アクティブビジョンによる移動物体の追跡とシーンの探索に関する研究

竹内義則

目 次

図目次

表 目 次

第1章　序論

1.1　本研究の目的

人間の眼球は，視野の中心に対応する網膜上の中心窩と呼ばれる部分で高解像度を持ち，周辺に行くにしたがって解像度が低下している [1, 2]．人間の有効視野は視角にして約 10deg であり，視野の周辺では，視力が低いかわりに広い範囲をとらえることができ，周囲の状況把握をして危険回避をしたり，移動物体をとらえたりする時に役立つ．その結果，人間の視覚は，広い視野範囲をもつと同時に，視野の中心で詳細な観察を行なうことができる．なぜ，眼球がこのような不均一の解像度を持っているのかについて考察すると，以下のような理由が考えられる [3]．

- 人間の日常行動において必要な情報は，視野の一部分であり，視野全体の詳細な情報は必要ない．
- 注視点とその周辺の解像度を変えることにより，情報量を大幅に減少させると同時に，脳に不必要な情報を持ち込まない．
- 情報量の削減によって，脳での処理時間を削減できる．
- 広い視野を持ち，周辺環境の的確な情報把握を行なえる．

人間の視覚が不均一な解像度を持っているため，人間がシーンを観測する時，高解像度の中心視野を有効に活用するために注視位置を移動し，シーンから必要な部分を順に選択し詳細な観測を行なう．このとき視点の移動は無秩序に行われているのではなく，興味のある対象に向けられると言われている．また，被験者に与えられた観察のタスクによって，注視点の分布が変化することも知られている [4]．さらに，被験者は，観察シーン中の情報

量の大きい部分を注視することも知られている [5]. このような人間の視覚の仕組みに基づき，画像全体を決められた順にスキャンするのではなく，興味のある部分を選択し，処理するアクティブビジョンの研究がさかんに行われている [6-15].

従来の受動的な視覚 (passive vision) は，以下の2つの問題をもつ.

- 受動的な視覚では，与えられた画像から必要とするすべての情報を獲得しなければならないが，与えられる画像の分解能やフォーカス，あるいはオクルージョンの存在などによって，与えられた画像が必要なすべての情報を含んでいることは少ない.
- 仮に，必要なすべての情報を含む画像が得られたとしても，画像全体にわたって詳細な解析を行うのは，コストの面から事実上不可能である.

これらの問題点を解決するためには，視覚センサの能動的な動きにより，タスクが必要とする情報を探索する必要がある. さらに，得られる画像中の情報を取捨選択する機能が必要である.

人間の視覚は色，視差，動きといったさまざまな特徴を統合し，眼球を回転することによって，環境と相互作用し，環境の認識などのタスクを達成している. アクティブビジョンの重要な問題点の一つに，どの部分に注視点を移動するかということが挙げられる. 本研究では，まず，移動物体に着目し，移動物体の追跡機能を実現する. さらに，静止したシーンにおいて，定量的な尺度であるシーンの情報量を提案し，情報量の大きい部分を選択し，注視するシステムを実現する.

移動物体の追跡システムは，防犯装置やテレビジョン放送の分野で幅広い応用が考えられる. また，シーン内の情報量の大きい部分を注視するシステムは，距離の離れた，あるいは，危険な場所で働くロボットの視覚として有用である. また，背景のテクスチャの撮影など，コンピュータビジョンだけでなくコンピュータグラフィクスの分野にも応用が考えられる.

1.2 本論文の概要

本研究では，移動物体の追跡システムを実現した．人間の移動物体追跡時に観測される眼球運動の saccade と pursuit に基づいて，移動物体の位置，速度を NTSC の 1 フレームに相当する 33ms 以内で計算する手法を汎用画像処理装置上に実現した．さらに，人間の予測制御の機能を簡単にモデル化し，運動が規則的で予測可能な移動物体を遅れ時間ゼロで追跡する手法を実現した．このように，人間の追跡機能に基づいていることが，システムの特徴である．

つぎに，シーンから得られる情報量を定義し，情報量の大きい部分を注視するシステムを実現した．従来のシステムでは，シーンの注目領域を表す saliency map をヒューリスティックに決定していた．その代わりに，本システムでは，情報理論に基づいた information map を用いている．このように，情報理論を基礎にしている点が，本システムの特徴であるといえる．

本論文の構成は以下のとおりである．

2 章では，人間の視覚情報処理のなかで本研究に関連深い部分と，従来のアクティブビジョンシステムについて述べる．視覚情報処理は，移動物体追跡時の眼球運動と予測制御，視覚探索をモデル化する特徴統合理論について述べる．さらに，注視を定式化することによって，その利点を明らかにする．アクティブビジョンの主要な機能の一つである移動物体の追跡について，従来に提案されたシステムを述べ，最後に従来のアクティブビジョンについてまとめる．

3 章では，眼球の saccade 運動と pursuit 運動，予測制御に示唆を得たシステムについて述べる．まず，移動物体の位置，速度を 33ms 以内で計算する手法を汎用画像処理装置上に実現し，予測制御をモデル化し実現する．実環境で実験を行ない，システムがノイズに強く，運動が規則的な時，遅れのない追跡を行なうことが可能であることを示す．

4 章では，いくつかの情報量について簡単に述べた後，シーンを定量化する情報量について述べる．その情報量をさまざまな画像について計算し，その特徴について考察する．さ

らに，特徴間の確率的な従属性として存在する相互情報量についても実験により考察する．

5章では，4章で定量化した情報量を用い，シーンから情報量の大きい部分を獲得するシステムについて述べる．シーンの特徴としては，シーンを撮影して得られる，明度，彩度，色相を用い，それぞれについて探索を行なった結果について考察する．また，シーンの探索順序についても，システムの効率の観点から考察する．

最後に，6章で本論文をまとめるとともに，今後の課題について考察する．

第2章　視覚情報処理とアクティブビジョンシステム

2.1　まえがき

本論文で取り扱う移動物体の追跡システムと情報量に基づくシーン画像の獲得システムは，人間の視覚情報処理に示唆を得たものである．この章では，眼球運動を伴う視覚情報処理とアクティブビジョンシステムについて述べる．以下，2.2で，移動物体を追跡するときの眼球運動とそのモデルについて，2.3で，視覚によってものを探す行動とそのモデルについて述べる．2.4では，注視を定式化することにより，その利点を明らかにする．2.5で，従来の移動物体追跡システムについて述べ，2.6で，従来のアクティブビジョンについてまとめる．

2.2　移動物体追跡時の眼球運動

2.2.1　指標追跡時の眼球運動の種類

指標の動きを追跡する眼球運動は2つの成分からなっている．一つは非常に速い階段状の成分で，saccade（凝視間，飛躍，飛越）運動と呼ばれる．もう一つは，比較的遅く滑らかに変化する成分で，pursuit（追従）運動と呼ばれる．**図2.1**に2つの成分の混在した眼球運動を示す[17]．図2.1では，指標の動きを検出してから130msほど後に，指標と同じ速度で眼球が動き出す．これはpursuit運動である．しかし，指標と眼球の位置がずれているのでそれを補正する動き(saccade)が生じる．その後，指標の変化に従って眼球の位置も変化していく．機能の上から見ると，saccade運動は指標と視線方向のずれを補正するのに役立ち，pursuit運動は指標と視線方向の相対速度をゼロにするのに役立っている．

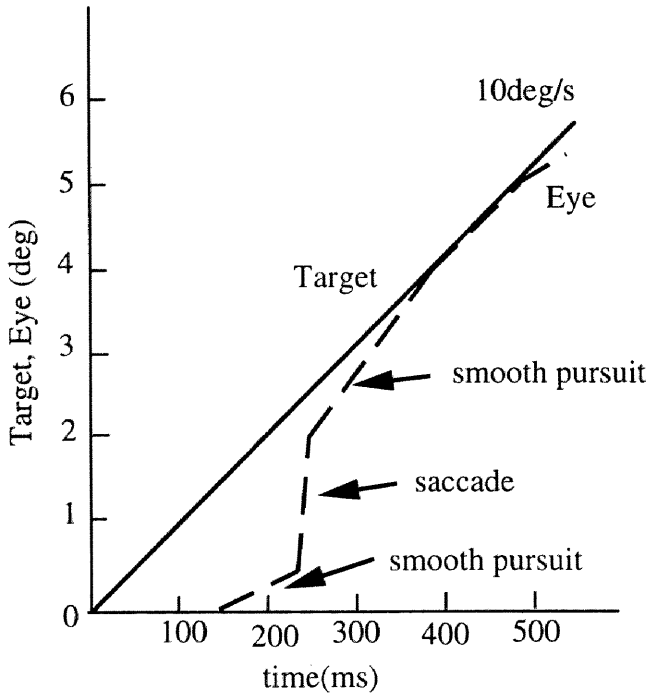


図 2.1: 眼球の saccade 運動と pursuit 運動 [17]

2.2.2 saccade 運動

saccade 成分の性質を調べるのに適した指標の動きは，ステップ状のものである．この時の応答を見ると，眼球運動系は約 200ms のむだ時間要素 ($T = 200ms$) と，二次系との縦列接続で表すことができる．なお，入力振幅が 20deg のとき，二次系の減衰係数ζは約 0.7，固有角周波数ω_nは約 120rad/s であることが報告されている [18].

しかし，入力がパルス状のときの応答は，系がむだ時間要素と二次系との縦列接続では表すことができない．そのときの応答の様子を**図 2.2** に示す．この図からわかるように，上向きの saccade が起きてから 200ms 以上もたってから下向きの saccade が起きる．これは，指標と眼球の位置のずれが離散的な時点でのみ補正されるシステム，すなわちサンプル値制御系を用いれば容易に説明することができる．**図 2.3** に saccade 運動のモデルを示す [19, 20]．ここで，sはラプラス演算子である．

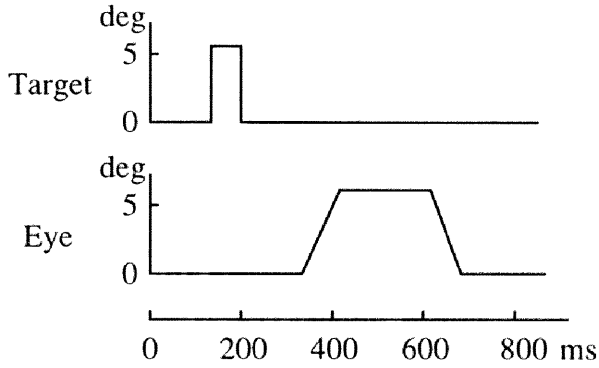


図 2.2: パルス応答

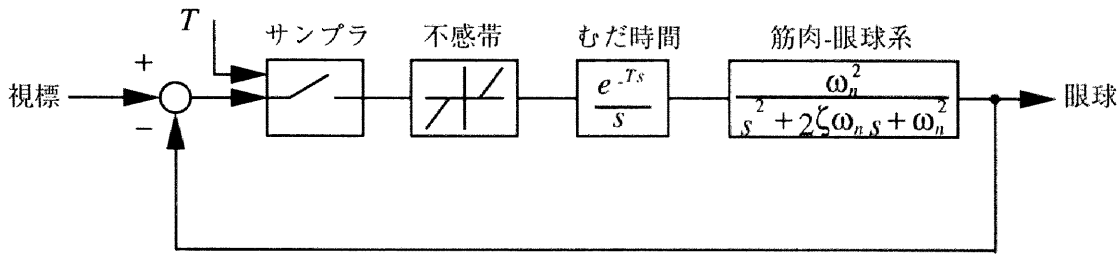


図 2.3: saccade 運動のモデル [20]

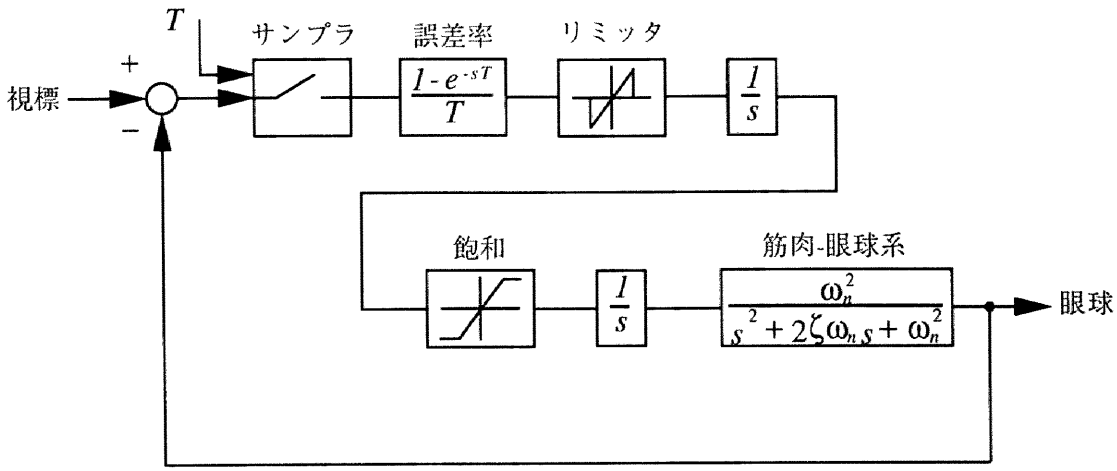


図 2.4: pursuit 運動のモデル [20]

2.2.3 pursuit 運動

Young らは，pursuit 運動もサンプル値制御系でモデル化できると主張し，**図 2.4** に示すような，pursuit 運動のモデルを提案した [20]．このモデルは，あるサンプル時点での位置の誤差と，その 1 つ前のサンプル時点での誤差との差から，速度の誤差を算出し，それにより眼球を駆動していることになる．このモデルは Young らも認めているように，いくつかの単純化がなされている．1 つは，pursuit 運動がサンプル値系であるという仮定はほとんど実験的裏付けがない．次に，pursuit 運動のむだ時間は saccade 運動とほぼ同じと仮定しているが，実際は約 125ms と短い．このように若干の問題点はあるが，2.2.4 節の統合モデルを考える場合に便利であり，また大筋では正しいモデルである．

2.2.4 統合モデル

saccade 運動と pursuit 運動とが別の信号経路によって生じるものであることは，多くの実験結果から，ほぼ間違いないものと思われる．したがって，眼球運動の統合的なモデルは図 2.3 の saccade 運動と図 2.4 の pursuit 運動をたし合わせた**図 2.5** のようなものになる [20]．このモデルを用いてシミュレーションした結果と，人間の眼球運動の実測値を種々の入力について比較した結果から，ランプ入力やステップ入力に対する応答はきわめてよく，

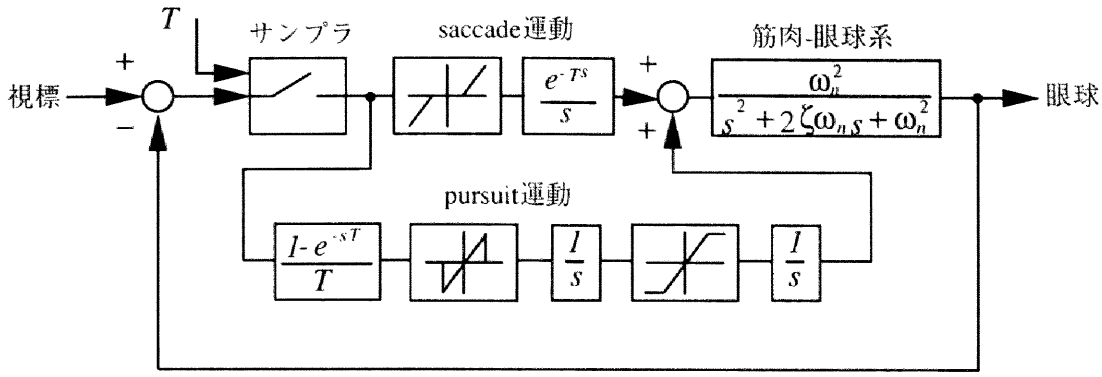


図 2.5: 眼球運動系の統合モデル [20]

パラボラ入力に対しても傾向はおおむね一致していることがわかっている．また，周波数応答についても，モデルと実測値がかなりよく一致することを示している [20].

2.2.5 予測制御

視標の運動が不規則で，予測できない時に起こるステップ入力に対する応答は，saccade であり，そのむだ時間は，約 200ms である．しかし，運動が規則的で，その未来値が予測できるときは，むだ時間の少ない応答を示す．例えば，規則的な矩形波入力に対しては，むだ時間は，呈示後のサイクル数が増加すると減少し，ある定常値に落ちつく．その定常値は，矩形波の繰り返し周波数が 0.4～1.0Hz ではほとんどゼロ，それより周波数が高くなっても，低くなっても，むだ時間は大きくなり，予測不能なステップ入力に対するむだ時間の値に近づく [19, 21, 22].

予測制御系の構成は，**図 2.6** のようなものと考えられる [23]．入力の過去から現在までの値を記憶し，入力信号のパターンを推定し，入力の未来値を予測し，何らかの評価関数を最小にするような制御信号を制御対象に加える．さらに，予測した値と実際の値を比較して，制御系全体のパラメータを調整する．

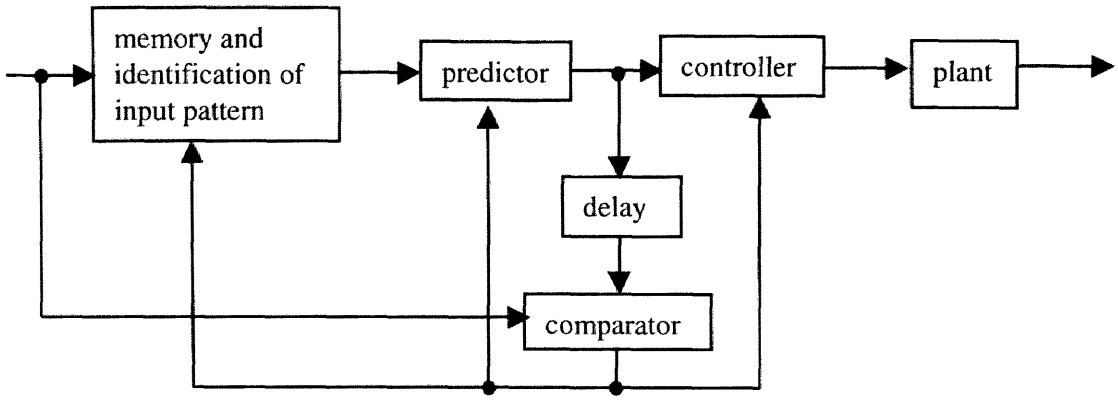


図 2.6: 予測制御系の構成 [23]

2.3 視覚探索と特徴統合理論

2.3.1 視覚探索

人間は，普段から視覚的に何かものを探すという行動をする．動物の場合も同じように，例えば餌を探すと言う行動をとる．ここでは，その視覚探索について考える．

視覚探索とは，様々な妨害刺激をもつ視覚場面から特定の目標刺激を検出する知覚的，認知的行動である．このとき，ポップアウトのように目標刺激が目立って知覚されることもあれば，電話帳を調べる時のように，目標刺激が見つかるまで順番に妨害刺激を確認しなければならないこともある．このような違いを解明することが視覚探索の研究の目的の一つであり，人間にさまざまな情報を探索させ，その時空間特性を分析することによって，視覚系の空間的並列処理限界や注意のメカニズムを明らかにしてきた [24, 25].

このような違いが出てくるのは，網膜からの視覚情報が常に並列に入力されていながら，そのすべてに注意を払うことができるわけではないことを意味している．そこで，視覚系には初期段階の並列処理とそれに続く逐次処理があると考えられている [26]．したがって，並列処理だけで目標が見つかる場合と，その後の逐次処理で目標を探さなければならない場合がある．逐次探索の場合は，多くの探索時間を必要とする．このとき，注意の機能を局所的な特徴統合ととらえ，並列処理では探索不可能な場合は，注意によって特徴統合しながら逐次処理が行なわれると考えられている [27]．このことについて，以下の節で詳し

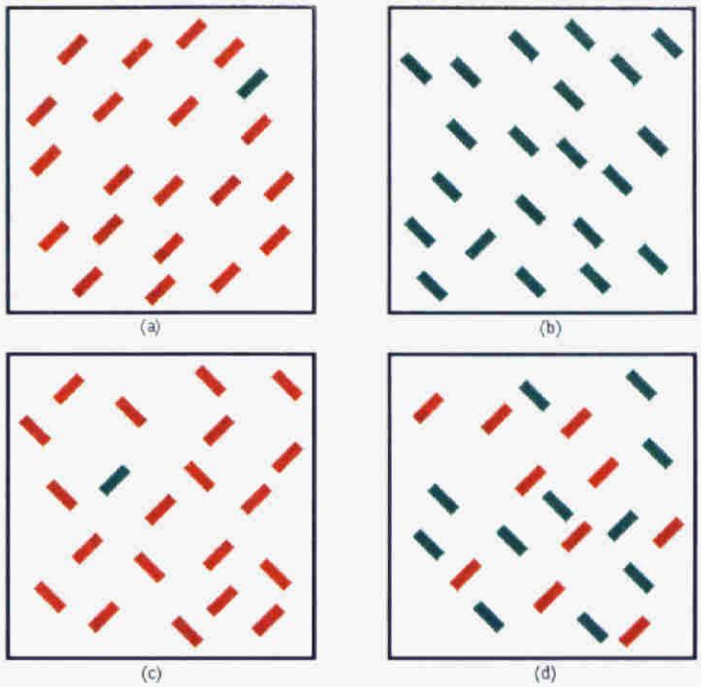


図 2.7: 視覚探索に用いられる刺激図形例 [28]

く述べる.

2.3.2 特徴統合理論

視覚探索の課題例を**図 2.7** に示す. 緑の右上斜線は, 図 2.7(a),(b),(c) ではすぐに見つかるが, 図 2.7(d) では時間がかかる. このように, 同じ目標刺激を探索する場合でも, それがすぐに見つかる場合と, 妨害刺激を順々に探索しなければならない場合がある.

特徴統合理論 [26] では, このような探索にかかる時間の違いを, 特徴マップを並列に生成する段階と, 注意を向けることによって特徴マップを局所的に結合する段階の 2 段階に分けることによって説明している.

第 1 段階では, 視覚領域に対して並列に特徴マップの集合が作られる. それぞれの特徴マップは, マップの特定の領域に赤の存在や, 斜線の方向などの特徴があるかないかを示している. これは, 生理学的知見に基づいており, サルの研究から, V1 野で形や色や運動などの成分に分け, これらの成分を直接あるいは V2 野経由で V3 野と V4 野と V5 野に分

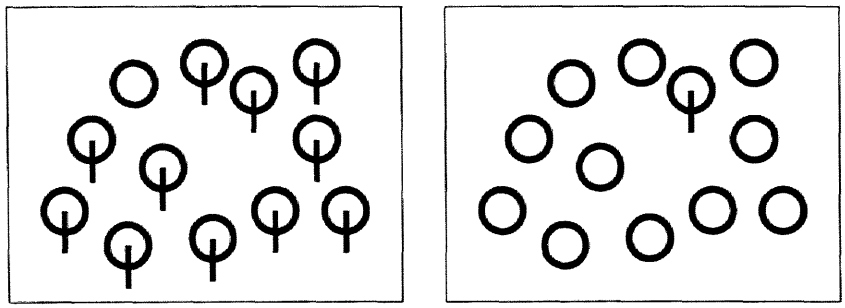


図 2.8: 探索非対称性の例

配していることが知られている [29]. V5野は運動成分を扱う領域，V4野は色成分を扱う領域，V3野はV4野とともに形成分を扱う領域である.

第2段階で，注意がある特定の領域に向けられることによって，それぞれのマップの情報を局所的に結合し，緑の右上斜線といった対象の認知が可能になる.

このように，視覚探索の難しさは第1段階の処理である特徴探索か，第2段階の処理である結合探索かによって決まる. 特徴統合理論によると，図2.7(a),(c) のように赤の妨害刺激から緑の目標を探索する特徴探索ならば，緑の特徴マップの活性度で目標の有無が分かる. このマップは並列に処理されるとしているため，探索時間は妨害刺激の数にほとんど影響を受けない. 一方，図2.7(d) のように，緑の左上斜線と赤の右上斜線の妨害刺激から目標である緑の右上斜線を探索する結合探索では，探索時間は妨害刺激の数に比例する.

あらかじめ目標の位置を与えておくと，注意の役割を直接的に調べることができる. 特徴探索では，あらかじめ目標の位置を与えておいても探索時間にあまり違いはないが，結合探索では非常に速くなる [30]. 特徴探索と結合探索の違いは，色，形，大きさ，動きなどの特徴次元でも支持されている [31, 32, 33].

2.3.3 特徴マップ

特徴マップは，それぞれ独立に存在すると考えられ，生理学的に明らかになってきた初期視覚の特徴モジュールに対応させることができる. 特徴モジュールでは，色，形，動きな

どの特徴を抽出し，特徴マップは，対応する特徴の有無を示すものと考えることができる．このとき，特徴マップにおける符号化を検討する心理学的知見として，探索非対称性があげられる．探索非対称性とは，視覚探索課題で一対のパターンが目標刺激と妨害刺激の役割を交代したときに，目標刺激の探索時間が著しく異なる現象をいう．例えば，**図2.8**(a)のような「円」の有無を複数の「垂直線を有する円」の中から探索するのは，図2.8(b)のような「垂直線を有する円」の有無を，複数の「円」の中から探索するよりも探索時間が長くなる．探索非対称性は,「垂直な線分」と「やや傾いた線分」,「暗い灰色」と「明るい灰色」,「原色」と「原色からずれた色」など，形，明るさ，色などの特徴で起こることが報告されている [32, 34].

