Pattern Recognition 47 (2014) 3265–3275

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Inexact Bayesian point pattern matching for linear transformations

J. Christmas [a,*], R.M. Everson [a], J. Bell [b], C.P. Winlove [c]

[a] Department of Computer Science, University of Exeter, Exeter EX4 4QF, UK
[b] Peninsula College of Medicine and Dentistry, University of Exeter, Exeter EX1 2LU, UK
[c] Department of Biophysics, University of Exeter, Exeter EX4 4QF, UK

ARTICLE INFO

Article history:
Received 16 March 2013
Received in revised form
11 February 2014
Accepted 26 April 2014
Available online 6 May 2014

Keywords:
Bayesian methods
Variational approximation
Point pattern matching
Iterative closest point algorithm
Linear transformation

ABSTRACT

We introduce a novel Bayesian inexact point pattern matching model that assumes that a linear transformation relates the two sets of points. The matching problem is inexact due to the lack of one-to-one correspondence between the point sets and the presence of noise. The algorithm is itself inexact; we use variational Bayesian approximation to estimate the posterior distributions in the face of a problematic evidence term. The method turns out to be similar in structure to the iterative closest point algorithm.

1. Introduction

Point pattern matching (also referred to as point set matching or point set registration) is a common pattern recognition problem that arises in many different fields, but perhaps particularly from the increasing use of automatic image processing techniques (e.g. [1–4]). A set of feature points is extracted from each of two similar images (possibly two frames of a video) and the aim is to determine correspondences between the two sets. Often it is assumed that the two sets are related through some linear transformation and any deviations from that are regarded as noise.

We denote the two sets of points as $\mathbf{Y}=\{\mathbf{y}_i\}$ and $\mathbf{X}=\{\mathbf{x}_j\}$, where each point is represented by its location in D-dimensional Euclidean space. The points do not have identities, or at least the identities are not known, and the sets of points are unordered, that is $\mathbf{y}_n$ does not necessarily correspond to $\mathbf{x}_n$ for any n.

In the simplest, *exact* case, each point set contains the same number of points and there is an exact one-to-one correspondence between them, with no noise. Thus we have the case that $\mathbf{Y}=f(\mathbf{X})$, where the function $f(\cdot)$ permutes the points in $\mathbf{X}$ and linearly transforms their coordinates so that they precisely coincide with the points in $\mathbf{Y}$. However, the nature of real problems and the automated processes by which features are often extracted, often result in the *inexact* case, where the point sets do not exactly correspond, both because of noise and because each set contains points with no counterpart in the other. In this case $\mathbf{Y}$ and $\mathbf{X}$ may contain different numbers of points and, with $\mathbf{Y}_s$ as a subset of the points in $\mathbf{Y}$ and $\mathbf{X}_s$ as a same-sized subset of the points in $\mathbf{X}$, we have $\mathbf{Y}_s=f(\mathbf{X}_s)+noise$. We refer to the points in $\mathbf{Y}_s$ and their counterparts in $\mathbf{X}_s$ as the *overlap* between the two sets.

The inexact problem has been shown to be NP-complete [5], that is the computation time required to find the global optimum increases exponentially with the number of points. Many methods therefore (including the one described in this paper) aim to find local optima in more acceptable time-frames.

Bottom-up approaches to this problem search directly for plausible point matches. In tree search algorithms with backtracking, for example, a partial (initially empty) set of mappings is progressively augmented with new mappings until a constraint is violated. The algorithm then backtracks, i.e. removes mappings, until some alternative route is available. Conte et al. [6] provide a useful overview of a number of important algorithms. In contrast, the top-down approach aims to determine the geometric transformation which relates the two point sets and uses that to find the point mappings. The iterative closest point (ICP) algorithm [1,7], for example, starts with an initial estimation of the point mappings, from which it estimates the parameters of a rigid transformation (rotation and translation) using a least squares method. The set of point mappings is recalculated based on this new estimate of the transformation. The process repeats the transformation–estimation and point-mapping steps iteratively until convergence.

We introduce an approximate Bayesian model for inexact point pattern matching which, due to the necessity of avoiding a problematic likelihood term, turns out to be similar in structure

* Corresponding author.
E-mail addresses: J.T.Christmas@exeter.ac.uk (J. Christmas), R.M.Everson@exeter.ac.uk (R.M. Everson), J.S.Bell@exeter.ac.uk (J. Bell), C.P.Winlove@exeter.ac.uk (C.P. Winlove).

http://dx.doi.org/10.1016/j.patcog.2014.04.022
0031-3203/

to ICP. We assume that the two point sets are related by a linear transformation and explicitly model each of its parameters, and the noise, as random variables. This allows us to incorporate prior knowledge about the transformation and provides estimates of the confidence intervals in the posterior distributions for each variable. We also end up with probabilities associated with every potential match; these provide a principled method for determining both the point mappings and which points in each set are unmatched. As with many Bayesian models, the integrals required for exact inference are intractable and so we use a variational approximation method [8–11]. This method minimises the Kullback–Leibler divergence [12,13] between the approximate and actual posterior distributions to determine the optimal hyperparameter values for the approximations. Interdependencies between the expressions for the posterior parameters in the variational scheme lead to an iterative update procedure which naturally results in an ICP-like update-remap process.

