

2011 Special Issue

Online classification of visual tasks for industrial workflow monitoring

Athanasios Voulodimos^{a,*}, Dimitrios Kosmopoulos^b, Galina Veres^c, Helmut Grabner^d, Luc Van Gool^{d,e}, Theodora Varvarigou^a^a School of Electrical and Computer Engineering, National Technical University of Athens, Greece^b Department of Computer Science & Engineering, University of Texas at Arlington, USA^c IT Innovation Centre, University of Southampton, UK^d Computer Vision Laboratory, ETH Zurich, Switzerland^e ESAT-PSI/IBBT, K.U. Leuven, Belgium

ARTICLE INFO

Keywords:

HMM
ESN
Genetic algorithm
Fusion
Workflow
Activity recognition

ABSTRACT

Modelling and classification of time series stemming from visual workflows is a very challenging problem due to the inherent complexity of the activity patterns involved and the difficulty in tracking moving targets. In this paper, we propose a framework for classification of visual tasks in industrial environments. We propose a novel method to automatically segment the input stream and to classify the resulting segments using prior knowledge and hidden Markov models (HMMs), combined through a genetic algorithm. We compare this method to an echo state network (ESN) approach, which is appropriate for general-purpose time-series classification. In addition, we explore the applicability of several fusion schemes for multicamera configuration in order to mitigate the problem of limited visibility and occlusions. The performance of the suggested approaches is evaluated on real-world visual behaviour scenarios.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Intelligent visual surveillance and classification of visual tasks are research fields that have rapidly gained momentum over recent years. Focusing on industrial plant smart monitoring, the aim is to recognise tasks happening in the scene, to monitor the smooth running of a workflow, and to detect any abnormal behaviour. Deviations from the workflow may cause severe deterioration of the quality of the product or may raise safety or security hazards.

An example of such an industrial scenario is shown in Fig. 1. By monitoring industrial scenes, one faces several challenges such as recording data in work areas (camera positions and viewing area), industrial working conditions (sparks and vibrations), cluttered background (upright racks and heavy occlusion of the workers), high similarity of the individual workers (nearly all of them wearing a similar utility uniform), and other moving objects (welding machines and forklifts). Furthermore, the dynamics of the workflow can be quite complex. Several tasks within a workflow can have very different lengths and can be permutable. The high intraclass and low interclass variances make the classification process significantly challenging. Moreover, the tasks can include

both human actions and motions of machinery in the observed process.

Related work. Behaviour and workflow recognition has attracted the interest of many researchers. In the computer vision and machine learning communities, this is mainly addressed in applications such as abnormal behaviour recognition or unusual event detection. Many approaches have been suggested over recent years—reviews can be found in (Poppe, 2010; Turaga, Chellappa, Subrahmanian, & Udrea, 2008). Typically they build a model of normality, and the methods can differ in (i) the model used, (ii) the algorithm employed for learning the model parameters, and (iii) the features used. Models might be previously trained and kept fixed (Antonakaki, Kosmopoulos, & Perantonis, 2009; Wang, Ma, Ng, & Grimson, 2008) or adapt over time (Breitenstein, Grabner, & Gool, 2009) to cope with changing conditions. A broad variety of extracted image features are used, such as global scene three-dimensional (3D) motion (Padoy, Mateus, Weinland, Berger, & Navab, 2009) or object trajectories (Antonakaki et al., 2009; Johnson & Hogg, 1996; Nguyen, Phung, Venkatesh, & Bui, 2005; Shi, Huang, Minnen, Bobick, & Essa, 2004), which require accurate detection and tracking. On the other hand, holistic methods, which define features at the pixel level and try to identify patterns of activity using them directly, can bypass the challenging processes of detection and tracking. Such methods may use pixel or pixel group features such as colour, texture, or gradient; see, for example, (Zelnik-Manor & Irani, 2006)

* Corresponding address: School of Electrical and Computer Engineering, National Technical University of Athens, Heron Polytechniou 9, 15780 Athens, Greece. Tel.: +30 210 772 2559.

E-mail address: thanosv@mail.ntua.gr (A. Voulodimos).

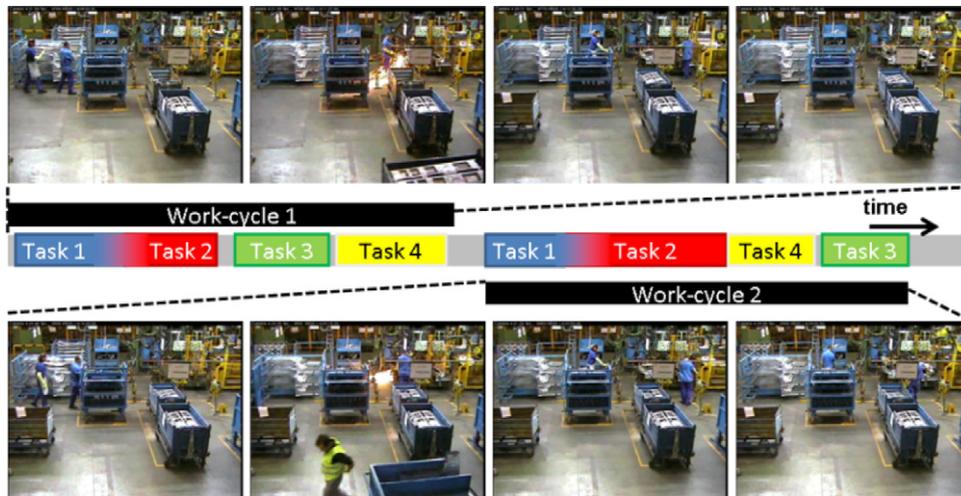


Fig. 1. Example of an industrial scenario: the workflow consists of several tasks of different length, which can be permutable (e.g., tasks 3 and 4 in this example), and the scene may present irrelevant motion and occlusions.

