

画像の弾性マッチング — パターン認識と画像工学の一接点

内田誠一

九州大学大学院 システム情報科学研究院

〒 812-8581 福岡市東区箱崎 6-10-1

E-mail : uchida@is.kyushu-u.ac.jp

あらまし 本チュートリアルでは、画像の弾性マッチング法について解説する。弾性マッチングは、画像工学やパターン認識の分野における基本技術である。例えば、画像工学の分野においては動画画像圧縮時の動き補償に利用される。一方、パターン認識の分野においては、2 画像間の類似度を評価する際に利用される。弾性マッチングは、ゴム膜マッチングとも呼ばれ、直感的には、一方の画像を変形させながらもう一方の画像に近づける方法である。数学的には、2 画像間の画素対応関係を定める 2 次元-2 次元写像関数の最適化問題として定式化される。この写像関数については様々なモデルが提案されている。これらモデルを概観すると共に、その最適化問題がどのように解かれるかについて解説する。

Elastic matching of images — a fundamental technique for both pattern recognition and image processing

Seiichi Uchida

Faculty of Information Science and Electrical Engineering

6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi 812-8581, JAPAN

E-mail : uchida@is.kyushu-u.ac.jp

Abstract: This tutorial is concerned with elastic matching, which is one of the most fundamental technique for both image processing and pattern recognition. For example, elastic matching is used in video compression for compensating motions between consecutive frames. Elastic matching is also used in pattern recognition for evaluating a similarity between two image patterns. Intuitively speaking, elastic matching is “rubber-sheet matching” where one image is nonlinearly/linearly fitted to another image. From a mathematical viewpoint, elastic matching is formulated as an optimization problem of a 2D-2D mapping function, called warping function, which specifies pixel-to-pixel correspondence between two images. The property of elastic matching is determined by the definition of the warping function and the algorithm for optimizing the warping function.

1 まえがき

本チュートリアルでは、弾性マッチング法について解説する。弾性マッチングは、画像工学やパターン認識の分野における基本技術である [11, 14, 19]。例えば、画像工学の分野では、動画画像圧縮において利用される動き補償方式 [8, 30] が弾性マッチングの一種として挙げられる。他にもステレオ対応付け [3, 27, 40]

やオプティカルフロー [2] も同様に弾性マッチング手法と考えることができる。一方、パターン認識の分野では、2 画像間の類似度を評価する際に弾性マッチングが利用される [36]。

ゴム膜マッチングとも呼ばれることから判るように、弾性マッチングは直感的には一方の画像を変形させながらもう一方の画像に近づける方法である。数学的には、2 画像間の画素対応関係を定める 2 次

元-2 次元写像関数の最適化問題として定式化される。この写像関数については様々なモデルが提案されている。以下本稿では、これらモデルを概観すると共に、その最適化問題がどのように解かれるかについて解説する。

2 弾性マッチングの定式化

2 つの $N \times N$ 画像 $A = \{a_{i,j}\}$ と $B = \{b_{x,y}\}$ を考える。ここで、 $a_{i,j}$ は画像 A の第 (i, j) 画素の特徴量、 $b_{x,y}$ は画像 B の第 (x, y) 画素の特徴量とする。このとき、 $F : (i, j) \mapsto (x, y)$ を A から B への 2 次元-2 次元写像関数とする。写像関数 F はワープ関数とも呼ばれ、2 画像間の画素対応を定める。

弾性マッチングは、次の目的関数 $J_{A,B}(F)$ のワープ関数 F に関する最小化問題として一般に定式化される。

$$J_{A,B}(F) = D(A, B_F), \quad (1)$$

ここで、 $B_F = \{b_{F(i,j)}\}$ であり、すなわち F を通して見た画像 B である。また、 $D(\cdot, \cdot)$ は、2 つの画像の“単純”マッチング距離である。例えば、 $D(\cdot, \cdot)$ としてユークリッド距離を用いた場合、

$$J_{A,B}(F) = \sum_{i,j} \|a_{i,j} - b_{F(i,j)}\| \quad (2)$$

となる。このように $D(A, B_F)$ はワープ関数 F を介して評価した 2 画像間の距離であり、従ってその最小化は F により 2 画像の最適整合を図ることに相当する。

式 (1) の目的関数は最小化に際しては、様々な制約条件が課されることが多い。この制約条件とワープ関数 F の定義の両者によって、実際に可能な画素対応付けの範囲が規定される。

