

MITにおけるコンピュータビジョン研究の現状

Computer Vision Research at MIT

長田 茂美

Shigemi Nagata

(株式会社 富士通研究所)

(FUJITU LABORATORIES LTD.)

池内 克史

Katsushi Ikeuchi

(電子技術総合研究所)

(ELECTROTRCHNICAL LAB.)

Abstract - This paper describes the computer vision research at MIT. Their work to date has concentrated primarily on the initial representation such as the primal sketch and reflectance maps, and the computation from them of depth, surface orientations. Their current emphasis is on the integration of the different sources of information, the analysis and representation of shape, the refinement and evaluation of the individual modules, the extension of their approach to deal with time varying images and moving objects. In this report we review their recent work on the analysis of edge detection, the measurement of visual motion, the refinement of stereo algorithms, the integration of surface maps, the shape description, and the object acquisition with photometric stereo.

1. まえがき

筆者らは1982年 9月～1983年 8月まで、MITの人工知能研究所で研究する機会を得た。人工知能研究所は大きく分けて五つの研究グループ (robotics and vision, expert system, learning and common-sense reasoning, natural language understanding, computer architecture) から成り、電気工学・計算機工学・数学・心理学・政治学科等に所属する研究者達が、オフィスおよびCHAOSNETと呼ばれるネットワークに結合された計算機を共用しながら研究を進めている。

MITにおけるコンピュータビジョンの研究は、この人工知能研究所が中心となって行われており、1975年に開始された DARPA (Defence Advanced Research Project Agency) の画像理解プロジェクトによって大きく推進された。これまでの研究は、プライマル・スケッチやレフレクタンス・マップ等の低レベル表現や、それらの表現からの奥行き情

報、表面の向きの抽出等に力を注いできたが、現在の研究は、異なった情報ソースの統合、形状の解析および表現、これまで開発してきた個々のモジュールの改善および評価、動画像処理への拡張、更に、これまでの成果の実時間ハードウェア・インプリメンテーションに重点を置いている。

以下、本報告ではエッジ検出の解析、視覚運動の測定、ステレオ・アルゴリズムの改善、表面マップの統合、形状の記述、照度差ステレオ (photometric stereo) を用いた対象物体の獲得に関する最近の成果を概観する。

2. エッジ検出の解析

MITにおけるエッジ検出に関する研究は、異なったサイズの $\nabla^2 G$ フィルタによって検出される画像のゼロ・クロッシング輪郭を使用してきた。実応用で使用されるエッジ検出器は、種々のノイズに対し強いものでなければならない。実時間ステレオ・マッチング・システムの開発を目指した

研究の中で、Nishihara は最近、 $\nabla^2 G$ コンボリューションとゼロ・クロッシング輪郭に対する画像ノイズの影響を検討した。このような標準的なエッジ検出法の発展とその信頼性の改善努力と並行して、現在、エッジ検出問題に対する新しいアプローチが追及されている。

Canny はエッジ検出と位置測定の正確な定式化から、最適なエッジ検出オペレータを決定する問題を検討した。¹⁾ 彼は最適形状はガウシアン的一次微分であることを見いだした。エッジ検出器の重要な性質は、正確に位置したエッジを生成し、エッジの誤分類の確率が低いものでなければならないということである。特に、オペレータは単一のエッジに対し複数の応答を生成すべきでない。

