

# Efficient Shape Matching using Vector Extrapolation

Emanuele Rodolà<sup>12</sup>  
emanuele.rodola@in.tum.de

Tatsuya Harada<sup>1</sup>  
harada@mi.t.u-tokyo.ac.jp

Yasuo Kuniyoshi<sup>1</sup>  
kuniyosh@isi.imi.i.u-tokyo.ac.jp

Daniel Cremers<sup>2</sup>  
cremers@tum.de

<sup>1</sup> Information Science and Technology  
The University of Tokyo  
Tokyo, Japan

<sup>2</sup> Computer Vision Group  
Technical University of Munich  
Garching, Germany

---

## Abstract

We propose the adoption of a vector extrapolation technique to accelerate convergence of correspondence problems under the quadratic assignment formulation for attributed graph matching (QAP). In order to capture a broad range of matching scenarios, we provide a class of relaxations of the QAP under elastic net constraints. This allows us to regulate the sparsity/complexity trade-off which is inherent to most instances of the matching problem, thus enabling us to study the application of the acceleration method over a family of problems of varying difficulty. The validity of the approach is assessed by considering three different matching scenarios; namely, rigid and non-rigid three-dimensional shape matching, and image matching for Structure from Motion. As demonstrated on both real and synthetic data, our approach leads to an increase in performance of up to one order of magnitude when compared to the standard methods.

## 1 Introduction

Shape matching is a pervasive problem in computer vision. It concerns, in its general form, the problem of determining a map  $f: X \rightarrow Y$  among two given shapes  $X$  and  $Y$  in such a way that their geometrical properties are preserved by the transformation. This problem has attracted the interest of many researchers over the years, and a large variety of approaches have been proposed. 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 of putative matches. This formulation is common in attributed graph matching literature [0, 0, 0, 0], but it also frequently arises in problems of inlier selection and object recognition [0, 0], and in minimum distortion correspondence problems under the notion of Gromov-Hausdorff distance between metric spaces [0, 0, 0]. In this wide variety of scope and objectives, it is not uncommon for existing methods to provide specifically tailored optimization techniques for the particular problem they attempt to solve, ranging from dual decomposition [0] to

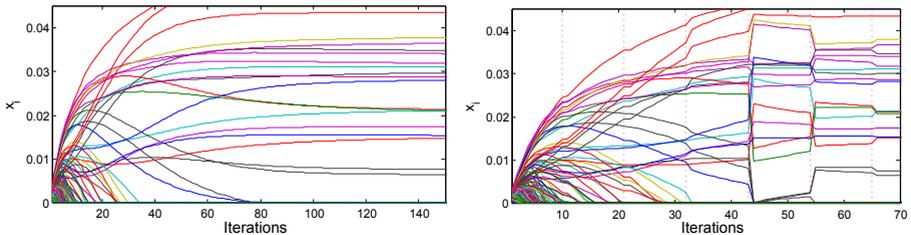


Figure 1: Evolution of a correspondence set via projected gradient (left figure) in a problem of non-rigid matching under elastic net constraints, and its accelerated version using vector extrapolation with cycling (right figure). Each curve represents the progress of a candidate match  $x_i \in [0, 1]$  during the optimization process (Eq. (6)). The vertical bars in the second plot indicate the time steps where extrapolation takes place; the jump occurring at iteration 44 corresponds to a large increase in the objective value (equal to 5% of the final optimum). In this case, the standard method (left) requires 333 iterations in total to reach a local optimum with the same level of accuracy of the accelerated solution (total 70 iterations).

first-order Taylor approximations of the objective [24] and even closed form solutions [9]. In most cases, though, it is difficult to devise a specific technique for reaching a solution, and thus more general optimization algorithms need to be considered. Our focus in this paper is on variants of the projected gradient method for convex optimization [10]. These algorithms tend to be slow, and accelerated approaches attempting to take into consideration the curvature of the objective surface (*e.g.*, parallel tangents [11]) usually bring only limited benefits in practical scenarios, making their application in large-scale problems often prohibitive.

In this paper, we propose the adoption of a vector extrapolation technique [21] to accelerate the numerical solution of QAP-based matching problems. In order to capture a broad family of matching scenarios, we look at a new family of constraints, namely the elastic net [22], which allow us to regulate the intrinsic trade-off between accuracy and size of the correspondence with continuity. 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 (see Figure 1). This suggests the possibility to infer the general direction of convergence from previous iterates. Since the main objective of such iterative process is to reach a fixed point, we propose to determine an approximate solution to it by estimating, in closed form, the limit value of the sequence as generated by the process. This approximation can be computed as a linear combination of the last few iterates, with coefficients uniquely determined as the solution to a linear system. While it can be proved that this approach allows to attain the global optimum in case the considered sequence is linearly generated, previous attempts at applying the method in nonlinear settings have demonstrated good convergence properties. Until now, limited attempts have been made to adopt such techniques in computer vision problems; the few examples include isometric embedding [18], blind deconvolution [9], and Beltrami filtering on images [10].

