画像を用いた三次元計測技術の動向

大 田 友 一 筑波大学電子情報工学系

特 集 解 説

1. ま え が き

視覚は、人間が外界の情報を得る上で最も重要な役割をもつ。このような視覚機能を機械で実現することがコンピュータビジョンと呼ばれる研究分野の目標である⁽¹⁾。特に最近では、移動ロボットの視覚という観点から、三次元の世界を対象とする研究が熱心に行われている⁽²⁾。

いうまでもなく、眼またはカメラにより得られるのは二次元の画像データであり、我々をとりまく三次元の世界が射影変換により縮退して記録されたものにすぎない。従って、1枚の画像データから元の三次元世界を復元することは数学的には不可能である。それにもかかわらず、我々は、視覚により外界の様子を知り、その中で行動することができ。これは、二つの眼を用いた両眼立体視、自分自身の動きや対象物の動きによる移動立体視、更には、我々をとりまく世界に関する知識を用いた推測などにより、総合的に三次元の情報を計測し復元しているものと考えられる。

本特集の、以下に続く3編の解説では、このような画像を用いた三次元計測を機械で実現する手法について、単眼視による方法、両眼立体視による方法、光源を制御する方法の三つに分類して紹介している。本稿では、これらの手法全体に対する見通しをよくすることを目的とし、各手法の基礎となる考え方や、相互の関連について述べることにする。

2. 単眼視による方法

人間は、1枚の写真からでも元の世界の三次元情報を推測することができる。これは、意識するか否かわからず、我々が住んでいる世界に関する一般的な

図 1 単眼視における共線条件

知識に基づいて、対象物に関する仮定を設定し、その仮定のもとに写真から得られる情報を解釈して三次元の知覚を得ているものと考えられる。

図1に示したように、レンズ中心 O をもつカメラにより撮像された画像が与える情報は、 O から画素 P_i に向かうベクトルの延長上に P_i に対応するシーン中の点 P_i が存在するということだけである。点 P_i の位置には1の自由度がある。 P_i を面素と考えれば、面素の傾きに関しても2の自由度をもつ。更に、 P_i が剛体上にあるとし、剛体の姿勢を知ろうとすれば、視線まわりの回転の自由度がもう一つ増加し、合計4となる。剛体の位置と姿勢の自由度は本来6であるから、 P_i を観測することにより、自由度を2だけ減らすことができたことになる。このように、射影の逆変換という観点のみからは、画像から三次元の情報はほとんどわからないといってよい。

単眼視による三次元計測手法の研究は、人間が通常用いているであろうと考えられる世界に関する知識の一つ一つを明確に切り分け、その特性を明らかにすることにより計算機上にアルゴリズムとして実現することを目標としている。

この知識として現在考えられているものは、位置に関するもの、明暗に関するもの、形に関するものに大別できる。

位置に関する知識とは、対象点が既知の面上に載っ

"Trends in the Techniques to Obtain Three Dimensional Information Using Image Data" By Yuichi Ohta (Institute of Information Sciences and Electronics, University of Tsukuba).

ているとか、幾つかの点の相対配置がわかっているような場合である。この場合に、最も基本となる幾何学的条件は『レンズ中心 O と画像上の点 P_i とシーン中の点 P_j は同一直線上になければならない (共線条件)』というものであり、1 個の点に付き三次元座標に関する方程式が 2 個得られる。先にも述べたように、三次元空間中に点を固定する自由度は 3、剛体を固定する自由度は 6 であるから、未知数の数に比べて方程式の数が不足である。対象点が既知の面上に載っていると仮定できる場合には、方程式がもう 1 個得られて、自由度を 0 にできる。幾つかの点の相対配置が既知であるとは、剛体上の数点が観測できる場合であり、未知数は 6 個のままで、点の数が増えただけ方程式が多く得られ、解けるようになる。