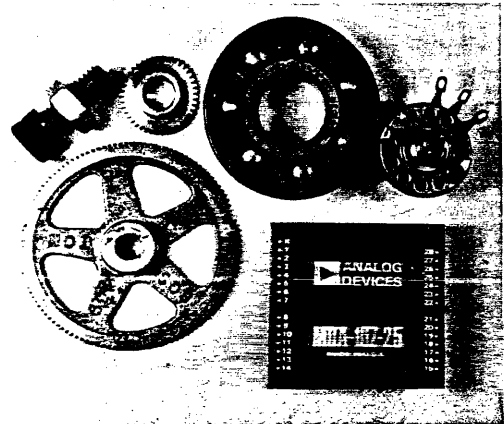
Canny の導出は次の三段階から成る。最初に、最適設計は線形オペレータのみに限定される。次に、その最適線形オペレータが、検出および位置測定の評価基準に関して最適である非線形な方法で結合される。最後に、非線形エッジ検出器からのエッジ点の出力が、輪郭のセグメントにエッジのタイプを表すパラメータを付与する線追跡手順により処理される。その結果得られるオペレータは、マイクロ・コードでLISPマシン上にインプリメントされ、後述する smoothed local symmetries および shape from contour に関する研究の前処理として使用されている。

線形オペレータはS/N比と位置測定の両者について直接最適化される。Canny は、これら二つに関係する不確定原理があり、ノイズのためにエッジが任意の正確さで、同時に検出、位置測定され得ないことを示す。この制限を達成するユニークなオペレータ形状（ほぼガウシアン的一次微分）がある。オペレータの幅は出力のS/N比と位置測定とのトレードオフを決定する。狭いオペレータはより良い位置測定を与えるがS/N比を落とす。画像のS/N比の変動を扱うために、数種類の幅のオペレータが使用される。許容できるS/N比を保存し、かつ、最適な位置測定を与えるアルゴリズムによって、数種類のオペレータが同一のエッジに反応する場所では、その一つが選択される。一次元のエッジ検出の定式化が二次元に拡張される時

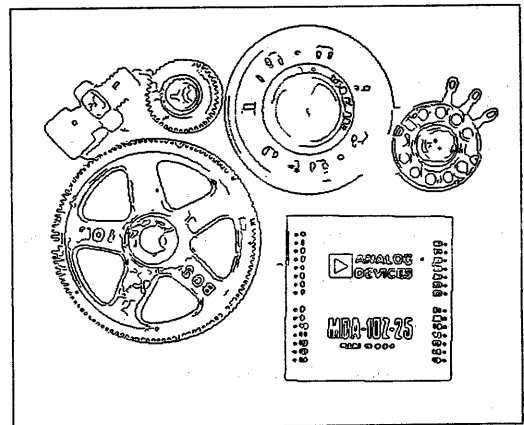
も、同一の最適化基準が用いられる。これは、そのノイズ評価およびエッジ検出がすべて独立に遂行される方向オペレータのシステムに導く。

このエッジ検出器は、現在Marr-Hildreth の $\nabla^2 G$ オペレータ²⁾やHaralickの二次微分方向検出器等の有名な検出器と比較・検討中である。

図1はCanny によって導出されたエッジ検出器の処理例を示す。



(a) 機械部品の実画像



(b) エッジ検出結果

図1. エッジ検出例

3. 動画の解析

視覚運動解析の分野で、Hildreth and Ullman はゼロ・クロッシングをベースにして動画から二次元の速度場を抽

出するアプローチを研究している。³⁾ この出発点は、図2に示すように、運動の初期検出を画像の $\nabla^2 G$ オペレータによるコンボルーション出力のゼロ・クロッシング位置で行う、Marr and Ullman の研究であった。初期の運動測定をゼロ・クロッシングで実行する理由は、ゼロ・クロッシングが明るさ変化の極値に対応し、最も信頼性のある測定を与えるということである。Hildreth and Ullman は、Marr and Ullmanの研究を拡張し、三次元物体の一般的運動から生じる投影二次元速度場を計算できるようにした。