## 2 Matching with pairwise constraints

Let  $X, Y \in \mathbb{R}^2$  be two given shapes, and let the binary function  $c : X \times Y \rightarrow \{0, 1\}$  represent a point-to-point correspondence among them, such that  $c(x, y) = 1$  if and only if  $x \in X$  is a corresponding point for  $y \in Y$ , and vice versa; mapping constraints may additionally be imposed over  $c$ , depending on the specific task at hand (*e.g.*, one-to-one, one-to-many). In

order to give a measure of quality to the correspondence, we evaluate the distortion induced by the mapping as measured on  $X$  and  $Y$  according to some predetermined distance function  $d : \mathbb{R}^z \times \mathbb{R}^z \rightarrow \mathbb{R}_+$ . Given two matches  $(x, y), (x', y') \in X \times Y$ , the absolute criterion

$$\varepsilon((x, y), (x', y')) = |d(x, x') - d(y, y')| \quad (1)$$

provides a measure of *cost* for the pair of matches. In the particular case where  $X$  and  $Y$  are compact metric spaces endowed with a metric  $d$ , Equation (1) directly quantifies to what extent the given correspondence deviates from isometry [12, 14, 15].

Equivalently, one may define a measure  $s$  of *similarity* among pairs of matches [10], assigning a large value to low cost pairs, and a small value otherwise. With this definition, given a correspondence set  $C \subset X \times Y$ , we can define its total similarity as the weighted sum

$$\sum_{(x,y),(x',y') \in C} c(x,y)c(x',y')s((x,y),(x',y')). \quad (2)$$

We are thus interested in determining the correspondence set  $C$  maximizing the score above. Note, however, that the binary constraints on  $c$  render this problem NP-hard. For this reason, it has been proposed [10] to relax the integral constraints by letting the correspondence function assume a continuum of values in  $[0, 1]$ , giving rise to a *fuzzy* notion of correspondence between shapes [15]. With these changes, the problem can be easily cast as a relaxed QAP:

$$\max_{\mathbf{x}} \mathbf{x}^T \mathbf{S} \mathbf{x} \quad \text{s.t. } \Pi(\mathbf{x}) \preceq \mathbf{b}, \quad (3)$$

where  $\mathbf{x} \in [0, 1]^{|C|}$  is a vector representation for the correspondence function,  $\mathbf{S} \in \mathbb{R}^{|C| \times |C|}$  is a symmetric matrix containing the similarity terms appearing in Equation (2),  $\preceq$  denotes element-wise inequality,  $\mathbf{b} \in \mathbb{R}^{|C|}$  is a fixed vector, and  $\Pi$  is a (possibly nonlinear) function specifying the mapping constraints on  $c$ .

## 2.1 Elastic net matching

Several attempts at finding good local optima for variants of the QAP have been proposed in computer vision literature, from graduated assignment [10] for attributed graph matching, to more recent proposals such as spectral matching [9], pseudo-boolean decomposition [23], and game-theoretic matching [10]. In [9], it was proposed to solve the modified problem

$$\max_{\mathbf{x}} \mathbf{x}^T \mathbf{S} \mathbf{x} \quad \text{s.t. } \|\mathbf{x}\|_2^2 = 1, \mathbf{x} \in [0, 1]^{|C|}, \quad (4)$$

which is then maximized, by Rayleigh's quotient theorem, by the principal eigenvector of matrix  $\mathbf{S}$ . Note that, in this case, the actual mapping constraints between the two shapes are lost, and they must be imposed *a posteriori* by a greedy method [9]. As a result, 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 [10, 15] replace the unit norm constraint from (4) by the  $L_1$  counterpart  $\|\mathbf{x}\|_1 = 1$ , with  $\mathbf{x} \succeq \mathbf{0}$ . The simplex constraint has the effect of promoting sparse, yet very stable solutions, a characteristic that makes the method particularly effective in tasks where strong selectivity is a major requirement [2, 16]; nevertheless, this selectivity may come as a disadvantage in a variety of tasks, and the high locality of the obtained solutions does not allow, in practice, densification methods to be applied [23].