明暗に関する知識を用いる場合⁽³⁾には、物体の反射特性に基づいて方程式がたてられる。この場合にも 1 点を独立に観測した場合には、未知数の数が方程式の数を上回るので、対象物の形状に対してなんらかの仮定を設定したり、光源を切り替えて複数の観測を行ったりして、方程式の数を増やすことが必要となる。

形に関する知識を用いる場合にも、同様の考え方が適用できる。対象物体の形に関して設定する条件としては、画像上で観測している直線が三次元空間中の 2 平面の交線であることが既知⁽⁴⁾、2 直線の三次元空間中での交角が既知⁽⁵⁾、2 直線の平行性が既知⁽⁶⁾などがあるが、いずれの場合にも、その条件に基づいて復元すべき三次元情報に関する方程式が得られる。

このように、単眼視により三次元計測を行う場合には、一つ一つではあいまい性の残る観測事象を、いかに組み合わせて三次元情報を復元するかが重要な点となる。更に、現実的な観点からは、方程式をたてるために導入する仮定は、通常容易に成立し得るものでなければならない。

3. 多重画像による方法

人間は、二つの眼により観測した 2 枚の画像を用いた両眼立体視により対象物体の凹凸を知ることができる。更には、移動する対象物を異なった時刻に観測したり、自らが移動して観測した複数の画像から三次元の情報を得ることもできる。これを、両眼立体視に対して「移動立体視」と呼んでいる。これらの、多重画像による方法では、複数枚の画像中で、互いに対応する点、すなわち、シーン中ではもともと同じであった点を抽出することにより、三次元情報の復元を行うのが基本である。

多重画像に基づく方式を考える場合に基本となる幾

図2 2枚の画像間の共面条件

何学的関係を図2に示す。「 O_1, O_2 は各々カメラ 1, カメラ 2 のレンズ中心、 P_1, P_2 はシーン中の点 P_i の像であるとき、 O_1, O_2, P_1, P_2 の 4 点は同一平面上になければならない」。これを共面条件と呼んでいる。

多重画像により三次元情報を求める方式のうち、原理的に最も簡単なものは、互いの位置関係が既知の 2 枚の画像を用いる場合、すなわち、両眼立体視の場合である。この場合には、図2で O_1, O_2 が固定され、2 枚の画像上で互いに対応する 2 点 P_1, P_2 を求めれば、 $O_1 P_1$ と $O_2 P_2$ の交点としてシーン中の点 P_i の座標が求められる。これは、三角測量の原理として古くから知られている。回転台上の物体を固定カメラで観測する場合のように、カメラが固定されていて、物体が既知の距離と角度だけ移動する場合も同様である。

両眼立体視により、三次元情報を得る原理自体はこのように簡単なものであるが、実際に適用する場合には、2 枚の画像上で対応点を捜す対応探索が問題になり、研究の一つの焦点となっている⁽²⁾。

この場合、上述の共面条件により、 P_1 に対応する点 P_2 は O_1, O_2, P_1 で決定される平面上になければならず、この平面と画像面 2 との交線上に拘束される。 P_1 についても同様のことがいえ、結局、対応点の対は 2 枚の画像上の一対の直線上に存在することになる。この直線対のことをエピポーラ線 (epipolar line) と呼んでいる。このように、両眼立体視では、2 枚の画像上の対応点の探索は一対のエピポーラ線上で行えばよい。

これに対して、移動するカメラや、互いの位置関係が未知の 2 台のカメラにより観測した画像上で求めた対応点により三次元情報を得る問題では、カメラ相互の位置関係、すなわち、カメラ座標系間の変換行列を求めることが必要である。これは、未知の移動を行う物体を固定カメラから観測し、物体の移動ベクトル・

回転パラメータを求める問題と等価であり、結局、単眼視の場合と同様に、未知パラメータの数と方程式の数に関する考察が必要になる。これらの問題は、両眼立体視におけるカメラのキャリブレーションの問題⁽⁷⁾や、動画解析による三次元情報の獲得の問題⁽⁸⁾として研究されている。

