Automated Detection and Tracking of Slalom Paddlers from Broadcast Image Sequences using Cascade Classifiers and Discriminative Correlation Filters

Ami Drory[a,*], Gao Zhu[a], Hongdong Li[a,b], Richard Hartley[a,b,c]

[a] *Australian National University, Canberra, Australia*
[b] *Australian Centre for Robotic Vision*
[c] *Data61, CSIRO, Australia*

Abstract

This paper addresses the problem of automatic detection and tracking of slalom paddlers through a long sequence of sports broadcast images comprised of persistent view changes. In this context, the task of visual object tracking is particularly challenging due to frequent shot transitions (i.e. camera switches), which violate the fundamental spatial continuity assumption used by most of the state-of-the-art object tracking algorithms. The problem is further compounded by significant variations in object location, shape and appearance in typical sports scenarios where the athletes often move rapidly. To overcome these challenges, we propose a *Periodically Prior Regularised Discriminative Correlation Filters* (PPRDCF) framework, which exploits recent successful Discriminative Correlation Filters (DCF) with a periodic regularisation by a prior that constitutes a rich discriminative cascade classifier. The PPRDCF framework reduces the corruption of positive samples during online learning of the correlation filters by negative training samples. Our framework detects rapid shot transitions to reinitialise the tracker. It successfully recovers the tracker when the location, view or scale of the object changes or the tracker drifts from the object. The PPRDCF also provides the race context by detection of the ordered course obstacles and their spatial relations to the paddler. Our framework robustly outputs the evidence base pre-requisite to derived race kinematics for analysis of performance. Experiments are performed on task-specific dataset containing Canoe/Kayak Slalom race image sequences with successful results obtained.

Keywords: Detection, Tracking, Cascade Classification, Discriminative Correlation Filter, Multi-class SVM, Canoe Kayak Slalom, Shot Transition, Sports Biomechanics, Performance Analysis

1. Introduction

In competitive Canoe/Kayak Slalom (CK Slalom), negotiation of obstacles through gates is the fundamental skill and key determinant of overall performance. In race context where the winner is commonly decided by fractions of a second, minimising task time-to-completion is paramount. Thus, developing an optimal strategy and techniques for negotiation of gates that minimises overall course time-to-completion is critical. However, there is currently little quantitative data that characterises the trajectory of gate negotiation in Slalom.

Through extensive literature survey, we have found but only one paper that attempted to characterize the strategy employed by slalom paddlers in negotiation of upstream gates [24]. It analyzed upstream gate negotiation strategies of 17 elite Slalom paddlers using manual extraction of spatial kinematic data of the boat and athletes' head from image sequences obtained by overhead camera. The utility of the methodology used by [24] is, however, limited by the use of a custom calibration rig when there is no water on the course, obtrusive attachment of markers to the boat and athlete, and laborious object labelling for extraction of trajectory kinematic information. In order to be relevant in elite sport training environment or competition and improve the likelihood of feedback driven technical or tactical amendments, an analysis method must provide near real-time results.

In this work, we investigate the challenging problem of simultaneous human detection and long-term tracking from readily available image sequences comprised of persistent view changes obtained from multiple uncalibrated cameras typical of broadcast image sequences. This task serves as a crucial evidence base, a pre-requisite to kinematic motion analysis of athletes aimed to optimise technique and performance in sport (see figure 2). We aim to tackle the limitations of existing visual object detection and object tracking algorithms especially for long term sequence with frequent view changes. We develop a new and unified framework for object detection and tracking from disparate multi-view image sequences that couples the advantages of each approach to overcome the limitations of the other. The method is applied to detection and tracking of CK Slalom paddlers through gate negotiation of a race course, which enables near real-time performance analysis.

*Corresponding author
Email address: ami.drory@anu.edu.au (Ami Drory)

Figure 1: Our PPRDCF framework outputs location and scale of the slalom paddler, and the location and order of the gates.

Contributions. A Periodically Prior Regularised Discriminative Correlation Filters (PPRDCF) framework is proposed for tracking fast moving objects in sport event using broadcast image sequences with possibly frequent shot transitions. Our framework exploits recent successfully applied Spatially Regularised Discriminative Correlation Filters (SRDCF) [7] with a periodic regularisation by a prior discriminative cascade classifier that is learnt offline. To overcome tracking failure associated with rapid shot transition, we introduce a robust adaptive shot transition detection algorithm that allows soft initialisation of the tracker. Finally, our framework provides race context through the detection of course obstacles and their spatial relations to the paddler. We perform experiments on task-specific dataset containing CK Slalom race image sequences and compare our results to state-of-the-art trackers. Our framework robustly outputs the evidence base pre-requisite to derived race kinematics for analysis of performance.