Treismanは，自らが発見した探索非対称性という現象を，彼女自信が提唱している特徴統合理論で解釈しようとした．それによると，目標刺激と妨害刺激を弁別する特徴に対応している特徴マップが，それらの刺激からくる入力によって活性化される際，目標刺激が多く特徴を含んでいるため強く活性化され，妨害刺激は特徴が少ないため弱く活性化されるとき，探索はし易く，その逆の場合には探索はしにくいという一般原則が成り立つという．図2.8の例では,「垂直線を有する円」が「円」よりも強く活性化されるために，図2.8(a)では探索しにくく，図2.8(b)では探索し易くなる．さらに彼女は，古典的な精神物理学的法則であるWeberの法則（二つの刺激を区別するのに必要とされる最小差異は刺激の強さに比例する）からも説明できるとしている．

2.3.4 注意と眼球運動

逐次処理を行なうときに生じる眼球運動は，saccade運動ではなく，注意によって連続的に行なわれると考えられている．なぜなら，短時間提示でも提示時間に制約がないときと同じ特性が得られ，眼球運動に依存しない結果が得られているためである [32]．さらに，視線方向が注視点から移動しなかった試行のみを分析しても，そのような制約のない通常の場合と等しい探索時間が得られている [35]．このように，眼球運動を伴わずに内在化したも

のが注意であるとする立場がある一方，注意を眼球運動のために不可欠な準備とみなす立場も存在する [36]．さらに，これらの立場の関係が独立でありえないことも示唆している．

2.3.5 注意の誘導

特徴統合理論は，複数の特徴が組み合わされたすべての結合探索には，逐次処理が必要であるとしている．しかし，大きさの著しい差，水平線分と垂直線分，十字と円，赤と緑のいずれかの結合探索の場合，探索時間が刺激数によらず高速であり，これを特徴統合理論で説明することは難しい．これを説明するため，逐次処理が並列処理による情報に導かれる誘導探索モデルが提案されている [37]．

このモデルは，並列処理の後，各々の特徴は活性化マップに統合される．逐次処理では，注意は活性化マップによって最も目標らしい位置へ移動することができる．もし，並列処理からの信号が付加ノイズに対して大きいならば，目標をすぐに見つけることができる．逆に，この信号が小さいならば，目標を見つける前に多くの妨害刺激を確かめることになる．

誘導探索モデルはニューラルネットワークのシミュレーションモデルによって，特徴統合理論では説明できないいくつかの実験と一致した結果を得ている [38]．このモデルでは，ボトムアップとトップダウンの2要素の活性化値に正規ノイズを加えて活性化マップが算出される．ボトムアップの活性化値は，ある位置の特徴とそのまわりの特徴との差異を指数関数で強調し，トップダウンの活性化値は，ある位置の特徴と目標値との類似度に基づいている．このように，注意は単一過程ではなく，少なくとも二つの過程で決定されると考えられている．

これに対して，Treisman は，特徴マップの抑制機構によって制御されるという，特徴統合理論の修正によって説明している [39]．誘導探索モデルが，目標の位置の活性化によって探索が容易になるとしているのに対して，抑制機構に基づくモデルは，妨害刺激特徴を含む位置の抑制によって探索が容易になるとしている．

2.3.6 視覚情報の階層表現

目標の形状を判断する課題と，目標の有無，色を判断する課題を比較すると，妨害刺激の数に対する探索時間の変化に違いがあることが報告されている [40]．この違いは注意の大きさと解像度のトレードオフの問題であると説明される．すなわち，形状の判断だけ高い解像度の注意が必要であり，それ以外の課題は分散的注意しか必要としない．この結果は，中心窩では逐次処理，近中心窩では比較的大きな受容野を持つ，解像度の低い検出器で並列処理が行なわれているという報告 [41] とも一致する．

これらの結果は，解像度ピラミッドモデルによって説明される [21, 42, 43]．解像度ピラミッドモデルでは，注意モデルとして，単一焦点，注意領域の大きさの連続変化を仮定するズームレンズ [44]，コントラストや色のエッジと運動成分による注意関数，移動の近接性，注意位置への復帰の抑制などを仮定している．また，シミュレーションによって高い解像度が必要となる心理実験の結果と一致していることが確認されている [28]．

2.4 注視の定式化

2.4.1 動きモデル

ここでは，ピンホールカメラモデルを元に注視を定式化し，その利点について明らかにする．そのために，まず座標系を設定し，環境内を移動する点の画像面での速度を計算する．

図 2.9 に示すとおり，カメラ座標はカメラの投影中心を原点として設定される．画像面は環境の画像が投影された位置にある．

この座標系では画像面は $X-Y$ 平面に平行で，$X-Y$ 平面からの距離が焦点距離と等しい．環境内のすべての点$\boldsymbol{P}$は画像面内の点$\boldsymbol{p}$に移される．

このような座標系で環境内の点は，

$$\boldsymbol{P} = (X\ Y\ Z)^t \tag{2.1}$$

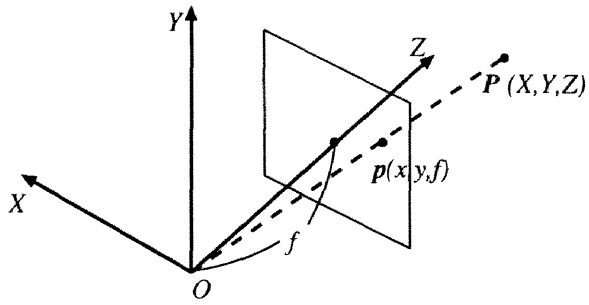


図 2.9: カメラの投影中心を原点とした座標系

それに対応する画像面の点は，

$$\boldsymbol{p} = (x\ y\ f)^t \tag{2.2}$$

と表される．f は焦点距離である．画像面は $Z = f$ の平面に置かれている．画像面の原点は Z 軸との交点で，x, y 軸は X, Y 軸に平行である．$\boldsymbol{p}$ と $\boldsymbol{P}$ の関係は，次の中心投影の関係式を満たす．

$$\boldsymbol{p} = (x\ y\ f)^t = \left(\frac{fX}{Z}\ \frac{fY}{Z}\ \frac{fZ}{Z}\right)^t = \frac{f\boldsymbol{P}}{\boldsymbol{P}\cdot\hat{\boldsymbol{z}}} \tag{2.3}$$

ここで，$\hat{\boldsymbol{z}} = (0\,0\,1)^t$ である．もし，環境内の点 $\boldsymbol{P}$ が並進速度 $\boldsymbol{v} = \left(\dot{X}\ \dot{Y}\ \dot{Z}\right)^t$，角速度 $\boldsymbol{\omega} = (\omega_x\ \omega_y\ \omega_z)^t$ で運動しているとすると，その点 $\boldsymbol{P}$ の速度は，

$$\frac{d\boldsymbol{P}}{dt} = \boldsymbol{v} + \boldsymbol{\omega}\times\boldsymbol{P} \tag{2.4}$$

と表される．$\boldsymbol{P}$ が移動することによって，$\boldsymbol{p}$ も移動する．その速度は，

$$\begin{aligned}\frac{d\boldsymbol{p}}{dt} &= \frac{d}{dt}\left(\frac{f\boldsymbol{P}}{\hat{\boldsymbol{z}}\cdot\boldsymbol{P}}\right)\\ &= f\frac{d\boldsymbol{P}}{dt}\frac{1}{\boldsymbol{P}\cdot\hat{\boldsymbol{z}}} - \frac{f\boldsymbol{P}}{(\boldsymbol{P}\cdot\hat{\boldsymbol{z}})^2}\left(\hat{\boldsymbol{z}}\cdot\frac{d\boldsymbol{P}}{dt}\right)\\ &= \frac{1}{\boldsymbol{P}\cdot\hat{\boldsymbol{z}}}\left\{f\frac{d\boldsymbol{P}}{dt} - \frac{f\boldsymbol{P}}{\boldsymbol{P}\cdot\hat{\boldsymbol{z}}}\left(\hat{\boldsymbol{z}}\cdot\frac{d\boldsymbol{P}}{dt}\right)\right\}\end{aligned}$$

である．さらに，式 (2.3) と $\hat{\boldsymbol{z}}\cdot\boldsymbol{p} = \hat{\boldsymbol{z}}\cdot\frac{f\boldsymbol{P}}{\boldsymbol{P}\cdot\hat{\boldsymbol{z}}} = f$ より，

$$\frac{d\boldsymbol{p}}{dt} = \frac{1}{\boldsymbol{P}\cdot\hat{\boldsymbol{z}}}\left\{(\hat{\boldsymbol{z}}\cdot\boldsymbol{p})\frac{d\boldsymbol{P}}{dt} - \boldsymbol{p}\left(\hat{\boldsymbol{z}}\cdot\frac{d\boldsymbol{P}}{dt}\right)\right\}$$

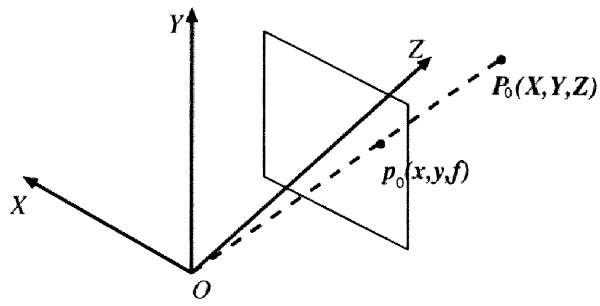


図 2.10: 注視点

公式$\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \cdot \boldsymbol{c})\,\boldsymbol{b} - (\boldsymbol{a} \cdot \boldsymbol{b})\,\boldsymbol{c}$ より，

$$\frac{d\boldsymbol{p}}{dt} = \frac{f\hat{\boldsymbol{z}} \times \left(\frac{d\boldsymbol{P}}{dt} \times \boldsymbol{p}\right)}{\boldsymbol{P} \cdot \hat{\boldsymbol{z}}} \tag{2.5}$$

が得られる．

2.4.2 注視の束縛

我々人間は，動物体を追跡する際に，目や頭を動かさずにいるわけではない [45, 46]．目や頭を物体が進む方向へ動かせて，網膜上で物体を静止させている．視野の中である物体に着目することを注視という [7, 47]．ここでは，注視によって得られた束縛条件について述べる．

図 2.10 のように，画面上の注視点を$\boldsymbol{p_o}$とすると，注視の束縛条件とは，注視点を動かさないこと，すなわち，

$$\frac{d\boldsymbol{p_o}}{dt} = 0 \tag{2.6}$$

となる．一方，式 (2.5) より，$\boldsymbol{p_o}$では次のようになる．

$$\frac{d\boldsymbol{p_o}}{dt} = \frac{f\hat{\boldsymbol{z}} \times \left(\frac{d\boldsymbol{P_o}}{dt} \times \boldsymbol{p_o}\right)}{\boldsymbol{P_o} \cdot \hat{\boldsymbol{z}}} \tag{2.7}$$

ここで，$\boldsymbol{P_o}$は環境内の注視点である．式 (2.6) と式 (2.7) より

$$\hat{\boldsymbol{z}} \times \left(\frac{d\boldsymbol{P_o}}{dt} \times \boldsymbol{p_o}\right) = 0 \tag{2.8}$$

が得られる．これから，$\frac{d}{dt}\boldsymbol{P_o} \times \boldsymbol{p_o}$が0か，$\hat{\boldsymbol{z}}$に並行であることがわかる．もし，$\frac{d}{dt}\boldsymbol{P_o} \times \boldsymbol{p_o}$が$\hat{\boldsymbol{z}}$に並行であるとすると，$\boldsymbol{p_o}$は$\hat{\boldsymbol{z}}$に垂直となる．これは，限られた視野の中では不可能である．よって，

$$\frac{d\boldsymbol{P_o}}{dt} \times \boldsymbol{p_o} = 0 \tag{2.9}$$

となる．$\boldsymbol{P_o}$は$\boldsymbol{p_o}$と同じ方向なので，式(2.9)は，

$$\frac{d\boldsymbol{P_o}}{dt} \times \boldsymbol{P_o} = 0 \tag{2.10}$$

とも表せる．式(2.4)より，$\frac{d}{dt}\boldsymbol{P_o} = \boldsymbol{v} + \boldsymbol{\omega} \times \boldsymbol{P_o}$であるから，これで，式(2.10)を置き換えると，

$$(\boldsymbol{\omega} \times \boldsymbol{P_o}) \times \boldsymbol{P_o} + \boldsymbol{v} \times \boldsymbol{P_o} = 0 \tag{2.11}$$

公式$(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} = (\boldsymbol{c} \cdot \boldsymbol{a})\,\boldsymbol{b} - (\boldsymbol{c} \cdot \boldsymbol{b})\,\boldsymbol{a}$より，

$$(\boldsymbol{P_o} \cdot \boldsymbol{\omega})\,\boldsymbol{P_o} - (\boldsymbol{P_o} \cdot \boldsymbol{P_o})\,\boldsymbol{\omega} + \boldsymbol{v} \times \boldsymbol{P_o} = 0 \tag{2.12}$$

となる．$\boldsymbol{v} \times \boldsymbol{P_o} \neq 0$すなわち，$\boldsymbol{v}$が0でもなく，$\boldsymbol{P_o}$に平行でもなければ，$\boldsymbol{\omega}$は$\boldsymbol{P_o}, \boldsymbol{v} \times \boldsymbol{P_o}, \boldsymbol{v}$の線形和で表せる．従って，

$$\boldsymbol{\omega} = \alpha \boldsymbol{P_o} + \beta\,(\boldsymbol{v} \times \boldsymbol{P_o}) + \gamma \boldsymbol{v} \tag{2.13}$$

となる．ここで，α, β, γはパラメータである．

式(2.13)を式(2.12)に代入して，

$$\{1 - \beta\,(\boldsymbol{P_o} \cdot \boldsymbol{P_o})\}\,(\boldsymbol{v} \times \boldsymbol{P_o}) + \gamma\,(\boldsymbol{P_o} \cdot \boldsymbol{v})\,\boldsymbol{P_o} - \gamma\,(\boldsymbol{P_o} \cdot \boldsymbol{P_o})\,\boldsymbol{v} = 0 \tag{2.14}$$

これから，βとγを求める．まず，式(2.14)と$\boldsymbol{v} \times \boldsymbol{P_o}$の内積をとると，

$$\{1 - \beta\,(\boldsymbol{P_o} \cdot \boldsymbol{P_o})\}\,|\boldsymbol{v} \times \boldsymbol{P_o}|^2 = 0 \tag{2.15}$$

$\boldsymbol{v} \times \boldsymbol{P_o} \neq 0$より，

$$\beta = \frac{1}{|\boldsymbol{P_o}|^2} \tag{2.16}$$

を得る．同様に式 (2.14) と$\boldsymbol{v}$の内積をとると，

$$\gamma\left(\boldsymbol{P_o}\cdot\boldsymbol{v}\right)\left(\boldsymbol{P_o}\cdot\boldsymbol{v}\right)-\gamma\left(\boldsymbol{P_o}\cdot\boldsymbol{P_o}\right)\left(\boldsymbol{v}\cdot\boldsymbol{v}\right)=0 \tag{2.17}$$

公式 $(\boldsymbol{a}\times\boldsymbol{b})\cdot(\boldsymbol{c}\times\boldsymbol{d})=(\boldsymbol{c}\cdot\boldsymbol{a})(\boldsymbol{b}\cdot\boldsymbol{d})-(\boldsymbol{b}\cdot\boldsymbol{c})(\boldsymbol{d}\cdot\boldsymbol{a})$ より，

$$\gamma\left|\boldsymbol{v}\times\boldsymbol{P_o}\right|^2=0 \tag{2.18}$$

よって，

$$\gamma=0 \tag{2.19}$$

式 (2.13) に，式 (2.16)，式 (2.19) を代入して，

$$\boldsymbol{\omega}=\alpha\boldsymbol{P_o}+\frac{1}{\left|\boldsymbol{P_o}\right|^2}\left(\boldsymbol{v}\times\boldsymbol{P_o}\right) \tag{2.20}$$

を得る．αは，まだ決めることができない．つまり，$\boldsymbol{P_o}$軸回りの回転は注視の束縛条件では求めることができない．そこで，αを$\boldsymbol{\omega_{P_o}}$とおくことにする．

さて，$\boldsymbol{v}\times\boldsymbol{P_o}=0$($\boldsymbol{v}$が 0 か$\boldsymbol{P_o}$に垂直) の場合，式 (2.12) は，

$$\left(\boldsymbol{P_o}\cdot\boldsymbol{\omega}\right)\boldsymbol{P_o}-\left(\boldsymbol{P_o}\cdot\boldsymbol{P_o}\right)\boldsymbol{\omega}=0 \tag{2.21}$$

となる．ここで，$\boldsymbol{v}$と平行でないベクトル$\boldsymbol{x}$を用いて，$\boldsymbol{\omega}$は，$\boldsymbol{P_o},\boldsymbol{x}\times\boldsymbol{P_o},\boldsymbol{x}$の線形和で表される．

$$\boldsymbol{\omega}=\lambda\boldsymbol{P_o}+\mu\left(\boldsymbol{x}\times\boldsymbol{P_o}\right)+\nu\boldsymbol{x} \tag{2.22}$$

ここで，λ,μ,νはパラメータである．先と同様に，これらのパラメータを求める．

式 (2.22) を式 (2.21) に代入して，

$$\nu\left(\boldsymbol{P_o}\cdot\boldsymbol{x}\right)\boldsymbol{P_o}-\mu\left(\boldsymbol{P_o}\cdot\boldsymbol{P_o}\right)\left(\boldsymbol{x}\times\boldsymbol{P_o}\right)+\nu\left(\boldsymbol{P_o}\cdot\boldsymbol{P_o}\right)\boldsymbol{x}=0 \tag{2.23}$$

式 (2.23) と$\boldsymbol{x}\times\boldsymbol{P_o}$の内積をとると，

$$-\mu\left(\boldsymbol{P_o}\cdot\boldsymbol{P_o}\right)\left|\boldsymbol{x}\times\boldsymbol{P_o}\right|^2=0 \tag{2.24}$$

$\boldsymbol{x} \times \boldsymbol{P_o} \neq 0$ より，

$$\mu = 0 \tag{2.25}$$

を得る．同様に式 (2.23) と $\boldsymbol{x}$ の内積をとると，

$$\nu (\boldsymbol{P_o} \cdot \boldsymbol{x}) (\boldsymbol{P_o} \cdot \boldsymbol{x}) - \nu (\boldsymbol{P_o} \cdot \boldsymbol{P_o}) (\boldsymbol{x} \cdot \boldsymbol{x}) = 0 \tag{2.26}$$

公式 $(\boldsymbol{a} \times \boldsymbol{b}) \cdot (\boldsymbol{c} \times \boldsymbol{d}) = (\boldsymbol{c} \cdot \boldsymbol{a}) (\boldsymbol{b} \cdot \boldsymbol{d}) - (\boldsymbol{b} \cdot \boldsymbol{c}) (\boldsymbol{d} \cdot \boldsymbol{a})$ より，

$$\nu |\boldsymbol{x} \times \boldsymbol{P_o}|^2 = 0 \tag{2.27}$$

よって，

$$\nu = 0 \tag{2.28}$$

式 (2.13) に，式 (2.25)，式 (2.28) を代入して，

$$\boldsymbol{\omega} = \lambda \boldsymbol{P_o} \tag{2.29}$$

同様に，λ も求めることができない．よって，λ を $\boldsymbol{\omega_{P_o}}$ とおくことにする．

このようにして $\boldsymbol{\omega}$ について得られた式，

$$\boldsymbol{\omega} = \boldsymbol{\omega_{P_o}} \boldsymbol{P_o} + \frac{1}{|\boldsymbol{P_o}|^2} (\boldsymbol{v} \times \boldsymbol{P_o}) \tag{2.30}$$

は，$\boldsymbol{\omega}$ が $\boldsymbol{P_o}$ 軸回りの回転成分 $\boldsymbol{\omega_{P_o}}$ と並進成分 $\boldsymbol{v}$ から求めることができることを意味する．回転運動の 2 方向の成分は，注視によって得られた束縛条件によって削除された．したがって，自由度が 6 から 4 へ低下していることがわかる．このように注視によって，計算コストを下げることが可能である [7, 47, 48].

2.5 従来の移動物体追跡システム

移動物体の追跡はアクティブビジョンシステムの基本的な能力の一つであり，これまでに，数多くの追跡システムが提案されている．以下で，今までに開発されたシステムについて述べる．ここで述べるシステムは，移動物体を視野の中央にとらえるため，カメラを

移動させるものである．カメラを固定して，視野内の移動物体を追跡するものは除外している．

Hwangらはズーミングの機能を持ったオンライン追跡システムを開発した[49, 50]．このシステムは移動物体を追跡すると同時に，移動物体を視野内で最適な大きさにするようにズーミングを行うことができる．しかし，システムのサンプリング間隔は1.6sで，非常にゆっくりした移動物体しか追跡することができない．

Uhlinらは両眼視のビジョンシステムを提案した[51, 52]．特に，静止している，または運動している他の物体による遮蔽に注目し，3つの手がかり(カメラと移動物体の相対運動，移動物体の動き，視差)を統合して，ロバストなふるまいを達成した．さらに，画像全体からカメラの運動を推定する方法として，アフィンフローモデルを提案した．このモデルは，カメラの回転運動だけでなく，並進運動も推定することができる．しかしながら，移動物体が視野に非常に小さく映っているという仮定が必要である．

FerrierらはRobinsonが考案した人間の眼球運動制御に基づき，カメラ制御のスキーマを提案し，それを実現した[14, 16]．しかし，実現したシステムは，実環境で移動物体と背景を分離することができないため，黒い背景中を移動する白い物体(あるいは，その逆)の場合にしか，追跡することができない．システムのサンプリング間隔は100msである．Fialaらのシステムも同様に，簡単な画像処理を用いている[53]．

D. Murrayらはパン，チルトステージにとりつけたアクティブカメラを用いた動きの検出法を提案した[9]．カメラの回転に伴うフローをキャンセルするために，カメラのモータのエンコーダの値を使い，画像をシフトすることによって背景の動きを補償する．画像間の差分値の大きい部分と，画像の勾配の大きい部分を組み合わせて，二値画像を得る．この二値画像に対して，ノイズ除去を行い，重心を求める．しかし，実時間で動作するシステムの実現は報告されていない．

Tölg[54]は，2-3Hzの制御周期で移動物体を追跡するシステムを提案した．オプティカルフローを計算し，それを分割することによって移動物体を検出する．移動物体の速度を

計算するために，カメラの動きを用いる．追跡は，saccadeとpursuitを順番に行う．

D. W. Murrayらは実時間で単眼の追跡をおこなうアクティブビジョンシステムを開発した[10, 11, 55, 56]．このシステムは，カメラのモータのエンコーダの値を使い，カメラの運動によるフローをキャンセルする．システムは，25Hzで動作し，処理の遅れ時間は110msである．カメラの移動は，saccadeとpursuitをモデルにしている．システムは，粗い周辺視野での動きの検出によってsaccadeを起こし，中心視野でのオプティカルフローによってpursuitを起こす．saccadeとpursuitは，有限オートマトンによって切り換えられている．このシステムは，角速度が18deg/sまでの移動物体まで追跡することができる性能をもつ．

CoombsとBrownは，垂直エッジを含む物体を7.5Hzの制御周期で追跡する両眼のシステムを開発した[57]．ゼロ視差フィルタ[58]を用いて移動物体の検出を行なっている．

國吉らは，Kaenelらが開発したゼロ視差フィルタ[58]を改良し，高速性と実現の容易さの向上した拡張ゼロ視差フィルタ法[59]を開発した．これを用いて，処理遅れ90msの追跡システムを開発した[60]．

Duは，カメラが静止している時に検出した物体を，時間相関を用いて追跡するシステムを開発した[61]．システムは，12.5Hzの制御周期で90msの遅れ時間を伴っている．

Diasらは，人間の歩く速さぐらいの移動物体を追跡する移動ロボットを開発した[62]．水平方向に移動する物体のみをフレーム間差分によって検出している．しかし，自分自身の動きによるフレーム間のずれは考慮していない．

Daniilidisらは，25Hzの制御周期で動作し，80msの遅れ時間をもつシステムを開発した[63]．カメラの回転軸の回転速度を測定し，自分自身の動きを補償して移動物体を検出する方法を実現した．このシステムは，10deg/sの移動物体を追跡することが可能である．

和田らは，カメラの回転中心とレンズ主点が一致しているカメラを用い，移動物体を追跡するシステムを開発した[64, 65, 66]．カメラの回転中心とレンズ主点が一致しているため，カメラの回転に伴う運動視差が生じない．この性質を利用して，背景をあらかじめ撮影しておき，入力画像とその差分をとることによって移動物体を検出することができる．さ

らに，環境にカメラを多く配置することによって，複数の移動物体を共同注視するシステムを開発した [67].

以上の研究は，視覚センサとしてCCDカメラを使用しているが，専用のビジョンチップによる高速なシステムもいくつか開発されている.

中坊らは，光検出器と汎用な処理回路を画素ごとに直結したものを1チップに集積化したビジョンチップデバイス [68, 69] を用いて，1msの制御周期で動作するアクティブビジョンシステムを開発した [70]. しかし，このシステムで行なわれている画像処理は単純なものであり，明るい部分のみを追跡することができる.

Horiuchiらは，時間遅れと相関演算をVLSIチップに組み込んだ動きの検出チップを開発し，1次元の方向のみの物体追跡システムを開発した [71, 72, 73].

2.6 従来のアクティブビジョン

能動的な視覚は，Bajcsy[74] によってその基本的な概念が提案され，Aloimonos[75] やBallard[76] らが定義した.

Aloimonosは，能動的な視覚を以下のように定義した.

> 観測者がセンサーの幾何的パラメータを制御するという目的を持つ，ある種の活動に従事するとき，その観測者は能動的 (active) である [75].

すなわち，Aloimonosは従来解きにくかったビジョンの問題 (shape from X，Xにはcontour, texture, shading, motion等が入る) を，解きやすくするものとして能動的なビジョンを定義した.

