解説

物理則に基づくコンピュータビジョンとその展望

Physics based Vision and its Future

和田 俊 和* *京都大学大学院情報学研究科

Toshikazu Wada* *Graduate School of Informatics, Kyoto University

1. はじめに

今回の特集記事執筆に当たり、編集委員会から、現実世界の光学的／幾何学的な拘束を利用して三次元世界の有用な情報を取り出そうとする数理的なアプローチ（“shape from X” など）について今後も王道でありつづけられるのかについて記事を書くよう依頼があった。shape from X のすべてについて書くには紙数が足りないため、「数理的」というよりも、「物理的」法則に基づく画像からの三次元形状復元手法、すなわち Physics based Vision に分類される shape from X について簡単に述べ、それらの問題点や可能性について検討することにする。

まず、shape from X によって復元しようとする三次元情報には次の 2 種類があることを明記しておかなければならない。

Depth 奥行き（距離）情報を含む三次元情報

Shape 奥行き情報を含まず、物体表面の傾きのみから成る三次元情報

これは、本来 depth from X とすべきところを語呂の良さから shape from X としてしまったものが多く存在するために生じた混乱である。本稿では、これらを区別して扱い、両者を特に区別しない場合には、単に「三次元情報」と表記することにする。

次に、復元に利用される visual cue を光学的情報と、幾何学的情報の二つに分類し、それぞれが形状復元に利用される際の制約条件について述べる。

2. 光学的制約に基づく復元

光学的制約に基づく三次元情報の復元では、画像の輝度値を手がかりとして形状復元が行われる。この場合、観測画像は、光源、物体、カメラの影響を受けるため、これら三つのうちの、いくつかは既知であるという条件を定めな

図1 物体表面における光の反射

れば復元問題を解くことができない。この条件設定によって、以下のように数種類の復元法が考えられる[†]。

2.1 陰影

これは、照明の入射方向と物体表面の法線がなす角の空間的变化によって生じる陰影パターンから、形状を復元する手法[1]であり、いわゆる「影」によって生じる陰影パターンは解析対象ではない。この問題を解くためには、物体表面の（1）反射率分布、（2）反射特性、照明光の（3）照射方向、（4）強度分布、が既知でなければならない。

最も簡単な問題設定では、（1）反射率一様、（2）物体表面は完全拡散面、（3）平行照明光、（4）照明光の強度分布は一様、と仮定される。以下、この最も簡単なケースについて説明する。

図1に示すように、照明光強度を I_s 、物体表面の反射率を ρ 、照明方向に対する物体表面のなす角を θ とすると、反射光強度 I_o は、

$$I_o = \rho I_s \cos \theta \quad (1)$$

と表される。反射光強度 I_o と画素値 I_p が $I_p = \alpha I_o$ (α は定数) という関係にあるとき、 $\alpha \rho I_s$ が既知であれば、 $\cos \theta$ すなわち θ の値が画素値から直接求められる^{††}。しかし、これだけでは図2(a)に示すように物体表面の法線方向は決定できないため、以下に示すような拘束条件を利用して物体表面の法線ベクトルを求める手法が知られている。

原稿受付 2002 年 3 月 18 日

キーワード: Physics based Vision, Shape from Shading, Depth from Focus/Defocus

*〒 606-8501 京都市左京区吉田本町

*Sakyo-ku, Kyoto-shi, Kyoto

[†]shape from texture は物体表面上の模様の規則性に基づく shape の復元法であり、物理現象に基づかない手法であるため、本稿では割愛する。

^{††} $\theta = 0$ となっている点が撮影されていることが仮定できれば、その点は最大輝度値を有するはずであり、その輝度値を $\alpha \rho I_s$ と見なせばよい。

図2 法線方向の曖昧さ

図3 薄レンズモデルにおける像のボケ

- 複数光源の下での複数の陰影パターンを用いて法線方向を決定する(図2(b)参照: Photometric Stereo法[2]).
- 見開きにした書籍表面のように物体の断面形状が一定であると仮定できる場合、その条件を用いる[3][6].
- 物体形状に関する滑らかさの拘束条件を与える[4].

以上のケース以外にも、完全拡散面ではない場合や、物体表面で相互反射が起きる場合[5][6]、光源が近接しているため物体までの距離によって照明光が減衰する場合[6][†]、などについても多くの研究がある。

いずれの場合も、対象および光環境に関する詳細な情報が必要であるため、この手法が単独で利用できる用途はかなり限定されているが、物体表面に模様がないう場面で使用できるのが、ほかの手法にはないこの手法の特長である。

2.2 Focus/Defocus

これは、カメラのレンズと撮像面間の距離(フォーカス)を変化させることによって生じる画像のボケを利用した距離計測法であり、以下の2手法に大別される。

- フォーカスを変化させながら、ボケが生じない「合焦フォーカス」の値を求める depth from focus 法[7][8]
- 数枚のボケた画像から、合焦フォーカス値を推定する depth from defocus 法[9][10]

[†]この場合、shapeではなくdepthが求められる。

^{††}ボケた画像を撮影した際のfocus値が合焦focus値と大きく異なる場合には距離推定精度が低下するという問題点がある。

合焦フォーカス値はカメラから物体までの距離に1対1に対応しており、図3のようにカメラから対象までの距離を l 、レンズの焦点距離を f 、合焦フォーカス値を v とすると、次式が成り立つ。

$$\frac{1}{f} = \frac{1}{v} + \frac{1}{l} \quad (2)$$

この関係を、事前に求めておけば、合焦フォーカス値から対象までの距離を求めることができる。

depth from focus 法は、フォーカスを連続的に変化させて画像の鮮明さの評価値が最大になる値を合焦フォーカス値とする方法であり、計測に時間がかかるという問題点がある。これに対して、depth from defocus は、数枚のボケた画像とそれらのフォーカス値から、合焦フォーカス値を求めるため計測に要する時間は少ない^{††}。いずれの手法も、フォーカス位置によるボケの変化が大きくなるようにカメラの「被写界焦点深度」を浅く設定しておく必要がある。これは、口径が大きく焦点距離が長いレンズを用いて、絞りを開放にすることによって実現できる。また、フォーカス値の変化による画像の拡大縮小が起きるため、telecentric 光学系を用いたり、画像の拡大縮小による補正を行うことが必要である。これらの手法は原理的に模様のない物体に対する距離計測は行えない。

3. 幾何学的制約に基づく復元

幾何学的制約に基づく三次元形状復元では、物体とカメラ間に成立する幾何学的な関係を利用して、物体の三次元情報を復元する方法である。この場合、物体あるいはカメラが移動した際に観測される

- 物体表面上の点の移動ベクトル
- 物体の輪郭形状の変化

など画像上で観測可能な情報を手がかりとして形状復元を行う。前者は、Structure from Motion と呼ばれ、移動前後のカメラを複数のカメラと見なせば、ステレオ視もこの範疇に分類される。また、いわゆる因子分解法などもこの範疇に分類される。後者は、Shape from Contour と呼ばれ、拡大解釈をすれば視体積交差法 (Volume Intersection) などと同じ visual cue を用いている点ではこの範疇に含まれると考えてよいだろう。

このように、視点移動による見え方の変化のうち、何に着目するかで、幾何学的手法は二つに大別することができる。これらはさらに、カメラの運動(複数カメラの場合にはカメラ間の位置関係)が既知の場合と未知の場合、対象の運動が既知の場合と未知の場合、対応点が既知と未知の場合、などに細分化されるが、これ以上の詳細な説明は、ここでは省略する。

4. 光学的情報から幾何学的情報へ

幾何学的制約に基づく形状復元において、手がかりとなる情報はやはり、移動ベクトルや、対応点などの幾何学的情報である。しかし、このような場合でも、幾何学的情報を画像から抽出する際には、必ずどこかで光学的情報から幾何学的情報への変換が行われる。このような変換のうち、分かりやすい例を以下に紹介する。

時刻 t において撮影された画像を $I(x, y, t)$ と表す。時刻 $t + \Delta t$ において、点 (x, y) が点 $(x + u, y + v)$ に移動し、そこで輝度値の変化がまったく起きなければ、次式が成立する。

$$I(x, y, t) = I(x + u, y + v, t + \Delta t) \quad (3)$$

この式の左辺を Taylor 展開し、一次の項までで近似することにより、移動ベクトルと輝度値の時空間的な変化の関係式が得られる [11]。

$$\frac{\partial}{\partial t} I(x, y, t) + \frac{\partial}{\partial x} I(x, y, t)u + \frac{\partial}{\partial y} I(x, y, t)v = 0 \quad (4)$$

これは、連続時刻で撮影された画像間で近似的に成立する光学的情報と幾何学的情報の間の制約式であり、

輝度勾配法に基づく optical flow の計算において利用されている。光学的情報から幾何学的情報への変換はこれ以外にも多くある。例えば、2 章で述べた例では、

- shape from shading では陰影から法線と光源方向のなす角を求めている。
- depth from focus/defocus ではボケの量から、合焦 focus 値を求めている。

ということを行っていたことになる。このように考えると、輝度値、もしくはその時空間的な分布パターンから物理則に基づいて被写体の幾何学的属性を求めることが、shape from X の本質であるといえる。

5. Physics based Vision の今後

コンピュータビジョンで扱う問題を物理則のみで説明しようとする極めて複雑な問題になることは、誰の目にも明らかである。それにもかかわらず、Physics based Vision の研究が盛んに行われてきた背景には、正しく動作する条件や保証できる精度などが明確なアルゴリズムを構成するためには、何らかの法則や規則が必要だったことが挙げられる。このための法則として、光学や射影幾何学が選ばれたことは自然な成り行きである。

しかし、Physics based Vision に基づくアルゴリズムの中には、極めて理想的な条件下でしか成立し得ないような前提条件のもとに構成されたものが多く、この種のアルゴリズムに対しては、「コンピュータビジョンの問題ではなく、光学や幾何学の問題を解こうとしている」、「役に立たない」という批判が浴びせられてきた。これらの批判は「役に立たないアルゴリズム」を「完成された手法」と見なせば妥当であるが、単に画像から shape や depth を獲得するための「知見」を与えてくれたのだと考えれば、的外れになる。私は、必要以上にアルゴリズムの詳細化を推し進め、何の新たな知見ももたらさなかった研究に対して、これらの批判をするべきであり、過去の知見から、より現実的な場面で適用できる新たなアルゴリズムを生み出した研究はこのような批判の対象とはならないと思う。

Physics based Vision のもう一つの問題点として、光学的情報と幾何学的情報の間に成立する「局所的な」制約条件のみを利用してきことが挙げられる。この結果、ノイズや誤差が結果に直接反映される脆弱なアルゴリズムしか得られない、あるいはこの脆弱性を覆い隠すため平滑化などの処理を施し、精度の低い結果しか得られないといった問題が生じていた。この状況を Physics based Vision の立場から解決するには、光源の分布やその特性、物体表面の反射率分布、反射特性などの推定に基づく、より「大域的な」制約条件を導入することが必要になる。しかし、事前の計測を行わずに、撮影した画像のみから、光源や物体表面のすべての情報を推定するという問題は明らかに条件数が不足しており、いわゆる「不良設定問題」になってしまう[†]。

こういった問題を回避する一つの方法として、画像の輝度値や色といった物理量そのものに影響を与え得るすべての光学現象の因果関係を明らかにするのではなく、ある一部の光学現象の影響しか受けない何らかの記述単位（画像特徴）を抽出し、この画像特徴に関する因果関係のみを解析するアプローチがある。典型的な例は、幾何学的特徴のみを用いたコンピュータビジョンである。この分野の研究は、射影幾何学に基づく有用な知見を数多くもたらしたが、研究成果が得られれば得られるほど、光学的手法との遊離が進み、結果的に「幾何学的特徴自体が常に正しく求められるわけではない」ということがクローズアップされてしまったように思われる。こういった状況を打破するには、幾何学と光学の中間的な別の記述単位が必要になるのかもしれない。

これら以外にも、Physics based Vision では、ある単一の原理に基づくアルゴリズムはよく研究されてきたが、複数の visual cue を利用した統一理論への見通しが立っていないという問題点もある。これは、個々の手法が適用できるための前提条件が成立しているか否かを画像のみから判

[†]例えば、縞が写された一枚の画像から、その縞が、照明の強度分布、物体の形状、物体表面の反射率分布、あるいは、これらの複合的な要因によるものかどうかを区別できないことから、これは自明である。

断し難いことが主な原因である。

以上のように、今後の Physics based Vision では、大域的制約の利用、複数手法の統合ということが模索されると思われるが、その過程で Physics 以外の制約を導入せざるを得ない状況になることは想像に難くない。例えば、積み木世界の解釈では、頂点と稜線の種別を表すラベルを無矛盾に付けるという規則 [12] [13] のみで、存在可能な積み木としての解釈を構成する方法を確立している。これに類する画像の記述単位とそれらを支配する法則が、より複雑な状況下でも成立するようにしていくことがコンピュータビジョンの将来にとって重要であると思われるが、その背後にもやはり物理則が厳然と横たわっているはずである。この意味では、Physics based Vision は永遠に不滅であるといえよう。

参 考 文 献

- [1] B. K. P. Horn: Obtaining shape from shading information. pp. 115-155, The Psychology of Computer Vision, P. H. Winston ed., McGraw-Hill Book Co., New York, 1975.
- [2] R. J. Woodham: "Photometric Method for Determining Surface Orientation from Multiple Images," Optical Engineering, vol.19, no.1, pp.139-144, 1980.
- [3] M. Asada: "Cylindrical Shape From Contour And Shading Without Knowledge Of Lighting Conditions Or Surface Albedo," Proc. Of ICCV87, pp.412-416, 1987.
- [4] K. Ikeuchi and B.K.P. Horn: "Numerical shape from shading and occluding boundaries," Artificial Intelligence, vol.17, pp.141-184, 1981.
- [5] S.K. Nayar, K. Ikeuchi and T. Kanade: "Shape from Interreflections," Int. J. Computer Vision, vol.6, pp.173-195, 1991.
- [6] T. Wada, H. Ukida and T. Matsuyama: "Shape from Shading with Interreflections under a Proximal Light Source: Distortion-free Copying of an Unfolded Book," IJCV, vol.24, no.2, pp.125-135, 1997.
- [7] B.K.P. Horn: "Focusing," Artificial Intelligence Memo no.160, MIT, 1968.
- [8] J.F. Schlag, A.C. Sanderson, C.P. Neuman and F.C. Wimberly: "Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control," CMU-RI-TR-83-14 Robotics Institute, Carnegie-Mellon University, 1983.
- [9] A.P. Pentland: "A new sense for depth of field," IEEE Transactions on PAMI, vol.9, no.4, pp.523-531, 1987.
- [10] M. Subbarao: "Parallel depth recovery by changing camera parameters," proc of ICCV, pp.149-155, 1988.
- [11] B.K.P. Horn and B.G. Schunck: "Determining Optical Flow," Artificial Intelligence, vol.17, pp.185-203, 1981.
- [12] D. Huffman: "Impossible Objects as Nonsense Sentences," Machine Intelligence 6, B. Meltzer and D. Michie ed., Edinburgh University press, Edinburgh, pp.295-324, 1971.
- [13] M. Clowes: "On Seeing Things," Artificial Intelligence, vol.2, no.1, pp.79-116, 1971.

和田俊和 (Toshikazu Wada)

1987 年東京工業大学大学院修士課程修了。1990 年同大学大学院博士課程修了。同年岡山大学工学部助手。1994 年同大学大学院自然科学研究科助手。1995 年同大学工学部講師。1997 年京都大学大学院工学研究科助教授。1998 年同大学大学院情報学研究科助教授。現在に至る。工学博士。画像理解、パターン認識の研究に従事。1995 年 David Marr 賞。1997 年情報処理学会山下記念研究賞。1999 年電子情報通信学会論文賞各受賞。情報処理学会、人工知能学会、電子情報通信学会会員。