次のように正則化項 P も併せた目的関数が用いられることも多い。

$$J_{A,B}(F) = D(A, B_F) + P(F) \quad (3)$$

正則化項 $P(F)$ は、求まるワープ関数に何らかの望ましい性質を与えるためのものである。制約条件と異なり、ワープ関数 F の範囲を厳密に規定することにはならないが、求まる最適ワープ関数に所望の傾向を与えることができる。例えば、 $P(F)$ としてワープ関数 F の 2 次微分の大きさの総和

$$\iint \left(\left(\frac{\partial^2 F}{\partial i^2} \right)^2 + 2 \left(\frac{\partial^2 F}{\partial i \partial j} \right)^2 + \left(\frac{\partial^2 F}{\partial j^2} \right)^2 \right) di dj$$

を用いれば、なるべく平滑なワープ関数を求めることができる。

動き推定法であるオプティカルフローも、時間的に連続した 2 画像間の各画素対応すなわち F を求めるという意味で一種の弾性マッチング法であると言える。ただし、オプティカルフローの場合、(1) の目的関数の最小化問題としてではなく、オプティカルフロー拘束式と呼ばれる方程式を解くことで F を求める。

3 弾性マッチングの利用先

弾性マッチングは 2 画像間の変動 (変形) 解析に利用される。これは、目的関数 $J_{A,B}(F)$ を最小化するワープ関数 $\tilde{F}$ が画像 A を基準とした際の画像 B の変形を表現するためである。例えば、 A, B が時間的に連続した 2 画像であれば、 $\tilde{F}$ は動きを表すことになる。また A, B がステレオ画像対であれば、 $\tilde{F}$ は視差を表す。さらに A が標準画像パターン、 B が入力画像パターンであったとすれば、 $\tilde{F}$ は入力画像パターンに生じた幾何歪みを表す。

弾性マッチングは 2 画像間の類似度評価にも利用される。すなわち弾性マッチングによって与えられる距離

$$D_{EM}(A, B) = \min_F J_{A,B}(F) = J_{A,B}(\tilde{F}). \quad (4)$$

は、ワープ関数 F により B を極力 A に近づけた後の 2 画像 A, B 間の距離であるためである。すなわち、弾性マッチング距離は画像パターン B に生じた変形を F によって補償した後での A, B 間の距離に相当する。従ってある種の変形不変距離となる。弾性マッチング距離は、パターン認識における識別関数としてよく用いられる。例えば変形の多い対象を扱う手書き文字認識の分野においては、様々な弾性マッチングが利用されている [36]。

4 弾性マッチングの種類

前述のように、弾性マッチングの性質は、ワープ関数 F と制約条件に依存する。特に、ワープ関数については、パラメトリックなものとノンパラメトリックなもの、およびそれら両者の混合系、の 3 種に大別される。

4.1 パラメトリックなワーブ関数に基づく手法

パラメトリックなワーブ関数とは、少数のパラメータで規定される関数 F を表す。アフィン変換 [37] などの線形変換や、2次元正弦波の加重和による F の表現 [15] はその一種である。なお後者については重みがパラメータとなる。このパラメータを制御変数として目的関数 J を最小化することになる。

前述のようにパラメータは一般に少ないので、この最小化は後述するノンパラメトリックな場合に比べて計算量的に少なく済む場合が多い。しかしながら、 F の自由度は低くなる。例えばアフィン変換の場合には、パラメータ数は6個で済むが、局所的な変形に対応できない。

目的関数の最小化問題は、たとえワーブ関数が線形変換であっても、非線形最適化問題となる。これは、制御変数が非線形関数 B に内包されているためである。従って、最小二乗問題のような閉じた解を求めるのは一般に不可能である。このため Taylor 展開 [29] や反復解法 [37] などの近似解法を利用して解くことが多い。また文献 [10] の手法では、各画素 (ブロック) (i, j) が、 (x, y) に対応した際の誤差を2次までの Taylor 展開近似を用いてモデル化しておき、それを用いて画像全体としての最適なアフィン変換パラメータを陽に求めている。

4.2 ノンパラメトリックなワーブ関数に基づく手法

ノンパラメトリックなワーブ関数とは、各 (i, j) における F の値 (すなわち (i, j) の対応点 (x, y)) を直接制御可能な関数である。幾つかのパラメータを用いて間接的に関数形状を制御する 4.1 の手法と異なり、非常に自由度の高い制御が可能である。すなわち、より「柔らかな」弾性マッチング法を実現できる。

このノンパラメトリックなワーブ関数は、以下のように連続的なものと離散的なものに細分される。

4.2.1 連続的な場合

連続的かつノンパラメトリックなワーブ関数を用いた場合、(1) の最小化問題は F を変関数とする変分問題となる。変分問題には、通常な最大最小問題の極値条件にあたるものとして、Euler-Lagrange 条