これに対し， BallardはAloimonosの能動的な視覚よりも広い研究の枠組みとして，人間の視覚行動からヒントを得たAnimate Visionを提唱した [7]. Animate Visionでは，タスクの文脈に応じて視線を制御する注視制御 (gaze control) に重点が置かれている.

ある目的のもとに環境内を探索する，観測者の注視点の動きを想像してみる. 人間は，与えられた目的を達成するため，環境内の様々な物体への注視を繰り返しながら，必要とす

る視覚情報を得る．このような視覚システムの利点は，必要な視覚情報を取捨選択することができることである．これらの利点により，複雑で動的に変化する環境においても，ロバストで高速な環境や物体の認識が可能となる．

Ballardは，このようなAnimate Visionには以下の3つが必要であるとしている．

- 中心窩を持つ網膜を利用するため視覚的順序づけによる視覚タスクの単純化
- 動きを補償するための注視制御
- 世界の予測不能性を補償するための学習

また，Aloimonosも先の定義に基づく研究の枠組みを拡張して，Purposive Visionを提案している[77, 78]．AloimonosのPurposive VisionもBallardのAnimate Visionも目指すものは同じと考えられるが，Purposive Visionでは，定性的な情報(正確な計測を必要とせず，非常に自由度の小さい空間上で表現される情報)を獲得するモジュールの集合としてシステムを構築することに重点を置いている．

他にも，Eklundh, Sandini等多くの研究者が能動的な視覚の研究を行っているが，その基本的な考えは，情報の量と質，空間的な広がりの点で制約された視覚センサに，能動的な視覚パラメータの制御を取り入れることにより，実時間でロバストな視覚情報処理を可能とするシステムを開発しようとするものであり，そのシステム開発における基本的なアイデアとして，どの研究も以下の2つを重視している．

- 視覚パラメータの能動的な制御により，問題を単純化して解く複数のモジュールによってシステムを構築する．
- 人間の視覚機能を模倣する（不均一な網膜を持つ視覚をタスクの文脈に応じて制御する）．

第3章　移動物体の追跡

3.1　まえがき

人間は容易に移動物体を発見し，移動する物体に視線を合わせて追跡することができる．こういった人間の視覚系と同じふるまいを持つ追跡システムを実現することは，コンピュータビジョンの重要なテーマの一つである．さらに，このような追跡システムは，防犯装置やテレビジョン放送の分野で幅広い応用が考えられる [79].

人間の視覚系の研究から，移動物体を追跡する場合の眼球運動の成分は，位置に依存したものと速度に依存したものの2種類があること [18] や，初期の段階で種々の特徴を検出し，それを統合して眼球の移動位置を決める機構があること [14] がわかっている．人間の視覚系と従来の受動的なビジョンシステムとの違いの一つは，眼球 (視覚センサ) が回転することである．このような視覚センサが能動的に行動し，環境と相互作用することによって，必要な情報を取捨選択する能動的な視覚 (active vision) の研究が，近年のコンピュータビジョンで注目されている [80].

移動する物体を追跡するビジョンシステムは，動きという情報を画像全体から抽出し，その物体を視野からはずれないようにするという点で能動的な視覚システムであると言える．このような追跡システムを実環境に適用する場合，次のような問題点が考えられる．

1. 移動する物体の速度にシステムが追いつかなければならない．
2. 実環境のノイズに対してロバストな動きの抽出を行なわなければならない．

本研究では，現在の技術水準で，実環境，実時間で動作可能なシステムの構築をめざす．そのために，まず，実時間で動作させるために汎用画像処理装置を用いて画像処理を高速

化し，高速な移動物体への追従を行なうために高速なパン・チルトステージを用いた．さらに，予測制御を用いて，規則的な運動を遅れ時間なしで追跡する手法を実現した．その結果，2.5で述べた従来のシステムよりも，追跡可能な最大速度を向上することが可能となった．また，ノイズに対して頑健にするために，移動物体の検出に時空間の微分を組み合わせた手法を用いた．以下，3.2で，移動物体の追跡方法，3.3で，実現したシステムについて述べる．システムを用いて行なった実験とその結果を3.4に述べ，3.5でその結果について考察する．最後に3.6で本研究をまとめるとともに，今後の課題について考察する．

3.2 移動物体の追跡方法

3.2.1 オプティカルフローの拘束方程式

画像中のある点(x,y)の速度（オプティカルフロー）を(u,v)，時刻tでの点(x,y)の輝度を$I(x,y,t)$とする．カメラ，あるいは物体がなめらかに移動しているとき，微小時間δtでは見え方は変わらない．すなわち，

$$I(x,y,t) = I(x+u\delta t, y+v\delta t, t+\delta t) \tag{3.1}$$

となる．式(3.1)では，照明の時間的，空間的変化や物体のオクルージョンはないことを仮定している．

輝度Iがx,y,tに関してなめらかに変化する，すなわち，x,y,tに関して$I(x,y,t)$が微分可能であると仮定すると，式(3.1)の右辺をテイラー展開することによって，

$$I(x,y,t) = I(x,y,t) + u\delta t I_x + v\delta t I_y + \delta t I_t + \varepsilon \tag{3.2}$$

ここで，I_x, I_y, I_tはそれぞれIのx,y,tによる偏微分，εは2次以上の項である．εを無視すると，

$$uI_x + vI_y + I_t = 0 \tag{3.3}$$

が得られる．この式(3.3)は，画像中の点(x,y)での速度の成分(u,v)の制約を表しているため，オプティカルフローの拘束方程式と呼ばれる[81].

3.2.2 移動物体の位置推定

環境中を移動する物体が存在した場合，その環境に固定されたカメラからその移動物体を含む領域を観測すると，移動物体は輝度値の時間的変化として検出することができる．したがって，入力画像を移動物体と背景に分離することは，入力画像の時間微分を利用することで容易に行なうことができる．時間微分の結果，その絶対値の大きさは，移動物体と背景のコントラストに依存している．しかし，この方法は照明の変化などの，実環境での変動（ノイズ）に敏感である．この問題を避けるため，エッジ画像を用いる．移動物体によって引き起こされる時間微分の変化は，その物体のエッジ付近で大きな値をとるためである．

$\mathbf{u} = (u, v)$，$\nabla I = (I_x, I_y)$ とおくと，式 (3.3) は，

$$\mathbf{u} \cdot \nabla I = -I_t \tag{3.4}$$

となる．任意のベクトル $\mathbf{a}, \mathbf{b}$ について，$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|$ であるから，式 (3.4) は，

$$|\mathbf{u}||\nabla I| \geq |I_t| \tag{3.5}$$

となる．したがって，移動物体の速度の大きさ $|\mathbf{u}|$ は，

$$|\mathbf{u}| \geq \frac{|I_t|}{|\nabla I|} \tag{3.6}$$

となるため，$|I_t|/|\nabla I|$ がある閾値よりも大きければ，移動物体が存在すると判定できる．

このように検出した移動物体の画像を I_m とおく．すなわち，

$$I_m(x, y) = \begin{cases} 1 & \text{if } |\nabla I(x, y)| > Th \text{ and} \\ & |I_t(x, y)| > K|\nabla I(x, y)| \\ 0 & \text{otherwise,} \end{cases} \tag{3.7}$$

である．ここで，Th と K は定数で，$|\nabla I(x, y)| > Th$ は，エッジを検出する閾値である．式 (3.7) はある画素 (x, y) において，高いコントラストと大きい時間微分値を持つとき，移動物体のエッジ画像 $I_m(x, y)$ が 1 になることを示している．

移動物体の位置は $I_m(x,y)$ の重心とする．すなわち，

$$\begin{aligned} d_x &= \frac{\sum_{(x,y)} x I_m(x,y)}{\sum_{(x,y)} I_m(x,y)} \\ d_y &= \frac{\sum_{(x,y)} y I_m(x,y)}{\sum_{(x,y)} I_m(x,y)} \end{aligned} \tag{3.8}$$

である．

3.2.3 移動物体の速度推定

輝度変化の条件式 (3.3) をもとに速度を推定する．移動物体は剛体で，視野内の狭い領域に存在し，視線軸まわりの回転運動を伴わないと仮定する．このとき，その物体のオプティカルフローは，位置によらずほぼ一定と近似できる．したがって，画像中の領域 $D \subseteq \mathbf{R}^2$ の速度を最小二乗法によって求めることができる．つまり，

$$\iint_D (uI_x + vI_y + I_t)^2 \, dxdy \tag{3.9}$$

を最小にする (u,v) を求める．その結果は以下のようになる．

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{bmatrix}^{-1} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{3.10}$$

ここで，

$$\begin{aligned} m_{11} &= \iint (I_x)^2 \, dx\, dy \\ m_{12} &= \iint I_x I_y \, dx\, dy \\ m_{22} &= \iint (I_y)^2 \, dx\, dy \end{aligned}$$

$$\begin{aligned} c_1 &= -\iint I_x I_t \, dx\, dy \\ c_2 &= -\iint I_y I_t \, dx\, dy \end{aligned}$$

である．この計算は，汎用画像処理装置上で高速に行なうことが可能である．

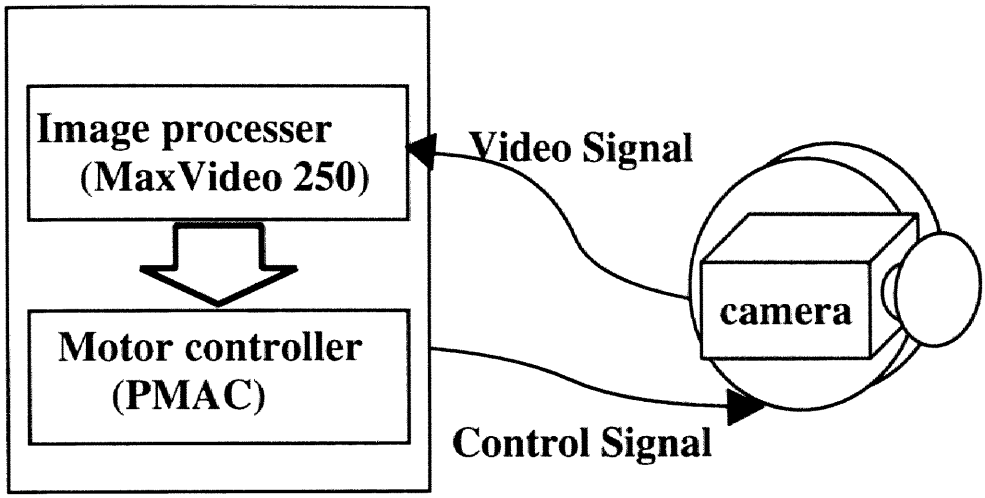


図 3.1: 移動物体追跡システムの概観

3.3 移動物体追跡システム

3.3.1 システムの概要

システムはカメラ，画像処理装置，パン・チルトステージからなる．**図 3.1** にシステムの概観を，**図 3.2** に，パン・チルトステージを示す．図 3.2 は，複眼のシステムであるが，移動物体の追跡は単眼のみで行なっている．カメラは小型 2/3" カラー CCD カメラ (SONY XC-711RR) を，レンズは 10 倍ズームレンズ (Fujinon H10x11E-X41) を用いた．カメラの出力 (NTSC) は，汎用画像処理装置に入力され，処理される．汎用画像処理装置は，ホストコンピュータ (MVME 167,Lynx OS) と VME スロットに接続された動画像処理 VME ボード (Datacube, MAXVIDEO 250) およびフルカラー画像入出力ボード (同, DIGI COLOR) からなる．この装置で移動物体を検出し，その予測位置を VME バスを介して，モータ制御装置 (DeltaTau PMAC) に送る．モータ制御装置は，接続されたパン・チルトステージの回転角，レンズのズーム，フォーカス，アイリスを制御する．使用したパン・チルトステージ (Helpmate, Bisight) の最高回転速度は，1,000 deg/s，最高回転加速度は，12,000 deg/s^2 である．

図 3.2: 高速パン・チルトステージ：単眼のみ使用

3.3.2 移動物体の検出

移動物体の検出は汎用画像処理装置で行なわれる．その概略図を**図 3.3** に示す．カメラからの信号はフルカラー画像入出力ボードで離散値化され，その輝度成分のみをフレームバッファに書き込む．このようにして取り込まれた画像に対して，移動物体の位置推定，速度推定の処理を行なう．以下にそれぞれの実現方法を述べる．

移動物体の位置の推定

3.2.2 節に述べた方法で，移動物体のエッジを検出する位置を推定する．**図 3.4** に実現した移動物体の位置推定部を示す．

フレーム間差分画像は 1 フレーム前のフレームバッファを用意しておき，現在のフレームと 1 フレーム前のフレームの差分を計算することで得られる．その際，同時に現在のフレームを 1 フレーム前のフレームバッファに書き込むことによって，次のフレームの入力に対応する．

エッジ画像は入力画像に対して，Sobel フィルタをかけることによって計算する．Sobel フィルタとして，**図 3.5**，**3.6** のような 3×3 のフィルタを用いた．この 2 つのフィルタを入

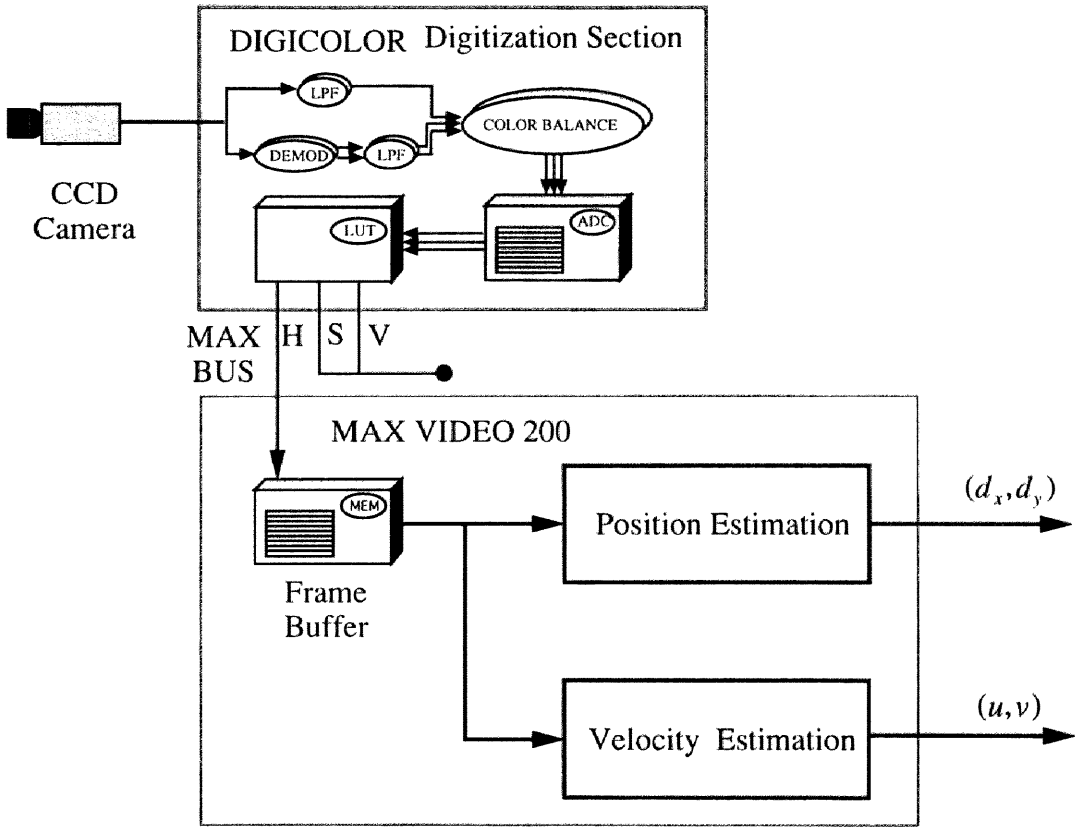


図 3.3: 移動物体の検出部の概略図

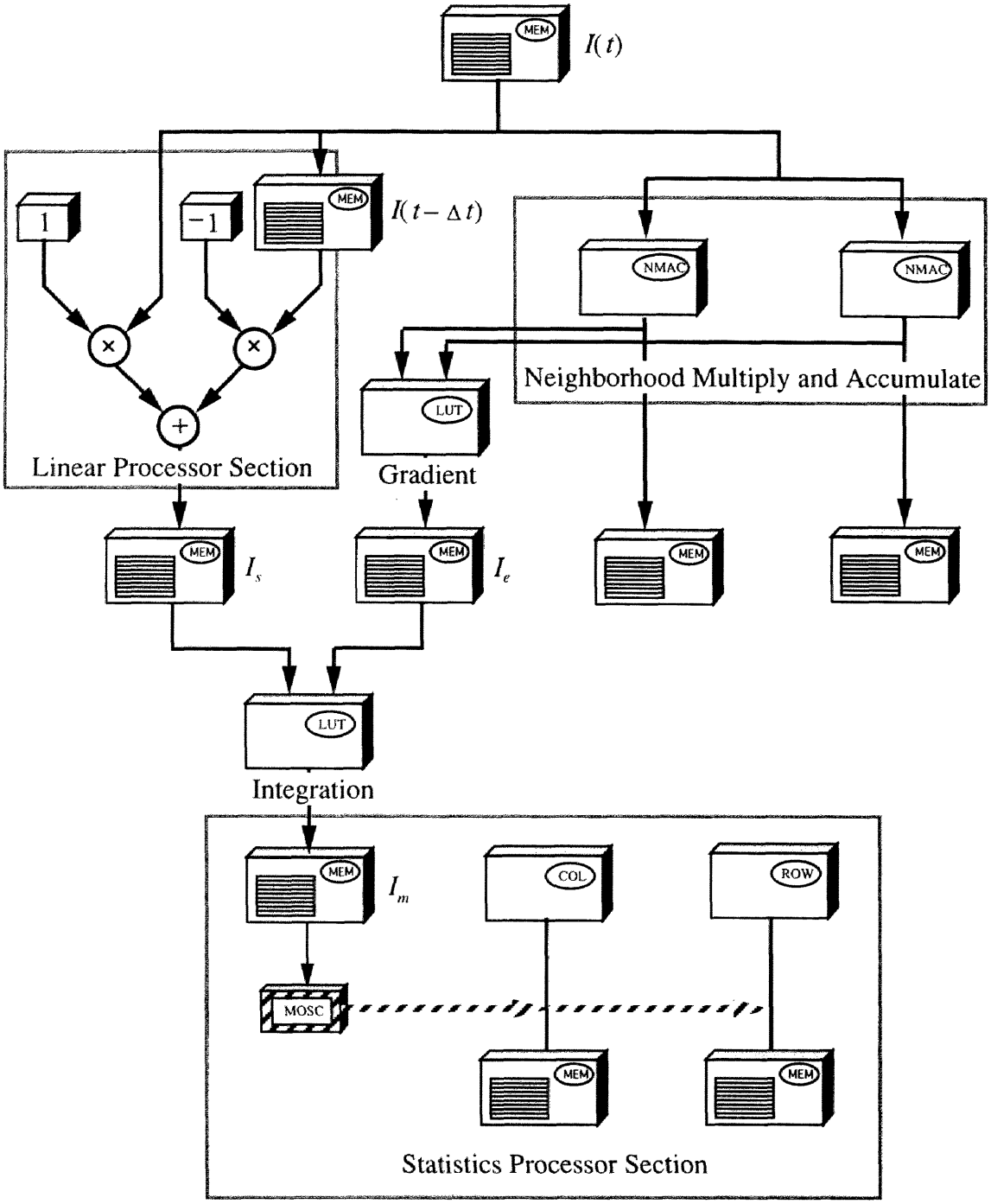


図 3.4: 移動物体の位置推定部の実現

1	0	-1
2	0	-2
1	0	-1

図 3.5: x 方向の Sobel オペレータ

1	2	1
0	0	0
-1	-2	-1

図 3.6: y 方向の Sobel オペレータ

力画像に対して並列にかけ，その結果の2乗和の平方根を計算する．2乗和の平方根の計算はルックアップテーブルにあらかじめすべての組合せに対する出力を記述しておき，ルックアップテーブルを引くことで，容易に計算される．また，x, y それぞれの方向のSobelフィルタをかけた画像は，速度の推定で使うので，これらを保存しておく．

得られたフレーム間差分画像とエッジ画像を統合することによって移動物体のエッジを検出する．この計算も2乗和の平方根の計算と同様に，式(3.7)をすべての入力に対して計算したルックアップテーブルを用意し，それを引くことで計算される．得られた移動物体のエッジから移動物体の位置を計算する．式(3.8)を書き直すと，以下の式を得る．

$$
\begin{aligned}
E &= \{(x,y)|I_i(x,y)=1\} \\
d_x &= \frac{\sum_{(x,y)\in E} x}{N(E)} \\
d_y &= \frac{\sum_{(x,y)\in E} y}{N(E)}
\end{aligned}
\tag{3.11}
$$

ここで，$N(E)$ は集合 E の要素数である．この計算は汎用画像処理装置内の演算ユニットの統計処理部で計算される．

本研究では，移動物体は一つしかないと仮定している．さらに，その移動物体のエッジ付近の速度は一定であると仮定する．このような条件の下で，式(3.10)から速度を推定する．この計算は，3.3.2節で得られた移動物体のエッジをマスクにして，x 方向の微分画像，y 方向の微分画像，フレーム間の差分画像を選択し，かけあわせて合計をとることによって計算される．これを実現すると**図3.7**のようになる．

画像座標系からカメラを中心とする極座標系への変換

図3.8のような射影系を想定すると，点 (x,y) に対応する光軸からの角度 $(\theta(t), \varphi(t))$ は，

$$
\begin{aligned}
\theta(t) &= \tan^{-1}\frac{\epsilon_x}{f}(x(t)-x_c) \\
\varphi(t) &= \tan^{-1}\frac{\epsilon_y}{f}(y(t)-y_c)
\end{aligned}
\tag{3.12}
$$

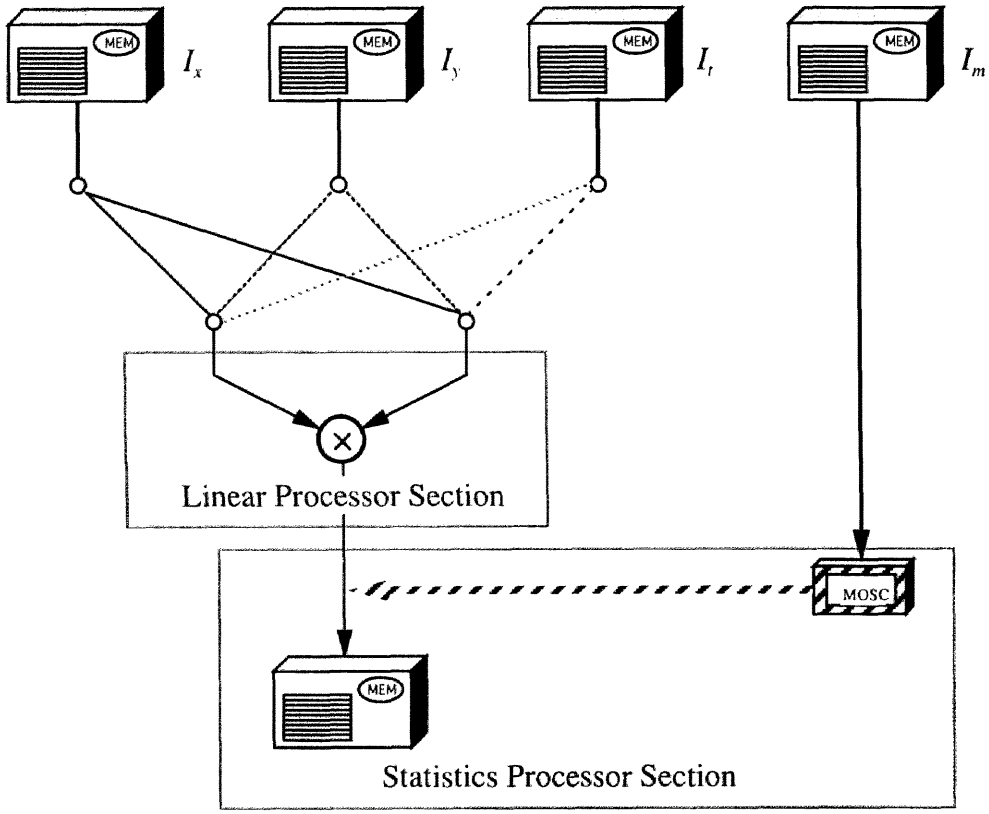


図 3.7: 移動物体の速度推定部の実現

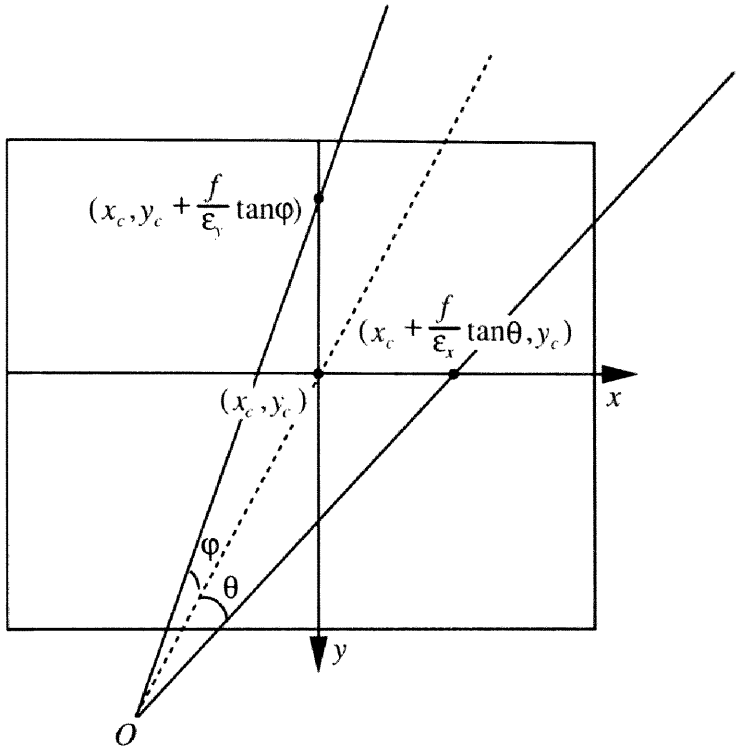


図 3.8: カメラのモデル

となる．ここで，(x_c, y_c) は光軸と画像面との交点の座標，ε_x と ε_y は画素の大きさ，f はカメラレンズの焦点距離である．式 (3.12) を t で微分すると，

$$\begin{aligned} \frac{d}{dt}\theta(t) &= \frac{\varepsilon_x}{f} u \cos^2\theta \\ \frac{d}{dt}\varphi(t) &= \frac{\varepsilon_y}{f} v \cos^2\varphi \end{aligned} \tag{3.13}$$