Most existing formulations for shape matching typically aim at striking a balance between the two. In order to provide a representative framework for a whole range of different

methods, we introduce a new family of constraints into the matching problem, namely the *elastic net* [24]. The elastic net is a regularization technique that arose in the context of statistical learning and regression analysis. Its penalty function is defined as a convex combination of  $L_1$  and  $L_2$  penalties; the first term enforces the sparsity of the solution, whereas the second term ensures democracy among groups of correlated variables. This particular behaviour (called the *grouping effect*) allows to select entire groups of correlated features when these groups are not known in advance. In the context of shape matching, we expect the grouping effect to manifest itself as a joint selection of entire clusters of highly similar matches, *i.e.*, groups of matches with high internal cohesion in terms of the adopted measure of similarity. 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 most correspondences. Conversely, incorrect matches will be unlikely to form strongly correlated groups, and thus they will be filtered out in a measure that depends on the penalty incurred by them under the elastic net constraint. This accuracy/sparsity trade-off is regulated by a single convexity parameter  $\alpha \in [0, 1]$ , leading to the following family of (non-convex) relaxations for the QAP:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{x}^T \mathbf{S} \mathbf{x} \\ \text{s.t.} \quad & (1 - \alpha) \|\mathbf{x}\|_1 + \alpha \|\mathbf{x}\|_2^2 = 1, \mathbf{x} \succeq \mathbf{0}. \end{aligned} \quad (5)$$

This family of problems varies from the highly selective game-theoretic behaviour for  $\alpha = 0$  to the more tolerant  $L_2$  solution for  $\alpha = 1$ . Similarly to the spectral approach, this formulation does not impose any mapping constraint on the correspondence. However, these may be imposed ex-post by using the same technique of [9], or by any other binarization method.

## 2.2 Optimization

We determine a local optimum to problem (5) by following a projected gradient approach [10]. The optimization process is governed by the equations

$$\mathbf{x}^{(t+1)} = P_\alpha \left( \mathbf{x}^{(t)} + \delta \mathbf{S} \mathbf{x}^{(t)} \right), \quad (6)$$

where  $\mathbf{S} \mathbf{x} = \frac{1}{2} \nabla \mathbf{x}^T \mathbf{S} \mathbf{x}$  is an ascent direction for the objective,  $\delta > 0$  is the step length taken in that direction, and  $P_\alpha : \mathbb{R}^{|C|} \rightarrow \mathbb{R}^{|C|}$  is a projection operator taking a solution back onto the feasible set. The computation of  $P_\alpha(\hat{\mathbf{x}})$  requires solving a Euclidean projection sub-problem onto the intersection of the non-negative cone with the boundary of the elastic net ball, amounting to the minimization:

$$\min_{\mathbf{x}} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \quad \text{s.t.} \quad (1 - \alpha) \|\mathbf{x}\|_1 + \alpha \|\mathbf{x}\|_2^2 = 1, \mathbf{x} \succeq \mathbf{0}. \quad (7)$$

This is a convex optimization problem, and it has been recently solved by Mairal *et al.* [10] in the context of sparse dictionary learning. Their simple algorithm guarantees to find the unique projection onto the set of constraints, and has expected linear time complexity.

## 3 Vector extrapolation

In this Section we introduce a technique for accelerating convergence of QAP-based methods for matching. Our motivations derive from the observation that, in typical non-convex matching scenarios, the vector sequence  $\{\mathbf{x}_i\}$  generated by the optimization process generally exhibits a certain degree of smoothness (see Figure (1), left plot). It seems, indeed, that the general direction of the convergence process may be inferred from past iterates.

To this end, we look at a family of techniques coming under the umbrella term of *vector extrapolation*. These techniques have found application in the context of fixed-point iterative methods for solving linear and nonlinear systems of equations, such as those arising in finite-difference schemes for large-scale computational fluid dynamics problems [20].

### 3.1 Reduced rank extrapolation

Consider the vector sequence  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$  as generated by the linear process

$$\mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i + \mathbf{b}, \quad (8)$$

where  $\mathbf{A}$  is a fixed  $N \times N$  matrix and  $\mathbf{b} \in \mathbb{R}^N$  is a fixed vector. We assume that neither  $\mathbf{A}$  nor  $\mathbf{b}$  are known, and only the sequence  $\{\mathbf{x}_i\}$  or the generating process are given. Starting from an initial point  $\mathbf{x}_0$ , the sequence either diverges to an *anti-limit* point  $\lim_{i \rightarrow \infty} \mathbf{x}_i$ , or converges to the unique solution to the system  $\mathbf{s} = \mathbf{A}\mathbf{s} + \mathbf{b}$ , provided that 1 is not eigenvalue of  $\mathbf{A}$ . In case of convergence, the fixed point will thus equal  $\mathbf{s} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}$ . It is often the case that the iterative process (8) requires many iterations to reach good accuracy, or that the individual terms  $\mathbf{x}_i$  themselves are expensive to compute. We are thus looking for a means to give an estimate of the limit point  $\mathbf{s}$  using as few terms as possible.