件がある。弾性マッチング問題の場合、この条件はそのまま解ける形にはならない。そこでワーブ関数を i, j について離散化して非線形連立方程式を定め、それを解くことになる [21]。なおこの場合、(i) あくまで Euler-Lagrange 条件の離散近似条件である、(ii) Euler-Lagrange 条件は最適解の必要条件であり十分条件ではない、(iii) 非整数座標における画素値が必要になる、(iv) 非線形連立方程式は数値的な手法で解くしかない、といった問題があるため、実用の際には妥当な解を得るための工夫が必要となる。例えば正則化項 P を導入した目的関数 (3) を利用したり、粗密探索の援用などが行なわれる [22, 39]。

オプティカルフローも、ノンパラメトリックかつ連続的なワーブ関数を用いた弾性マッチング法の一つと言える。オプティカルフローの場合、前述のように、(1) の目的関数の最小化問題としてではなく、オプティカルフロー拘束式と呼ばれる方程式を解くことで F を求める。ただし、解を求める際に、離散近似を行ったり、また正則化を採用する点 [13] など、上述の変分ベースの方法との共通点も多い。

4.2.2 離散的な場合

離散的なワーブ関数を用いた場合、ワーブ関数は画素の離散的な対応関係そのものとなり、目的関数の最小化問題はワーブ関数を表現する $2N^2$ 個の変数 $((x_1, y_1), \dots, (x_i, y_j), \dots, (x_N, y_N))$ の最適化問題に帰着する。デジタル画像を対象とする場合、こうした離散的なワーブ関数を用いるのは、ある意味合理的である。

変数 $\{(x_i, y_j)\}$ の間に制約条件が無ければ、(1) の最小化問題は、

$$\min_{\{x_i, y_j\}} \sum_{i,j} \|a_{i,j} - b_{x_i, y_j}\| = \sum_{i,j} \min_{x_i, y_j} \|a_{i,j} - b_{x_i, y_j}\| \quad (5)$$

のように各 (x_i, y_j) 独立の最適化問題になる。すなわち A 上の各画素 (i, j) 毎に、最も適合する B 上の画素を探索する問題になる。この場合の弾性マッチングは摂動法と呼ばれることもある。この方法は広く検討されているが、多くの場合、誤対応を防ぐために、粗密探索手法や高次元画素特徴などが援用されている [4, 28, 38]。

一方、変数 $\{(x_i, y_j)\}$ の間に制約があると、問題は一気に難しくなる。厳密な最適解 $\hat{F}$ を求めようとしても、上のように各変数毎の独立な最適化を行な

うことはできない [20] . 特に, F を同相写像とすべく, 上下左右の画素間に次のような単調連続性 [33] の制約

$$\left. \begin{aligned} 0 &\leq x_i - x_{i-1} \leq 2 \\ 0 &\leq y_j - y_{j-1} \leq 2 \\ |x_j - x_{j-1}| &\leq 1 \\ |y_i - y_{i-1}| &\leq 1 \end{aligned} \right\} \quad (6)$$

を課すると, 目的関数 J の最小化問題は NP-hard 問題になり [16], 厳密解を求めるのは事実上不可能になる. このため, 第 $i-1$ 行の対応関係を制約条件として第 i 行の対応関係を定め, 次にその i 行の対応関係を制約条件として第 $i+1$ 行の対応関係を定めるといった逐次的な近似解法が採られる [5, 9, 25, 26, 31, 34]. また, A の i 行の画素は B 上ですべて同じ x 行内に対応付けられるといった制約を用いると, F は疑似 2 次元的なものとなるが, 計算量的には相当軽減された問題に帰着できる [1, 6, 7, 12, 17].

4.3 パラメトリックとノンパラメトリックの混合系

動画像圧縮の動き補償で採用されるブロックマッチングは, ブロックを画素のように見るとノンパラメトリックなワーブ関数を摂動法で最適化している方法になる. 一方, 1 ブロックに着目すると平行移動のみであり, これは線形ワーブ関数である. 従って, ブロックマッチングはパラメトリックとノンパラメトリックの混合系であると言える.

メッシュベースの動き補償 [32] でも同様に混合系が使用される. すなわち, メッシュを構成する三角形パッチの頂点に着目すると摂動法的かつ逐次的に最適化されるされるノンパラメトリックなワーブ関数であり, パッチ内に着目するとアフィン変換すなわち線形ワーブ関数である [23, 24]. なお, 同様の手法は顔画像マッチングにも適用されている [18].

