



HAL
open science

Adaptation of multiscale function extension to inexact matching. Application to the mapping of individuals to a learnt manifold

Nicolas Duchateau, Mathieu de Craene, Marta Sitges, Vicent Caselles

► To cite this version:

Nicolas Duchateau, Mathieu de Craene, Marta Sitges, Vicent Caselles. Adaptation of multiscale function extension to inexact matching. Application to the mapping of individuals to a learnt manifold. SEE International Conference on Geometric Science of Information (GSI 2013), 2013, Paris, France. pp.578-586, 10.1007/978-3-642-40020-9_64 . hal-02320250

HAL Id: hal-02320250

<https://hal.science/hal-02320250>

Submitted on 18 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptation of multiscale function extension to inexact matching. Application to the mapping of individuals to a learnt manifold.

Nicolas Duchateau^{*1}, Mathieu De Craene², Marta Sitges¹, and Vicent Caselles³

¹ Hospital Clínic - IDIBAPS - Universitat de Barcelona, Spain

² Philips Research, Medisys, Suresnes, France

³ Universitat Pompeu Fabra, Barcelona, Spain

Abstract. This paper targets the specific issue of out-of-sample interpolation when mapping individuals to a learnt manifold. This process involves two successive interpolations, which we formulate by means of kernel functions: from the ambient space to the coordinates space parametrizing the manifold and reciprocally. We combine two existing interpolation schemes: (i) inexact matching, to take into account the data dispersion around the manifold, and (ii) a multiscale strategy, to overcome single kernel scale limitations. Experiments involve synthetic data, and real data from 108 subjects, representing myocardial motion patterns used for the comparison of individuals to both normality and to a given abnormal pattern, whose manifold representation has been learnt previously.

1 Introduction

In this paper, we address the specific issue of out-of-sample interpolation when mapping individuals to a learnt manifold. This process aims at estimating the element of the manifold that shares the same coordinates than a tested individual, possibly out of this manifold. This approach shares similar objectives with the “pre-image” problem [10], but differs in that it explicitly looks for a formulation of the correspondence functions between the ambient space and the manifold [9] [8].

The estimation of these functions consists of two separated interpolation problems, from the ambient space to the space of coordinates parametrizing the manifold, and reciprocally [6]. Each of them can be formulated by means of kernel functions. In particular, the framework of reproducible Kernel Hilbert spaces (RKHS) allows setting the mathematical foundations to solve this problem. Its formulation is similar to the Nyström extension, and was used as the out-of-sample extension in the field of manifold learning in [2]. Some analogies to the Ridge regression and Kriging technique [13] were also reported in [4].

* Corresponding author. Address: Arrhythmia, resynchronization and cardiac imaging unit, c/ Villarroel 170, E08036 Barcelona, Spain. Email: n.duchateaucistib@gmail.com

Variants of this interpolation problem including regularization were also proposed in [1]. In the context of our application, this means assuming that the manifold may not pass exactly by the training set, namely that some dispersion of the data around the manifold can be tolerated. The interpolation can therefore be written as an exact or inexact matching problem, or a combination of both, depending on the assumptions made on the studied data [6].

However, such formulations only consider a unique scale of extension (a unique kernel bandwidth), which may be set as the average distance between the k -nearest neighbors (k -NN) over the whole dataset [9]. In fact, the data distribution and its local density (when non-uniformly distributed) may turn the issue of choosing an appropriate kernel bandwidth critical. This is likely to be the case when processing real data. Some proposed alternatives consisted in locally adapting the neighborhood size (number of k -NN) and using equal weights instead of using a kernel with a certain bandwidth [14], or locally adapting the kernel bandwidth to the neighborhood size [6]. The latter has the advantage of maintaining a kernel-based formulation of the interpolation problem, but results in a loss of symmetry of the considered kernel.

Alternatively, a multiscale scheme for kernel-based interpolation was proposed in [4], which relied on a family of functions called geometric harmonics. The algorithm was recently detailed in [3], but the technique was only described for the exact matching procedure.