図2で、カメラ1を基準にとった場合のカメラ2の姿勢に関する自由度は、 O_2 の位置に関して3、 O_2 を中心とする回転に関して3の合計6である。しかし、シーン中の点間の距離と基線長 O_1O_2 とは比例の関係にあり、シーン中の絶対距離が不明の場合には、基線長は決定できず、ベクトル O_1O_2 の向きが決定できるのみである。従って、求め得るパラメータ数は5である。上述の共面条件を用いれば、1組の対応点対について1個の方程式が得られる。求めるべきパラメータは5個なので、5組の対応点対があれば自由度は0となる。

カメラの配置が未知だが制限がある場合には対応点対の個数は少なくともよい。カメラ間に回転がない場合には、レンズ中心の移動ベクトルの方向だけが未知なので求めるべきパラメータは2個のみである。従って、2組の対応点対があればよい。これは、固定した1台のカメラから平行移動する物体を観測するとみなせば別の解釈も可能である。すなわち、2個の点の軌跡は平行線を描き無限遠点を形成する。単眼視により平行線を観測した場合と同様に、無限遠点の画像上での座標と、平行線の方法、すなわち移動ベクトルの方向は1対1に対応するので、移動ベクトルの方向が求まることになる。この場合の無限遠点を、特に「動きの無限遠点」と呼んでいる。

物体やカメラの微小な移動によって生じる2枚の画像間の変化を、像の流れ、オプティカルフロー(optical flow)としてとらえ、上述のような対応点対を求めずに、直接、移動ベクトルを推定することも試みられている。画面上の点 (x, y) の時刻 t における明るさを $I(x, y, t)$ とし、物体自身の明るさは変化しないものとすれば、

$$I_x \cdot u + I_y \cdot v + I_t = 0$$

なる関係が成立する。 (u, v) は、点 (x, y) の画像上での移動ベクトルを表わす。上式は、 (u, v) が点 (x, y) を原点とする座標系で点 (x, y) における明るさの空間こう配に直交する1本の直線上に拘束されることを表わす。すなわち、固定物体を移動カメラから観測する場合を考えれば、図3に示したように、時刻 t から単位時間後に観測される画像上での、点 (x, y) の対応点 $(x+u, y+v)$ はある直線上に存在しなければな

図3 オプティカルフローによる拘束の幾何学的解釈

らない。カメラの移動パラメータを未知数としてエピソードを設定すれば、2直線の交点として対応点が固定され、三角測量の原理によりシーン中に点が固定できる。すなわち、上式は未知の移動パラメータと画像上の明るさの変化に関する条件とをシーン中の点の座標に関する条件に読み替える働きをもつといえる。従って、対象物体の構造に関して何らかの制約を導入して方程式をたてることができる。この方式の問題点は、明るさの微分値を用いているので精度が上がり難いことであり、一般の三次元計測に適用するのはあまり容易ではない。

4. 光源を制御する方法

単眼視による方法では、それ単独では対象物体の三次元パラメータを完全には固定できない場合が多い。また、多重画像による方法では、画像間で互に対応する点を探す問題が困難である。これらの問題を回避し、実用的に三次元情報を得るために、撮像する環境を制御することが考えられる。この結果として、アルゴリズムが単純化されハードウェア化が容易になるので、実用性は高い。

例えば、単眼視により明暗を利用する手法においては、1枚の画像を観測するだけでは、各画素に対応する画素の向きは1になるのみで、固定できない。2個の光源を順次切り替えて撮像した2枚の画像を用いれば、自由度は0になるが、一義には決まらない。3個の光源を用いれば一義に決定できる。これをホトメトリックステレオ (photometric stereo) 法⁽⁹⁾と呼んでいる。ホトメトリックステレオ法は、「ステレオ」という名前は付いているが、三次元情報を得る原理は単眼視によるものである。