となる．したがって，画像座標系で点 (x, y) を速度 (u, v) で運動している物体のカメラ中心からの角度，角速度は，式 (3.12)，(3.13) のようになる．

3.3.3 移動物体位置の予測

線形予測

モータの回転中，移動物体の角速度 $(\frac{d}{dt}\theta(t), \frac{d}{dt}\varphi(t))$ は不変であると仮定する．説明を簡単にするため，θ成分のみで考える．すると，移動物体の位置 x は時間 t の 1 次関数，

$$x = \theta + \dot{\theta} t \tag{3.14}$$

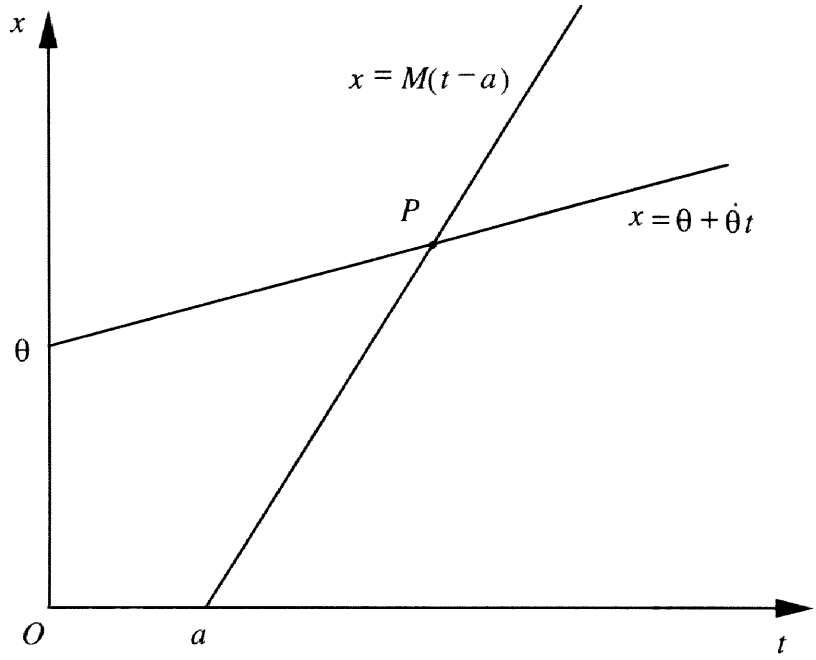


図 3.9: 移動物体の位置とカメラの回転角

と表すことができる．これに対し，モータの回転を以下のように t の 1 次式で近似する．

$$x = M(t-a) \ \ (t \geq a) \tag{3.15}$$

ここで， M はモーターの最高速度， a は画像処理やモータの加減速にともなう遅れを表す時間で，定数である．直線 (3.14)，(3.15) の位置関係を**図 3.9** に示す．移動物体をカメラの視野の中央にとらえるためには，図 3.9 の交点 P の位置にカメラを移動すればよい．したがって，モータの回転角度は式 (3.14)，(3.15) を解いて，

$$x = \begin{cases} \dfrac{M(\theta + a\dot{\theta})}{M - \dot{\theta}} & (\theta + a\dot{\theta} \geq 0) \\ \dfrac{M(\theta + a\dot{\theta})}{M + \dot{\theta}} & (\theta + a\dot{\theta} < 0) \end{cases} \tag{3.16}$$

となる． φ 成分についても同様である．

自己回帰モデルによる予測

物体の運動が規則的なとき，予測制御によって遅れ時間の少ない追跡ができる．ここでは，物体の移動パターンが未知の場合について考える．このとき，物体の移動パターンの

時系列を記憶し，モデルのパラメータを同定する必要がある．これを自己回帰モデルによって実現する．入力時系列を $\{x(t)\}$ とすると，予測値は以下の式によって計算される．

$$\hat{x}(t) = a_1 x(t-1) + a_2 x(t-2) + \cdots + a_m x(t-m) \tag{3.17}$$

$a_1, a_2, \ldots, a_m$は，予測位置$\hat{x}(t)$と，観測位置 $x(t)$ の差に基づいて更新される．

カルマンフィルタによる予測

物体の移動パターンがあらかじめわかっている場合，そのモデルをカルマンフィルタ [82] の状態方程式に表すことによって，予測制御が可能となる．例えば，移動パターンが正弦波である場合 $(x(t) = A\sin\omega t + x_c)$，以下の方程式で表現が可能である．

$$\begin{pmatrix} x(t) - x_c \\ \dot{x}(t) \end{pmatrix} = \begin{pmatrix} \cos\omega & \frac{\sin\omega}{\omega} \\ -\omega\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} x(t-1) - x_c \\ \dot{x}(t-1) \end{pmatrix} \tag{3.18}$$

ここで，x_c, ωは，別の方法で同定しなければならない．x_cは，時系列 $\{x(t)\}$ の最大値と最小値の中間に設定し，ωは，超過予測のときは値を小さくし，不足予測のときは値を大きくするという方法をとる．

3.4 実験とその結果

3.4.1 キャリブレーション

追跡を行なう前に，式 (3.12) の$\frac{\epsilon_x}{f}, \frac{\epsilon_y}{f}$を求める，すなわち，カメラの回転に対して画像がどう変化するかを求める実験を行なった．

方法

便宜上，２次元の画像平面を１次元にして考える．カメラの投影中心を原点に合わせ，光軸を y 軸に合わせる．**図 3.10** にその様子を示す．図 3.10 で，原点と画像平面との距離は f である．従って，このときある特徴点 P の画素の座標が x_0であったとすると，その投影点 A の座標は $(\epsilon_x x_0, f)^t$となる．カメラをθだけ回転させた場合，画像中心 B は

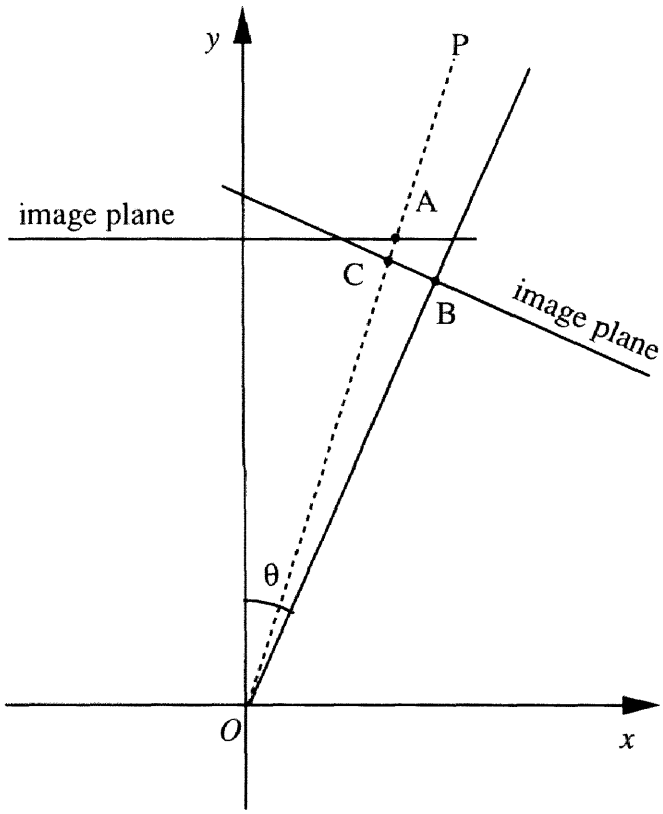


図 3.10: カメラの回転による画素値の変化

$(f\sin\theta, f\cos\theta)^t$に移る．このとき，ある特徴点$P$の画素の座標が$x_1$に移ったとする．すると，$\overrightarrow{BC} = (\varepsilon_x x_1\cos\theta, -\varepsilon_x x_1\sin\theta)^t$となり，点$C$の座標は$(f\sin\theta + \varepsilon_x x_1\cos\theta, f\cos\theta - \varepsilon_x x_1\sin\theta)^t$となる．点$A, C$はともに，原点を通る直線上にあるから，

$$t\begin{pmatrix} \varepsilon_x x_0 \\ f \end{pmatrix} = \begin{pmatrix} f\sin\theta + \varepsilon_x x_1\cos\theta \\ f\cos\theta - \varepsilon_x x_1\sin\theta \end{pmatrix} \tag{3.19}$$

を満たす．これを$\frac{\varepsilon_x}{f}$について解いて，

$$\frac{\varepsilon_x}{f} = \frac{-(x_0 - x_1) \pm \sqrt{(x_0 - x_1)^2 - 4\tan^2\theta x_0 x_1}}{2\tan\theta x_0 x_1} \tag{3.20}$$

を得る．複号はθの符号によって変化する．$\frac{\varepsilon_y}{f}$についても同様に計算することができる．

この方法は，参照点が1つでよいという利点がある．従って，参照点の検出や対応付けといった処理は必要なく，キャリブレーションを自動化できるという利点がある．

カメラを回転ステージにとりつけ，その20cmほど前にホワイトボードを置く．そのホワイトボード上に参照点となる黒点を1つ描く．システムは入力画像を2値化し，黒い部分の重心(x_0, y_0)を求める．そこから適当な角度(θ, φ)だけ移動し，同様に黒い部分の重心を(x_1, y_1)求める．$(x_0, y_0), (x_1, y_1), (\theta, \varphi)$から，式(3.20)により，$(\frac{\varepsilon_x}{f}, \frac{\varepsilon_y}{f})$を求める．

結果

1000サンプルについて$(\frac{\varepsilon_x}{f}, \frac{\varepsilon_y}{f})$を求め，その平均を求めた．**表3.1**にその一部を示す．その結果，$(\frac{\varepsilon_x}{f}, \frac{\varepsilon_y}{f}) = (0.0014, 0.0010)$となった．

3.4.2 追跡実験

追跡実験は屋内(蛍光灯照明下)で行った．移動物体として，人間とビーチボールを用い，追跡実験を行った．その際，追跡中のカメラから見たシーンをビデオテープレコーダに録画し，それを評価した．また，環境の変化に対する頑健性を調べるため，先の追跡の途中で照明を切った場合の応答性を調べた．さらに，比較のため，位置情報のみを使った場合と，速度情報のみを使った場合について，人間の往復運動を追跡し，そのときのカメラの回転角を記録した．

表 3.1: カメラのキャリブレーション結果

サンプル	x_0	x_1	θ	$\frac{\varepsilon_x}{f}$
1	418.7	369.4	4.00	0.001357
2	368.0	319.3	4.00	0.001407
3	319.3	245.2	6.00	0.001416
4	245.0	180.1	5.00	0.001347
5	180.1	269.0	-7.00	0.001380
6	268.9	217.4	4.00	0.001358
7	217.3	306.3	-7.00	0.001380
8	306.3	257.8	4.00	0.001438
⋮	⋮	⋮	⋮	⋮
1000	181.3	255.0	-6.00	0.001429
サンプル	y_0	y_1	φ	$\frac{\varepsilon_y}{f}$
1	104.3	163.4	-3.00	0.000882
2	163.6	222.0	-3.00	0.000897
3	221.9	280.6	-3.00	0.000894
4	280.9	139.6	8.00	0.001000
5	139.3	243.0	-6.00	0.001014
6	243.0	105.1	8.00	0.001022
7	105.1	222.9	-7.00	0.001042
8	223.0	314.8	-5.00	0.000953
⋮	⋮	⋮	⋮	⋮
1000	228.3	138.7	5.00	0.000977

参考のため複数の移動物体(二人の人間)が移動するシーンでもシステムの応答を調べた．環境は先の実験と同じである．

結果

図 3.11 に人間を追跡した結果を，**図 3.12** にビーチボールを追跡した結果を示す．表示したフレームの時間間隔は，図 3.11 が 1s で，図 3.12 が 0.25s である．また，**図 3.13** に画像処理の結果を示す．図 3.13 では，左上が入力画像 I，右上が差分画像 I_s，左下がエッジ画像 I_e，右下が統合画像 I_mである。

図 3.14，3.15 に照明を変化させた場合の追跡結果を示す．図 3.14 では 10 の時点で，図 3.15 では 8 の時点で照明を暗くしている．表示したフレームの時間間隔は，先の実験と同様に，図 3.14 が 1s で，図 3.15 が 0.25s である．

図 3.16，3.17，3.18 はそれぞれ，位置と速度情報を使った追跡結果，位置情報のみを使った追跡結果，速度情報のみを使った追跡結果である．図中，縦軸は角度で横軸はサンプル時刻である．サンプル時間間隔は約 300ms である．

図 3.19 に 2 つの移動物体が存在する環境での追跡結果を示す．フレーム間間隔は 1s である．

3.4.3 規則的な運動の追跡実験

線形予測では移動物体は等角速度運動していると仮定している．したがって，移動物体の角加速度が大きくなると，追跡に遅れが生じたり，物体を見失ったりすることが生じる．そこで，予測制御を用いて遅れのない追跡を行なう実験を行なった．移動物体は，小型の電動模型自動車を用い，軌道レールに沿って走らせることによって，運動を規則的にした．自動車の速度は約 10km/h で，軌道レールから約 3m 離れた地点から観測すると，その運動は，振幅 28deg，周期 1.4s の正弦波に近い．**図 3.20，3.21** に，それぞれ，自己回帰モデルによる予測制御，カルマンフィルタによる予測制御に基づく追跡結果を示す．図中，黒

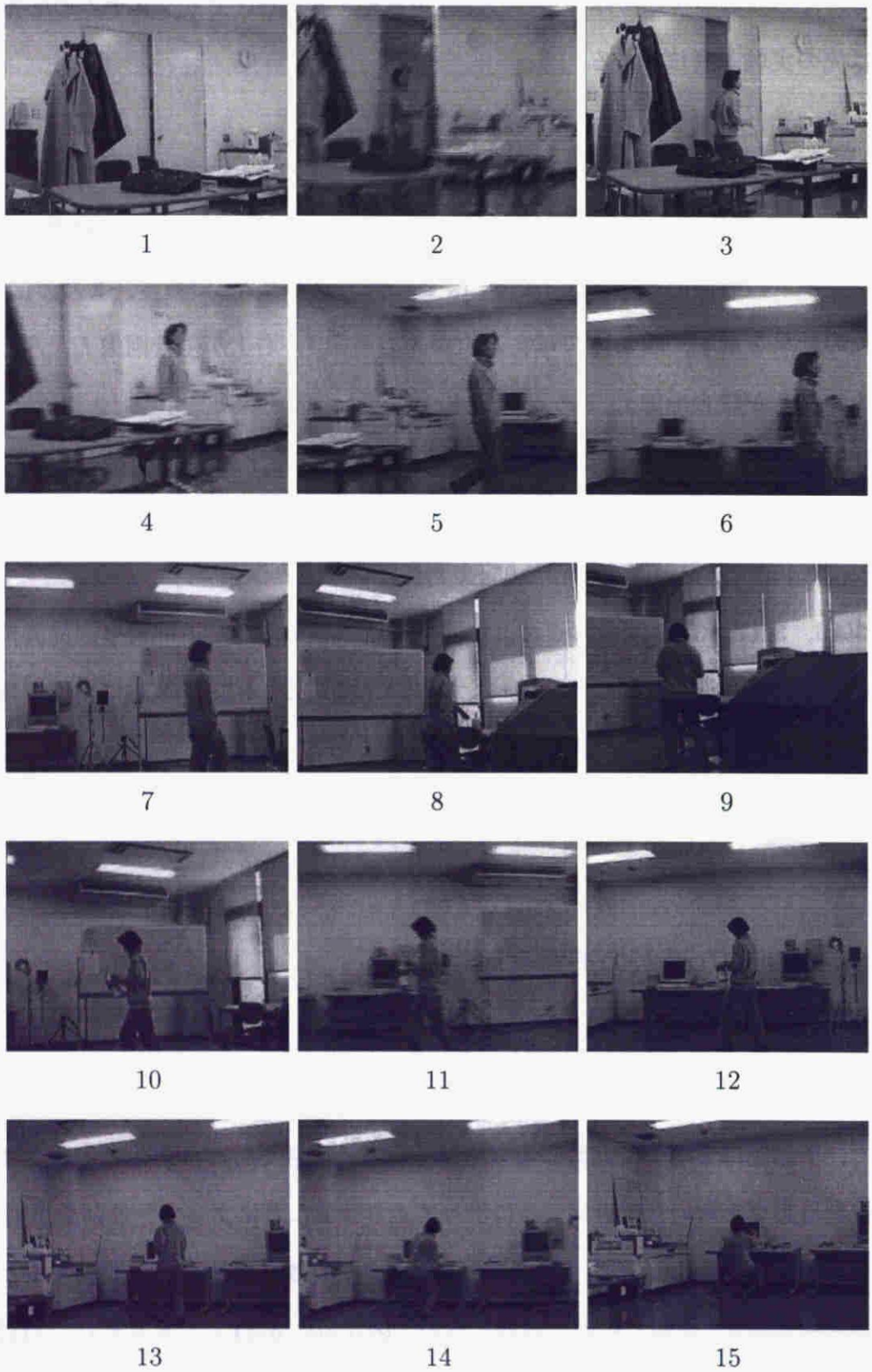


図 3.11: 人間の追跡結果（1秒間隔）

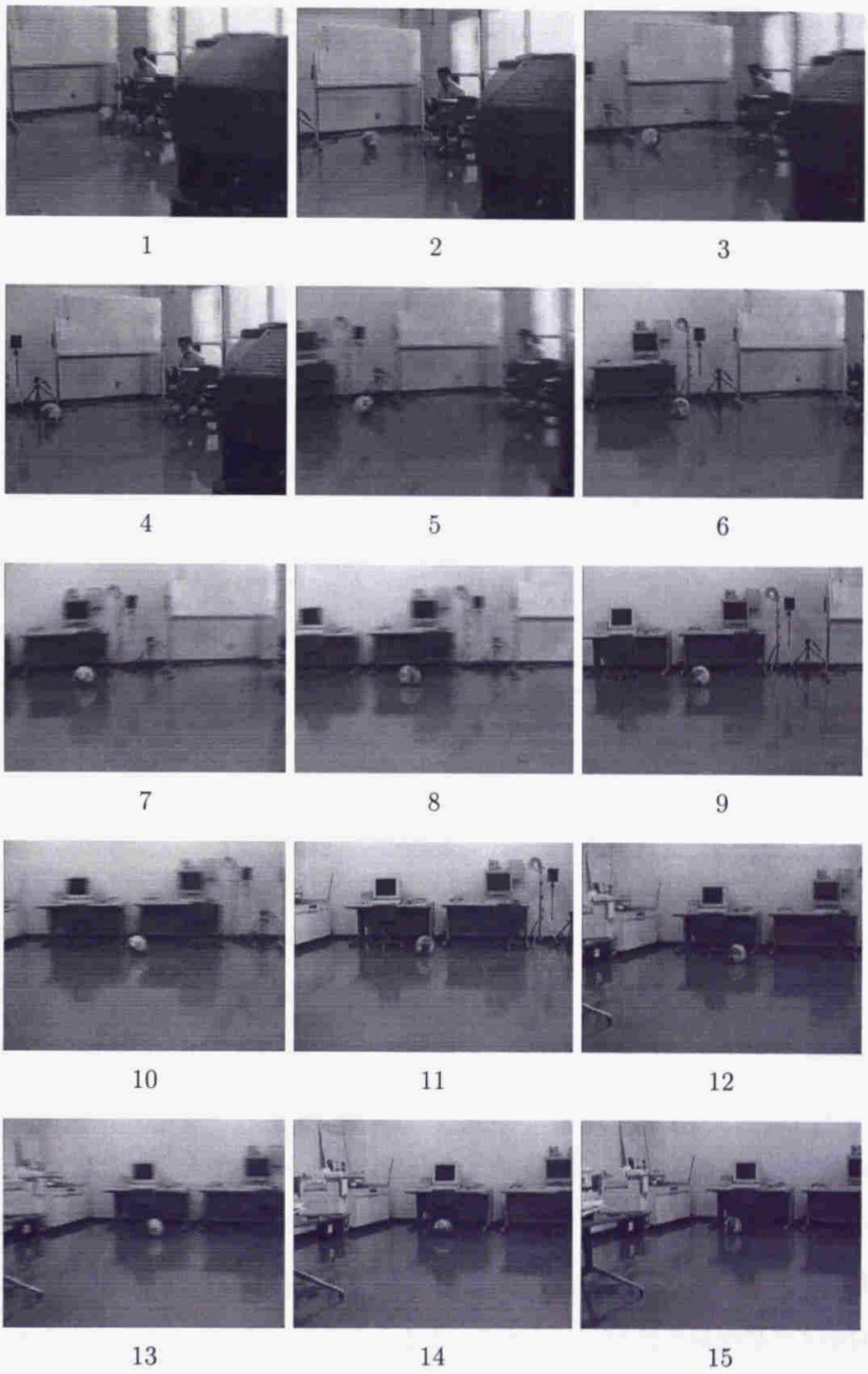


図 3.12: ビーチボールの追跡結果（0.25 秒間隔）

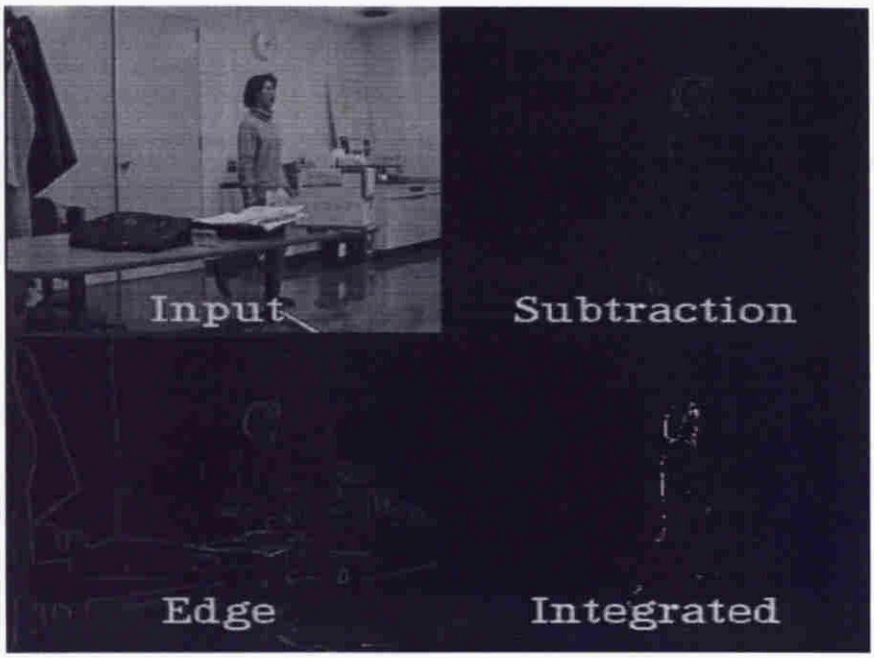


図 3.13: 画像処理の結果

線は，移動物体の位置を表し，灰線は，カメラの位置を表す．比較のため，予測制御なしで追跡した結果を，**図 3.22** に示す．

3.5 考察

図 3.11，3.12 から，移動物体の追跡に成功していることがわかる．しかし，図 3.11 の 5，6，7，図 3.12 の 3，4，5，6，7 については追跡に遅れが生じている．これは，速度の推定誤差が大きいためであると考えられる．経験的に，移動物体の速度が速い場合，推定値は実際の速度よりも小さい値になることが多い．速度の推定の誤差が大きい点に関しては，図 3.18 からも明らかになる．図 3.18 は移動物体の速度情報のみを使って追跡したものである．速度の推定の誤差が小さければ，移動物体の位置とカメラの位置は最初の移動物体とカメラの位置のずれを保って移動する．従って，誤差は一定になることが期待される．しかし，結果は誤差の標準偏差が 2.9 deg となり，かなりの変化が見られた．このように，速度推定にはかなりの誤差が含まれていると考えられる．

