (X(i) \cdot X(i)) \quad (4)$$

(2) . 使用メモリーを小さくする方法として、背景のエッジ強調像を、通常の画像記憶メモリーに形成する方法が考えられる。これは入力画像と等しい大きさのメモリーを用意して、そこに背景像をすでに形成していくものである。この方法は、撮像素子の移動による影響を補償することはできない。以下に、図4を用いて通常の画像記憶を使用した、移動物体像抽出装置に関して述べる。移動物体像と、背景像を含む画像は、I . T . V . によって取り込まれ、D/A変換された後に画像記憶装置に取り込まれる。その後、図5に示す空間フィルタによってエッジ強調が行なわれる。このフィルタは、ラプ

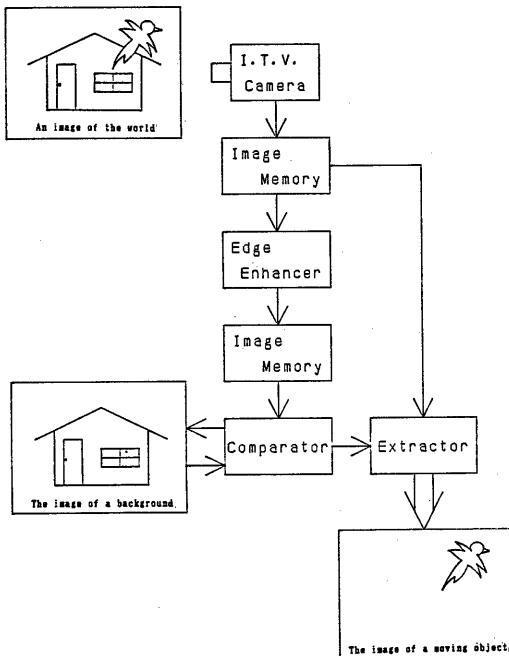
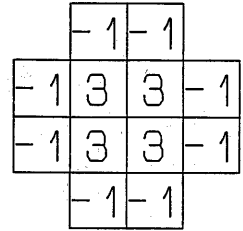


Fig.4 System diagram for extraction

ラシアン・フィルタよりもノイズに強く、またエッジ以外の情報(エッジに囲まれた像の、内部の値)を出力するものである。このエッジ強調には、



当社のImage Signal Processorを4個使用し Fig.5 Waiting value

ている。次に、蓄えられ

ている背景に関する像と、処理された入力画像とを比較し、その値が大きく異なっている画素に関して入力画像から対応する画素を切り取り、出力する。背景像は、下式によって更新される。

$$Wij(t) = a \cdot Wij(t-1) + (1-a) \cdot Xij(t) \quad (5)$$

- a : 記憶定数 (0 < a < 1)
- Wij(t) : t時点に於ける背景のエッジ強調像を構成する座標(i,j)の画素の値
- Xij(t) : t時点に於ける入力画像のエッジ強調像を構成する座標(i,j)の画素の値

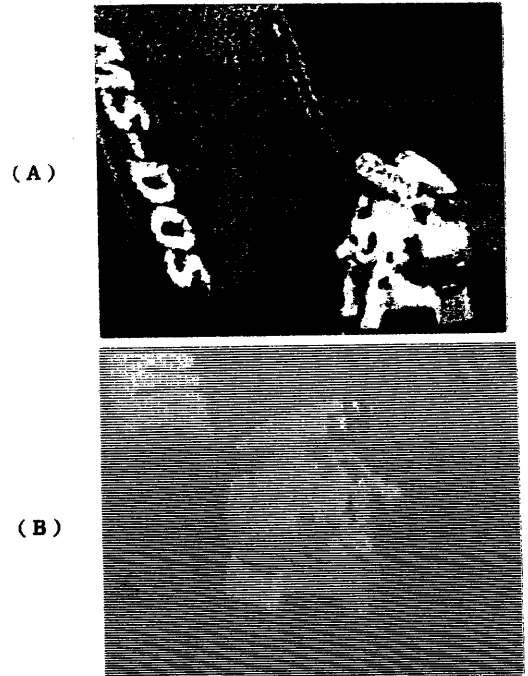
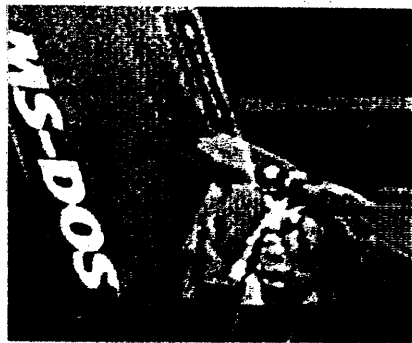
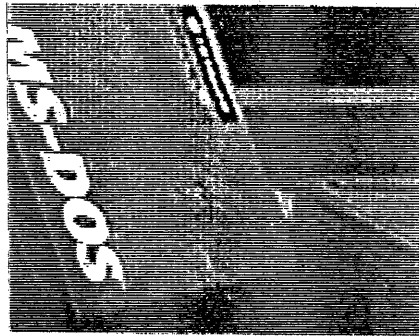