In the present paper, we combine both the multiscale extension and inexact matching schemes and apply this to the above-mentioned problem of mapping individuals to a learnt manifold. Special attention is paid to the estimation of the involved parameters (the weighting term in the regularization part of the interpolation and the smallest scale performing the analysis). Experiments involve synthetic data and the real data used in [6]. Such data was previously used for the characterization of the myocardial motion pattern of a given individual with respect to normality and to a given (abnormal, in our case) pattern, whose representation had been learnt using manifold learning techniques.

2 Methods

2.1 Inexact matching problem

We formulate the interpolation problem as follows. We denote by $f : \mathcal{A} \rightarrow \mathcal{C}$ the function to be interpolated, whose values are only known at the sample locations $\{\mathbf{I}_i\}_{i \in [0, N]} \subset \mathcal{A}$, namely $\{\mathbf{x}_i = f(\mathbf{I}_i)\}_{i \in [0, N]} \subset \mathcal{C}$. The dimensionalities of \mathcal{A} and \mathcal{C} are P and M , respectively. The metrics on these spaces are denoted by $S_{\mathcal{A}} : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}^+$ and $S_{\mathcal{C}} : \mathcal{C} \times \mathcal{C} \rightarrow \mathbb{R}^+$.

We may want to constrain the interpolation to be exact at some specific locations (Fig.1a). Without loss of generality, we assume in the following that this subset is made of the first J points. Thus, we look for a function f solution

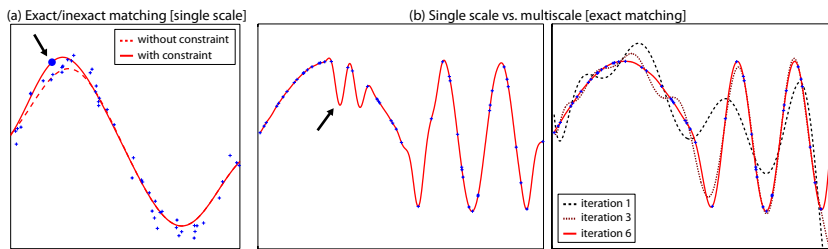


Fig. 1: (a) Synthetic dataset interpolated using single scale interpolation (inexact matching), without/with the addition of a constraint forcing the curve to pass by the point indicated by the arrow. (b) Synthetic dataset with varying density due to varying frequencies of the function parametrizing the dataset. Interpolation (exact matching for all points) using either single scale extension, the choice of a given kernel bandwidth being not robust to local variations in the density of the dataset (black arrow), or the multiscale extension scheme, which provides such a robustness.

of:

$$\begin{cases} \operatorname{argmin}_{f \in \mathcal{F}} \left(\frac{1}{2} \|f\|_{\mathcal{F}}^2 + \frac{\gamma_f}{2} \sum_{i=J+1}^N S_C(f(\mathbf{I}_i), \mathbf{x}_i)^2 \right), \\ \text{under the constraint } \forall j \in [0, J], f(\mathbf{I}_j) = \mathbf{x}_j. \end{cases} \quad (1)$$

Here, γ_f is a weighting coefficient balancing the smoothness of the interpolation and the adherence to the data. $\|\cdot\|_{\mathcal{F}}$ is the norm equipping the RKHS \mathcal{F} of functions $\mathcal{A} \rightarrow \mathcal{C}$, and is defined as $\|f\|_{\mathcal{F}} = \sum_{i=0}^N \sum_{j=0}^N \mathbf{c}_i^t \cdot K_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j) \cdot \mathbf{c}_j$, where t is the transposition operator, $K_{\mathcal{F}} : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{M}_{M,M}$ (set of $M \times M$ -dimensional real-valued matrices) is the reproducible kernel of \mathcal{F} , and $\mathbf{c}_i \in \mathcal{C}$.