Let us define the difference vectors  $\mathbf{u}_i = \Delta\mathbf{x}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$  and  $\mathbf{w}_i = \Delta^2\mathbf{x}_i = \mathbf{u}_{i+1} - \mathbf{u}_i$  for  $i = 0, 1, 2, \dots$ . By induction, we obtain  $\mathbf{u}_{i+1} = \mathbf{A}\mathbf{u}_i = \mathbf{A}^{i+1}\mathbf{u}_0$  and then  $\mathbf{w}_i = (\mathbf{A} - \mathbf{I})\mathbf{u}_i$ . Construct the  $N \times k$  matrices

$$\mathbf{U}_k = [\mathbf{u}_0 | \mathbf{u}_1 | \dots | \mathbf{u}_{k-1}], \quad \mathbf{W}_k = [\mathbf{w}_0 | \mathbf{w}_1 | \dots | \mathbf{w}_{k-1}], \quad (9)$$

where  $k$  is the total number of terms to use for the estimation of  $\mathbf{s}$ . If  $\mathbf{W}_k$  is invertible and  $k = N$ , then it follows from the above relations that

$$(\mathbf{I} - \mathbf{A})^{-1} = -\mathbf{U}_k \mathbf{W}_k^{-1}, \quad (10)$$

and then, since  $\mathbf{u}_i = (\mathbf{A} - \mathbf{I})(\mathbf{x}_i - \mathbf{s})$ , as it is easy to verify from the definitions, it holds

$$\mathbf{s} = \mathbf{x}_i + (\mathbf{I} - \mathbf{A})^{-1}\mathbf{u}_i = \mathbf{x}_i - \mathbf{U}_k \mathbf{W}_k^{-1}\mathbf{u}_i, \quad i = 0, 1, \dots, k-1. \quad (11)$$

We have thus obtained an expression for  $\mathbf{s}$  that does not depend on the explicit knowledge of  $\mathbf{A}$ , as desired. However, it requires the inversion of a full rank matrix, a very costly operation in most cases. Following [21], we rewrite (11) as the pair of simultaneous equations

$$\mathbf{s} = \mathbf{x}_0 + \mathbf{U}_k \boldsymbol{\xi}, \quad \mathbf{0} = \mathbf{u}_0 + \mathbf{W}_k \boldsymbol{\xi}, \quad (12)$$

with  $\boldsymbol{\xi} \in \mathbb{R}^k$ . Suppose now we choose  $k < N$ . Then, according to the equations above, we seek an approximation  $\mathbf{s}^* = \mathbf{x}_0 + \sum_{i=0}^{k-1} \xi_i \mathbf{u}_i$  such that the norm of the residual  $\|\mathbf{u}_0 + \mathbf{W}_k \boldsymbol{\xi}\|_2$  is uniquely minimized by  $\boldsymbol{\xi}$ . Now define  $\gamma \in \mathbb{R}^{k+1}$  to be the unique vector of coefficients such that  $\gamma_0 = 1 - \xi_0$ ,  $\gamma_i = \xi_{i-1} - \xi_i$  for  $1 \leq i \leq k-1$ , and  $\gamma_k = \xi_{k-1}$ . With this choice, it is easy to verify that  $\sum_{i=0}^k \gamma_i = 1$ , and we obtain an alternative expression for our estimate:

$$\mathbf{s}^* = \sum_{i=0}^k \gamma_i \mathbf{x}_i, \quad (13)$$

from which we see that we can determine an approximation to the limit point as a linear combination of  $k+1$  past linearly generated iterates. Now, noting that

$$\mathbf{u}_0 + \mathbf{W}_k \boldsymbol{\xi} = \mathbf{u}_0 + \sum_{i=0}^{k-1} \xi_i \mathbf{w}_i = \Delta \left( \mathbf{x}_0 + \sum_{i=0}^{k-1} \xi_i \mathbf{u}_i \right) = \Delta \left( \sum_{i=0}^k \gamma_i \mathbf{x}_i \right) = \sum_{i=0}^k \gamma_i \mathbf{u}_i, \quad (14)$$

we can rewrite Equations (12) as the least-squares problem (Mešina’s algorithm [13])

$$\min_{\gamma \in \mathbb{R}^{k+1}} \left\| \sum_{i=0}^k \gamma_i \mathbf{u}_i \right\|_2 \quad \text{s.t.} \quad \sum_{i=0}^k \gamma_i = 1. \quad (15)$$