(histograms of spatiotemporal gradients) and (Laptev & Perez, 2007) (spatiotemporal patches). Pixel change history (PCH) is used in (Xiang & Gong, 2006) to represent each target separately after frame differencing. However, the representation of objects in PCH images is very simplistic (through ellipses), and cannot cope with realistic environments. A popular feature to use for action recognition is optical flow (see, e.g., Efros, Berg, Mori, & Malik, 2003), where a relatively small region of interest is extracted around a single human actor. In our case we need a much more efficient method, since our goal is online classification at high frame rates. Furthermore, in real applications, the targets may be partially occluded, so action recognition as defined in works such as (Efros et al., 2003) would not be feasible.

Various machine learning and statistical methods have been used for activity recognition, such as clustering (Boiman & Irani, 2005) and density estimation (Johnson & Hogg, 1996). A very popular approach is hidden Markov models (HMMs) (Ivanov & Bobick, 2000; Lv & Nevatia, 2006; Padoy et al., 2009), due to the fact that they can efficiently model stochastic time series at various time scales. However, the HMMs assume that the input data are already segmented, an assumption which significantly limits their application in realistic applications. For this purpose, more complex HMM-based methods have been proposed such as hierarchical HMMs (HHMMs) (Fine, Singer, & Tishby, 1998; Padoy et al., 2009) and layered HMMs (LHMMs) (Oliver, Garg, & Horvitz, 2004). However, the applicability of these methods assumes that the Markovian assumption holds for the tasks to be recognised, in other words the probability for the appearance of a task depends only on the previous one; this is not true in structured applications, where the execution of a task may influence the appearance of a series of following tasks. In such cases, the Markovian assumption would be an oversimplification, which would violate the application constraints. The use of higher-order models would result in very high complexity (Rabiner, 1989) and would raise issues such as “how many previous states do we have to consider?”. With a small number of tasks the problem could be still tractable; however, such approaches are not scalable to large numbers of tasks.

In (Shi et al., 2004), the feasible task paths in a glucose calibration process were defined, using the so-called P-net to encode possible paths. The goals in our work are similar, but here we aim to show how to employ the HMM framework for recognising tasks in workflows, because of its very important extension possibilities (for example, with fusion (Zeng, Tu, Pianfetti, & Huang, 2008) or robustness (Chatzis, Kosmopoulos, &

Varvarigou, 2009)); furthermore, we are going to encode possible paths as solutions provided by a genetic algorithm to cover a huge search space efficiently.

An alternative approach to the HMM for the analysis of complex dynamical systems is echo state networks (ESNs) (Jaeger, 2001). ESNs offer several benefits, such as (i) fast and simple learning of many outputs simultaneously, (ii) the possibility of both offline and online learning, (iii) the capability of directly dealing with high-dimensional input data, and (iv) the ability to learn complex dynamic behaviours without any explicit Markovian assumption. On the other hand, there are two main limitations involved: (i) they can only recognise repetitive dynamics and (ii) all significant variations of task order in a given workflow have to be learnt to provide the best classification results. Previously, ESNs have been successfully used for time-series classification in speech recognition (Skowronski & Harris, 2007), human-robot interactions (Hellbach, Strauss, Eggert, Komer, & Gross, 2008), emotion recognition (Scherer, Oubbati, Schwenker, & Palm, 2008), and medicine (Verplancke et al., 2010). Recently, we examined the effectiveness of ESNs for workflow recognition from a single camera (Veres, Grabner, Middleton, & Gool, 2010).

Nevertheless, the target visibility of specific tasks can be limited due to camera configuration and self-occlusions; therefore efficient ways to fuse observations from multiple cameras are necessary. Several fusion schemes for HMMs have been presented in the past, such as synchronous HMMs (Dupont & Luetin, 2000), parallel HMMs (Vogler & Metaxas, 1999), and multistream fused HMMs (Zeng et al., 2008). However, their applicability in multicamera systems has been examined only to a limited extent, for example in (Voulodimos, Grabner, Kosmopoulos, Van Gool, & Varvarigou, 2010), a previous work that is extended in this paper to address online behaviour and workflow recognition in continuous data streams, and in (Kosmopoulos & Chatzis, 2010), where offline classification of segmented sequences was examined. As far as ESNs are concerned, to our knowledge no fusion techniques have been employed for similar applications.

Contribution. To our knowledge, no state-of-the-art tracking-based approach is able to cope with the significant particular challenges (as described above) of workflow analysis in continuous streams within industrial environments. We tried state-of-the-art methods for person detection/tracking (Felzenszwalb, McAllester, & Ramanan, 2008; Grabner & Bischof, 2006)¹; however, none

¹ The code was downloaded from the authors' webpages.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات