Pattern Recognition Letters 33 (2012) 1820–1827

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Robust re-identification using randomness and statistical learning: Quo vadis

Michele Nappi [a,*], Harry Wechsler [b,1]

[a] Università di Salerno, via Ponte don Melillo, 84084 Fisciano, Italy
[b] George Mason University, Fairfax, VA 22030, USA

ARTICLE INFO

Article history:
Available online 18 February 2012

Keywords:
Biometrics
Evidence-based management
Face recognition
Identity management
Re-identification
Statistical learning theory

ABSTRACT

The re-identification problem is to match objects across multiple but possibly disjoint fields of view for the purpose of sequential authentication over space and time. Detection and seeding for initialization do not presume known identity and allow for re-identification of objects and/or faces whose identity might remain unknown. Specific functionalities involved in re-identification include clustering and selection, recognition-by-parts, anomaly and change detection, sampling and tracking, fast indexing and search, sensitivity analysis, and their integration for the purpose of identity management. As re-identification processes data streams and involves change detection and on-line adaptation three complementary statistical learning frameworks, driven by randomness for the purpose of robust prediction, are advanced here to support the functionalities listed earlier and their combination thereof. The intertwined learning frameworks employed are those of (a) semi-supervised learning (SSL); (b) transduction; and (c) conformal prediction. The overall architecture proposed is data-driven and modular, on one side, and discriminative and progressive, on the other side. The architecture is built around autonomic computing and W5+. Autonomic computing or self-management provides for closed-loop control. W5+ answers questions related to *What* data to consider for sampling and collection, *When* to capture the data and from *Where*, and *How* to best process the data. The *Who* (is) query is about identity for biometrics, and the *Why* question for explanation purposes. The challenge addressed throughout is that of evidence-based management to progressively collect and add value to data in order to generate knowledge that leads to purposeful and gainful action including active learning for the overall purpose of re-identification. A venue for future research includes adversarial learning when re-identification is possibly "distracted" using deliberate corrupt information.

1. Introduction

The re-identification problem is to correctly match objects across multiple but possibly disjoint fields of view. Note that detection and seeding for initialization do not presume known identity and they allow for re-identification of objects and/or faces whose identity might remain unknown. Towards that end, once an object and/or her biometrics have been detected and initialized at one location, one, be it human and/or machine, seeks to re-identify it over space and time. This requires flexible registration to establish correspondences between pairs of images and their parts thereof using relatively invariant template representations suitable for flexible matching. The invariance requirement comes from the inherent variability in the data capture process with respect to both sensors and subjects, e.g., PIE (pose, illumination, and expression). We describe in the sequel a modular architecture that includes specific modules for the purpose of re-identification (see Fig. 1). Specific functionalities for the modules described include clustering and selection, recognition-by-parts, anomaly and change detection, sampling and tracking, fast indexing and search, sensitivity analysis, and their integration for the purpose of identity management. Additional modules, specific to task and domain, are presumed, e.g., preprocessing and image analysis. They generate suitable image representations for re-identification, e.g., using SIFT and/or Gabor filters, which are fed to the above architecture. Feedback from the complex of modules leads to further exploration and image analysis that bears on tracking and prediction.

The three learning frameworks interface along the assumptions they make and employ vis-à-vis similarity pseudo-metrics. These assumptions are, for SSL, the *smoothness assumption*, characteristic of supervised learning, where similar examples share similar labels; the *cluster assumption*, where samples in the same cluster are likely to be of the same class; and the *low density separation assumption*, to seek for decision boundaries in low-density regions. Semi-supervised learning spans the divide between supervised and unsupervised learning (Chapelle et al., 2006). Supervision is sparse and covers only some of examples, possibly using constraints, while the unsupervised aspect refers to clustering. Transduction,

* Corresponding author. Tel.: +39 (0)89 963334; fax: +39 (0)89 963303/3307.
E-mail addresses: mnappi@unisa.it (M. Nappi), wechsler@gmu.edu (H. Wechsler).
[1] Tel.: +1 703 993 1533; fax: +1 703 993 1710.

