Efficient Shape Matching using Vector Extrapolation

Emanuele Rodolà[1,2]
emanuele.rodola@in.tum.de
Tatsuya Harada[1]
harada@mi.t.u-tokyo.ac.jp
Yasuo Kuniyoshi[1]
kuniyosh@isi.imi.i.u-tokyo.ac.jp
Daniel Cremers[2]
cremers@tum.de

[1] Information Science and Technology,
The University of Tokyo

[2] Computer Vision Group,
Technical University of Munich

Shape matching is a pervasive problem in computer vision. It concerns, in its general form, the problem of determining a map $f: X \to Y$ among two given shapes X and Y in such a way that their geometrical properties are preserved by the transformation. Arguably one of the most adopted formulations for shape matching takes form as a NP-hard quadratic assignment problem (QAP), where a quadratic term in the objective function encodes a measure of pairwise association among a set $C \subseteq X \times Y$ of putative matches. In particular, let $\mathbf{x} \in [0,1]^{|C|}$ be a (soft) indicator vector representing a set of point-to-point matches among the two shapes, and let $\mathbf{S} \in \mathbb{R}^{|C|\times|C|}$ be a *similarity matrix* among pairs of matches. Then the matching problem can be cast as the quadratic program

$$\max_{\mathbf{x}} \ \mathbf{x}^{\mathrm{T}}\mathbf{S}\mathbf{x} \quad \text{s.t. } \Pi(\mathbf{x}) \preceq \mathbf{c}\,, \tag{1}$$

where $\preceq$ denotes element-wise inequality, $\mathbf{c} \in \mathbb{R}^{|C|}$ is a fixed vector, and Π is a (possibly nonlinear) function specifying the mapping constraints on the correspondence (*e.g.*, one-to-one, one-to-many).

Several attempts at finding good local optima for variants of the QAP have been proposed in computer vision literature. A well-known approach consists in solving a modified problem in which $\|\mathbf{x}\|_2^2 = 1$ replaces the mapping constraints appearing in (1). The method manifests a tendency to assign matches to each point, therefore bringing incorrect correspondences in the final solution even in presence of moderate noise. The more recent techniques based on game theory [1] replace this unit norm constraint by the L_1 counterpart $\|\mathbf{x}\|_1 = 1$, promoting stable, yet very sparse solutions. In order to strike a balance between the two, and therefore capture a broader family of matching scenarios, in this paper we first propose to consider the following family of relaxations for the QAP:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^{\mathrm{T}}\mathbf{S}\mathbf{x} \\ \text{s.t.} \quad & (1-\alpha)\|\mathbf{x}\|_1 + \alpha\|\mathbf{x}\|_2^2 = 1\,,\ \mathbf{x} \succeq \mathbf{0}\,,\ \alpha \in [0,1]\,. \end{aligned} \tag{2}$$

The convex combination of L_1 and L_2 penalties appearing above takes the name of *elastic net* [4]. This particular formulation gives rise to the so called *grouping effect*, *i.e.*, the joint selection of entire clusters of highly similar matches. This is a desirable feature in most relaxed QAP scenarios, in which one typically seeks only high-precision correspondences in a situation where there is huge ambiguity in the candidate set. A local optimum to problem (2) can be determined by following a projected gradient approach. The optimization process is governed by the equations

$$\mathbf{x}^{(t+1)} = P_\alpha\left(\mathbf{x}^{(t)} + \delta\mathbf{S}\mathbf{x}^{(t)}\right), \tag{3}$$

where P_α is a projection operator taking a solution back to the feasible set. Projected gradient is a natural choice for this family of problems, nevertheless its application in real-world settings is often rendered prohibitive by the large scale of the data.

The major contribution of this paper derives from the observation that, in most non-convex matching scenarios, the intermediate sequence of solutions generated by the optimization algorithm has a tendency to exhibit a certain degree of smoothness. This suggests the possibility to infer the general direction of convergence from previous iterates. To this end, we look at a family of techniques coming under the umbrella term of *vector extrapolation* [3]. These techniques have found application in the context of fixed-point iterative methods for solving linear and nonlinear systems of equations; until now, limited attempts have been made to adopt such methods in computer vision problems [2].

Consider the vector sequence $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ as generated by the non-linear process of Eq. (3). It is often the case that the iterative process

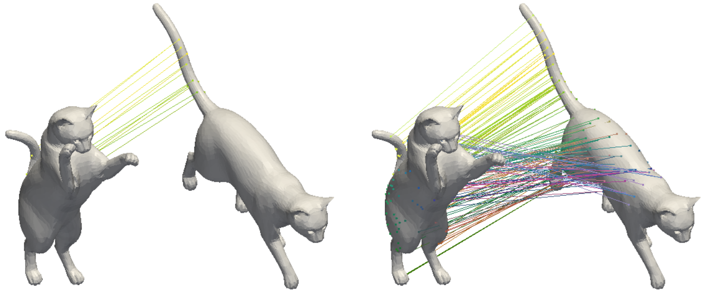

Figure 1: Solutions obtained via vector-extrapolated projected gradient in a problem of non-rigid matching under elastic net constraints. A very selective correspondence is given by $\alpha = 0.1$ (left); a denser, but similarly accurate solution can be obtained with $\alpha = 0.85$ (right).

requires many iterations to reach good accuracy, or that the individual terms $\mathbf{x}^{(t)}$ themselves are expensive to compute due to the action of P_α. We are thus looking for a means to give an estimate of the limit point $\mathbf{s} = \lim_{t\to\infty}\mathbf{x}^{(t)}$ using as few terms as possible. As we show in the paper, such an estimate can be expressed as the linear combination

$$\mathbf{s}^* = \sum_{i=0}^{k} \gamma_i \mathbf{x}_i\,, \tag{4}$$

with $\sum_{i=0}^{k}\gamma_i = 1$. In other words, an approximation to the limit point can be given as a linear combination of $k+1$ past non-linearly generated iterates. In practice, solving for the optimal set of weights γ_i simply amounts to minimizing a least-squares problem subject to linear constraints.

In order to assess the validity of the method, we performed a wide range of experiments on three related computer vision tasks, namely rigid matching of point clouds, non-rigid matching of three-dimensional shapes, and feature matching for multiple-view stereo. From an optimization point of view, these matching scenarios differ in their specific definitions of the similarity term, and typically exhibit energy landscapes with rather different characteristics. This allows us to assess to what extent vector extrapolation may be adopted as a means to accelerate convergence in difficult settings, and whether its introduction into the matching process may lead to premature convergence and thus poor local optima. We note that the general improvement is almost one order of magnitude if compared with an optimization carried out with no vector extrapolation at all. This large increase in performance suggests that there is little reason not to adopt vector extrapolation for this class of problems, and even in those cases where the advantage is limited, vector extrapolation may still be adopted whenever a higher degree of accuracy is required.

[1] A. Albarelli, S. Rota Bulò, A. Torsello, and M. Pelillo. Matching as a non-cooperative game. In *Proc. IEEE Intl. Conf. on Comput. Vis.*, pages 1319 –1326, 29 2009-oct. 2 2009.

[2] G. Rosman, M. M. Bronstein, A. M. Bronstein, and R. Kimmel. Nonlinear dimensionality reduction by topologically constrained isometric embedding. *Int. J. Comput. Vision*, 89(1):56–68, August 2010.

[3] D. A. Smith, W. F. Ford, and A. Sidi. Extrapolation methods for vector sequences. *SIAM Review*, 29(2):pp. 199–233, 1987.

[4] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.