2. Related Work

A comprehensive survey of visual object tracking is outside of the scope of this paper. Instead, this section presents a brief survey of recent techniques relevant to our task, to provide the context for our new method.

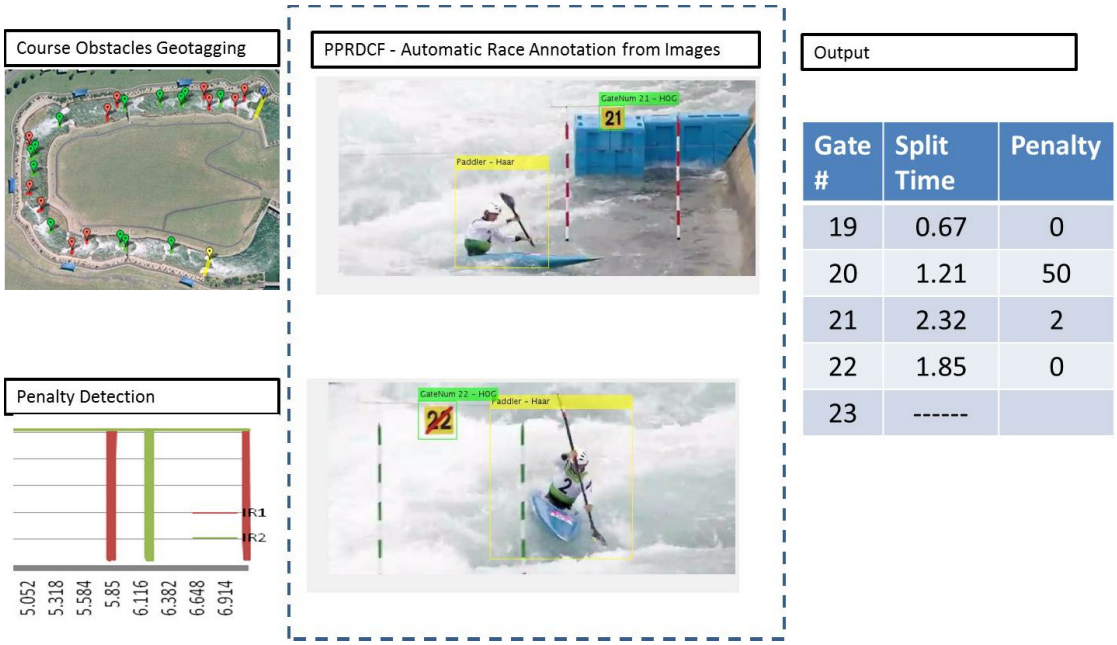
Visual Tracking

Visual tracking is an important computer vision problem of estimating an object's kinematics from an image sequence. State-of-the-art tracking algorithms generalise the object's appearance from a small set of training samples. The tracker then performs temporal search for probabilistically matching candidates in the spatial vicinity of the object's previous image location under the assumption of trajectory smoothness [3, 19]. Online learning trackers update the model with the selected candidate [2, 16, 8].

Human movement tracking is challenging due to the varied pose and appearance caused by severe occlusions induced by the articulated body motions. The challenge is compounded in typical broadcast of sporting races, where an athlete rapidly changing pose, occlusion and appearance. In our CK Slalom task, the pose can rapidly change form front to rear view, and from top view of paddler and boat to bottom view of the boat and no visibility of the paddler. Further, the paddler is often partially or fully submerged or severely occluded from view by obstacles or water. These present critical challenges to tracking algorithms that assume small changes in the object's pose or appearance.