The analytical solution for this problem is written:

$$f(\mathbf{I}) = \sum_{i=0}^N k_{\mathcal{F}}(\mathbf{I}, \mathbf{I}_i) \cdot \mathbf{c}_i \quad \text{with} \quad \mathbf{C} = \left(\mathbf{K}_{\mathcal{I}} + \frac{1}{\gamma_f} \mathbf{M} \right)^{-1} \cdot \mathbf{X}, \quad (2)$$

where $\mathbf{C} = (\mathbf{c}_0, \dots, \mathbf{c}_N)^t$, $\mathbf{X} = (\mathbf{x}_0, \dots, \mathbf{x}_N)^t$ are both in $\mathbb{M}_{N+1,M}$, and $\mathbf{K}_{\mathcal{I}} = (k_{\mathcal{F}}(\mathbf{I}_i, \mathbf{I}_j))_{(i,j) \in [0,N]^2}$, $\mathbf{M} = (M_{i,j})_{(i,j) \in [0,N]^2}$ are both in $\mathbb{M}_{N+1,N+1}$, with $M_{i,i} = 1 \forall i \in [J+1, N]$ and 0 otherwise. The scalar function $k_{\mathcal{F}}$ defining the kernel $K_{\mathcal{F}}$ is arbitrarily chosen as $k_{\mathcal{F}}(\mathbf{I}, \mathbf{J}) = \exp(-S_{\mathcal{A}}(\mathbf{I}, \mathbf{J})^2 / \sigma_{\mathcal{F}}^2)$, with $(\mathbf{I}, \mathbf{J}) \in \mathcal{A}^2$, $\sigma_{\mathcal{F}}$ being its bandwidth.

2.2 Multiscale extension

The multiscale extension algorithm consists in iterating such a process across scales s (from large to small scales, the kernel bandwidth being set to $\sigma_{\mathcal{F}} = T/2^s$, with $T > 0$).

The function to be interpolated at scale s is $f - F^{(s-1)}$, the algorithm being initialized with $F^{(-1)} = 0$. The application of the previously described interpolation scheme results in an approximation $f^{(s)}$ of $f - F^{(s-1)}$, and therefore to an

approximation $F^{(s)}$ of the original function f by setting: $F^{(s)} = F^{(s-1)} + f^{(s)}$. Details about up to which scale iterate this procedure are given in Sec.2.4.

The algorithm is fully detailed in [3]. It features a pre-processing step selecting the columns of $\mathbf{K}_{\mathcal{I}}$ to work at the effective rank of this matrix, as the original function f may be oversampled in some regions. The tests performed in Sec.3 of the present paper do not include such a pre-processing, the reason for it being that this involves a random selection of the effective samples. In such a case, the “optimal” parameters selected would be specific to this random selection, which is not desirable.

The results obtained by single or multiscale extension schemes are illustrated in Fig.1b.

2.3 Manifold learning and mapping new subjects to it

The interpolation scheme described above is inserted in our method for matching individuals samples, which belong to an ambient space $\mathcal{A} \subset \mathbb{R}^P$, to a learnt manifold $\mathcal{M} \subset \mathcal{A}$ of estimated dimensionality $M \leq P$. The input for the learning process is made of $N + 1$ samples $\{\mathbf{I}_i\}_{i \in [0, N]} \subset \mathcal{A}$. Learning is performed through the isomap algorithm [12], which consists in building a k -NN graph that defines a geodesic distance between any pair of samples, and an Euclidean embedding of this graph that provides a set of coordinates $\{\mathbf{x}_i\}_{i \in [0, N]} \subset \mathcal{C}$, where $\mathcal{C} \subset \mathbb{R}^M$ is the coordinate space of the surface defining the manifold.

Any new sample $\mathbf{I} \in \mathcal{A}$ is then associated to another sample $\hat{\mathbf{I}}$ laying on the manifold, using the composition of the correspondence functions $f : \mathcal{A} \rightarrow \mathcal{C}$ and $g : \mathcal{C} \rightarrow \mathcal{A}$,

$$\hat{\mathbf{I}} = g \circ f(\mathbf{I}). \quad (3)$$

Both functions are known at the samples locations $\{(\mathbf{I}_i, \mathbf{x}_i)\}_{i \in [0, N]}$, and interpolated over the whole sets \mathcal{A} and \mathcal{C} as described in Sec.2.1 and 2.2, the only difference between them being the interchanged roles of images and coordinates.

2.4 Parameters estimation

Assuming that the manifold learning parameters have been estimated earlier (for the isomap algorithm: the number of k -NN and the dimensionality M , mainly, see [6] for details), the interpolation part of the method requires a few additional parameters to be estimated: the regularization weights γ_f and γ_g , and the range of scales to run the multiscale extension algorithm. This process is illustrated in Fig.2.

We use an initialization similar to the one described in [3], setting that the starting scale $s = 0$ corresponds to the interval spanned by the set of known samples: $T = D^2/2$, where D is the distance between the most distant pair to be interpolated. The procedure is then iterated across scales, and stopped once the resolution of the known samples is reached, namely when $T/2^s \leq 2\hat{d}$, where \hat{d} is the average 1-NN distance over this dataset.

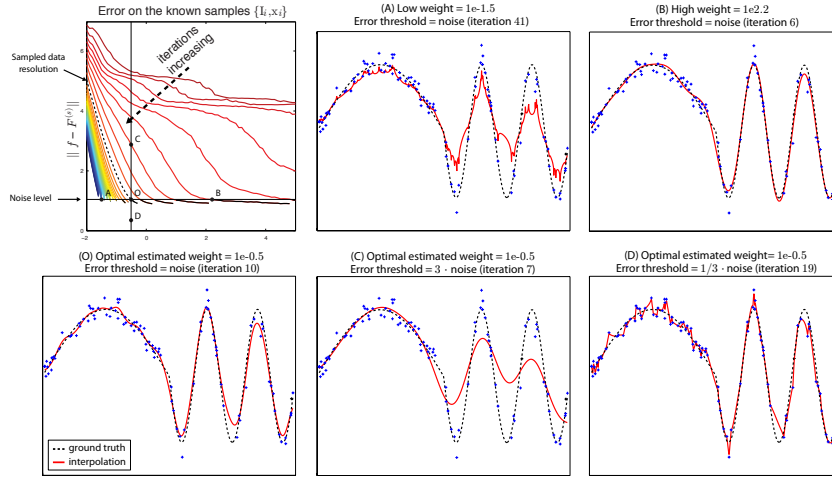


Fig. 2: Heuristic tests to determine the multiscale extension parameters: final resolution and retained regularization weight. The Euclidean distance was used for both S_A and S_C .

Heuristic tests are performed to determine the optimal regularization weights γ_f and γ_g . In case the noise level affecting the samples is known (Sec.3.1), the retained regularization weights correspond to the smallest pair (γ_f, γ_g) above which the reconstruction error $\|\mathbf{I} - g \circ f(\mathbf{I})\|$ is within this noise level. Note that this error is calculated on the known samples only, and there may be a more optimal weight with respect to the noiseless version of this data, but such a ground truth is usually unknown and cannot therefore be considered in this procedure.

In case the noise level affecting the samples is unknown (which is the case for the real data tested in the present paper, Sec.3.2), an alternative is to determine the weights to be used as the ones minimizing the generalization ability over the available dataset (the reconstruction error for points included within the range of noise of the available set, this being obtained by leave-one-out over the available samples).

3 Results

3.1 Synthetic data

We first applied the “mapping” part of the proposed method to a synthetic dataset similar to the one tested in Figs.1 and 2. Note that, here, the spaces \mathcal{A} and \mathcal{C} are respectively 2D ($x - y$ position) and 1D (parametrization by the curve length). On the contrary, Figs.1 and 2 only served for illustrating the interpolation part of the method, which was therefore 1D to 1D (x position to y position).

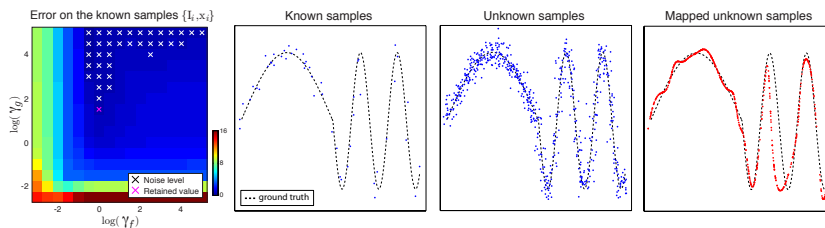


Fig. 3: Mapping individual samples to the learnt manifold: synthetic data. Heuristic tests to determine the regularization weights γ_f and γ_g involved in the multiscale extension. The Euclidean distance was used for both S_A and S_C .