両眼立体視におけるカメラの一方を光源に置き換え

ることにより対応点探索の問題を回避することを狙った手法として、光投影法、または光切断法と呼ばれる手法がある⁽¹⁰⁾。これは、一方のカメラの視線を光線に置き換えたものと考えてよく、三次元情報を得るための幾何学的関係は両眼立体視の場合と全く同じである。しかし、光源から照射された光線により、対象とするシーン中にマークを付けるので、カメラにより撮像した画像中で、その点をユニークに検出することは容易である。

光源としてスポット光を用いる場合には、空間を走査するのに時間がかかるので、スリット光によりシーンを線状に照射するのが一般的である。スリット光を用いた場合でも、それを撮像した画像のエピポーラ線上で光っているのは一点だけなので対応のあいまいさは生じない。1枚のスリット光しか用いない場合には、エピポーラ線に沿う標準化点の数だけスリットを移動し、画像を入力する必要がある。入力画像数を減らすために、各スリット光に色や符号で印を付け、同時に複数照射したスリット光を識別する方式も考えられている。

光切断法では、光源からの光線の方向を制御・測定できることがポイントである。光源から細い縞模様を照射し、カメラ側に置いた縞模様との干渉により生ずる差の周波数を観測するモアレ法は、個々の光線を識別して照射方向を知ることではできないので、原理としては単眼視に属する。

画像から距離情報を得るという枠組みからははずれるが、光源から発したレーザ光が対象物体で反射して戻ってくるまでの時間を計測することにより、距離を求める手法を time-of-flight 法と呼んでいる。レーザパルスを照射し、遅れ時間を直接計時する方式と、変調したレーザ光の位相のずれから遅れ時間を計測する方式とがある⁽¹¹⁾。

この方式は測距装置にも用いられており、「点」としての距離測定は高精度で実行できる。しかし、距離画像を得るというように、「面」としての距離測定を行う必要がある場合には、測定速度と精度との兼ね合いが問題になる。最近では、距離画像を比較的高速に得ることができる装置も開発されている。アメリカの DARPA (Defence Advanced Research Projects Agency) が推進している ALV (Autonomous Land Vehicle) プロジェクトにおいて採用している ERIM 社のレーザレンジファインダ⁽¹²⁾ では、64 行 256 けたの距離画像を毎秒 2 枚の速度で得ることができる。測定誤差は、50 フィートの距離で 0.4 フィート程度である。

表 1 各方式の比較

	単眼視	両眼立体視	光源制御	time-of-flight
適用対象	×	△	○	○
適用環境	○	○	×	△
共存性	○	○	×	×
観測時間	○	○	△	×～△
処理時間	×～△	×～△	○	○
距離精度	×～△	○	○	○～◎

5. 各方式の比較

以上に述べた単眼視、両眼立体視、光源制御、time-of-flight の 4 方式の比較を表 1 に示す。但し、単眼視、両眼立体視、光源制御の各手法とも、種々のアルゴリズムが存在するので、厳密な比較は困難であり、大体の傾向を示したものに過ぎない。また、time-of-flight 方式は、DARPA の ALV で用いている ERIM のレンジファインダを想定して比較を行った。

表 1 の適用対象と適用環境はそのセンサの汎用性を示すものである。ある決められた仮定を満たす対象にのみ有効な、単眼視による方法が適用対象という点では最も劣る。両眼立体視では、シーン中で特徴点として検出できる箇所の距離情報が求まるのみであるのに対し、能動的な光源を用いる方式ではすべての点で距離情報を得ることができる。しかし、能動光源を用いる方式は、一般的に、屋外などではかの光源が存在する場合や、遠距離の対象に対する使用が困難である。上述の ERIM センサではこの点はかなり改善されており、屋外での使用も可能である。

共存性とは、2 個以上の同種のセンサが共存可能かどうかを示す。これは、同種のセンサを搭載した多数の移動ロボットが接近して存在する場合には必須の条件である。能動的な光源を用いる手法では、互いに干渉する可能性がある。