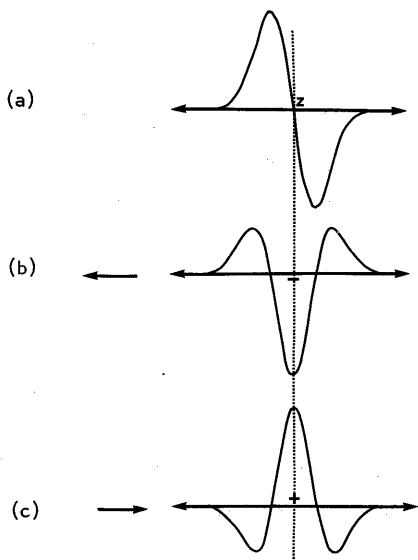


図2. Marr-Ullman のスキーマ

- (a) ステップ状の明るさ変化の $\nabla^2 G$ によるコンボルーション
- (b) 左方向への運動に対する時間的な明るさの微分
- (c) 右方向への運動に対する時間的な明るさの微分

口径問題 (aperture problem) によって、動画像の局所的な運動の測定は、ただ単にゼロ・クロッシング輪郭に垂直な方向の速度成分を与えるだけである。輪郭部分の速度場はアーク・レンクスを s を用いて $V(s)$ で表現できる。この時、図3に示すように、 $V(s)$ は輪郭線に垂直な成

分と水平な成分に分解される。

$$V(s) = v^{\perp}(s) u^{\perp}(s) + v^{\parallel}(s) u^{\parallel}(s)$$

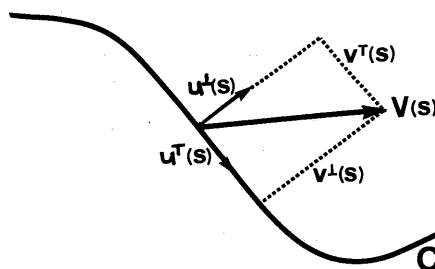


図3. 速度 $V(s)$ の水平・垂直成分への分解

但し、 $u^{\perp}(s)$ および $u^{\parallel}(s)$ は輪郭に対し垂直方向と接線方向の単位ベクトルであり、 $v^{\perp}(s)$ と $v^{\parallel}(s)$ は二つの成分の大きさである。上記表現において、第一項は動画像から直接測定されるが、第二項は測定できず、速度場 $V(s)$ を計算するために、復元されなければならない。

この復元の理論的問題は、 $V(s)$ を動画像において有効な情報から一意に決定できないということである。従って、ユニークな速度場を計算するためには、何らかの拘束条件が必要となる。彼等はオプティカル・フローの計算に関するHorn and Schunk の研究⁴⁾ から、滑らかさの拘束を引用し、使用する。通常、運動条件下では、物体は滑らかに変化する速度場を生成する。彼等はユニークな速度場を計算するために、動画像に一貫しており、かつ、最も変化の少ない速度場を検出する。

数学的な解析を通して、上述の滑らかさの拘束条件は、ユニークな速度場の解を保証する形で定式化されることがわかった。彼等は、まず速度場 $V(s)$ の局所的な変化を図4に示すように dV/ds で定義し、全体の輪郭線についての全変動を、この局所的な変化値の大きさ $|dV/ds|$ の積分によって定義する。次に、速度場の計算は、動画像に一貫した、輪郭線に沿った速度の全変動を最小化する速度場を探す。画像から得られる $v(s)$ の測定値と一致し、 $\int |dV/ds|^2 ds$ で与えられる全変動の測定値を最小化する速度場の存在が解析的に示される。

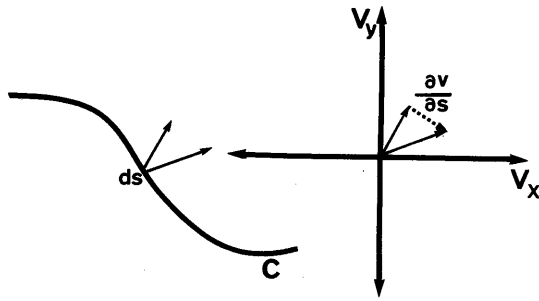


図4. 速度場の変動の測定値

要約すると、二次元速度場の計算は二段階から成る。(1)初期運動測定値がゼロ・クロッシング輪郭に沿って得られ、輪郭に垂直な速度成分を供給する。(2)測定値 $\int |dV/ds|^2 ds$ で与えられる全変動を最小化する二次元速度場 $V(s)$ を計算するために、局所的な運動測定値が輪郭に沿って積分される。このような定式化により、空間内で一般的運動を行っている剛体および非剛体についての二次元速度場が計算される。この計算は標準的最適化アルゴリズムを用いてインプリメントされた。