We start by examining different approaches to probabilistic modelling in inexact point pattern matching and the more generalised problem of graph matching, and Bayesian approximation. The new model is described in Section 2 in terms of 3-dimensional point sets, though it is easily extended to lower or higher dimensionalities. Section 3 describes the results obtained from synthetic data and in Section 4.2 the method is demonstrated on a real problem of matching cartilage cells in image stacks captured before and after a stretch is applied to the cartilage and the position and orientation of the sample in the microscope's viewing window is changed.

1.1. Probabilistic approaches in point set matching

Many inexact matching algorithms relax the tight constraints imposed on exact matching by calculating a cost associated with that relaxation; the larger the deviation the higher the cost and hence the aim is to minimise the total cost. Cost calculations often explicitly define different types of constraint violation and specify heuristically established costs with each of them. The tree search Attributed Relational Graphs algorithm [14], for example, bases the cost on graph edit operations of node and edge substitution.

An intuitive alternative to cost minimisation is probability maximisation. Continuous optimisation approaches to point pattern matching start with an initial guess at the mappings which is then refined over successive iterations. One such method is relaxation labelling [15,16], where each point in one set is assigned a vector containing the probabilities that the point is mapped to each of the points in the other set. These probabilities are initialised heuristically and then refined by taking into account the probabilities associated with adjacent points. At the end the maximum probability mapping is selected.

Relaxation labelling only enforces one-to-one correspondence in one direction. Weighted Graph Matching (e.g. [17,18]) is a quadratic optimisation method that allows two-way enforcement by way of a matching matrix of probabilities. The graduated assignment graph matching algorithm [19] gradually increases the constraints on the matching matrix to avoid poor local optima.

Although these models use probability measures, they might not be considered to be probabilistic models. A number of different probabilistic modelling approaches have been considered, using iterative expectation maximisation (EM) algorithms to find maximum likelihood solutions. Luo and Hancock [20] consider one set of points to be latent variables and the other to be observations, casting the problem as a Markov random field. Granger and Pennec [21] define a probabilistic ICP model based on a rigid transformation and a binary matching matrix, which is considered to be a latent variable. They use an annealing scheme to improve the reliability with which the global optimum is found. Jian and Vemuri [22,23] and Myronenko and Song [24] represent the two sets of points as Gaussian mixture models and maximise the likelihood of the point mappings. Xiao et al. [25] use a hidden Markov model to model the distribution of points in each of the sets and minimise the dissimilarity between the two models by minimising the Kullback–Leibler divergence between them. Serradell et al. [4] use a tree search algorithm with backtracking to learn an affine transformation that approximately aligns the two point sets as a starting point for modelling the localised perturbations as Gaussian Processes. The update of the affine transformation estimate is performed using a process similar to the Kalman filter.

A fully Bayesian technique avoids the pitfalls associated with the maximum likelihood method: integrating (averaging) over all possible values of the parameter variables guards against overfitting and posterior probability distributions (rather than point estimates) are calculated for each of them, from which we obtain a measure of confidence in the inference.

Zhu et al. [26] note that although ICP has been widely used for problems where the transformation is rigid, it does not work well if the transformation is, for example, affine. Du et al. [27] incorporate the affine transformation into ICP and use an iterative quadratic programming method to converge on a local optimum. They decompose the transformation matrix into three using singular value decomposition and then constrain these matrices to try and avoid the problem that the most likely transformation maps all of the points in one set onto a very small subset (often a single point) of the other set. Zhu et al. [26] avoid this problem by defining the mappings bidirectionally.

We represent each of the parameters of the linear transformation as separate random variables and use prior probability distributions to constrain them, both so that we may incorporate our prior knowledge about the likely transformation and to avoid the degenerate case described above. Point mappings are derived from a matching matrix containing probabilities for all possible mappings and from this we may also estimate which points in each set are unmapped.

As is often the case, calculation of the evidence or marginal likelihood for our Bayesian model is intractable, so we must resort to some approximation scheme. With a large number of variables, numerical methods, such as quadrature [28], are not feasible and sampling methods such as the Markov chain Monte Carlo algorithms of Metropolis–Hastings [29,30] and Gibbs sampling [31] (e.g., [2]) or particle filtering [32] (e.g., [3]) are too computationally expensive. Instead we estimate the posterior distributions using variational Bayesian approximation, which we describe in Section 2.2.

2. The model

We describe the model here in terms of 3-dimensional space; it is easily extended to spaces of lower or higher dimensionality.

Without loss of generality let $\mathbf{Y}$ be the smaller of the two sets and each set be independently mean-centred such that

$$\sum_{i=1}^{N_y} \mathbf{y}_i = \sum_{j=1}^{N_x} \mathbf{x}_j = \mathbf{0} \tag{1}$$

where N_y and N_x are the numbers of points in $\mathbf{Y}$ and $\mathbf{X}$ respectively. We denote a match between point $\mathbf{y}_i$ and point $\mathbf{x}_j$ as $\mathbf{y}_i \leftrightarrow \mathbf{x}_j$. In our scheme we assume that every point in $\mathbf{Y}$ is matched to a unique point in $\mathbf{X}$ and that the relationship between each pair of matched points is that of a linear transformation plus noise, encapsulated in the following expression:

$$\mathbf{y}_i = \mathbf{W}\mathbf{x}_j + \mathbf{t} + \boldsymbol{\epsilon}_{i,j} \tag{2}$$

The translation component of the linear transformation is modelled by $\mathbf{t}$; other components are captured in $\mathbf{W}$. These are considered to be