This amounts to solving an inconsistent linear system  $\mathbf{U}_k \gamma = \mathbf{0}$  subject to a linear constraint. A numerically stable solution to this system can be obtained in an efficient manner via QR factorization. In particular, let  $\mathbf{U}_k = \mathbf{Q}_k \mathbf{R}_k$  be the QR decomposition of  $\mathbf{U}_k$ . Then, the  $\gamma_i$  can be computed by first solving the linear system  $\mathbf{R}_k^* \mathbf{R}_k \mathbf{d} = \mathbf{1}$ , and finally setting  $\gamma = \mathbf{d} / (\sum_{i=0}^k \mathbf{d}_i)$  [20]. This optimal estimate for  $\mathbf{s}^*$  can equivalently be obtained [20] by solving a reduced rank equivalent of Equation (11), namely

$$\mathbf{s} = \mathbf{x}_0 - \mathbf{U}_k \mathbf{W}_k^\dagger \mathbf{u}_0, \quad (16)$$

where  $\dagger$  denotes the Moore-Penrose pseudoinverse. For this reason, Mešina’s method is usually referred to as *reduced rank extrapolation* (RRE).

Another method, called *minimal polynomial extrapolation* (MPE) [6], is frequently employed along with the reduced rank method in many practical problems. The two techniques are very similar, the essential distinction being how the coefficients  $\gamma_i$  of Equation (13) are determined. It has been found that RRE and MPE behave much alike in most circumstances [18, 20, 21]. In our experiments, we found MPE to be less numerically stable than RRE, and thus employed the latter throughout our analysis.

### 3.2 Nonlinear equations and cycling

The extrapolation technique presented in the previous Section provides a solution when the input sequence is linearly generated, which is not the case in our present setting (Equation (6)). Note, however, that this nonlinearity is caused by the presence of norm constraints on the solution. We thus expect the extrapolation method to still give reasonable estimates of the limit points, as long as we provide a safeguard to avoid generating unfeasible solutions for our optimization problem. To this end, we follow an approach known as *cycling* [21]. Cycling amounts to generating a sequence  $\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2, \dots$  of approximations to  $\mathbf{s}$  according to a simple procedure, which we now describe. A *cycle* is defined by the following steps: Given an initial vector  $\mathbf{x}_0$  and integers  $n, k$ , perform  $n+k+1$  standard projected gradient iterations (6), and save the last  $k+1$  vectors as  $\mathbf{x}_i$  for  $i=0, 1, \dots, k$ ; then, apply RRE to the  $\mathbf{x}_i$  by constructing matrix  $\mathbf{U}_k$  and solving (15), as described in the previous Section, and obtain the estimate  $\mathbf{s}_{n,k}$ ; finally, due to the mixed norm constraints,  $\mathbf{s}_{n,k}$  will not in general be a feasible direction for the relaxed QAP, thus we need to perform an additional projection step  $P_\alpha(\mathbf{s}_{n,k})$ . If the final solution attains a greater objective value than the standard iterate, set  $\mathbf{x}_0 = P_\alpha(\mathbf{s}_{n,k})$  and start another cycle, keeping  $\mathbf{x}_0 = \mathbf{x}_k$  otherwise.

The cycling method has been applied with success to accelerate the convergence of nonlinear problems [18, 20]; while an accurate convergence analysis has not been carried out to date, there is empirical evidence in favour of quadratic convergence [21]. Together with the limited storage requirements and minimal overhead incurred by its application, this renders the cycling procedure an attractive solution. In the following section, we will analyse the behaviour of this approach under different choices of the two parameters  $n$  and  $k$ .