パターン認識の分野では, こうした局所分割以外の混合系も提案されている. 文献 [35] の手法では, 学習フェーズとして, あるカテゴリの複数の学習パターンについてそれぞれノンパラメトリックなワーブ関数を最適化し, そうして得られた最適ワーブからそのカテゴリに発生しやすいワーブ関数 $F_1, F_2, \dots, F_k$ を統計的に推定しておく. 実際にマッチングを図る際には, F を線形和 $\sum_k \alpha_k F_k$ として定義し, α_k をパラメータとして最適化する. この方法では, ノン

パラメトリックなワーブ関数のもつ自由度を F_k により利用しながら, 少数のパラメータ α_k で最適化が行なえる.

5 まとめ

画像工学やパターン認識の分野における基本技術である画像の弾性マッチング法について概観した. 弾性マッチングの実体である 2 画像間の画素対応を定める 2 次元-2 次元写像 (ワーブ関数) について着目し, ワーブ関数がパラメトリック表現されるか, ノンパラメトリック表現されるかによって, 従来の弾性マッチング法を分類した. その際, こうしたワーブ関数の表現に応じて, その最適化手法 (アルゴリズム) に違いがあることも指摘した.

弾性マッチングでは結果として, (i) 最適化されたワーブ関数すなわち画素対関係, および (ii) 最小化された目的関数すなわち弾性マッチング距離, の 2 つが求まる. 前者は 2 画像間の変動 (変形) 解析に利用される. 後者はパターン認識において変形不変な識別関数として利用される.

本チュートリアルでは, 弾性マッチングの具体的な応用例については触れていない. 応用に際しては, マッチングの対象の性質に応じてワーブ関数や制約条件ならびに目的関数をチューニングすることが重要である. また応用に際して計算量的に妥当な手法を選ぶことも肝要と思われる.

参考文献

- [1] O. Agazzi, S. Kuo, E. Levin, and R. Pieracini, “Connected and degraded text recognition using planar hidden Markov models,” Proc. ICASSP, pp. V113–116, 1993.
- [2] S. S. Beauchemin and J. L. Barron, “The computation of optical flow,” ACM Computing Surveys, vol. 27, no. 3, pp. 433–467, 1995.
- [3] M. Z. Brown, D. Burschka, and G. D. Hager, “Advances in computational stereo,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 8, pp. 993–1008, 2003.

- [4] D. J. Burr, "A dynamic model for image registration," *Computer Graphics, & Image Proc.*, vol. 15, pp. 102–112, 1981.
- [5] M. C. Chen and A. N. Willson, Jr., "Motion-vector optimization of control grid interpolation and overlapped block motion compensation using iterated dynamic programming," *IEEE Trans. Image Proc.*, vol. 9, no. 7, pp. 1145–1157, 2000.
- [6] W. Cho, S.-W. Lee, and J. H. Kim, "Modeling and recognition of cursive words with hidden Markov models," *Pattern Recog.*, vol. 28, no. 12, pp. 1941–1953, 1995.
- [7] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs, "A maximum likelihood stereo algorithm," *Comput. Vis. & Image Und.*, vol. 63, no. 3, pp. 542–567, 1996.
- [8] F. Defaux and F. Moscheni, "Motion estimation techniques for digital TV: a review and a new contribution," *Proc. IEEE*, vol. 83, no. 6, pp. 858–876, 1995.
- [9] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 16, pp. 39–55, 1987.
- [10] M. Etoh, "Promotion of block matching: parametric representation for motion estimation," *Proc. ICPR*, vol. 1 of 2, pp. 282–285, 1998.
- [11] C. A. Glasbey and K. V. Mardia, "A review of image-warping methods," *J. Applied Statistics*, vol. 25, no. 2, pp. 155–171, 1998.
- [12] 原 学, 内田誠一, 迫江博昭, "画素を単位とした視差補償に基づくステレオ画像圧縮の検討," *情報処理学会研究報告*, 2004-AVM-46(1), pp. 1–7, 2004.
- [13] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.
- [14] A. K. Jain, Y. Zhong, and M. -P. Dubuisson-Jolly, "Deformable template models: A review," *Signal Processing*, vol. 71, no. 2, pp. 109–129, 1998.
- [15] A. K. Jain, Y. Zhong, and S. Lakshmanan, "Object matching using deformable templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 267–278, 1996.
- [16] D. Keysers and W. Unger, "Elastic image matching is NP-complete," *Pattern Recog. Lett.*, vol. 24, no. 1–3, pp. 445–453, 2003.
- [17] S. Kuo and O. E. Agazzi, "Keyword spotting in poorly printed documents using pseudo 2-D hidden Markov models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 842–848, 1994.
- [18] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Malsburg, P. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, no. 3, pp. 300–311, 1993.
- [19] H. Lester and S. R. Arridge, "A survey of hierarchical non-linear medical image registration," *Pattern Recog.*, vol. 32, no. 1, pp. 129–149, 1999.
- [20] E. Levin and R. Pieraccini, "Dynamic planar warping for optical character recognition," *Proc. ICASSP*, pp. III 149–152, 1992.
- [21] R. March, "Computation of stereo disparity using regularization," *Pattern Recog. Letters*, vol. 8, pp. 181–187, 1988.
- [22] Y. Mizukami, "A handwritten Chinese character recognition system using hierarchical displacement extraction based on directional features," *Pattern Recog. Lett.*, vol. 19, no. 7, pp. 595–604, 1998.
- [23] Y. Nakaya and H. Harashima, "Motion compensation based on spatial transformations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, no. 3, pp. 339–356, 1994.