1 2 3

4 5 6

7 8 9

10 * 11 12

13 14 15

図 3.14: 人間の追跡結果:照明変化 (*) あり

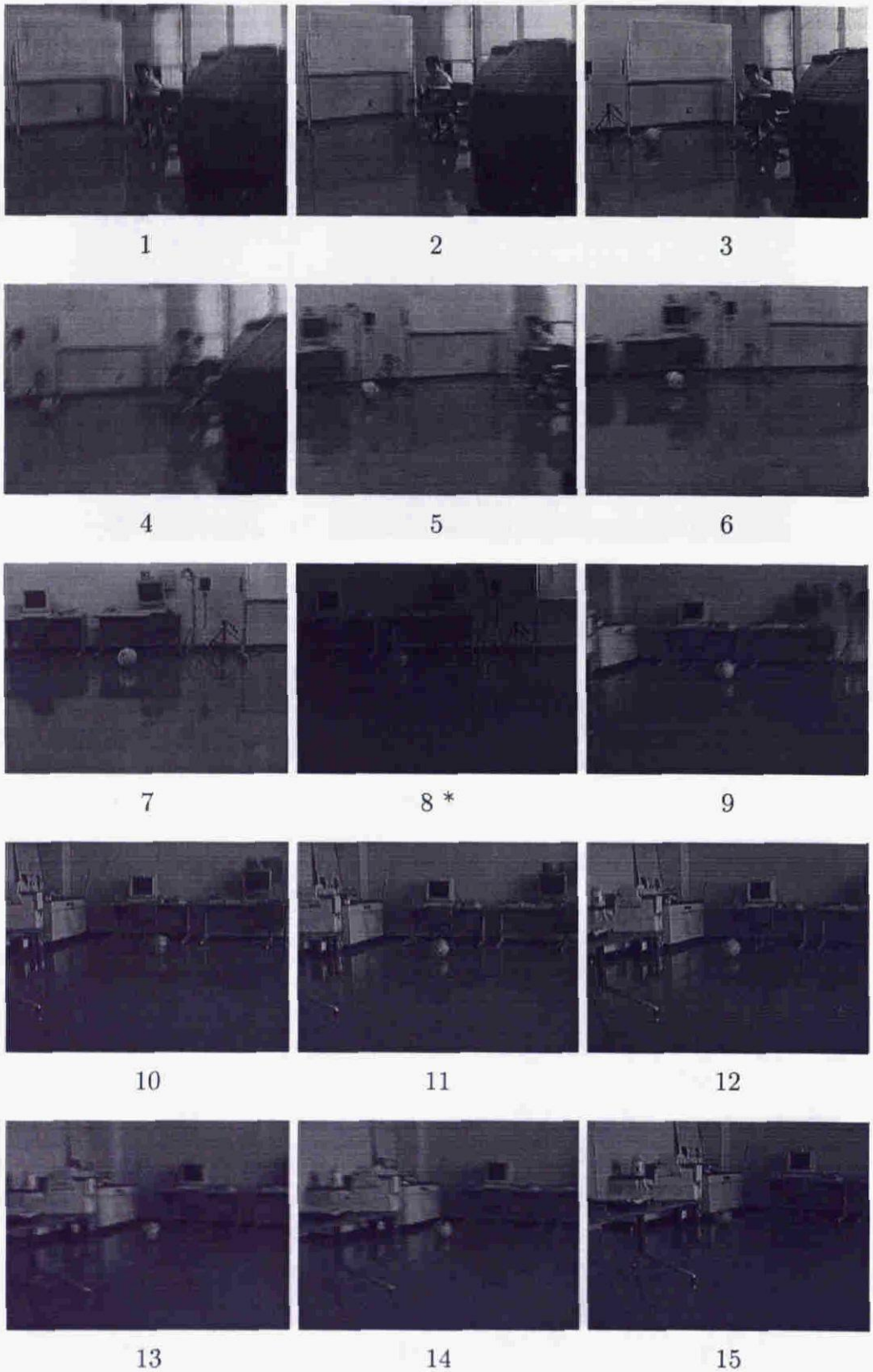


図 3.15: ビーチボールの追跡結果:照明変化 (*) あり

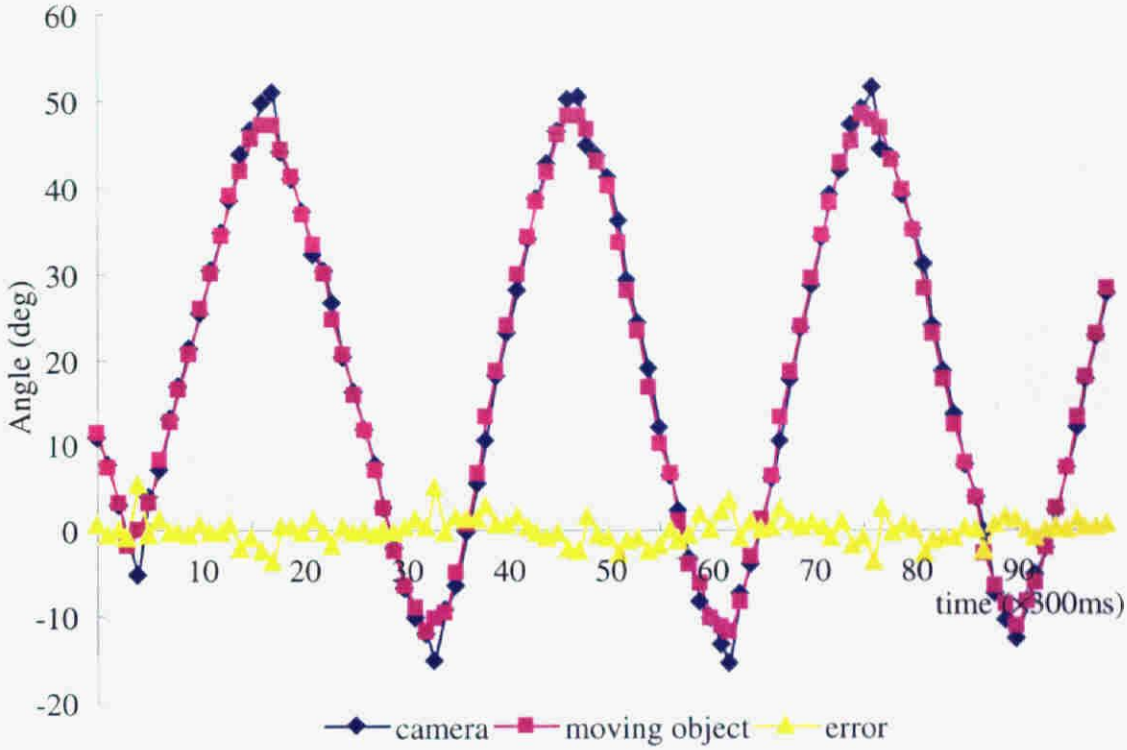


図 3.16: 移動物体の位置と速度を使った追跡結果

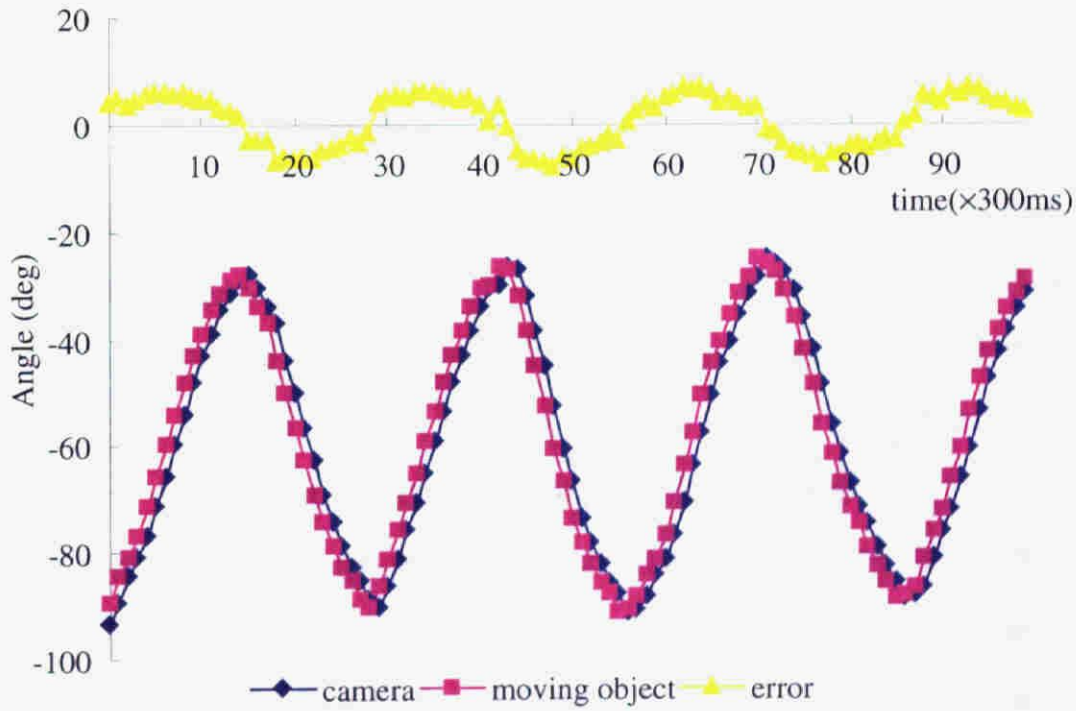


図 3.17: 移動物体の位置を使った追跡結果

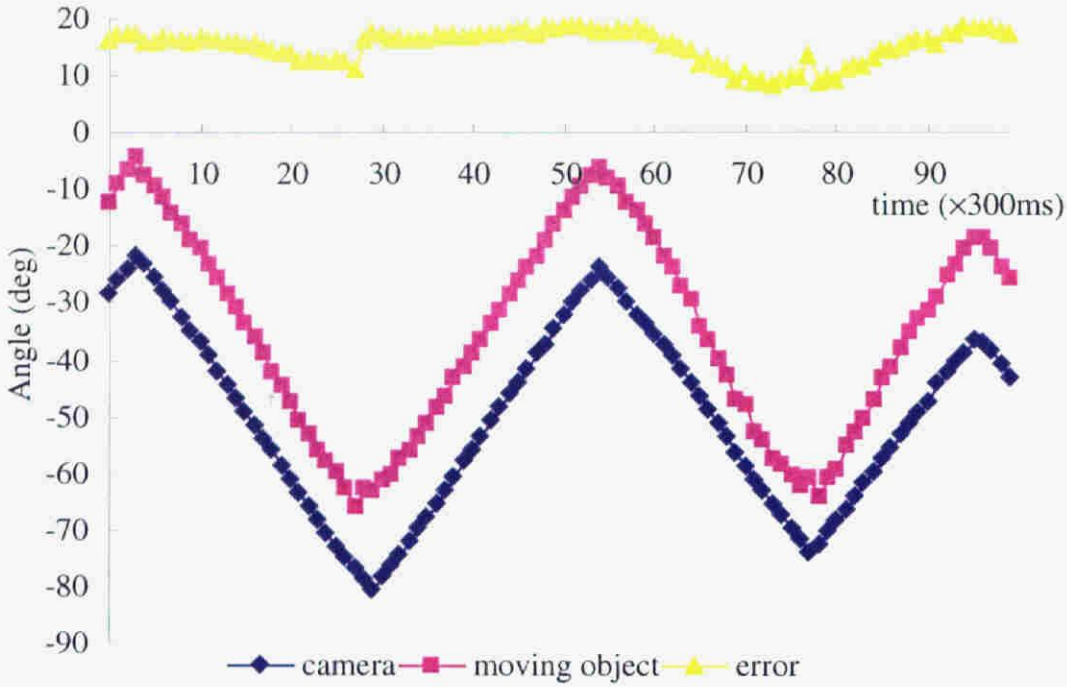


図 3.18: 移動物体の速度を使った追跡結果

図 3.13 は画像処理結果を示したものである。右下の統合画像を見ると移動物体のエッジ部分のみが抽出されている様子が分かる。移動物体のエッジは環境の変化が大きい状況でも容易に抽出されるため，図 3.14，3.15 に示すとおり，照明の変化に対してロバストな追跡を行うことが可能である．

図 3.16 と図 3.17 を比較すると，位置と速度情報を用いた方が，位置情報のみの場合よりも遅れの少ない追跡が可能であることがわかる．これは，位置と速度情報を用いた場合の誤差の標準偏差が 1.5 deg，位置情報のみの場合の誤差の標準偏差が 4.8 deg であることからも明らかである．

また，本システムは視野内に複数の物体が存在する場合は，図 3.19 を見て分かるように，複数の物体の重心を追跡する．複数の物体から 1 つを選択し，それを追跡することはできない．複数の移動物体から一つを選択するためには，移動物体の認識が必要となる．しかし，移動物体の認識処理は，本システムのようなパイプライン型の画像処理には不向きである．したがって，別の画像処理装置，認識アルゴリズムの開発が必要となる．

図 3.19: 二人の人間の追跡結果

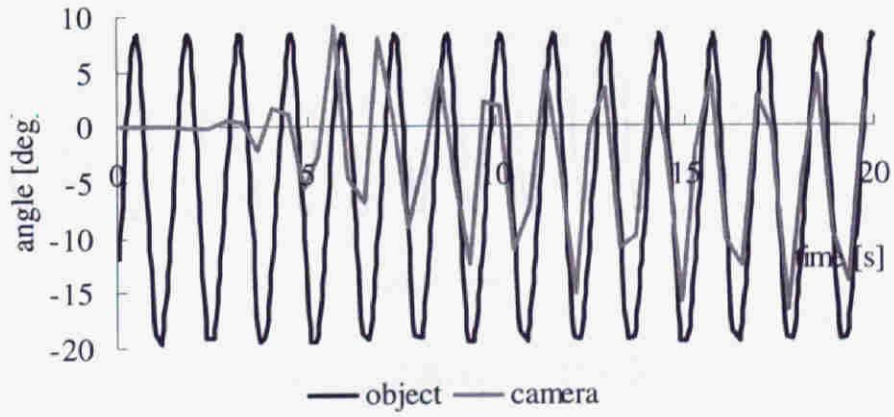


図 3.20: 自己回帰モデルによる予測を用いた追跡結果

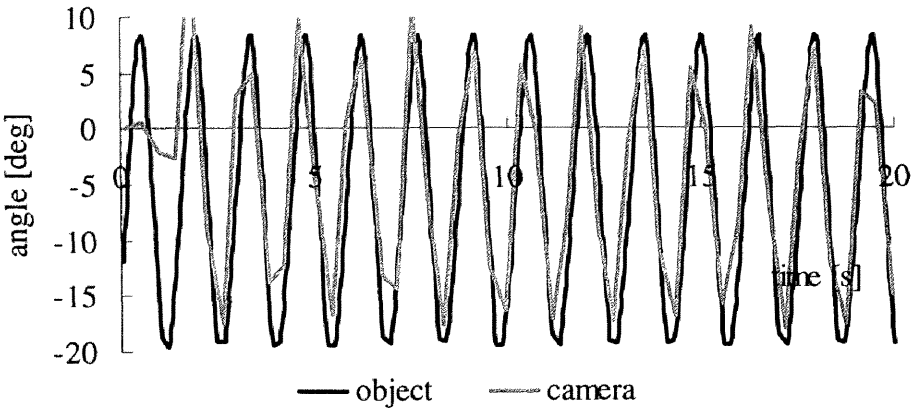


図 3.21: カルマンフィルタによる予測を用いた追跡結果

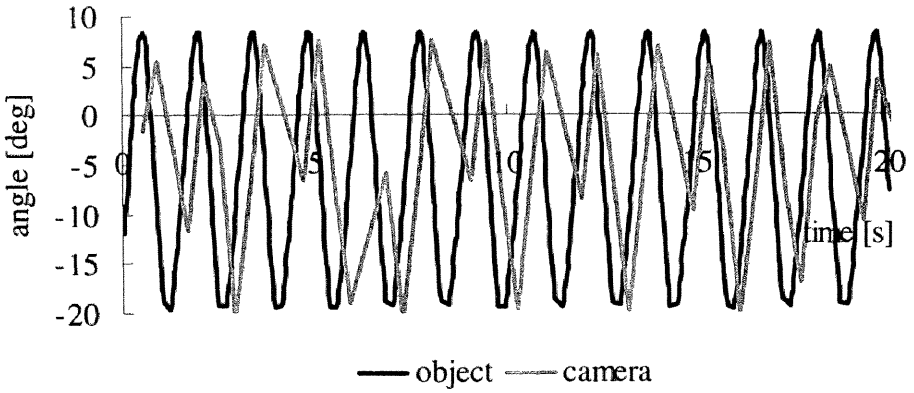


図 3.22: 予測制御なしでの追跡結果

表 3.2: 他のシステムとの比較

	the number of cameras	the degree of freedom	process time / latency	trackable speed	control rate
Hwang	1	3	— / —	7 deg/s	0.63 Hz
Uhlin	2	—	— / —	—	25 Hz
Ferrier	2	3	— / 100 ms	—	10 Hz
D.W.Murray	1	2	50 / 110 ms	18 deg/s	25 Hz
Coombs	2	3	92 / 150 ms	—	7.5 Hz
Kunieda	2	3	36 / 90 ms	—	30 Hz
Du	2	4	— / 90 ms	—	12.5 Hz
Daniilidis	1	2	40 / 80 ms	10 deg/s	25 Hz
Nakabo	1	2	— / —	—	1000 Hz
this work	1	2	33 / 67 ms	200 deg/s	2.5 Hz

— は不明を意味する.

規則的な運動をする移動物体は，予測制御を行なうことで遅れのない追跡が可能である．図 3.20 を見ると，自己回帰モデルによる予測では，移動パターンのモデルの同定に，時間がかかり，また，推定されたモデルの振幅は，実際のものよりも小さい．これは，移動物体が高速に運動しているため，正弦波の山の部分でのサンプリングが行なわれにくく，誤ったパターンを学習しているためだと考えられる．一方，カルマンフィルタでは（図 3.21），モデルの学習は早く，振幅の値も実際のものとほぼ等しく同定されている．線形予測の場合では（図 3.22），追跡に遅れが生じており，移動物体をよく見失っている．運動の規則性は，予測誤差を評価することによって判断が可能である．

3.6 まとめ

本章では移動物体を追跡するシステムについて述べた．人間の眼球運動における知見から，移動する指標を追跡する場合，位置情報を入力とした駆動系 (saccade) と速度情報を入力とした駆動系 (pursuit) の 2 種類がある．これを利用してシステムの制御を位置情報を扱う部分と，速度情報を扱う部分の 2 つに分ける手法を提案した．さらに，入力した画像を

時空間の微分値という特徴に分解し，それを統合することによって移動物体の位置，速度を推定する方法について述べた．

開発したシステムは，高速に回転することのできるパン・チルトステージと，汎用画像処理装置による画像処理の高速化によって，実時間での追跡が可能となった．移動物体の位置と速度を推定し，それらを統合してカメラの回転角を決定するという追跡アルゴリズムを実現した．システムを性能の面から評価すると，システムは実環境を移動する1つの不特定の物体を追跡することができる．実験から200deg/sの速度までの物体を追跡することができることを示した．また，移動物体が1フレームの時間(33ms)に1ピクセル以上移動するとその物体を検出することができることから，移動物体の検出可能な最低速度は，2.4deg/sである．

実験から，移動物体の追跡に関してはほぼ良好な結果が得られた．線形予測の場合，速度の推定に誤差が大きく，特に移動物体の速度が大きい場合は追跡に遅れが生じやすいことが明らかとなった．しかし，移動物体の運動が規則的な時，自己回帰モデルやカルマンフィルタを用いることによって，遅れのない追跡が可能であることを示した．

表3.2に，従来のシステムとの比較を示す．本システムは，処理時間がもっとも短く，追跡可能な速度の最大値ももっとも大きいことが分かる．これは，予測制御による移動物体の位置の正確な予測と高速なアクティブカメラを用いたことによる．

システムは，実環境を移動する1つの不特定の物体を追跡することができる．しかし，視野内を移動する複数の物体から主要な物体を選択し，それを追跡することはできない．このように複数の物体から1つを選択することを，人間は容易に行っている．この機能を実現するためには，本研究で開発したシステムが用いた時空間の微分値の特徴だけでなく，色，視差などの特徴を統合し，視野に含まれる物体を認識する必要がある．このような移動物体の認識を含めた追跡システムの実現が今後の課題である．

第4章　情報量によるシーンの定量化

4.1　まえがき

人間は網膜に写し出された2次元の網膜像から，色，形，動き，視差，テクスチャといった様々な手掛りをもとに，3次元の環境を理解することができる．また，人間が一度の観察で得られるシーンの範囲には限りがあり，注視位置の移動と観察を繰り返すことにより，シーン全体を把握している．このとき，注視点の移動は無秩序に行われているのではなく，興味のある対象に向けられると言われている．また，被験者に与えられた観察のタスクによって，注視点の分布が変化することも知られている [4].

このような人間の視覚の仕組みに基づき，画像全体を決められた順にスキャンするのではなく，興味のある部分を選択し，処理するアクティブビジョンの研究がさかんに行われている．アクティブビジョンにとって，解決しなければならない重要な問題点の一つは，どのようにして興味のある部分を選択し，その部分に注視点を移動するかということである [15]．注視点の移動方法として，データ駆動型ボトムアップ方式とモデル駆動型トップダウン方式の2種類がある [83]．ここでは，色，線の方向といった低次の特徴から，saliency map，すなわち，各特徴の注目点を統合したマップを構成するデータ駆動型ボトムアップ方式について考える．

データ駆動型ボトムアップ方式では，シーンの顕著な部分を選択するため，さまざまな特徴から saliency map を構成する必要がある．視覚的注意や視覚探索の分野では，saliency map の研究が盛んに行なわれ，特徴統合理論 [26] やそれを修正したモデルがいくつか提案されている [21, 28]．このモデルをアクティブビジョンシステムに実現する場合，カメラからの入力画像に対して，いくつかの特徴を抽出し，その特徴量をもとに saliency map を計

算する方法が取られている．しかし，特徴を抽出するとき，ヒューリスティックに特徴量を計算していた．例えば，画像の目立つ部分として，赤や黄色の部分や，高い空間周波数をもつ部分を抽出していた．

そこで，本論文では，特徴に理論的な裏付けを与えるため，情報理論，特にシャノンの情報量を用い，特徴量を統一的に扱い，注視点を選択する手法を提案する．この手法によって，特徴を情報量によって定量化することができる．

情報理論に基づき，画像を解析する研究は，これまでにもいくつかの研究が行われている．Jägersandは，解像度に対しての情報量を定義した[84]．画像の解像度に関するモデルをもとに，解像度を変化させたときのKullbackの情報量の最大値を見つけることによって，その画像に対する最適な解像度を知ることができる．また，画像の部分領域や画素ごとにこの情報量を計算することにより，画像のsaliencyとその部分のscaleを同時に知ることができる．ViolaとWells IIIは，画像中の物体の姿勢を推定する，情報理論に基づくアプローチを提案した[85, 86]．これは，画像とモデルの対応を表す式を，モデル画像と対象画像の相互情報量の極大点を与えるように推定する方法である．これは，物体の表面の特性や照明条件などと独立しているという利点がある．これらの研究は相互情報量を最大にするというアプローチをとっているが，ボトムアップ方式の本研究では，シーンを情報源とみなし，観察することによって得られる情報を最大にすることを目的としている．

本研究で扱う特徴量には，画像から得られるもっともプリミティブな情報である明度，彩度，色相を用いる．シーンからそれぞれの特徴に対して情報量を計算し，その性質について考察する．また，特徴間のオーバーラップとして存在する相互情報量について考察する．以下，4.2で，情報理論の中から本研究に関連深い，情報量，相互情報量，Kullback情報量について簡単に述べる．4.3で，シーンの情報量を情報理論をもとに定量化する．さまざまなシーンで，定量化した情報量を計算した実験結果について4.4に述べ，4.5でその結果について考察する．最後に，4.6で提案したシーンの情報量についてまとめる．

4.2 情報量

4.2.1 情報量の定義

シャノンは，離散確率変数 X に対し，その情報量を以下のように定義した．

$$H(X) = -\sum_{x \in A} p(x) \log p(x) \tag{4.1}$$

ここで，A は X のとる値の集合，$p(x)$ は X の確率をあらわす．対数の底は，主に 2 が使われ，底が 2 のとき，単位はビット (bit) となる．なお，$E_X[\cdot]$ を X に関する期待値とすれば，情報量は次のようにも書ける．

$$H(X) = E_X[-\log p(X)] \tag{4.2}$$

この情報量に対して，一般に，不等式

$$0 \leq H(X) \leq \log |A| \tag{4.3}$$

が成立し，X が 1 つの値しかとらないときにゼロ，X が等確率で A のすべての値をとるときに最大となる．ただし，$|A|$ は A が含む要素の数を表す．同様にして，離散確率変数の組 $X^N = X_1 \cdots X_N$ の情報量は，

$$H(X^N) = E_X[-\log p(X^N)] \tag{4.4}$$

で定義される．$p(X^N)$ は X^N の確率分布である．このとき，確率変数の組 (X, Y) に対して，次の不等式が成立する．

$$H(Y) \leq H(X, Y) \leq H(X) + H(Y) \tag{4.5}$$

ここで，$H(X, Y)$ は，X が Y の関数であるときに $H(Y)$ と等しく，X と Y が独立の時，おのおのの情報量の和に等しい．式 (4.5) により，$H(X) \leq H(X, Y)$ であるから，X が Y の関数であるときには，

$$H(X) \leq H(Y) \tag{4.6}$$

が成り立つ.

離散確率変数 X と任意の確率変数 Y の組に対して，X の Y に関する条件付きエントロピーは，次式で定義される.

$$H(X|Y) = E_{(X,Y)}[\log p(X|Y)] \tag{4.7}$$

ただし，$p(x|y)$ は $Y = y$ であったときの X の条件付き確率分布である．もし，X と Y が，ともに離散確率変数であるなら，

$$H(X,Y) = H(X|Y) + H(Y) \tag{4.8}$$

なる恒等式が得られる．したがって，式 (4.5) と式 (4.8) より，不等式

$$0 \leq H(X|Y) \leq H(X) \tag{4.9}$$

が導かれる．ここで，$H(X|Y)$ は X が Y の関数であるときゼロ，X と Y が，独立の時に $H(X)$ と一致する.