## 4 Experimental results

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

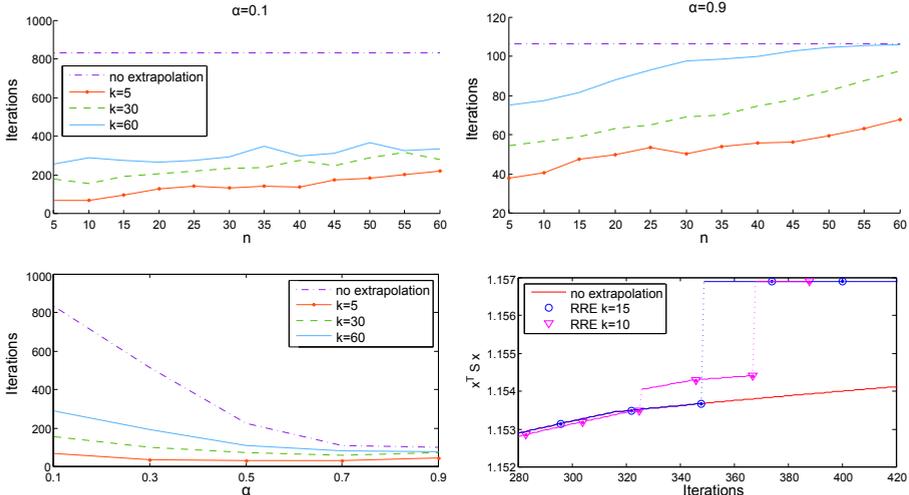


Figure 2: Sensitivity analysis to parameters  $n$  and  $k$  for the rigid matching experiments. We verified the final solutions obtained with and without extrapolation to be equivalent in all the cases. The bullets in the bottom-right graph indicate the beginning of a new cycle.

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 function  $s$ , 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.

## 4.1 Sensitivity analysis

The first set of experiments is aimed at analysing the behaviour of the matching process under different choices of parameters in the extrapolation step (namely,  $k$  and  $n$ ).

The analysis was performed in a rigid setting, with synthetic data. We generated a model point cloud  $M$  of 50 random points in  $\mathbb{R}^3$ , and produced the corresponding data cloud  $D$  by transforming a replica of the model according to a random rigid transformation. The data points were further perturbed with additive positional noise according to a Gaussian distribution with standard deviation  $\sigma = 10^{-3}$ . Finally, 10 additional random points were introduced in both model and data, thus making approximately 80% of the data points matchable with low error. This setup simulates a simple, yet realistic scenario for rigid matching. Data generation was repeated 50 times, producing as many matching instances. In all the cases, we defined the space of assignments as the whole Cartesian product  $M \times D$ , leading to a ground-truth inlier ratio of 1.4%. The similarity function was defined as

$$s((m, d), (m', d')) = \frac{\min\{\|m - m'\|_2, \|d - d'\|_2\}}{\max\{\|m - m'\|_2, \|d - d'\|_2\}}, \quad (17)$$

with  $m, m' \in M$  and  $d, d' \in D$ . Note that this definition for  $s$  enforces rigid isometries between model and data.

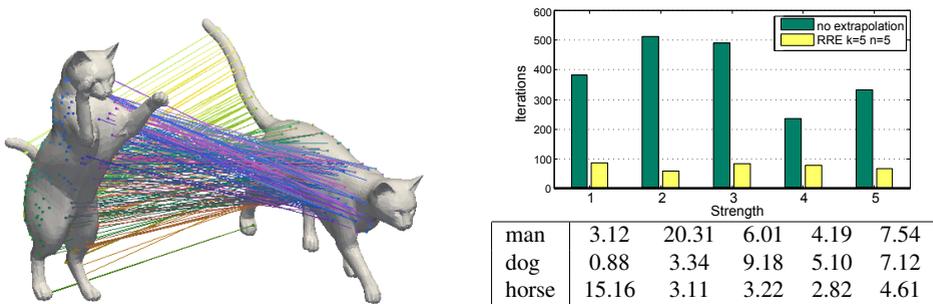


Figure 3: Example of correspondence obtained with elastic net matching on deformable shapes (left). On the right, quantitative results of the non-rigid matching experiments on the SHREC’10 dataset. The table shows the ground-truth geodesic error induced by the final solutions, for each shape across all deformation strengths. Average number of matches is 50.

For each matching instance, we varied the convexity parameter  $\alpha$  at discrete steps in the range  $[0.1, 0.9]$ , and solved the corresponding QAPs without using vector extrapolation, and by using extrapolation with cycling for different values of  $n$  and  $k$ . The results of these experiments are shown in Figure 2. The graphs in the first row plot the number of iterations (averaged over the 50 runs) required by the optimization process to reach convergence as a function of  $n$ . We chose as stopping criterion the relative change in the objective value (set to  $10^{-7}$ ). We report these results for  $\alpha = 0.1$  (first graph) and  $\alpha = 0.9$  (second graph); as expected from the elastic net formulation, the former case turned out to be more selective (30 exact matches on average) than the latter (45 matches). We note that, in general, smaller values of  $k$  attain faster convergence for fixed values of  $n$ , and the improvement is almost one order of magnitude if compared with an optimization carried out with no vector extrapolation at all. Parameter  $n$  seems to have a smaller influence over the convergence rate, with smaller values yielding faster solutions. This seems to be true for nonlinear problems in general [20]. These results indicate that vector extrapolation is a very inexpensive operation as far as both time and core memory requirements are concerned. The first graph of the second row shows the number of iterations versus  $\alpha$ , for  $n = 10$ . We notice how, for larger values of  $\alpha$ , the standard method becomes much faster as it approaches a situation in which the optimal solution is close to the principal eigenvector of  $\mathbf{S}$ . In this extreme situation the advantage of the extrapolation method is less evident, and approximately amounts to a two-fold increase in performance. The last graph of Figure 2 shows the change in objective value for the QAP across 140 iterations of the matching process. In this graph the extrapolated solutions have reached convergence, whereas the standard method is still far from the local optimum, requiring 600 more iterations to reach the same extremum with the same level of accuracy.