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.02.005

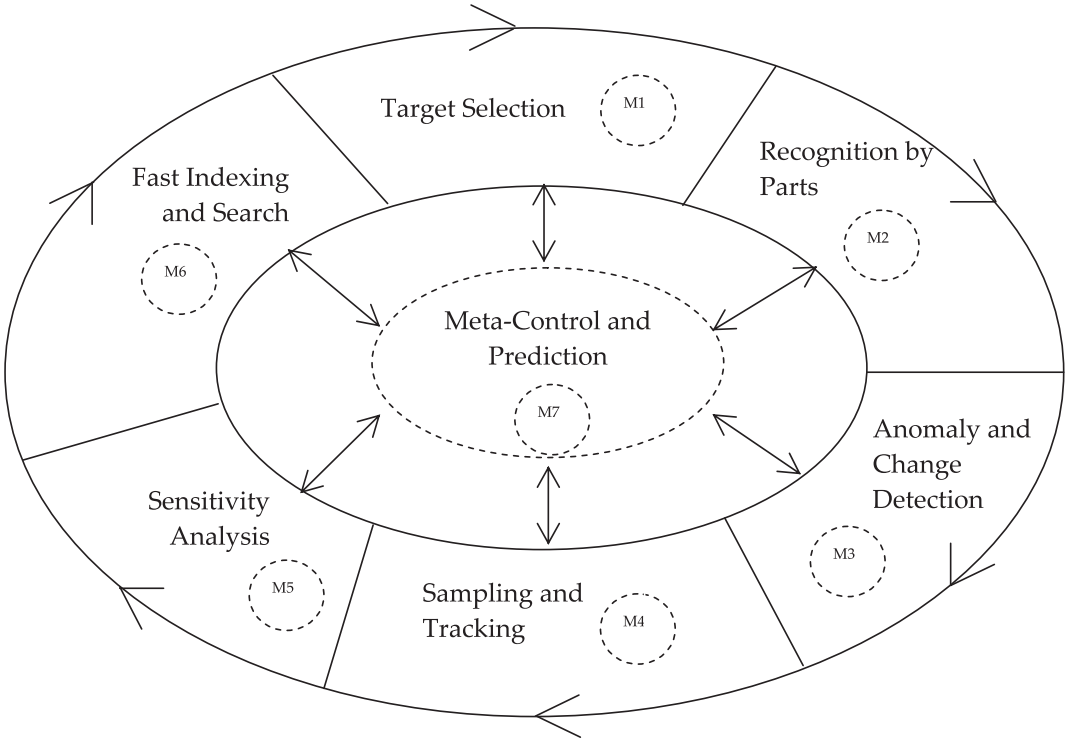

Fig. 1. Prediction for Re-Identification On-Line Using Semi-Supervised Learning and Transduction (PROUST).

which makes use of strangeness and typicality of samples, links Kolmogorov complexity and minimum description length (MDL). Last, it is the role of conformal prediction to hedge and punt on putative labels and to rank them for re-identification. This goes beyond bare predictions and includes reliability indexes for the specific choices made regarding re-identification. Reliability indexes include credibility and confidence for the predictions made, learning consistency and stability in modeling targets of interest and their recognizers, and validity and efficiency for the predictions made. Reliability indexes are derived using sensitivity analysis driven by semi-supervised learning, transduction, and conformal prediction.

The SSL framework invokes spectral clustering. Transduction, driven by randomness, seeks for learning stability using both training and test data (Vapnik, 1998). Conformal prediction (Vovk et al., 2005) ranks predictions according to their validity and efficiency. The use of the three learning frameworks is illustrated in the sequel using biometrics, in general, and face recognition, in particular. The outline of the paper is as follows. Section 2 describes target selection using spectral clustering. Section 3, on randomness and complexity, introduces strangeness and p-values for ranking and typicality purposes. Section 4, on transduction and boosting, is about data fusion and evidence accumulation. Section 5 is about recognition-by-parts using multi-layer and multi-level data fusion. Section 6 is on anomaly and change detection, while Section 7 is on sampling and tracking. Section 8 is about sensitivity analysis, while Section 9 is about fast indexing and search. Section 10 is about meta-control and meta-prediction, while Section 11 concludes the paper.

2. Module 1: target selection using spectral clustering

Target selection, an intermediate step toward re-identification and tracking, expands on the traditional but usually static and limited reach of identification and recognition. In particular, we consider here face selection, where multiple still image sets and/or video sequences for each enrollee are available during training, while a data streaming video sequence of face images, usually acquired from CCTV, becomes available during surveillance. The goal is to identify the subset of (CCTV) frames, if any, where each enrolled subject, if any, shows up. Subjects can appear and disappear as time progresses and the presence of any face is not necessarily continuous across (video) frames. Faces belonging to different subjects can appear in a sporadic fashion across the video sequence. Some of the CCTV frames could actually be void of any face, while other frames could include occluded or disguised faces from different subjects.

Spectral clustering (Dhillon et al., 2004) is a recent methodology for segmentation and clustering. The inspiration for spectral clustering comes from graph theory (minimum spanning trees (MST) and normalized cuts) and the spectral (eigen decomposition) of the adjacency/proximity ("similarity") matrix and its subsequent projection to a lower dimensional space. This describes in a succinct fashion the graph induced by the set of biometric data samples ("patterns"). Minimizing the "cut" (over the set of edges connecting k clusters) yields "pure" (homogeneous) clusters. Similar to decision trees, where information gain is replaced by gain ratio to prevent spurious fragmentation, one substitutes the "normalized cut" (that minimizes the cut while keeping the size of the clusters large) for "cut." To minimize the normal cut (for $k = 2$) is equivalent to minimize the Raleigh quotient of the normalized graph Laplace matrix L^* where $L^* = D^{-1/2}LD^{-1/2}$ with $L = D - W$; W is the proximity ("similarity") matrix and the (diagonal) degree matrix D is the "index" matrix that measures the "significance" for each node. The Raleigh quotient (for $k = 2$) is minimized for the eigenvector z corresponding to the second smallest eigenvalue of L^*. Given n data samples and the number of clusters expected k, spectral clustering (for $k > 2$) employs the Raleigh–Ritz theorem and leads to algorithms such as Ng et al. (2002) where one (i) computes W, D, L, and L^*; (ii) derives the largest k eigenvectors z_i of L^*; (iii) forms the matrix $U \in R^{n \times k}$ by normalizing the row sums of z_i to