This experiment is summarized in Fig.3. It confirms that the combination of the correspondence functions f and g results in mapping individual samples (originally belonging to the “high-dimensional” ambient space) to the expected “low-dimensional” structure. In this concrete example, the main errors made with respect to the ground truth (oscillations in the low frequency part of the curve, and first high frequency wave) are only due the reduced number of samples at these locations.

3.2 Real data

The whole method was then applied to real data from 108 subjects (21 healthy volunteers and 87 patients before their treatment through cardiac resynchronization therapy [CRT]).

Each individual underwent a 2D echocardiographic examination. Motion (myocardial velocities) was estimated along the sequences in a 4-chamber view, using temporal diffeomorphic free-form deformation (TDDFD [5]). These velocities were spatiotemporally aligned to a common reference using the framework of [7]. This allowed the computation of a 2D spatiotemporal abnormality map for each individual, the range for normality being defined by the subpopulation of 21 healthy volunteers.

56 patients shared a specific abnormal motion pattern of intra-ventricular dyssynchrony at baseline, also referred to as septal flash (SF) [11], the presence of which was shown to be a high determinant of CRT response.

Manifold learning was done on a subgroup of 50 SF patients, to which a synthetic image \mathbf{I}_0 having value 0 at every pixel was added, thus corresponding to a perfectly normal motion pattern. On this data, this learning process can be seen as a way to characterize a given pathological pattern as a deviation from normality along a manifold structure. The interpolations leading to f and g were set to pass exactly by this point $\{\mathbf{I}_i, \mathbf{x}_i\}$, while inexact matching was retained for the rest of the samples (Eq.1).

The composition of f and g (Eq.3) allows defining two distances $\mathcal{A} \rightarrow \mathbb{R}^+$: d_P , between any image $\mathbf{I} \in \mathcal{A}$ and the manifold [9], and d_M , which compares

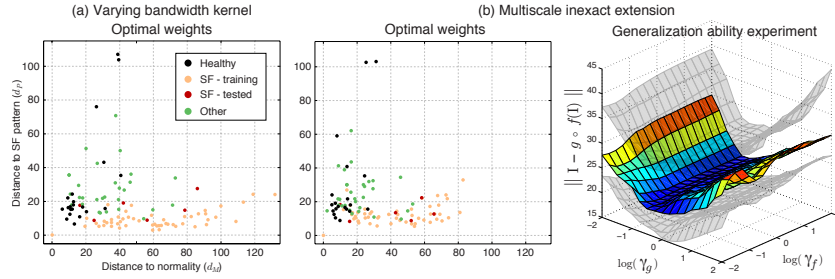


Fig. 4: Mapping individual samples to the learnt manifold: real data. Each sample corresponds to the motion pattern of a given individual (subpopulations being: healthy volunteers, CRT candidates with SF [one subgroup being used to learn the manifold representation], and CRT candidates with another abnormal pattern). Exhaustive tests to determine the regularization weights involved in the multiscale extension (generalization ability, Sec.2.4, median and first/third quartiles), and comparison with the results obtained using a kernel with locally varying bandwidth, as proposed in [6].

individuals to normality along the manifold structure:

$$d_P(\mathbf{I}) = S_A(\hat{\mathbf{I}}, \mathbf{I}) \quad \text{and} \quad d_M(\mathbf{I}) = S_C(f(\mathbf{I}), f(\mathbf{I}_0)). \quad (4)$$

In practice, we used the Euclidean distance for both S_A and S_C . The two distances d_P and d_M provide a 2D space into which any image $\mathbf{I} \in \mathcal{A}$ is mapped (Fig.4). On this specific data, they can be interpreted as distance to the SF pattern, and distance to normality along the estimated manifold, respectively.

Dimensionality and number of k -NN were estimated to be $M = 4$ and $K = 5$ [6]. Optimal values for the regularization weights were found to be $\log(\gamma_f) = -1.25$ and $\log(\gamma_g) = -0.75$ (Fig.4), and were determined as explained in Sec.2.4.