4. ステレオ・アルゴリズム

MIT における成果として、Marr-Poggio⁵⁾ のステレオ・アルゴリズムがある。これは、 $\nabla^2 G$ オペレータによりコンボリューションされた画像のゼロ・クロッシング輪郭を検出し、左右画像上におけるゼロ・クロッシング輪郭の位置から、視差を計算するというものである。

Nishihara はこのステレオをロボットの眼として実用化することを目的とし、高速かつノイズに強いステレオ・マッチャーであるプリズムステレオを開発した。基本的なマッチングは、同様にゼロ・クロッシング輪郭を基にして実行されるが、彼はノイズによる信号歪に対するマッチャーの感度を最小化する手法を研究し、ノイズに対する感度は、Marr and Poggio が視差レンジと解像度とのトレードオフを提案したのと同様な方法で、信頼性に対する感度とのトレードオフにより減少できることを見出した。

彼は、以前に開発した実時間コンボリューション・ハード

ウェア⁶⁾ 上に、これらの結果を用いてプロトタイプのマッチャーをインプリメントした。このシステムは画像を粗く分割し、各分割領域の奥行き情報として、分割領域内のゼロ・クロッシング輪郭における奥行き情報の平均値を生成するものである。一般に、分割領域内に多くのゼロ・クロッシングが存在する方が精度は向上するため、ランダムドット・パターンを対象物に投げ掛けて使用する。現在、10～20%のノイズ・レベルを持つ画像から、2～30秒で32～28分割の配列で表現される奥行き情報を生成する。LISPのマイクロコード化により、マッチング時間はかなり減少されるであろう。

このシステムは、マニピュレータが障害物を避けながら物体を移動させる、経路検出プログラムの入力として使用された。また、後述する照度差ステレオを用いて対象物体を獲得するハンド・アイ・システムにも結合された。

5. 表面マップの統合

コンピュータビジョンは、表面形状のリッチな記述の構築を必要とする。Marr and Nishihara の 2-1/2D スケッチやシーンにおける可視表面の観測者中心座標系による記述は、表面解析、そして究極的には対象認識への道程の重要な中間レベルの記述形式である。

Terzopoulos による現在の研究は、表面の視覚解析に関する四つの問題を調査した。その四つとは、(1)拘束条件の統合問題；(2)不連続性の問題；(3)内挿の問題；(4)計算効率の問題である。計算効率の研究は、表面形状に関する疎らな拘束条件からの滑らかな表面の内挿⁷⁾、効率的なマルチ・レベルの表面再構成アルゴリズムの開発 等の成果を産み出した。ここでは、拘束条件の統合についての最近の成果を述べる。

各視覚様式は表面形状を拘束する情報の異なったソースを構成する。ステレオや運動解析のようなプロセスは局所的な奥行き情報の拘束を生成するが、shape from shading, texture, contours のようなプロセスは局所的な面の向きの拘束を供給する。表面再構成の問題は、幾つかのソース

について、これらの疎らな拘束条件の統合を必要とする。

表面の再構成の問題は、物理モデル（拘束条件下で、薄い柔軟な板の均衡を記述する問題）で定式化された。それは次の板エネルギー関数を含む。

$$\mathcal{E}_p(v) = \iint_{\Omega} \frac{1}{2} (\Delta v)^2 - (1 - \sigma)(v_{xx}v_{yy} - v_{xy}^2) dx dy.$$