## 4.2 Non-rigid matching

We now turn to the considerably more challenging scenario in which the two shapes to be matched are allowed to undergo non-rigid deformations (elastic bendings, topological changes, scale, and so forth). For this set of experiments, we employed the SHREC’10 [6] and “Non-rigid world” [4] public datasets. Our similarity function in this case is defined as in Equation (17), where we replaced the Euclidean metric by an intrinsic diffusion metric [15]. We performed our analysis on quasi-isometric deformations at 5 intensity levels over three different shapes (man, dog, horse). The set of candidate matches  $C \subset X \times Y$  was reduced to

include only points that are similar enough in terms of heat kernel signature [12].

The main objective of these experiments is to assess how vector extrapolation adapts to problems of increasing difficulty due to different shapes and deformation strengths. In fact, it has been recently demonstrated [14] that the *difficulty* of matching in a non-rigid setting is directly related to the intrinsic properties of the shapes, resulting in a much more complex energy landscape if compared to a simpler rigid scenario. Figure 3 gives a qualitative example of matching obtained with  $\alpha = 0.92$  on two cat models, yielding 340 total matches. Smaller values for  $\alpha$  yielded increasingly sparse correspondences, while larger values brought wrong matches into the final solution. The bar plot on the right shows the total number of iterations of the optimization process averaged over the three shapes of the SHREC’10 dataset, while in the table below we report the corresponding average geodesic error with respect to the ground-truth matches (this is the same measure of error adopted in the SHREC contest [5]). The large increase in performance evident in the first plot suggests that there is little reason not to adopt vector extrapolation for this class of problems. In these experiments we used  $\alpha = 0.65$ , and the best values for  $k$  and  $n$  as determined in the previous Section. Note, from the table, that this choice of  $\alpha$  allows us to obtain state-of-the-art results if compared with methods yielding the same number of matches (50 on average) on the same data [5, 15].

### 4.3 Multiple-view stereo

Finally, we considered the problem of multi-view feature matching for Structure from Motion [8]. This is a scenario in which matching performance has a very direct influence over the practical applicability of 3D reconstruction methods in real-world tasks. In these experiments we used data from the Middlebury multi-view stereo dataset [16], specifically the “dino sparse ring” and “temple sparse ring” sequences, and adopted the locally-affine similarity function of [2] to populate the similarity matrix  $\mathbf{S}$ . In order to meet the strong selectivity requirements of this class of problems, we set  $\alpha = 0.2$ . Note that setting  $\alpha = 0$  would yield an equivalent formulation to the one proposed in [2], and in particular we expect comparable sparsity in both cases; since our main focus is on the increased efficiency of the matching step, we refer to Albarelli *et al.* for quantitative results on the same dataset.

In these experiments, we ran a matching process operating on the 16 successive pairs of images constituting each dataset, and accumulated the total number of iterations required by the optimization to reach a local optimum across all matching instances. The total number of iterations needed to match the whole dino sequence without extrapolation amounted to 14,328 iterations; using RRE with cycling and  $k = 5$ ,  $n = 25$  reduced the total number of iterations to 9,424. Further, we verified the final 3D reconstructions resulting from the matching process followed by a bundle adjustment step to be geometrically equivalent with and without vector extrapolation. Similarly, the temple sequence required 15,076 iterations without using extrapolation, and 10,122 iterations by using RRE with the same parameters (a detailed breakdown of this experiment is omitted for space reasons). Observe how, in this case, vector extrapolation brings comparatively little advantage if compared to the previous experiments. In particular, choosing  $n = 5$  as in the previous cases lead to premature convergence of the matching process, an effect that disappeared as we increased the cycling period to  $n = 25$ . This difference in behaviour is probably due to the very nature of this problem, where the presence of repeated structure and very similar features (as in the temple sequence) as well as of relatively feature-less objects (the dino sequence) render the whole setting more unstable and prone to getting stuck at poor local minima. Nevertheless, vector extrapolation may still be adopted whenever a higher degree of accuracy is required.