観測時間とは、撮像に要する時間であり、移動物体の観測には十分に短いことが要求される。単眼視、両眼立体視では、テレビカメラの 1 画面の撮像時間でよいのに対し、能動光源を用いる場合にはスポットの走査が必要となり、原理的に、ある瞬間の情報を固定することが困難である。

一方、観測データから距離を求めるための処理をほとんど必要としない光源制御法・time-of-flight 法では、処理時間の点では非常に有利である。いうまでもなく、1 枚の距離画像を得るのに要する時間は、観測時間と処理時間の和である。

距離精度に関しては、距離情報が直接に得られないことが多い単眼視が最も劣ると考えられる。両眼立体視と光切断法は測距原理は同じである。時間を測定す

ることにより距離を得る time-of-flight 法が原理的には最も優れていると考えられるが、測定時間と精度の兼ね合いがある。

6. あとがき

本稿では、画像から三次元情報を得るための各種方式について、各方式の基礎となる考え方や、相互の関連に重点において述べた。

表1からもわかるように、いずれの方式も万能ではない。実用性という観点からは、現時点では能動光源を用いる方式に分がある。しかし、共存性や観測時間の問題から、これを通常の環境下で移動するロボットや自律走行車の視覚として利用することには困難があり、処理時間や適用対象に関する問題を解決して、単眼視や両眼立体視による手法の実用化を図ることが必要であろう。

処理時間の問題に関しては、ハードウェア技術の急速な進歩により、専用のプロセッサを作成することが容易になっており、いずれ解決されるものと考えられる⁽¹³⁾。

個々のアルゴリズムの適用対象が限定されているという問題は、時と状況に応じて有効なアルゴリズムを選択し、使い分けることにより解決しなければならない。このためには、各種の方式のセンサを統合した統合化センサシステムの開発が必要であり、センサ統合

あるいはセンサ融合 (sensor fusion) と称して研究が進められている⁽¹⁴⁾。(昭和62年4月10日受付)

文 献

- (1) 辻:「コンピュータビジョンへの道」情報処理 24, 12, 1414 (昭 58-12)
- (2) 大田・池内:「3次元情報の再構成とモデル化」O plus E, 71, 86 (昭 60-10)
- (3) B. K. P. Horn: "Obtaining shape from shading information" *Artificial Intelligence* 8, 201 (1977)
- (4) A. K. Mackworth: "Interpreting pictures of polyhedral scenes" *ibid.* 4, 121 (1973)
- (5) T. Kanade: "Recovery of the three-dimensional shape of an object from a single view" *ibid.* 17, 409 (1981)
- (6) J. Kender: "Shape from texture" *Proc. 6th Int. Joint Conf. on Artificial Intelligence* p. 475 (1979)
- (7) 日本写真測量学会(編):「写真による三次元測定」(昭 58) 共立出版
- (8) 北橋・浅田:「動画像処理」O plus E, 69, 72 (昭 60-8)
- (9) R. J. Woodham: "Photometric method for determining surface orientation from multiple images" *Optical Engineering* 19, 1, 139 (1978)
- (10) 大島・高野:「3次元物体認識システムのための特殊ハードウェア」電総研彙報 37, 5, 493 (昭 48)
- (11) R. A. Jarvis: "A Perspective on range finding techniques for computer vision" *IEEE Trans. Patt. Anal. Mach. Intel. PAMI-5*, 2, 122 (1983)
- (12) M. Hebert, et al.: "Outdoor scene analysis using range data" *Proc. Int. Conf. Robotics and Automation* p. 1426 (1986)
- (13) 大田, 他:「ステレオ画像における区間対応法の高速ハードウェア処理」情報処理学会論文誌 28, 2 (昭 62-2)
- (14) S. A. Shafer, et al.: "An architecture for sensor fusion in a mobile robot" *Proc. Int. Conf. Robotics and Automation* p. 2002 (1986)