一般に，離散確率変数の組 X^N に対して，式 (4.8) より次式が導かれる.

$$H(X^N) = \sum_{n=2}^{N} H(X_n|X^{n-1}) + H(X_1) \tag{4.10}$$

ここで，もし X^N が定常的な確率過程 $X = X_1X_2\cdots$ の部分列であれば，$H(X_n|X^{n-1})$ は，n について減少列であり，極限

$$\bar{H}(X) = \lim_{N\to\infty} \frac{1}{N} H(X^N) = \lim_{N\to\infty} H(X_N|X^{N-1}) \tag{4.11}$$

が存在する．この $\bar{H}(X)$ を X の（平均）エントロピーとよぶ．X が m 次の定常マルコフ過程であれば $\bar{H}(X)$ は

$$\bar{H}(X) = H(X_{m+1}|X^m) \tag{4.12}$$

となり，特に X が互いに独立なとき（0次のマルコフ過程）次式が成立する.

$$\bar{H}(X) = H(X_1) \tag{4.13}$$

4.2.2 相互情報量

離散確率変数の組 (X,Y) に対し，その相互情報量は次式で定義される．

$$I(X,Y) = -\sum W(X,Y)\log\frac{W(X,Y)}{P(X)Q(Y)} \tag{4.14}$$

ここで，$P(X),Q(Y),W(X,Y)$ はそれぞれ $X,Y,(X,Y)$ の確率分布である．単位は，2 を底とすれば，情報量と同じくビットとなる．

相互情報量は，以下の性質を持つ．

$$I(X,Y) = I(Y,X) \tag{4.15}$$

$$I(X,Y) = H(X)+H(Y)-H(X,Y) \tag{4.16}$$

$$I(X,Y) = H(Y)-H(Y|X) \tag{4.17}$$

さらに，式 (4.9),(4.17) より，不等式

$$0 \le I(X,Y) \le H(Y) \tag{4.18}$$

を得る．ここで，$I(X,Y)$ は X と Y が独立なときにゼロ，Y が X の関数であるときに $H(Y)$ に等しい．

4.2.3 Kullback 情報量

Kullback は，2 つの確率分布 $P(X),Q(Y)$ がどのくらい異なっているかをあらわす尺度として，以下の量を定義した．

$$K(X,Y) = \sum_{X,Y} P(X)\log\frac{P(X)}{Q(Y)} \tag{4.19}$$

これは，Kullback の情報量と呼ばれ，2 つの分布が等しいときにゼロになる．

4.3 シーンの情報量

4.3.1 シーン情報量の定義

シーンを情報源と見なし,その情報量について考える．シーンを撮影したとき，$I(x,y)$という画像が得られる確率は，$p(I(x_1,y_1),I(x_2,y_2),\ldots,I(x_n,y_n))$となる（$n$は画素数）．しかし，この確率を計算するのは困難であるため，画素は相互に独立であり，各々の画素での確率分布はすべて等しいと仮定する．すると，この確率は，$\Pi_{i=1}^{n}p(I(x_i,y_i))$と表すことができ，これを式(4.1)に代入すると，その画像全体の情報量は，

$$H=-n\sum_{I}p(I(x_i,y_i))\log p(I(x_i,y_i)) \tag{4.20}$$

となる．したがって，1画素あたりの情報量は，

$$H=-\sum_{I}p(I(x_i,y_i))\log p(I(x_i,y_i)) \tag{4.21}$$

である．ここで，確率$p(I(x,y))$は，シーンを観測した画像のヒストグラムを正規化して以下のように推定する．

$$p(k)=\frac{Hist(k)}{n} \tag{4.22}$$

ここで，$Hist(k)$は，画素$I(x,y)$がkとなる度数である．

4.3.2 複数の特徴からの情報量

視覚探索は被験者に与えたタスクに大きく依存している．従来のアクティブビジョンでは，複数の特徴を足し合わせてsaliency mapを生成する際の重み係数を変更することによって，タスク依存型のシステムを実現していた．しかし，その重み係数は，ヒューリスティックな方法で決定していた．

そこで本研究では，複数の特徴の同時確率を用いることによってこの問題を解決する．例えば，2つの特徴X,Yから情報量マップを計算する場合，まず，2つの特徴の同時確率$p(X,Y)$を，X,Yの2次元ヒストグラムを正規化することによって推定する．すると，式

(4.21) によって情報量を計算することが可能である．情報量マップは，与えられたタスクにしたがって扱う特徴を決定するだけで，理論的に計算することが可能であり，扱う特徴の決定方法以外には，ヒューリスティックな手段をとる必要はない．

4.3.3 特徴間の相互情報量

複数の特徴を統合した情報量の計算は，単純にそれらの特徴の情報量を足し合わせればよいわけではない．実際，式 (4.17) より，

$$H(X,Y) = H(X) + H(Y) - I(X,Y) \tag{4.23}$$

となる．ここで，$I(X,Y)$ は，X,Y の相互情報量で式 (4.14) によって定義される．これは，X,Y が独立の時に 0 になる．確率分布として，明度，彩度，色相を考えたとき，一般のシーンにおいて，それらが独立となることはほとんどなく，それぞれの特徴間の確率的な従属性としてある値をもつと考えられる．このとき，その値は，それぞれの特徴間の相関の強さを表している．たとえば，グラデーションのように，明度，彩度，色相がなだらかに変化するような部分では，大きな相互情報量を持つと考えられる．

4.4 実験とその結果

シーンの情報量を様々な画像に対して計算する実験を行なった．シーンの特徴として，明度を用いた．このとき，明度の情報量は，シーンを撮影した画像から明度情報 $I(x,y)$ を抽出し，式 (4.21) より，計算することが可能である．明度の情報量を計算した例を**図 4.1** に示す．図中，上の 2 枚は合成画像で，下の 2 枚は実画像である．また，左から順に，入力画像，明度ヒストグラム，計算した情報量である．

図 4.2 に，屋外のシーンについて，明度，彩度，色相を組み合わせた情報量と，その相互情報量を計算した結果を示す．図中，Inf. は，3 つの特徴の情報量 $H(X,Y,Z)$ を，M.I. は，相互情報量 $I(X,Y,Z)$ を示す．

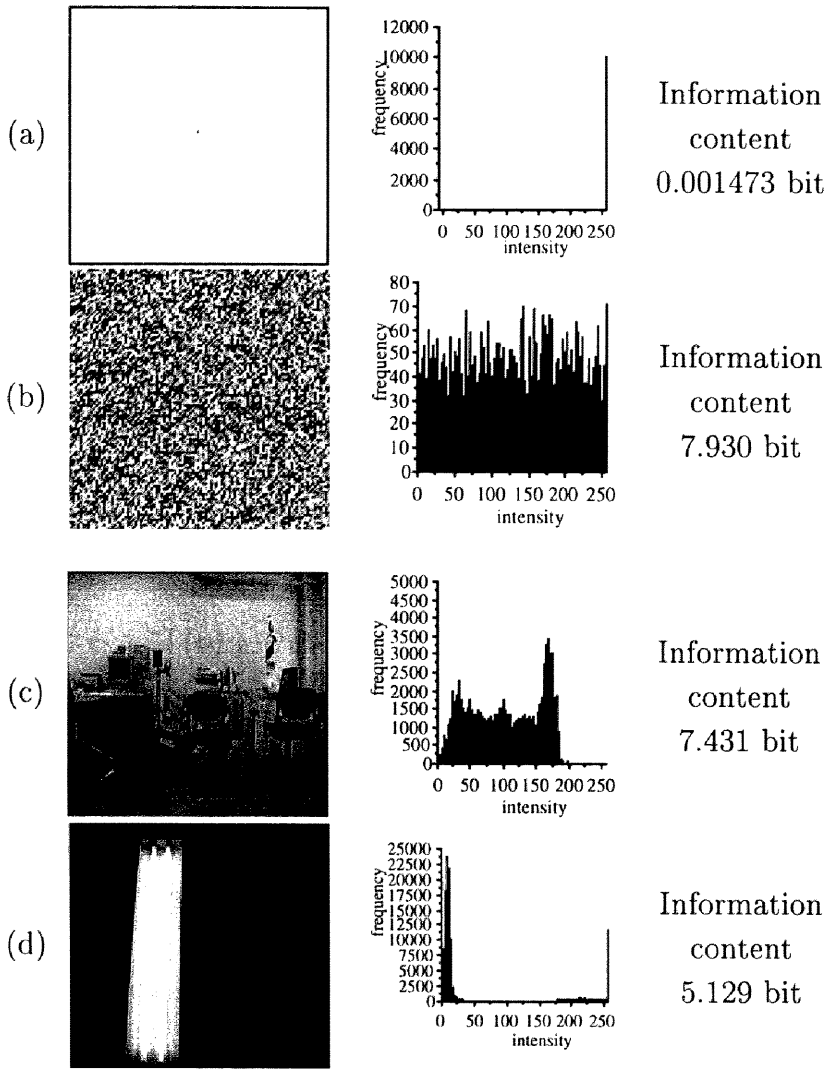


図 4.1: 明度の情報量をさまざまな画像に適用した例

4.5 考察

図 4.1 において，(a) の画像は，白い背景の中央に 1 点のみ黒点が存在する画像である．この画像は非常に単調なものであり，情報量は非常に小さい．(b) は，輝度が 0 から 255 までの一様なノイズである．この画像は，さまざまな明度を持つ画素を含んでおり，複雑に見える．そして，この画像の情報量は，大きくなっている．実画像の場合，(c) と (d) を比較すると，画像 (c) の方が画像 (d) よりも複雑に見える．情報量を比較しても，(c) の方が大きくなっている．

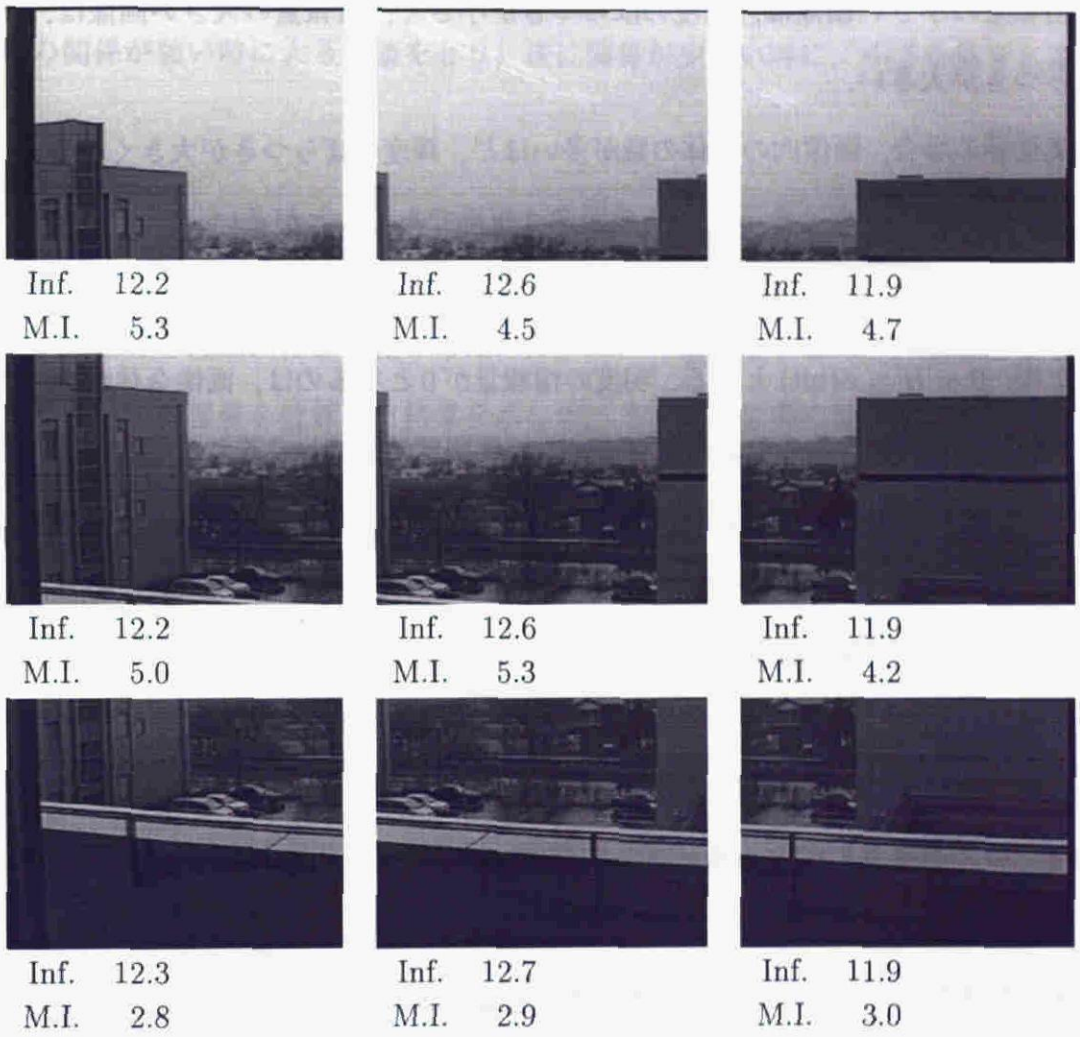


図 4.2: シーンの相互情報量

これらの結果と情報量の定義式から，明度の情報量は，定性的に以下の特徴を持つことがわかる．

- ノイズによる若干の輝度変化に対してロバストである．なぜなら，ヒストグラムをとる時点でノイズ分が平滑化されるからである．
- 情報量の小さい画像は，明度のばらつきが小さく，情報量の大きい画像は，明度のばらつきが大きい．
- 実世界の場合，画像内の物体の数が多いほど，輝度のばらつきが大きくなる．したがって，情報量が大きくなると，その画像は複雑であることが多い．

画像の階調数を N とすると，式(4.3)より，$0 \leq H \leq \log N$ である．実験では $N = 256$ としたため，$0 \leq H \leq 8$(bit) となる．明度の情報量が0となるのは，画像全体が単一の明度値をもつ時で，最大となるのは，画像中にすべての明度値が等しく現れる時である．このことからも，情報量が大きい画像は，複雑であり，逆に，情報量が小さい画像は，単純であることが分かる．ただし，この情報量の計算において，空間情報を捨てているので，空間的な複雑さとは一致しない．空間情報を扱うためには，4.3.1の確率 $p(I(x_1, y_1), I(x_2, y_2), \ldots, I(x_n, y_n))$ を扱う必要がある．しかし，この確率を推定するためには，多くの画像サンプルを必要とするため，現実的には不可能である．

彩度，色相の場合も明度の情報量と同様に，それぞれ，さまざまな彩度をもつ画像，さまざまな色相をもつ画像に対して大きな値をとることが分かる．

図4.2を見ると，これらの部分画像の情報量は，ほぼ同じであるが，相互情報量は，上の6枚の画像で大きく，下の3枚の画像で小さくなっていることが分かる．上下の画像での大きな違いは，上の画像には空のようなグラデーションの部分を多く含むのに対して，下の画像にはその部分が存在しないことである．グラデーションの部分は，明度，彩度，色相が一様に変化しており，その結果，それらの相関が大きくなっている．したがって，相互情報量が大きくなっていると考察される．このように，相互情報量は，特徴間の変化の様子を表す尺度として用いることができる．

式 (4.18) を，3 つの確率変数の組 X, Y, Z の場合に拡張すると，

$$0 \leq I(X,Y,Z) \leq \min(H(X), H(Y), H(Z)) \tag{4.24}$$

が得られる．$I(X,Y,Z)$ は，X, Y, Z が独立の時に最小値 0 となる．最大となるのは，たとえば，X が Y, Z の関数の時で，最大値は，$H(X)$ となる．このことからも，相互情報量は，特徴間の関係が強い時に大きな値をとり，逆に関係が少ない時に，小さな値をとることがわかる．

4.6 まとめ

シーンの画像を理論的に定量化するためにシーンの情報量を導入し，さまざまな画像に対してシーンの情報量を計算した結果を示した．また，特徴間の確率的な従属性を表す相互情報量について述べた．扱う特徴を理論的に裏付けることによって，シーンからもっとも多くの情報を効率よく獲得することが可能であると考えられる．

第5章　情報量に基づくシーン画像の獲得システム

5.1　まえがき

アクティブビジョンの工学的研究では，色，線の方向といった低次の特徴から，saliency map，すなわち，各特徴の注目点を統合したマップを構成し，そのマップにしたがって注視点を移動させている．一方，視覚的注意や視覚探索の心理学的研究では，saliency mapの研究が盛んに行なわれ，特徴統合理論[26]やそれを修正したモデルがいくつか提案されている[21, 28]．このモデルをアクティブビジョンシステムに実現する場合，カメラからの入力画像に対して，いくつかの特徴を抽出し，その特徴量をもとにsaliency mapを計算する方法が取られている．しかし，特徴を抽出するとき，ヒューリスティックに特徴量を計算していた．例えば，画像の目立つ部分として，赤や黄色の部分や，高い空間周波数をもつ部分を抽出していた．

4章で，特徴に理論的な裏付けを与えるため，シャノンの情報量によって特徴を定量的に表現する手法を提案した．ここでは，シーンのsaliencyを記述するのにヒューリスティックな方法ではなく，理論的なinformation mapを提案する．さらに，これ用いてシーンから情報の大きい部分を選択的に注視するアクティブビジョンシステムを開発する．以下，5.2で，開発したシステムについて述べる．そのシステムを用いて行なったシーン画像の獲得実験とその結果を5.3に述べ，5.4で考察する．最後に5.5で，本研究をまとめるとともに今後の課題について考察する．

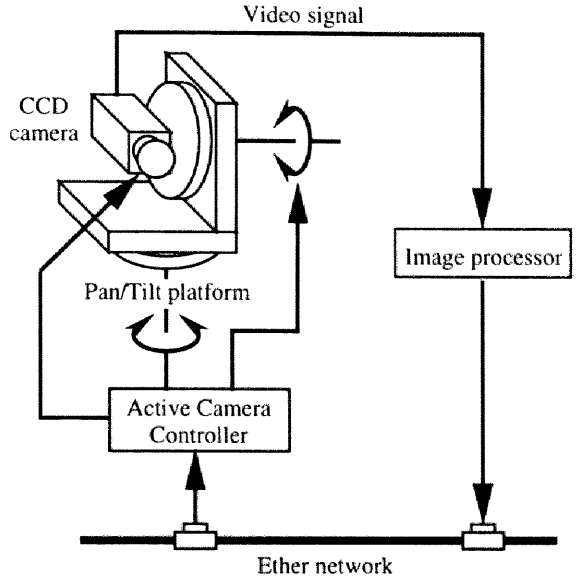


図 5.1: 画像の獲得システムの概観

5.2 画像の獲得システム

5.2.1 システムの概要

システムはカメラ，画像処理装置，アクティブ・カメラ制御装置，パン・チルトステージからなる．**図5.1**にその概観を，**図5.2**にパン・チルトステージを示す．カメラは小型一体型1/3″カラーCCDカメラ (SONY EVI-310) を用いた．カメラは汎用画像処理装置に接続されている．汎用画像処理装置はホストコンピュータ (MVME167，Lynx OS) とVMEスロットに接続された動画像処理VMEボード (Datacube，MAXVIDEO 200) とフルカラー画像入出力ボード (同，DIGI COLOR) からなる．これは，512×484(pixels) の画像を入力し，それをリアルタイムで処理することが可能である．汎用画像処理装置とアクティブ・カメラ制御装置はEther Network経由で通信を行う．アクティブ・カメラ制御装置はパン・チルトステージ (駿河精機製，K44/K45-110) の各モータ (UPD534M-B) を駆動し，カメラのズーム，フォーカスなどのパラメータを制御する．パン・チルトステージの最高回転速度は50[deg/s]，1パルスあたりの移動量は0.002[deg]である．カメラのズームレンズは，8倍 (f=5.9～47.2mm) まで制御可能である．

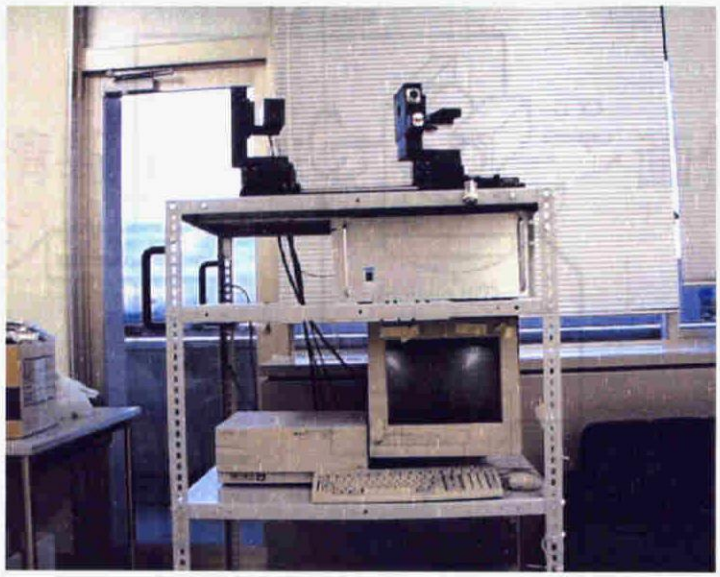


図 5.2: 高精度パン・チルトステージ（上）とその制御装置

5.2.2 探索に使用する特徴

シーンの特徴として，カラーCCDカメラからの画像をHSV色空間[87]に分割して得られる明度，彩度，色相を用いた．この変換は，フルカラー画像入出力ボードでフレームレートで行なうことができる．それぞれの特徴に対して，情報量を計算する．例えば，明度の情報量を計算する場合，まず，確率$p(k)$を画像の明度のヒストグラムから以下のように推定する．

$$p(k) = \frac{\text{Number of pixels with } V(x,y) = k}{\text{Total number of pixels}} \tag{5.1}$$

ここで，$V(x,y)$は，画素(x,y)での明度値を表す．したがって，明度の情報量H_Vは，

$$H_V = -\sum_{k=0}^{N-1} p(k) \log p(k) \tag{5.2}$$

と計算される．ここで，Nは量子化のレベル数$(N = 256)$である．彩度，色相の特徴についても，それぞれのヒストグラムを用いて同様に計算することができる．複数の特徴を組み合わせる場合も，多次元のヒストグラムを正規化し，確率分布を推定することによって，同様に計算が可能である．

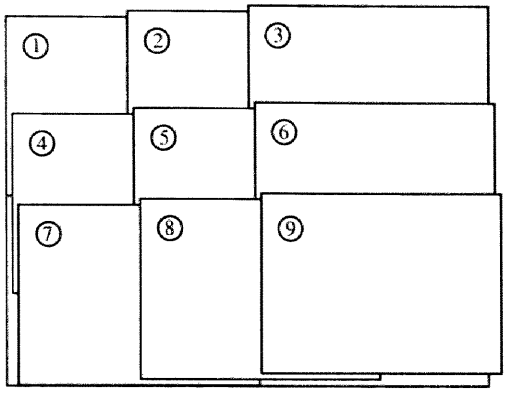


図 5.3: オーバーラップした画像の分割

5.2.3 情報量マップの生成

本研究では，システムに複雑な部分を注視するというタスクを与えている．このタスクを達成するため，画像中のどの部分がどれくらい複雑であるかという情報量マップを生成する．

この情報量マップは，画像を部分領域に分割し，その領域ごとの輝度の情報量を計算することで容易に得ることができる．しかし，複雑な部分が領域の境界にまたがって存在した場合，その部分の正確な情報量を計算することができない．そこで本研究では，**図 5.3** に示すように，領域をオーバーラップさせることによってこの問題を解決した．

本研究で使用した情報量マップの大きさは 3×3 と非常に小さい．これは，情報量の計算に時間がかかるためである．情報量マップの大きさが 3×3 であれば，情報量の計算時間は，汎用画像処理装置を用いた場合，43[ms/frame] となり，ほぼフレームレートでの処理が可能となる．しかし，情報量マップを大きくすると，その大きさに比例して計算時間がかかるようになる．

5.2.4 ズームを用いた注視点の移動

人間の目は，中心窩を持ち，視野の中心で高解像度を持ち，周辺に行くにしたがって解像度が低下している．中心付近の高解像度を有効に利用するため，人間は眼球を高速に移

動させ，シーン全体を探索している．この機能をアクティブビジョンシステムにもたせるため，カメラのズーム機能を利用する．つまり，情報量が大きく高解像度での撮影が必要な部分は，カメラをズームインし，情報量が小さく，低解像度でも十分な部分は，カメラをズームアウトする．これによって，人間の視覚のような不均一解像度を疑似的に実現することができる．

5.2.5 シーンの探索方法

本システムの目的は，シーンを高解像度で効率的に探索することである．ズームレンズを積極的に利用し，情報量の大きい部分を拡大して観察することによって高解像度の探索を行なう．詳細を以下に述べる．

システムは，まず，シーンを低解像度で撮影する．その画像 (512×484) を 1/4 の領域 (256×242) に，オーバーラップするように分割する（図 5.3 参照）．分割したそれぞれの領域について，扱う特徴の情報量を 5.2.2 で述べた方法で計算し，情報量マップを構成する．情報量マップの大きさは，3×3 である．情報量の大きい部分へ視点を移動し，ズームによって詳細に観察する．ズームの拡大率は，2 倍とする．つまり，低解像度での部分領域を画面全体に拡大する．さらに，2 倍に拡大した画像についても，同様に情報量マップを計算し，情報量の大きい部分を拡大することにより，4 倍のズームで観察することができる．もしズームの範囲が許せば，8, 16, ...倍の拡大が可能である．

したがって，システムは，低解像度の画像を 9 つの部分領域に分割し，その部分領域を 9 つの部分領域に分割するといった動作を繰り返す．これは，9 分木の探索問題と等価である．木の探索方法には，深さ優先探索，幅優先探索，最良探索があり，探索方法によって，システムの効率が変化することが考えられる．その効率を，シーンから得られる情報量の大きさとシステムの動作コスト（視点の移動距離，ズームの回数）の 2 つの点から評価する．

探索結果を，ズームに応じた解像度で合成して表示する．使用したズームレンズの制約から，1，2，4 倍のズームを用いて探索を行う．すなわち，1，2，4 倍のズームに対して，