## 5 Conclusions

In this paper, we proposed the adoption of a vector extrapolation technique for correspondence problems in computer vision. Instances of these problems frequently arise in the areas of shape analysis and three-dimensional reconstruction, where the matching problem is often formulated as a quadratic assignment subject to mapping constraints. In order to assess the validity of the approach, we introduced elastic net constraints as regularizers of the QAP; this allowed us to regulate the sparsity/complexity tradeoff, and thus to evaluate the acceleration method over a family of problems of varying difficulty. Experiments on both synthetic and real data confirmed the effectiveness of the approach, leading to an increase in performance of up to one order of magnitude when compared to standard projected gradient optimization.

## Acknowledgments

We acknowledge the support of the Japan Society for the Promotion of Science under fellowship program FY2012.

## References

- [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. doi: 10.1109/ICCV.2009.5459312.
- [2] A. Albarelli, E. Rodolà, and A. Torsello. Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: A game-theoretic perspective. *Int. J. Comput. Vision*, 97(1):36–53, March 2012. ISSN 0920-5691. doi: 10.1007/s11263-011-0432-4.
- [3] D. S. C. Biggs. *Accelerated iterative blind deconvolution*. PhD thesis, The University of Auckland, 1998.
- [4] A. M. Bronstein, M. M. Bronstein, and Ron Kimmel. Efficient computation of isometry-invariant distances between surfaces. *SIAM J. Sci. Comput.*, 28(5):1812–1836, September 2006. ISSN 1064-8275. doi: 10.1137/050639296.
- [5] A. M. Bronstein, M. M. Bronstein, U. Castellani, et al. Shrec 2010: Robust correspondence benchmark. In *Eurographics Workshop on 3D Object Retrieval*, 2010.
- [6] S. Cabay and L. W. Jackson. A polynomial extrapolation method for finding limits and antilimits of vector sequences. *SIAM Journal on Numerical Analysis*, 13(5):734–752, 1976.
- [7] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Trans. Patt. Analysis and Machine Intelligence*, 18:377–388, 1996. ISSN 0162-8828. doi: <http://doi.ieeecomputersociety.org/10.1109/34.491619>.
- [8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [9] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *Proc. IEEE Intl. Conf. on Comput. Vis.*, volume 2, pages 1482–1489, 2005. doi: 10.1109/ICCV.2005.20.

- [10] D. G. Luenberger and Y. Ye. *Linear and nonlinear programming*. Springer, third edition, 2010.
- [11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.*, 11:19–60, 2010.
- [12] F. Méoli. On the use of Gromov-Hausdorff Distances for Shape Comparison. pages 81–90, Prague, Czech Republic, 2007. Eurographics Association. ISBN 978-3-905673-51-7. doi: 10.2312/SPBG/SPBG07/081-090.
- [13] M. Mešina. Convergence acceleration for the iterative solution of the equation  $x=ax+f$ . *Comput. Methods Appl. Mech. Engrg.*, 10(2):165–173, 1977.
- [14] M. Ovsjanikov, Q.-X. Huang, and L. J. Guibas. A condition number for non-rigid shape matching. *Comput. Graph. Forum*, pages 1503–1512, 2011.
- [15] E. Rodolà, A. M. Bronstein, A. Albarelli, F. Bergamasco, and A. Torsello. A game-theoretic approach to deformable shape matching. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012.
- [16] E. Rodolà, A. Albarelli, F. Bergamasco, and A. Torsello. A scale independent selection process for 3d object recognition in cluttered scenes. *International Journal of Computer Vision*, 102(1-3):129–145, 2013.
- [17] G. Rosman, L. Dascal, A. Sidi, and R. Kimmel. Efficient Beltrami image filtering via vector extrapolation methods. *SIAM J. Img. Sci.*, 2(3):858–878, August 2009. ISSN 1936-4954. doi: 10.1137/080728391.
- [18] 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. ISSN 0920-5691. doi: 10.1007/s11263-010-0322-1.
- [19] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Computer Vision and Pattern Recognition (CVPR), 2006 IEEE Conference on*, volume 1, pages 519–528, 2006. doi: 10.1109/CVPR.2006.19.
- [20] A. Sidi. Efficient implementation of minimal polynomial and reduced rank extrapolation methods. *J. Comput. Appl. Math.*, 36(3):305–337, September 1991. ISSN 0377-0427. doi: 10.1016/0377-0427(91)90013-A.
- [21] D. A. Smith, W. F. Ford, and A. Sidi. Extrapolation methods for vector sequences. *SIAM Review*, 29(2):pp. 199–233, 1987. ISSN 00361445.
- [22] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum*, volume 28, pages 1383–1392. Wiley Online Library, 2009.
- [23] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios. Dense non-rigid surface registration using high-order graph matching. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 382–389, 2010.
- [24] 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.