一般的な定式化では、種々の拘束条件が表面を内挿する板に及ぼす影響は、ペナルティ関数によって支配される。⁸⁾

奥行き拘束条件は次式で表現され、

$$\mathcal{E}_d(v) = \frac{1}{2} \sum_{(x_i, y_i) \in D} \beta_{(x_i, y_i)} [v(x_i, y_i) - d_{(x_i, y_i)}]^2,$$

面の向きの拘束条件は次式で表現される。

$$\mathcal{E}_o(v) = \frac{1}{2} \sum_{(x_i, y_i) \in P} \alpha_{p(x_i, y_i)} [v_x(x_i, y_i) - p_{(x_i, y_i)}]^2 + \frac{1}{2} \sum_{(x_i, y_i) \in Q} \alpha_{q(x_i, y_i)} [v_y(x_i, y_i) - q_{(x_i, y_i)}]^2,$$

従って、最小化されるべき全エネルギー関数は、次式で与えられる。

$$\mathcal{E}(v) = \mathcal{E}_p(v) + \mathcal{E}_d(v) + \mathcal{E}_o(v).$$

再構成面は、疎らな奥行き拘束条件あるいは面の向きの拘束条件の影響下で、薄い板表面を表す最小化された関数 $v = u(x, y)$ である。こういう方法で、種々のソースから生成される有効な拘束条件が、統合された形で採り入れられ、再構成面は有効情報の観点から最適なものとなる。

6. 形状記述

二次元および三次元形状の記述は認識にとって非常に重要である。Brady は smoothed local symmetries (SLS) と呼ぶ二次元形状の表現形式を導入した。smoothed local symmetries は、symmetric axis transform (SAT) と一般化円筒 (generalized cylinders) の概念の統合として展開された。⁹⁾

smoothed local symmetries の定義は次の三段階で成される。まず最初に、図5はlocal symmetryのジオメトリを示す。AとBは形状の輪郭線上の点である。点Aにおける線分BAと外向き法線ベクトル n_A とのなす角 α は、点Bにお

けるBAと内向き法線ベクトル n_B とのなす角と等しい。図5bは、一般にある点Aについてlocal symmetriesを形成する数点のBがあり得ることを説明する。十分な正確さで形状の輪郭線が抽出され、タンジェント角が計算されるという要請がlocal symmetryの定義において暗黙裡にある。彼はCannyによって開発されたエッジ検出器を使用する。

二番目に、local symmetriesの滑らかな軌跡を形成する。図6aはコーナを含む輪郭線の一部に対する結果である。その結果である候補軸は形状の輪郭および領域の一部を記述するための尤もらしい道である。彼は、この形状の一部分を軸のカバーと呼ぶ。

最後に、包摂 (subsumption) と呼ぶプロセスで、そのカバーが他の軸のカバーに包含される候補軸を消去する。図6bは矩形に対する候補軸を示す。斜めの軸は、global symmetriesである水平・垂直の軸に包摂される。しかし、その形状が正方形であるならば、その斜めの軸は包摂されない (図6c)。

図7は smoothed local symmetriesの幾つかの例を示す。smoothed local symmetries は図には描かれていないが、多くのパラメータによって記述される。最終的に、形状は smoothed local symmetriesが個々に計算されるサブ形状に分解される。軸は局所的な座標フレームの如く振る舞い、サブ形状の記述を結合することにより全体形状の記述の生成を拘束する。