解像度を1，2，4倍に変化させて合成する．その結果，探索が進むにつれ，探索した場所の解像度が高くなり，詳細な情報が得られる．

5.3 実験とその結果

まず，明度の情報量のみを用いて，屋内で探索実験を行なった結果について述べる．探索方法は，深さ優先探索を用い，情報量が6.8ビット以上の部分のみを探索した．**図5.4**に探索で得られた画像列を示す．紙面の都合上，図5.4では6.9ビット以上の部分のみを表している．画像の下の左の数字は，システムが探索した順序を表し，右上は情報量，右下はズームした画像の左上の画素が画像1のどこにあたるかを表している．また，画像1は1倍のズームで撮影した画像2, 6, 7, 8は2倍のズーム，それ以外は4倍のズームで撮影されたものである．システムは，まず，初期の位置，1倍のズームで画像を撮影した（図5.4の1）．この画像について情報量マップを計算すると，図5.3の③の部分がもっとも大きい値を示した．そこで，システムはその部分にカメラを向け，2倍のズームで画像を撮影した（図5.4の2）．手続きは再帰的に繰り返されるため，この画像について同様に情報量マップを計算する．その結果，図5.3の⑧の部分がもっとも大きい値を示した．同様にその部分にカメラを向け，$2 \times 2 = 4$倍のズームで画像を撮影した（図5.4の3）．さらに，手続きを再帰的に繰り返すが，これ以上ズームインする事ができないので，もとの状態（図5.4の2）に戻る．システムは，次に情報量が大きい部分（図5.3の③）を拡大した（図5.4の4）．同様に繰り返して，画像5の撮影を行った．ここで，図5.4の2の中で情報量が6.8ビット以上の部分の探索は終わり，zoom:2での探索へ戻り，図5.4の6を注視した．以下，同様にして，図5.4の7, 8, 9, 10へ注視点を移し，撮影を行った．

さらに，探索実験で得られた画像を，1枚の大きな画像に合成した．その結果を**図5.5**に示す．ただし，入力画像 (512×484) を4倍した時，画像サイズが 2048×1936 となり，それを誌上で表現するのは困難となる．そこで，入力画像のサイズを1/4の 128×121 とし，オーバーラップしている部分は，情報量の大きい方を優先している．合成に用いた画像は，

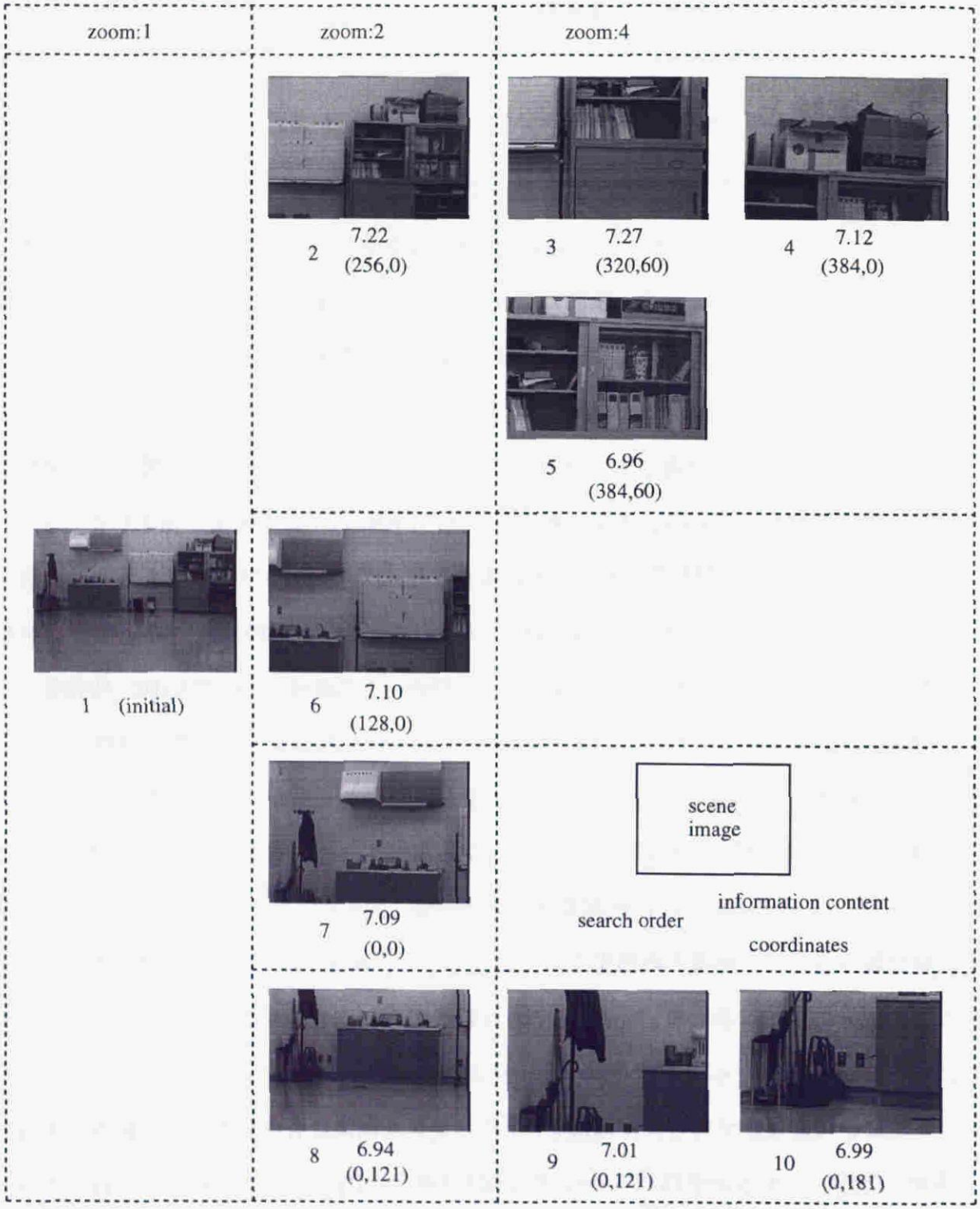


図 5.4: シーンの探索結果

図 5.5: 屋内のシーンの合成結果

図 5.4 に示した情報量が 6.9 ビット以上の部分の 10 枚の他に，6.8 ビット以上 6.9 ビット未満の 6 枚（zoom:2 での画像 2 枚，zoom:4 での画像 4 枚）を追加した．図 5.5 での解像度の変化の様子を**図 5.6** に示す．同様に，屋外のシーンで合成した画像を**図 5.7** に示す．

次に，屋外，屋内のシーンを対象にし，明度，彩度，色相の情報量を切り換えて行なった実験結果を示す．**図 5.8** は，1 倍のズームで撮影した屋外のシーンである．探索結果を比較しやすくするため，画像サイズを 32×30 に小さくしているので，モザイク画像になっている．**図 5.9**，**5.10**，**5.11** に，明度，色相，彩度の情報量を用いた探索結果を示す．それぞれの図中，(a) は，探索の途中の画像，(b) は，最終の画像を示す．また，システムが高い解像度で探索した部分を黒枠で示す．探索過程がよくわかるように，深さ優先探索での結果を示す．

同様に，屋内のシーンを**図 5.12** に，探索結果を**図 5.13**，**5.14**，**5.15** に示す．

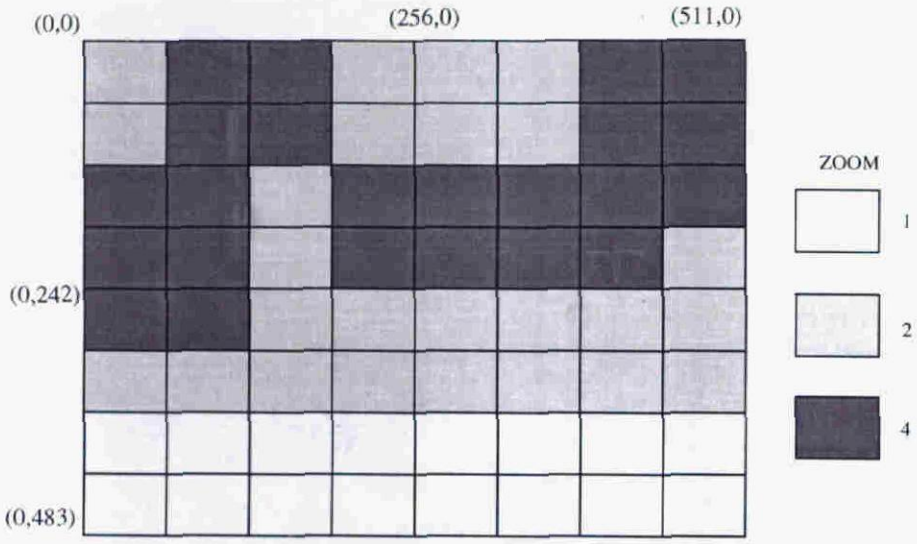


図 5.6: 解像度の変化の様子

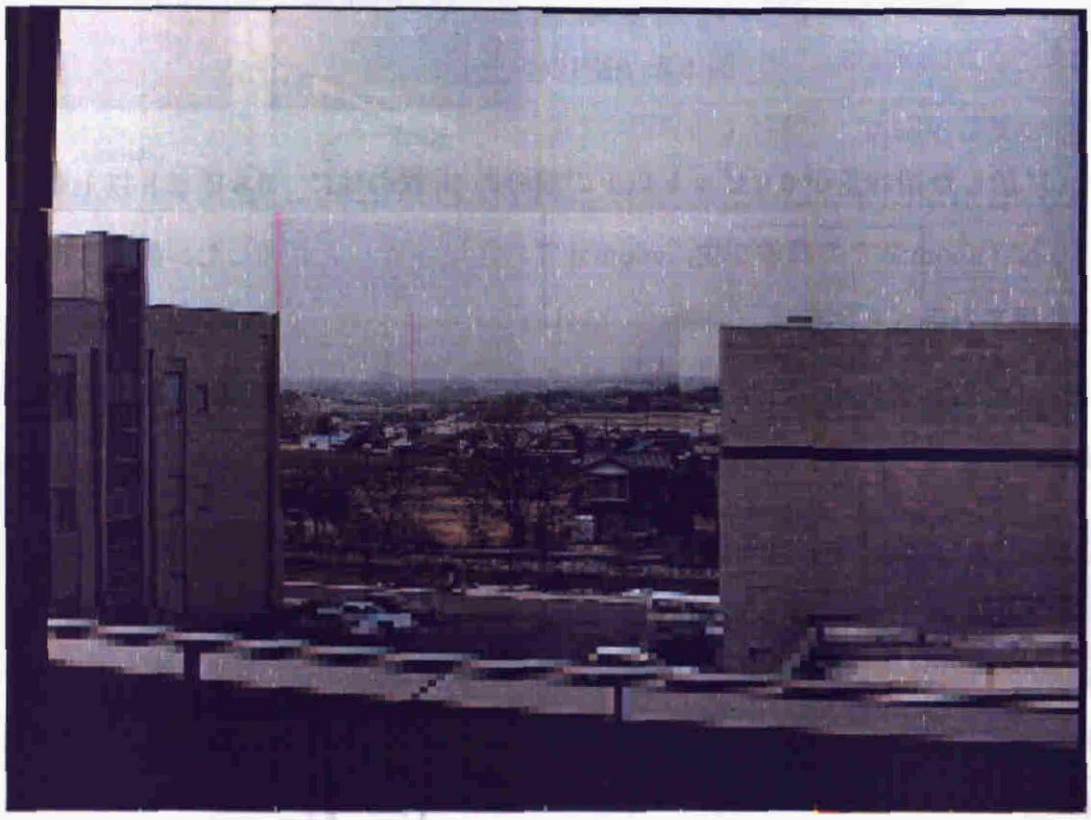


図 5.7: 屋外のシーンの合成結果

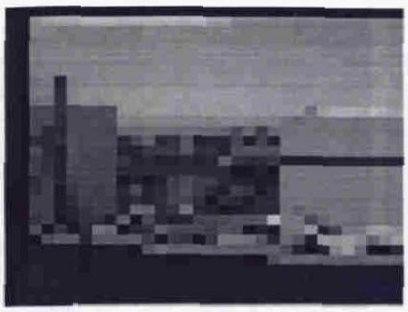
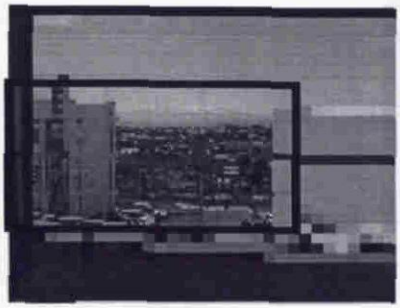
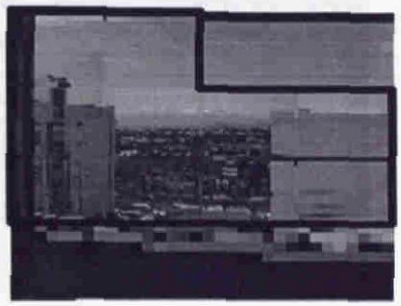


図 5.8: 屋外のシーン（1 倍ズーム）

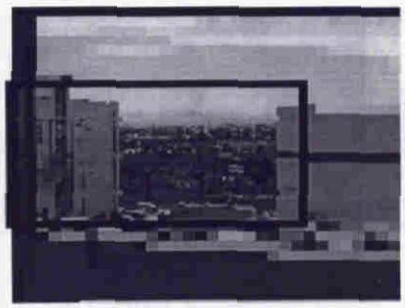


(a) 途中

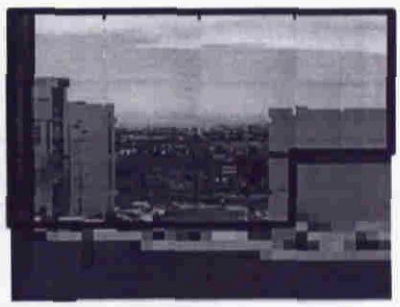


(b) 最終

図 5.9: 屋外のシーンの探索結果：明度の情報量

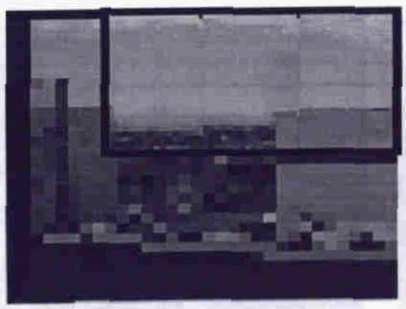


(a) 途中



(b) 最終

図 5.10: 屋外のシーンの探索結果：色相の情報量



(a) 途中

(b) 最終

図 5.11: 屋外のシーンの探索結果：彩度の情報量

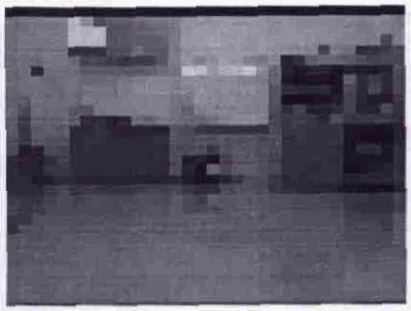
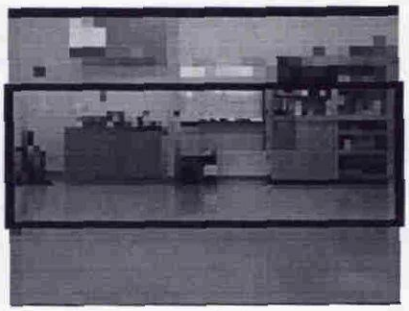
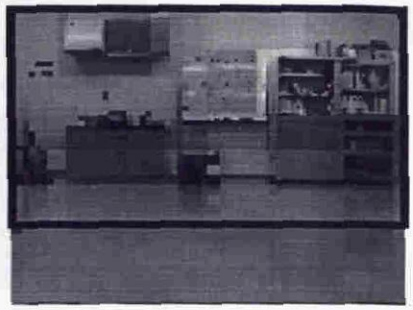


図 5.12: 屋内のシーン（1倍ズーム）

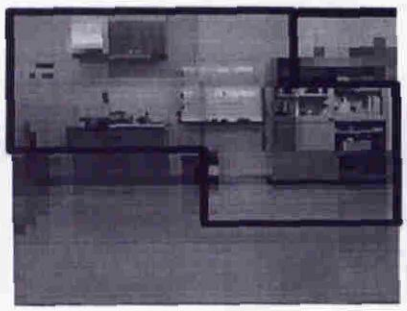


(a) 途中

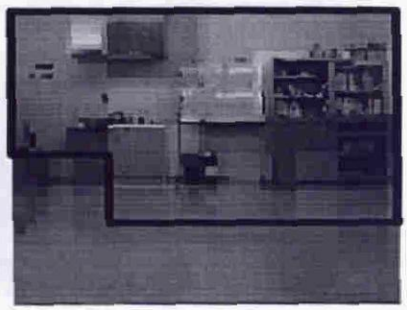


(b) 最終

図 5.13: 屋内のシーンの探索結果：明度の情報量

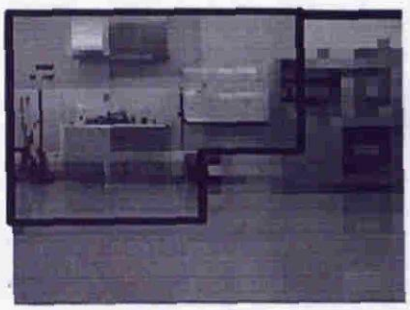


(a) 途中

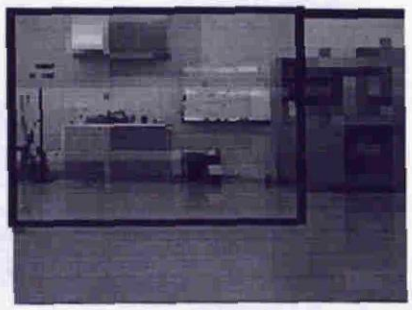


(b) 最終

図 5.14: 屋内のシーンの探索結果：色相の情報量



(a) 途中



(b) 最終

図 5.15: 屋内のシーンの探索結果：彩度の情報量

5.4 考察

明度の情報量を用いた場合，さまざまな明度を含む部分を先に探索した（図5.4，図5.9 (a)，図5.13 (a)）．また，画像下部の明度の単調な部分は，注視しなかった．このことから，明度の情報量は，画像の複雑な部分で大きな値になることがわかる．図5.4の5よりも，図5.4の6の方が情報量が大きいが，5の方を先に探索した．これは，深さ優先探索を行なったためである．このように深さ優先探索では，周囲に情報量の大きい部分が存在してもその部分の探索が後にまわされる欠点がある．色相の情報量を用いた場合，明度と色相を置き換えて考えることにより，さまざまな色を含む部分を注視することがわかる．これは，実験結果の示すものと等しい．例えば，図5.14 (a)では，探索した部分には，調度品，本といったさまざまな色をもつものが含まれている．彩度の情報量を用いた場合も同様に，さまざまな彩度をもつ部分を注視する．図5.11では，空と雲（白（無彩色）から青への色のグラデーションを含む）の部分をよく注視している．

明度の情報量を用いた場合と，色相の情報量，彩度の情報量を用いた場合を比較すると，探索過程の違いが見られる．屋外のシーンでは，明度の情報量では，建物の部分をよく注視しているのに対し（図5.9），色相の情報量，彩度の情報量では，空と雲の部分をよく注視している（図5.10, 5.11）．特に，彩度の情報量では，探索の初期の段階で，その部分を注視している．同様に，屋内のシーンでも情報量を変えることによって，システムは異なるふるまいをする．このように，システムに与えられたタスクに応じて，情報量を変更して使うことによって，システムに異なったふるまいをもたせることができる．

本システムと人間の視覚系との類似度を評価するため，人間の眼球運動を測定し，その注視点の分布と比較した．被験者の前に置かれたディスプレイに観測するシーンを表示し，被験者の眼球運動をアイマークレコーダ (EMR-7, ナック) で測定した．被験者にはシステムと同様な，複雑な部分を注視する，というタスクを与えた．3名の被験者に対し，30秒間の計測を行なった．

図5.16に測定した注視点の分布を示す．注視点は，200ms以上の間，視点が1deg以内

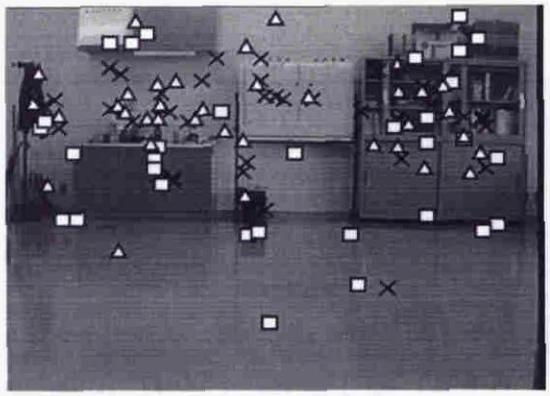


図 5.16: 注視点の分布

の範囲であった部分の重心として抽出した.

どの被験者もほぼ同じところに注視点を移動させているが，被験者間の差異も存在する．例えば，被験者 CN は他の被験者に対して，ホワイトボートをあまり見ておらず，棚の上や換気扇のあたりをよく注視してる.

システムが計算したシーンの情報量は，**図 5.17** のようになる．図 5.16 と図 5.17 を比較すると，情報量の少ない部分は，どの被験者もほとんど注視せず，また，情報量の大きい部分は，被験者によって注視したり，しなかったりすることがわかる．被験者全体として，シーンの情報量と領域を注視した回数との相関係数は，0.78 となり，相関があることが分かる．したがって，眼球運動の観点から，輝度の情報量は人間の複雑さの尺度として妥当である.

明度の場合は，ある程度の相関があることが分かったが，彩度，色相の場合は，様子が変わってくる．色についての心理学的な研究結果から，人間は，赤に近い色や彩度の高い色に目を引かれるという傾向があり，さらに，色の社会的意味（赤は危険，黄色は注意，緑は安全など）によっても，注意を引くかどうかが変化することが知られている [102]．このような色の誘目性は，情報理論の枠組みの中でボトムアップに情報量を計算する本システムでは考慮していない．しかし，別の枠組みとしてこのようなトップダウンの手法を取り入れることによって，人間に近いふるまいをもたせることができると考えられる.

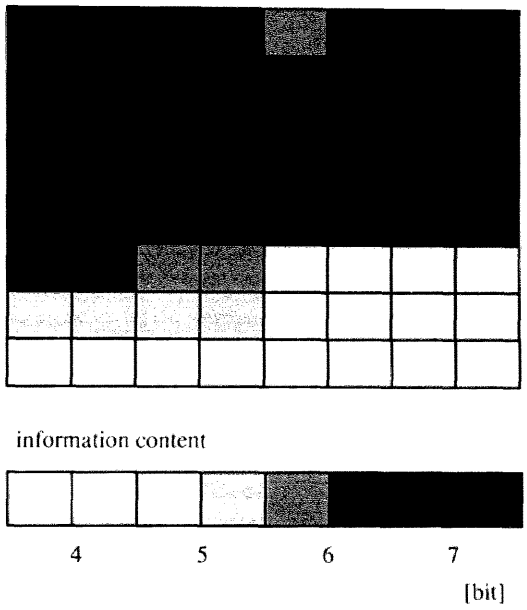


図 5.17: 情報量の計算結果

探索方法を変更したとき，シーンから得られる情報量の変化を**図 5.18** に示す．図中，横軸は探索の回数を，縦軸は，最良探索で得た総情報量とそれぞれの手法で得た総情報量の差分を表す．差分を用いた理由は，最良探索が探索の初期でシーンからもっとも多くの情報を得ることは自明であるからである．depth first, width first は，最良探索で得た総情報量から，それぞれ，深さ優先探索，幅優先探索で得た総情報量を引いたものであり，sequential は，情報量マップの左上から順番に探索する手法で得た総情報量を最良探索で得た総情報量から引いたものである．

どの探索手法でも最終的にシーンから得られる総情報量は同じであるが，初期の段階で得られる情報量に違いがみられる．幅優先探索は，最良探索と比べほぼ同程度の情報量を得ることができた．深さ優先探索は，最良探索，幅優先探索よりも初期の段階で得られる情報量は少なかった．また，最良探索，幅優先探索，深さ優先探索は，単純に左上から順番に探索する手法と比較して，初期の段階でかなり多くの情報量を得ることができた．

この 3 つの手法を情報量以外の点で比較すると，深さ優先探索は，カメラの移動量が，幅優先探索は，ズームの変化量がもっとも少なかった．これを，**表 5.1** に示す．したがって，

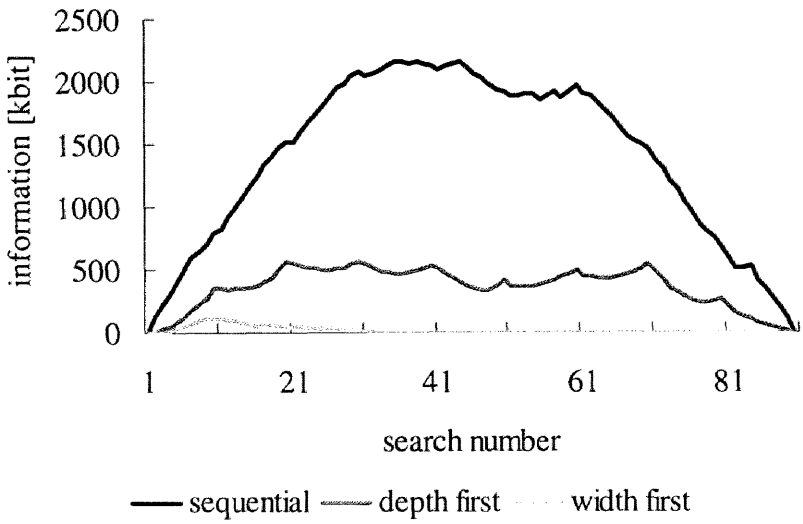


図 5.18: 得られる情報量の探索方法による比較

表 5.1: 探索方法による性能の比較

	information(*)	movements	zooms
best	3.8 Mbits	large	18
depth	3.5 Mbits	small	18
width	3.7 Mbits	large	2

(*) information obtained by ten searches.