Fig.6 Input image

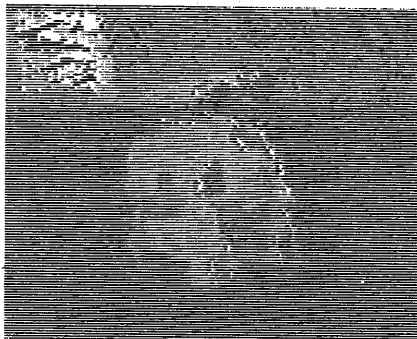


(A)

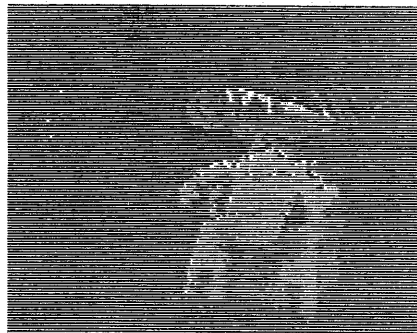


(B)

Fig.7 Recalled background by method (1)



(A)



(B)

Fig.8 Extracted image by method (2)

より正確に、背景を構成するためには、入力画像のエッジ強調像に、時間微分をほどこし、その値が、小さい画素のに関してのみ、(5)式を適用する方法が、考えられる。

[4]. 実験結果

前節で示した2つの方法の実験結果を示す。(1)の方法は、計算機シミュレーションで、(2)の方法は、ハードウェアと一部ソフトウェアを使用した装置で実験を行なった。図6は入力画像、図7は(1)の方法により、形成された背景像を示しており、図8は(2)の方法により抽出された、移動物体像を示している。図7、図8の両方とも(A)の方は、時刻 t が小さいもの、つまり処理があまり進んでいない段階のもので、(B)は t が十分に大きいものを示してある。

[5]. おわりに

本報告では、人間の視覚系に関する生理学的実験、及び心理学的実験により得られた知見に基づく、移動物体を抽出する、新しい生体の情報処理モデルを提案した。このモデルでは、視覚系の多くの情報処理が、脳の視覚野及び連合野で行なわれていると、仮定しており従来からあった網膜のY型細胞を使用した、移動物体抽出モデルとは、かなり異なった反応を示す。また、生体の情報処理モデルに基づいた移動物体抽出アルゴリズム、および装置を2つ提案し、有効に動作することができた。

参考文献

- (1). 清水 豊：感覚機能センサの現状と動向、昭和58年度製品科学研究所研究講演会資料、生体計測と感覚模倣センサー、pp. 1-10 (1983)。
- (2). 辻 三郎：動画像処理、情報処理、Vol. 21, No. 6, pp. 633-638 (June 1980)。
- (3). 森 見徳：運動物体の視覚情報処理に関する心理学的研究の現状、電子技術総合研究所調査報告第207号、ISSN 0366-9084 (1982. 10)
- (4). T. Kohonen and E. Oja: Fast Adaptive Formation of Orthogonalizing Filters and Associative Memory in Recurrent Networks of Neuron-Like Elements, Biol. Cybernetics, Vol. 21, pp.85-95(1976)。