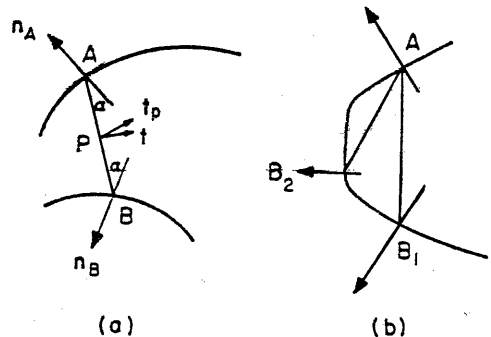


図5. (a) local symmetryのジオメトリ
(b) 点Aは点B1, B2 とlocal symmetriesを形成する

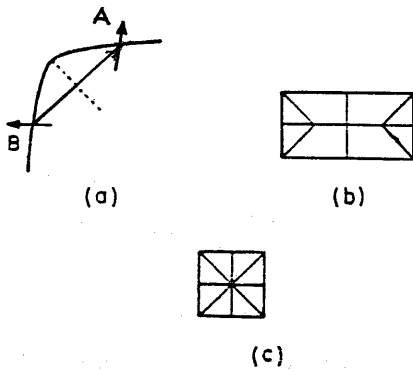


図6. (a) コーナを含む輪郭線の smoothed local symmetries
 (b) 矩形の候補軸
 (c) 正方形の smoothed local symmetries

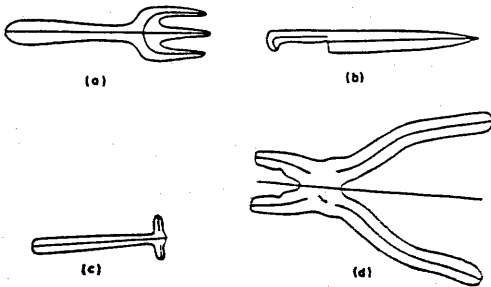


図7. smoothed local symmetries の例

smoothed local symmetries のパイロット・インプリメンテーションは、Brady により報告された。それは接線に対してある角度で入る線分が、その形状から脱する点を決定するための中間値低利に基づくアルゴリズムを繰り返して使用した。このようにして、繰り返しのよってある点での local symmetries が検出される。

最近、Brady and Asada は smoothed local symmetries の近似を計算するアルゴリズムを開発した。¹⁰⁾ 最初に、一連の特徴値が形状を包む輪郭線に対して計算される。特徴点は高い曲率率あるいは屈曲点であり、それらはエッジ検出に似たプロセスであるが輪郭線の方角（アーク・レングスの一次関数）に適用されるプロセスで検出される。次に、

第一段階で検出された特徴点を基にして、形状が直線と円弧による最適当てはめで近似される。彼等は、一定の曲率を持つ二つの輪郭線により生成される smoothed local symmetries を算出した。

最終的には smoothed local symmetries は認識および自動検査のための形状モデルのデータベースとの照合に使用される。

7. 対象獲得と shape from shading

Horn, Woodham and Silver によって開発された照度差ステレオは、同一シーンを光源の方向のみを変化させ、同一位置より写した複数枚の画像から、物体各点の面素の方向を得る手法である。この手法はレフレクタンス・マップに理論的根拠を置いている。¹¹⁾

一般に、被写体表面上の見掛けの明るさは光源の方向・観測者の方向・面素の方向・表面の材質の四つの要素に左右される。光源の方向・観測者の方向、更に材質が既知の場合には、見掛けの明るさは面素の方向を従属変数として、

$$E = R(p, q)$$

と表すことができる。(p, q) は観測者を基準とした座標系で表した面素の方向である。この E を (p, q) 平面にプロットしたものをレフレクタンス・マップと呼ぶ。分布 E は特に光源の方向に左右されることに注意を要する。

ある光源下で写した画像の画素で、見掛けの明るさ E1 を得たとする。この画素に対応する被写体上の面素の方向は、

$$E1 = R1(p, q) \quad (1)$$

を満たす (p, q) 平面上の軌跡上に存在する。第二の光源下で、同一の画素から E2 の明るさを得たとすると (p, q) は更に、

$$E2 = R2(p, q) \quad (2)$$

を満たすことがわかる。第一・第二の画像を取り込む際に、被写体とカメラの位置は変化してないので、同一画素は、(p, q) 平面で同一位置として表せる。よって式(1)と(2)の交点が面素の方向となる。また、実際は両式の非線形性のため、第三の光源下における画像が必要になる。