アクティブカメラの性能やアプリケーションによって，どの探索手法を用いると総合的に良い性能が得られるかが変化すると考えられる．カメラの移動やズームの駆動に時間のかかるシステムでは，それぞれ，深さ優先探索，幅優先探索を用いた方がシーン全体の探索時間は短くなる．また，遠隔地での操作など，画像の伝送に時間のかかるアプリケーションでは，カメラの移動やズームの変化の時間は相対的に小さくなり，最良探索によって最大の情報量を得ることが望ましいと考えられる．

5.5 まとめ

シーンの情報量の大きい部分を階層的に探索するアクティブビジョンシステムについて述べた．シーンの特徴を，情報理論に基づいて定量化する情報量を用い，入力画像に対して領域ごとに情報量を計算し，情報量マップを構成することによって，アクティブビジョンシステムを開発した．システムが扱う特徴を理論的に裏付けることによって，シーンからもっとも多くの情報を効率よく獲得することが可能である．

開発したアクティブビジョンシステムは，カメラレンズのズーム機能を積極的に利用し，人間の網膜のような不均一の解像度を実現することができる．それによって，シーンの高解像度での探索を効率的に行うことができることを示した．この手法は，距離の離れた，あるいは，危険な場所で働くロボットの視覚として有用である．なぜなら，この手法によって常に最大の情報量が得られることが保証されるからである．また，背景のテクスチャの撮影など，コンピュータビジョンだけでなくコンピュータグラフィクスの分野でも有用である．

人間の視覚は，明度，彩度，色相だけでなく，形，動き，テクスチャなど多くの特徴を扱っている．これらの手がかりを用いて，さらに洗練された情報量マップを生成することが今後の課題である．

第6章　結論

本論文では，人間の視覚系に示唆を得た移動物体の追跡システムと，シーンの情報量の定義とそれを用いた画像の獲得システムの実現について述べた．

1章では，本研究の背景と目的を明らかにし，本論文の特徴について簡単に述べた．

2章では，人間の視覚情報処理，特に移動物体追跡時の眼球運動と予測制御，視覚探索をモデル化する特徴統合理論について述べた．さらに，アクティブビジョンの主要な機能の1つである移動物体の追跡について，従来に提案されたシステムを述べ，最後に従来のアクティブビジョンについてまとめた．

3章では，眼球のsaccade運動とpursuit運動，予測制御に示唆を得たシステムについて述べた．まず，輝度変化の条件式をもとに移動物体の位置，速度をフレームレートで計算する手法を提案し，汎用画像処理装置上に実現した．予測制御を線形予測，自己回帰モデルによる予測，カルマンフィルタによる予測の3種類にモデル化し，システムに実装した．また，簡易なカメラキャリブレーション方法について述べ，実環境で実験を行ない，システムがノイズに強く，運動が規則的な時，遅れのない追跡を行なうことが可能であることを示した．

4章では，シャノンの情報量，相互情報量，Kullback情報量について簡単に述べた後，シーンを定量化する情報量について述べた．その情報量をさまざまな画像について計算し，以下のような特徴があることについて述べた．

- ノイズによる若干の輝度変化に対してロバストである．
- 情報量が大きくなると，そのシーンはさまざまな物体を含んでいることが多い．

さらに，特徴間の確率的な従属性として存在する相互情報量についても実験により考察し，

特徴間の変化の様子を表す尺度として用いることができることについて述べた．

5章では，4章で定量化した情報量を用い，シーンから情報量の大きい部分を獲得するシステムについて述べた．シーンの特徴としては，シーンを撮影して得られる，明度，彩度，色相を用い，それぞれについて探索実験を行なった．その結果，扱う特徴を変えるとシステムのふるまいが変化することが明らかとなった．また，探索方法を変えた時，最良探索では，探索の初期の段階でもっとも大きい情報量を得ることができ，深さ優先探索では，カメラの移動量をもっとも少なくすることができ，幅優先探索では，ズームの変化量がもっとも少なくすることができることについて述べた．

最後に本研究で得られた知見を総括すると次のようにまとめられる．

- 移動物体の追跡に関して
  - 汎用画像処理装置による画像処理の高速化と高速なパンチルトステージによって，システムは200deg/sの速度までの物体を追跡することが可能である．
  - 線形予測の場合，速度の推定に誤差が大きく，特に移動物体の速度が大きい場合は追跡に遅れが生じやすいが，移動物体の運動が規則的な時，自己回帰モデルやカルマンフィルタを用いることによって，遅れのない追跡が可能である．
- 画像の獲得システムに関して
  - システムが扱う特徴を理論的に裏付けることによって，シーンからもっとも多くの情報を効率よく獲得することが可能である．
  - カメラレンズのズーム機能を積極的に利用し，人間の網膜のような不均一の解像度を実現することができる．それによって，シーンの高解像度での探索を効率的に行うことができる

今後の課題としては，以下のようなものがあげられる．

- 移動物体の追跡は，1つの物体のみ可能である．しかし，視野内を移動する複数の物体から主要な物体を選択し，それを追跡することはできない．このように複数の物体

から1つを選択することを，人間は容易に行っている．この機能は，本研究で用いた時空間の微分値の特徴だ けでなく，色，視差などの特徴を統合し，視野に含まれる物体を認識する ことによって実現されると考えられる．

- 画像の獲得システムは，明度，彩度，色相の特徴を扱っているが，人間はそれらだけではなく，形，動き，テクスチャなど多くの特徴を扱っている．これらの特徴を統合し，さらに洗練された情報量マップを生成することによって，よりタスクに応じたふるまいをもたせることが可能であると考えられる．
- さらに，この2つのシステムを組み合わせ，動的なシーンでの環境理解などへの応用が考えられる．

最後に，本研究が人間の視覚情報処理の解明や，よりすぐれたアクティブビジョンシステムの開発の基礎となることを期待して，本論文の結びとする．

謝辞

はじめに，名古屋大学名誉教授 杉江 昇 先生（現在，名城大学理工学部教授）と名古屋大学大学院工学研究科 大西 昇 教授に深く感謝致します．両先生には，指導教官として本研究を進める機会を与えていただきました．また，両先生から，明確な研究指導と，熱心かつ的確な御教示，時宜を得た激励，そして絶え間ない援助をいただいたことにより，本論文をまとめることができました．両先生の御指導，御鞭撻は，今後の研究活動を行う上での支えとなるものです．ここに，心から感謝の意を表します．

本論文の作成にあたり，本論文を丁寧に読んでいただき，貴重な御意見と激励をいただいた，名古屋大学大学院工学研究科 鳥脇 純一郎 教授，末永 康仁 教授に深く感謝致します．それぞれの分野で最先端の研究をされていらっしゃる両先生からの御意見，御教示はとても興味深く，本研究をより幅広い視点から再認識することができました．さらに今後の研究の参考にさせていただきたいと思います．

本研究を進めるにあたり，多くの方々から有益な助言と討論をいただいたことに感謝します．名古屋大学大型計算機センター 田中 敏光 助教授には，熱心な討論をいただいたとともに，広範囲の研究分野から見た本研究の位置づけを示唆していただきました．また，名古屋大学大学院工学研究科 佐川 雄二 講師，山村 毅 講師（現在，愛知県立大学情報科学部助教授），皆川 洋喜 助手（現在，筑波技術短期大学電子情報学科助手），工藤 博章 助手，松本 哲也 助手，並びに大西研究室の皆様には熱心な討論と有益な意見をいただきました．さらに，名古屋大学工学部情報工学教室 松崎 規子 事務官に感謝致します．さまざまな心遣をいただくとともに，迅速かつ丁寧な事務を行なっていただいたことにより，滞りなく研究を進めることができました．

本研究の一部は，筆者が平成9年度 理化学研究所 ジュニア・リサーチ・アソシエイトとして携わったものであります．機器を使わせていただきました理化学研究所バイオ・ミメティックコントロール研究センター 故 伊藤 正美 センター長，佐田 登志夫 センター長をはじめ，熱心な討論と有益な意見をいただきました，生体ミメティックセンサー研究チームの研究員の皆様に感謝致します．また，さまざまな心遣をいただくとともに，迅速かつ丁寧な事務を行なっていただいた，近藤 亜紀 様に感謝致します．

また，本研究の一部は，筆者が平成10年度日本学術振興会特別研究員として携わったものであり，平成10年度文部省科学研究費補助金（特別研究員奨励費）(#3378)の援助を受けました．さらに，本研究を国際会議で発表するための渡航費は，C&C振興財団の助成によりました．

最後に，長期間にわたる教育を受けることを支えてくれた両親と家族に心から感謝致します．

参考文献

[1] L. N. Thibos, F. E. Cheney and D. J. Walsh, "Retinal limits to the detection and resolution of gratings," Journal of the Optical Society of America, vol.A4, pp.1524–1529, 1987.

[2] R. J. Jacobs, "Visual Resolution and Contour Interaction in the Fovea and Periphery," Vision Research, vol.19, pp.1187–1195, 1979.

[3] 末松良一, 山田宏尚, "中心窩を有する新しい視覚センサの開発 — 広角高歪曲レンズの開発 —", 計測自動制御学会論文集, vol.31, no.10, pp.1556–1563, 1995.

[4] A. L. Yarbus, "Eye Movements and Vision," Plenum Press, 1967.

[5] G. R. Loftus and N. H. Mackwartb, "Cognitive Determination of Fixation Location during Picture Viewing," Journal of Experimental Psychology: Human Perception and Performance, vol.4, pp.565–572, 1978.

[6] E. P. Krotkov: "Active Computer Vision by Cooperative Focus and Stereo," Springer-Verlag, 1989.

[7] D. H. Ballard, "Animate Vision," Artificial intelligence, vol.48, pp.57–86, 1991.

[8] R. Desimone: "Neural Circuits for Visual Attention in the Primate Brain," G. A. Carpenter and S. Grossberg (eds), Neural Networks for Vision and Image Processing, MIT Press, pp.343–364, 1992.

[9] D. Murray and A. Basu, "Motion Tracking with an Active Camera," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.16, no.5, pp.449–459, 1994.

[10] D. W Murray, P. F McLauchlan, L. D Reid and P. M. Sharkey, "Reactions to Peripheral Image Motion using a Head/Eye Platform," Proc. of ICCV, pp.403–411, 1993.

[11] K. J. Bradshaw, P. F. McLauchlan, I. D. Reid and D. W. Murray, "Saccade and Pursuit on an Active Head-Eye Platform," Image and Vision Computing, no.12, pp.155–163, 1994.

[12] K. Pahlavan, T. Uhlin and J-O. Eklundh, "Dynamic Fixation," Proc. of ICCV, pp.412–419, 1993.

[13] T. Viéville, "A Few Steps Towards 3D Active Vision," Springer, 1997.

[14] J. J. Clark and N. J. Ferrier: "Attentive Visual Servoing," Active Vision, A. Blake and A. Yuille (eds), pp.137–154, The MIT Press, 1993.

[15] A. L. Abbott, "A Survey of Selecting Fixation Control for Machine Vision," IEEE Control Systems, vol.12, no.4, pp.25–31, Aug. 1992.

[16] N. J. Ferrier and J. J. Clark: "The Harvard Binocular Head," International Journal of Pattern Recognition and Artificial Intelligence, no.7, pp.9–31, 1993.

[17] D. A. Robinson, "The Mechanics of Human Smooth Pursuit Eye Movement," Journal of Physiology, vol.180, pp.569–591, 1965.

[18] 杉江 昇 著, 南雲 仁一 編, "生体システム", 日刊工業新聞社.

[19] L. Stark, G. Vossius and L. R. Young, "Predictive Control of Eye Movements," IEEE Trans. on HFE, vol.3, 1962.

[20] L. R. Young and L. Stark, "Variable Feedback Experiments Testing a Sampled Data Model for Eye Tracking Movements," IEEE Trans. on HFE, vol.4, pp.38–51, 1963.

[21] C. Koch and S. Ullman, "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry," Human Neurobiology, 4, pp.219–227, 1985. pp.195–202, 1995.

[22] A. F. Fuchs, "Periodic Eye Tracking in the Monkey," Journal of Physiology, vol.193, 1967.

[23] "A Model of Predictive Control in Visual Target Tracking," IEEE Trans. on SMC, vol.1, no.2, 1971.

[24] D. Brogan, "Visual Search," Taylor and Francis, 1990.

[25] D. Brogan, A. Gale and K. Carr, "Visual Search 2," Taylor and Francis, 1990.

[26] A. Treisman and A. Gelade, "A Feature Integration Theory of Attention," Cognitive Psychology, 12, pp.97–136, 1980.

[27] 横澤一彦, "視覚的注意と視覚探索", 日本認知学会第10回大会発表論文集, pp.104–105, 1993.

[28] 横澤一彦, "多解像度モデルにおける視覚的注意と視覚探索の分析", 認知科学, vol.1, no.2, pp.64–82, 1994.

[29] S. Zeki, "The Visual Image in Mind and Brain," Scientific American, 1992-09.

[30] A. Treisman, "Features and Objects in Visual Processing," Scientific American, vol.254, no.11, pp.114–124, 1986.

[31] A. Treisman, "Preattentive Processing in Vision," Computer Vision, Graphics, and Image Processing, vol.31, pp.156–177, 1985.

[32] A. Treisman and S. Gormican, "Feature Analysis in Early Vision: Evidence from Search Asymmetry," Psychological Review, vol.95, no.1, pp.15–48, 1988.

[33] K. Nakayama and G. H. Silverman, "Serial and Parallel Processing of Visual Feature Conjunctions," Nature, vol.320, pp.264–265, 1986.

[34] A. Treisman and J. Souther, "Search Asymmetry: A Diagnostic for Preattentive Processing of Separable Features," Journal of Experimental Psychology: General, vol.144, pp.285–310, 1985.

[35] R. Klein and M. Farrell, "Search Performance without Eye Movements," Perception and Psychophysics, vol.46, no.5, pp.476–482, 1989.

[36] 下條信輔，彦坂興秀，“注意の心理学 — 視空間的注意を中心に —”，生体の科学，vol.43，pp.30–36，1992.

[37] J. M. Wolfe, K. R. Cave and S. L. Franzel, “Guided Search: An Alternative to the Feature Integration Model for Visual Search,” Journal of Experimental Psychology, vol.15, no.3, pp.419–433, 1989.

[38] K. R, Cave and J. M. Wolfe, “Modeling the Role of Parallel Processing in Visual Search,” Cognitive Psychology, vol.22, pp.225–271, 1990.

[39] A. Treisman and S. Sato, “Conjunction Search Revisited,” Journal of Experimental Psychology: Human Perception and Performance, vol.16, pp.459–478, 1990.

[40] M. J. Bravo and K. Nakayama, “The Role of Attention in Different Visual-Search Tasks,” Perception and Psychophysics, vol.51, no.5, 1992.

[41] A. Fiorentini, “Differences between Fovea and Parafovea in Visual Search Processes,” Vision Research, vol.29, pp.1153–1164, 1989.

[42] K. Nakayama, “The Iconic Bottleneck and the Tenuous Link between Early Visual Processing and Perception,” C. Blakemore (Ed.), Vision: Coding and Efficiency, Cambridge University Press, 1990.

[43] H. Zabrodsky and S. Peleg, “Attentive Transmission,” Journal of Visual Communication and Image Representation, vol.1, no.2, pp.189–198, 1990.

[44] C. W. Eriksen and J. D. St. James, “Visual Attention within and around the Field of Focal Attention: A Zoom lens Model,” Perception and Psychophysics, vol.40, no.4, pp.225–240, 1986.

[45] N. Sugie, “Investigation of Visual Perception of Position Based on the Reafferent Theory”, Biological Cybernetics, 21, pp.17–22, 1976.

[46] リレー連載，“視覚探索 (1)-(13)”，数理科学，no.344–356，1992–1993.

[47] M. Ali Taalebinezhaad: “Robot Motion Vision by Fixation”, MIT AI Lab., Technical Report 1384, 1992.

[48] E. Grosso and D. H. Ballard: “Head-Centered Orientation Strategies in Animate Vision”, Proc. of ICCV, pp.395–402, 1993.

[49] J. Hwang, Y. Ooi and S. Ozawa, “An Adaptive Sensing System with Tracking and Zooming a Moving Object,” IEICE Trans. Information and Systems, vol.E76-D, no.8, pp.926–934, 1993.

[50] J. Hwang, Y. Ooi and S. Ozawa, “An Advanced On-Line Visual Tracking System,” Trans. of the SICE, vol.30, no.12, pp.1427–1435, 1994.

[51] T. Uhlin, P. Nordlund, A. Maki and J.-O. Eklundh, “Towards an Active Visual Observer,” Proc. of ICCV, pp.679–686, 1995.

[52] P. Nordlund and T. Uhlin: “Closing the loop: pursuing a moving object by a moving observer,” Proc. of Int. Conf. CAIP, pp.400–407, 1995.

[53] J. C. Fiala, R. Lumia, K. J. Roberts and A. J. Wavering, "TRICLOPS: a tool for studying active vision," International Journal of Computer Vision, vol.12, pp.231–250, 1994.

[54] S. Tölg: "Gaze control for an active camera system by modeling human pursuit eye movement," Proc. SPIE on Intelligent Robots and Computer Vision, vol.1825, pp.585–598, 1992.

[55] K. J. Bradshaw, I. D. Reid and D. W. Murray, "The Active Recovery of 3D Motion Trajectories and Their Use in Prediction," IEEE Trans. on PAMI, vol.19, no.3, pp.219–234, 1997.

[56] D. W. Murray, K. J. Bradshaw, P. F. McLauchilan, I. D. Reid and P. M. Sharkey, "Driving Saccade to Pursuit using Image Motion," International Journal of Computer Vision, vol.16, pp.205–228, 1995.

[57] D. J. Coombs and C. M. Brown, "Real-Time binocular smooth pursuit," International Journal of Computer Vision, vol.11, pp.1439–1470, 1993.

[58] P. von Kaenel, C. M. Brown and D. J. Coombs, "Detecting Regions of Zero Disparity in Binocular Images," Technical Report 388, University of Rochester, 1991.

[59] 喜多伸之, Sebastien Rougeaux, 國吉康夫, 坂根茂幸, "仮想ホロプタを用いた実時間両眼追跡", 日本ロボット学会誌, vol.13, no.5, pp.683–690, 1995.

[60] 國吉康夫, "ステレオトラッキング視覚を搭載した小型移動ロボット", 日本ロボット学会誌, vol.13, no.3, pp.343–346, 1995.

[61] F. Du and M. Brady, "A four degree-of-freedom robot head for active vision," International Journal of Pattern Recognition and Artificial Intelligence, vol.8, pp.1439–1469, 1994.

[62] J. Dias, C. Paredes, I. Fonseca, H. Araujo, J. Batista and A. de Almeida, "Simulating Pursuit with Machines," Proc. of ICRA, vol.1, pp.472–477, 1995.

[63] K. Daniilidis, C. Krauss, M Hansen and G. Sommer, "Real-Time Tracking of Moving Objects with an Active Camera," Real-Time Imaging, vol.4, pp.3–20, 1998.

[64] 和田俊和, 浮田宗伯, 松山隆司, "Appearance Sphere — パン・チルト・ズームカメラのための背景モデル —", MIRU'96, vol.2, pp.103–108, 1996.

[65] 和田俊和, 浮田宗伯, 松山隆司, "視点固定型パンチルトズームカメラとその応用", 電子情報通信学会論文誌 D-II, vol.J81-D-II, no.6, pp.1182–1193, 1998.

[66] 和田俊和, 松山隆司, "視覚・行動機能の統合による柔軟・頑健な能動視覚システムの開発 — 視点固定型パン・チルト・ズームカメラを用いた実時間対称検出・追跡 —", MIRU'98, vol.1, pp.359–364, 1998.

[67] 松山隆司, 和田俊和, 丸山昌之, "能動視覚エージェントによる移動対象の協調的追跡", MIRU'98, vol.1, pp.365–370, 1998.

[68] M. Ishikawa, A. Morita and N. Takayanagi, "High Speed Vision System using Massively Parallel Processing," Proc. IEEE Int. Conf. on Intelligent Robotics and Systems, pp.373–377, 1992.

[69] 石川正俊, "超並列・超高速ワンチップビジョンとその応用", 日本ロボット学会誌, vol.13, no.3, pp.335–338, 1997.

[70] 中坊嘉宏，石井抱，石川正俊，“超並列・超高速ビジョンを用いた1msターゲットトラッキングシステム”，日本ロボット学会誌，vol.15，no.3，pp.417–421，1997.

[71] T. Horiuchi, J. Lazzaro, A. Moore and C. Koch, “A delay line based motion detection chip,” Advances in Neural Information Processing 3, pp.406–412, 1991.

[72] T. Horiuchi, W. Bair, B. Bishofberger, A. Moore, C. Koch and J. Lazzaro, “Computing motion using analog VLSI vision chips: an experimental comparison among different approaches,” International Journal of Computer Vision, vol.8, no.3, pp.203–216, 1992.

[73] T. Horiuchi, B. Bishofberger and C. Koch, “Building an analog VLSI saccadic eye movement system,” Advances in Neural Information Processing Systems 6, pp.582–589, 1994.

[74] R. Bajcsy, “Active Perception,” IEEE Proceedings, vol.76, no 8, pp.996–1006, 1988.

[75] Y. Aloimonos and I. Weiss: “Active Vision,” International Journal of Computer Vision, pp.333–356, 1988.

[76] D. H. Ballard: “Reference Frame for Animate Vision”, Proc. Int. Joint Conf. Artificial Intelligence, pp.1635–1641, 1989.

[77] Y. Aloimonos: “Active Vision Revisited,” Active Perception, Laurence Erlbaum Associates Publishers, pp.1–18, 1993.

[78] Y. Aloimonos, E Rivlin and L. Huang: “Designing Visual Systems: Purposive Navigation,” Active Perception, Laurence Erlbaum Associates Publishers, pp.47–102, 1993.

[79] Y. Takeuchi, Z.F. Wang, N. Ohnishi and N. Sugie: “Real Time Visual Tracking System Mimicking Saccadic Movements,” Proc. of ACCV, vol.1, pp.131–135, 1995.

[80] 石黒 浩，“CVCV-WG 特別報告:コンピュータビジョンにおける技術評論と将来展望”，情報処理学会研究報告，vol.94，no.6，pp.45–52，1995.

[81] B. K. P. Horn, “Robot Vision,” The MIT Press, 1986.

[82] 有本卓，“カルマン・フィルター”，産業図書，1977.

[83] R. Milanese, H. Wechsler, S. Gil, J. Bost and T. Pun, “Integration of Bottom-Up and Top-Down Cues for Visual Attention Using Non-Linear Relaxation,” Proc. of CVPR, pp.781–785, 1994.

[84] M. Jägersand, “Saliency Maps and Attention Selection in Scale and Spatial Coordinates: An Information Theoretic Approach,” Proc. of ICCV, pp.195–202, 1995.

[85] P. Viola and W. M. Wells III, “Alignment by Maximization of Mutual Information,” Proc. of ICCV, pp.16–23, 1995.

[86] P. Viola and W. M. Wells III, “Alignment by Maximization of Mutual Information,” International Journal of Computer Vision, vol.24, no.2, pp.137–154, 1997.

[87] H. Levkowitz, “Color Theory and Modeling for Computer Graphics, Visualization, and multimedia applications,” Kluwer Academic Publishers, 1997.

[88] L. Birnbaum, M. Brand and P. Cooper: “Looking for Trouble: Using Causal Semantics to Direct Focus of Attention”, Proc. of ICCV, pp.49–56, 1993.

[89] P. Whaite and F. P. Ferrie: "Active Exploration: Knowing when we're Wrong", Proc. of ICCV, pp.41–48, 1993.

[90] W.-S. Ching, P.-S. Toh, K.-L. Chan, M.-H. Er: "Robust Vergence with Concurrent Detection of Occlusion and Specular Highlights", Proc. of ICCV, pp.384–394, 1993.

[91] P. J. Dallos and R. W. Jones, "Learning Behavior of the Eye Fixation Control Systems," IEEE Trans. on AC, vol.8, 1963.

[92] A. T. Bahill and J. D. McDonald: "Model Emulates Human Smooth Pursuit System Producing Zero-Latency Target Tracking", Biological Cybernetics, 48, pp.213–222, 1983.

[93] デビッド・マー，乾 敏郎 安藤 広志 訳，"ビジョン"，産業図書.

[94] 樋渡 涓二，"眼球運動の個人差"，テレビジョン学会誌， vol 45，no.3，pp.384–386，1991.

[95] 顧，浅田，白井，"動き情報に基づくエッジセグメントの最適分割"，電子情報通信学会論文誌 D-II，vol.J76-D-II，no.8，pp.1544–1553，1993.

[96] K. Fukui, H. Nakai, Y. Kuno, "Multiple Object Tracking System with Three Level Continuous Process", IEEE Workshop on Applications of Computer Vision, pp.19–27, 1992.

[97] S. Yalamanchili, W. N. Martin and J. K. Aggarwal, "NOTE Extraction of Moving Object Descriptions via Differencing", Computer Graphics and Image Processing, 18, pp.188–201, 1982.

[98] T. J. Olson and D. J. Coombs, "Real-time Vergence Control for Binocular Robots", The University of Rochester Computer Science Department, Technical Report 348, 1990.

[99] 横澤一彦，"一目でわかること"，科学，vol.62，no.6，pp.356–362，1992.

[100] 吉田，豊田，佐藤，"視点移動機能を備えた視覚システム"，電子情報通信学会技術研究報告，PRU90-9，1990.

[101] 竹内義則，大西昇，杉江昇，"情報理論に基づいたアクティブビジョンシステム，" 電子情報学会論文誌 D-II，vol.J81-D-II，no.2，pp.323–330，1998.

[102] 財団法人日本色彩研究所編，"色彩ワンポイント 5 色彩と人間"，日本規格協会，1993