The distribution of the different subpopulations of subjects is very similar to the one obtained using the method of [6], up to a scaling factor. The arrangement of patients with SF (both training set and tested subjects) according to d_M corresponds to the amount of abnormality observed in each map, when looked at individually. They also have low d_P , the value of which (for the training ones) reflects the adherence to the data retained in the inexact matching problem). Patients with an abnormal pattern different from SF are farther from the manifold (higher d_P) and out of the normality range.

4 Conclusion

We presented a way to achieve multiscale interpolation within an inexact matching scheme, and provided insights into how to choose its intrinsic parameters. The technique was used in a specific application targeting the mapping of individuals to a learnt manifold, and involving two successive interpolations. Experiments on synthetic and real data confirmed the soundness of the method. Results were comparable to the ones obtained by the use of a varying bandwidth kernel [6],

but with the advantage that the problem of adding robustness to variations in the local density of the available samples is now well-posed.

Acknowledgements. The first author acknowledges his funding institution (Fundació Clínic, Barcelona, ES) and his collaborators on the previous steps this work is based on (G Piella [Universitat Pompeu Fabra, Barcelona, ES]; A Doltra, E Silva, MA Castel, L Mont, J Brugada [Hospital Clínic, Barcelona, ES]; AF Frangi [University of Sheffield, UK]). VC acknowledges partial support by MICINN (ref MTM2012-30772), by GRC (ref 2009-SGR-773), and by the Generalitat de Catalunya (“ICREA Acadèmia” prize for excellence in research).

References

1. M Belkin, P Niyogi, and V Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J Mach Learn Res*, 7:2399–434, 2006.
2. Y Bengio, JF Paiement, P Vincent, O Delalleau, N Le Roux, and M Ouimet. Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering. *Adv Neural Inf Process Syst*, 16:177–84, 2004.
3. A Bermanis, A Averbuch, and RR Coifman. Multiscale data sampling and function extension. *Appl Comput Harmon Anal*, 34:15–29, 2013.
4. RR Coifman and S Lafon. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. *Appl Comput Harmon Anal*, 21:31–52, 2006.
5. M De Craene, G Piella, O Camara, N Duchateau, E Silva, A Doltra, J D’hooge, J Brugada, M Sitges, and A F Frangi. Spatiotemporal diffeomorphic free-form deformation: Application to motion and strain estimation from 3D echocardiography. *Med Image Anal*, 16:427–50, 2012.
6. N Duchateau, M De Craene, G Piella, and AF Frangi. Constrained manifold learning for the characterization of pathological deviations from normality. *Med Image Anal*, 16:1532–49, 2012.
7. N Duchateau, M De Craene, G Piella, E Silva, A Doltra, M Sitges, BH Bijnens, and AF Frangi. A spatiotemporal statistical atlas of motion for the quantification of abnormalities in myocardial tissue velocities. *Med Image Anal*, 15:316–28, 2011.
8. P Etyngier, R Keriven, and F Segonne. Projection onto a shape manifold for image segmentation with prior. *Proc. IEEE Int Conf Image Process*, pages IV361–4, 2007.
9. S Gerber, T Tasdizen, PT Fletcher, S Joshi, and R Whitaker. Manifold modeling for brain population analysis. *Med Image Anal*, 14:643–53, 2010.
10. JTY Kwok and IWH Tsang. The pre-image problem in kernel methods. *IEEE Trans Neural Netw*, 15:1517–25, 2004.
11. C Parsai, BH Bijnens, GR Sutherland, A Baltabaeva, P Claus, M Marciniak, V Paul, M Scheffer, E Donal, G Derumeaux, and Anderson L. Toward understanding response to cardiac resynchronization therapy: left ventricular dyssynchrony is only one of multiple mechanisms. *Eur Heart J*, 30:940–9, 2009.
12. JB Tenenbaum, V De Silva, and JC Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–23, 2000.
13. H Wackernagel. *Multivariate Geostatistics*. Springer-Verlag, 3rd edition, 2003.
14. Z Zhang, J Wang, and H Zha. Adaptive manifold learning. *IEEE Trans Pattern Anal Mach Intell*, 34:253–65, 2012.