- [24] J. Niewęglowski, T. G. Campbell, and P. Haavisto, "A novel video coding scheme based on temporal prediction using digital image warping," *IEEE Trans. Consumer Electronics*, vol. 39, no. 3, pp. 141-150, 1993.
- [25] G. M. Quénou, "The orthogonal algorithm for optical flow detection using dynamic programming," *Proc. ICASSP*, pp. III-249-252, 1992.
- [26] A. L. Ratan, W. E. L. Grimson, and W. M. Wells III, "Object detection and localization by dynamic template warping," *Proc. CVPR*, pp. 634-640, 1998.
- [27] A. Redert, E. Hendriks, and J. Biemond, "Correspondence estimation in image pairs," *IEEE Signal Proc. Mag.*, vol. 16, no. 3, pp. 29-46, 1999.
- [28] Y. Shinagawa and T. L. Kunii, "Unconstrained automatic image matching using multiresolutional critical-point filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 9, pp. 994-1010, 1998.
- [29] P. Y. Simard, Y. Le. Cun, J. S. Denker, and B. Victorri, "An efficient algorithm for learning invariances in adaptive classifiers," *Proc. ICPR*, vol. 2, pp. 651-655, 1992.
- [30] C. Stiller and J. Konrad, "Estimating motion in image sequences," *IEEE Sig. Proc. Mag.*, vol. 16, no. 4, pp. 70-91, 1999.
- [31] 杉村昌彦, 飯国洋二, 足立紀彦, "ツェルニケモーメントを特徴量とする2次元動的計画法を用いたイメージマッチング," *信学論 (D-II)*, vol. J80-D-II, no. 1, pp. 101-108, 1997.
- [32] A. M. Tekalp, P. Van Beek, C. Toklu, B. Günsel, "Two-dimensional mesh-based visual-object representation for interactive synthetic/natural digital video," *Proc. IEEE*, vol. 86, no. 6, pp. 1029-51, 1998.
- [33] 内田誠一, 迫江博昭, "動的計画法に基づく単調連続2次元ワープ法の検討," *信学論 (D-II)*, vol. J81-D-II, no. 6, pp. 1251-1258, 1998.
- [34] S. Uchida and H. Sakoe, "An approximation algorithm for two-dimensional warping," *IEICE Trans. Info. & Syst.*, vol. E83-D, no. 1, pp. 109-111, 2000.
- [35] 内田誠一, 迫江博昭, "カテゴリ固有変形の線形結合モデルに基づく弾性マッチング法," *信学論 (D-II)*, vol. J87-D-II, no. 2, pp. 639-648, 2004.
- [36] S. Uchida and H. Sakoe, "A survey of elastic matching techniques for handwritten character recognition," *IEICE Trans. Inf. & Syst.*, vol. E88-D, no. 8, pp. 1781-1790, 2005.
- [37] T. Wakahara, Y. Kimura, and A. Tomono, "Affine-invariant recognition of gray-scale characters using global affine transformation correlation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 4, pp. 384-395, 2001.
- [38] 山田博三, 斉藤泰一, 森 俊二, "類似度法の一改良 - ずらし類似度 -, " *信学論*, vol. J64-D, no. 10, pp. 970-976, 1981.
- [39] N. Yokoya, "Dense matching of two views with large displacement," *Proc. ICIP*, vol. 1 of 3, pp. 213-217, 1994.
- [40] N. Yokoya, T. Shakunaga, M. Kanbara, "Passive range sensing techniques: depth from images," *IEICE Trans. Inf. Syst.*, vol. E82-D, no. 3, pp. 523-33, 1999.