照度差ステレオは各画像間に視差が存在せず、マッチングが高速に行なえる。更に、高速化を図るため、参照表法を使用することが多い。このため、観測者と被写体、ならびに、光源と被写体との距離は物体の大きさに比して充分遠いとする。この仮定下では被写体の各面素において、光源ならびに観測者の方向は一定と見なされる。従って、同一のレフレクタンス・マップが適用できる。換言すれば、明るさの三つ組と面素の関係は、画面上の位置に不変である。故に、予め三つ組と面素の方向の関係を計算し、三つ組から面素の方向が参照できる表として、計算機内部に蓄えておくものである。

最近、Ikeuchi, Horn and Nagataは、この手法を困難な^{1 2)}ビン・ピッキングの問題に適用した。産業用ロボットの幅広い普及を阻害している原因は、正確に位置決めされていない物体を取り扱えないことである。彼等は、自動的にマニピュレータに積み上げられた物体の山から、一つずつ物体を把み上げることが指示するシステムを実現した。

このシステムは、まず照度差ステレオを用いて、シーンのいがり表現像を得、得られたいがり表現像により、領域分割を行った後、一定の評価関数に従って、把むべき最適な領域を得る。次に、この領域内の面素分布から拡張ガウス像を得、これと計算機内部の拡張ガウス像とを照合することにより、物体の空間における姿勢を検出する。最後に、決定された姿勢から、把み点 (x, y) 位置を決定し、その情報をマニピュレータに渡す。この情報が渡されると、マニピュレータは、把み点とカメラを結ぶ視線に沿ってグリッパを降下させ、近接センサによって把み点の奥行き情報を決定し、得られた位置・姿勢情報を基にして物体を把み上げる。

図8のドーナツ形状物体を対象とした実験結果を図9～図13に示す。

更に、Ikeuchi らはこのシステムとNishihara のプリズムステレオとを結合し、プリズムステレオから得られる物体表面までの奥行き情報を基にした衝突回避の実験を行った。

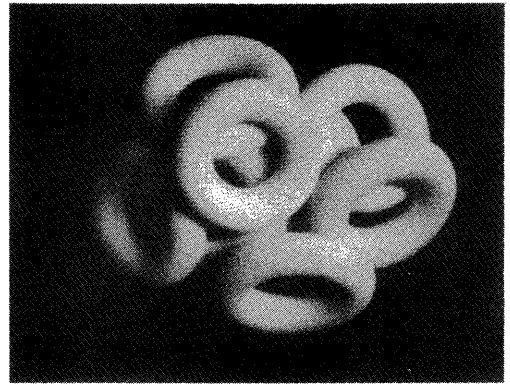


図8. 入力画像 (ドーナツ状曲面物体の山)

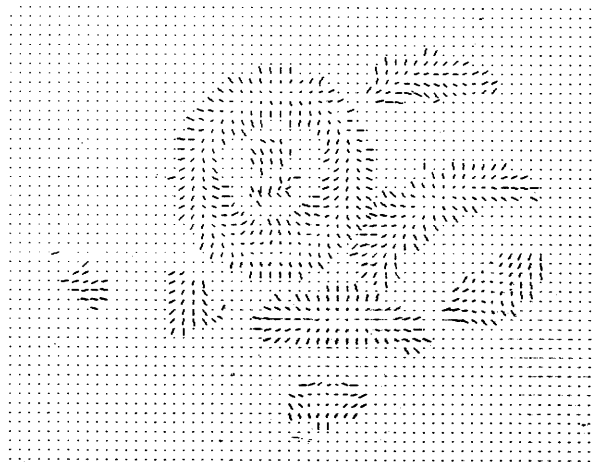


図9. 照度差ステレオにより得られたいがり表現像

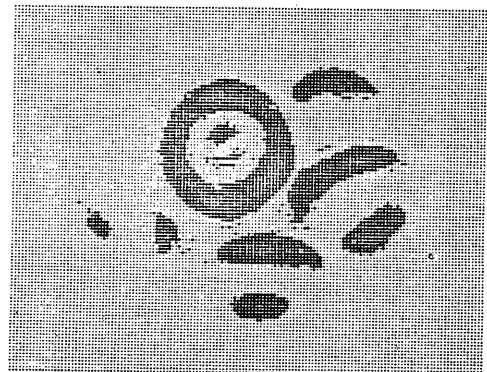


図10. 領域分割の結果

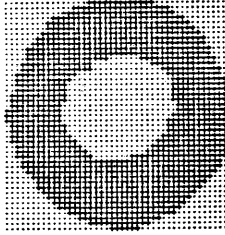


図11. 目標領域の決定

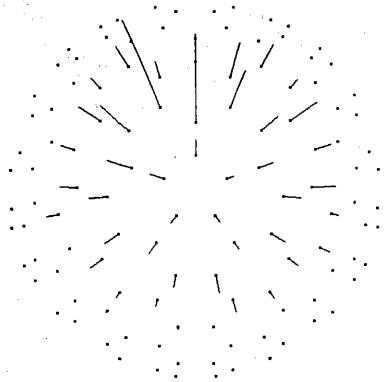


図12. 正規化拡張ガウス像

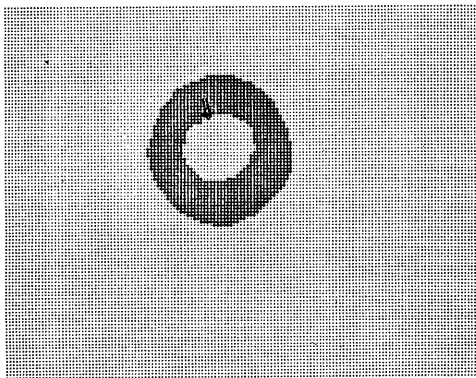


図13. 目標点

7. むすび

本稿では筆者らが MIT 滞在中に行われたコンピュータビジョン研究を概観した。最後に、このような機会を与えて下さった富士通研・棚橋部長、吉田室長、電総研・石井前部長、若松部長、白井室長、MIT・Winston 所長に深謝致します。

参考文献

- 1) Canny, J. F. "A variational approach to edge detection", AAAI Conf., Sept., 1983.
- 2) Marr, D., Hildreth, E. "Theory of edge detection", Proc. R. Soc. Lond. B, 207, 187-217, 1980.
- 3) Hildreth, E. C., Ullman, S. "The measurement of visual motion", MIT A.I. Memo 699, 1982.
- 4) Horn, B. K. P., Schunk, B. G. "Determining optical flow", Art. Intell., 17, 1981.
- 5) Marr, D., Poggio, T. "A theory of human stereo vision", Proc. R. Soc. Lond. B, 204, 301-328, 1979.
- 6) Nishihara, H. K., Larson, N. G. "Towards a real-time implementation of the Marr-Poggio stereo matcher", Proc. DARPA Image Understanding Workshop, April, 1981.
- 7) Grimson, W. E. L. From images to surfaces : a computational study of the human early visual system. MIT Press, Cambridge, 1981.
- 8) Terzopoulos, D. "The role of constraints and discontinuities in visible-surface reconstruction", Proc. Int. Jt. Conf. Art. Intell., Karlsruhe, 1983.
- 9) Brady, J. M. "Criteria for representation of shape", In Human & Machine Vision, Rosenfeld and Beck, eds., Academic Press, 1982.
- 10) Brady, M., Asada, H. "Shape description for inspection", Proc. 1st Int. Symp. Robotics Res., August, 1983.
- 11) Horn, B. K. P., Woodham, R. J., and Silver W. M. "Determining shape and reflectance using multiple images", MIT A.I. Memo 490, 1978.
- 12) Ikeuchi, K., Horn, B. K. P., Nagata, S., Callahan, T., Feingold, O. "Picking up an object from a pile of objects", Proc. 1st Int. Symp. Robotics Res., August, 1983.