Many tracking algorithms model the image background [33] or extract a temporal flow field [4, 10] to aid the object tracking under the strong assumption that the object differs from the background in either appearance or motion. For example, most tracking benchmark datasets constitute image sequences of a stationary background (e.g. roads, streets or buildings) and a moving object (e.g. cars, bicycles or pedestrians). In our CK Slalom task, however, the background water often flows in the same direction as the paddler and rapidly changes in appearance due to illumination and reflections, essentially eliminating the realistic option of background modelling. Moreover, image sequences with rapid shot transition from multiple moving cameras remain the majority of available race and training data. This severely violates the global smoothness and brightness assumptions for dense correspondences requirements of tracking algorithms [22, 4]. Hence, shot transition detection and regularisation or re-initialisation of the tracker model is paramount to enhance the chance of recovery from occlusion or loss of track, contamination by negative samples, or rapid change in pose or appearance due to fast motion, or sudden change in view.

Recent best performing tracking algorithms use a Fast Fourier Transform (FFT) based Discriminative Correla-



Gate #	Split Time	Penalty
19	0.67	0
20	1.21	50
21	2.32	2
22	1.85	0
23	------	

Figure 2: An illustrative schematic overview of a CK Slalom annotation system for the daily training environment and competition. The system includes global race course and obstacles geotagging, penalty detection and race annotations from image sequences, and outputs a detailed comprehensive race annotations including split times and penalties. This paper focuses on the race annotations from image sequences (encompassed by the dashed line).

tion Filter (DCF) approach (see section 3.1). The approach, however, accumulates errors during online learning and typically drifts from the object, as detection recovery after occlusion is poor [7]. Consequently, even the current state-of-the-art tracker is not robust in long-term tracking of rapidly moving and deforming objects (see fig. 3).

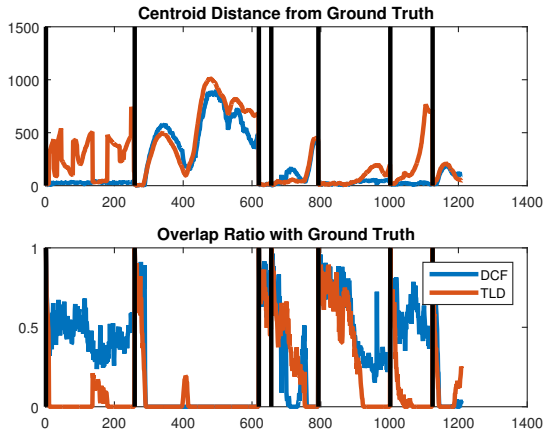


Figure 3: Results of state of the art trackers on our test data demonstrating rapid tracking deterioration, manifested by rapid loss of overlap with ground truth (bottom figure; value approaching 1 is indicative of good performance) and increase in precision score (top figure; value approaching 0 is indicative of good performance) within camera view (shot). Vertical black lines indicate shot transition (a sudden change of camera view). This violates the continuity assumption of the tracker and requires reinitialisation (here, by ground truth).

Object Detection

Similar to visual tracking, object detection, often a prerequisite to tracking algorithms, is a challenging computer vision problem due to the variable appearance and pose that may be present. Robust discriminative methods extract an extensive set of features at multiple scales from positive and negative image samples to train a classifier. Due to the very large feature set, learning is computationally expensive and typically performed offline. The resultant classifier is comprised of rich feature descriptors that capture the object of interest. In inference, a multi-scale sliding window search scheme scores candidate patches and the best scored patches are selected as detections. While this approach has been very successful [9, 15, 13, 6, 39], the computational cost of feature extraction and exhaustive search restricts the scheme from being used for tracking at every time step of an image sequence. Moreover, a detection score does not guarantee temporal continuity of an object detection. Post-hoc regimes are required to enforce spatial continuity [21, 36]. Furthermore, the scheme is susceptible to candidate proposal multiplicity. A heuristic, adaptive [29] or learnt score threshold influences the number of detections selected and subsequently the incidence of false positive and false negative detections.

3. Periodically Prior Regularised DCF (PPRDCF)

In this section, we introduce our Periodically Prior Regularised Discriminative Correlation Filters (PPRDCF) unified framework that couples the advantages of each technique to overcome the limitations of the other. Our

framework exploits recent successfully applied Spatially Regularised Discriminative Correlation Filters (SRDCF) [7] with a periodic regularisation by a prior discriminative cascade classifier that is learnt offline. We introduce the tracking framework in section 3.1 followed by the object detection using a cascade of rejectors classifiers in section 3.2. Section 3.3 introduces our shot transition detection algorithm that enables soft re-initialisation of the tracking framework. In addition, our framework provides race context by detecting the ordered course obstacles and their spatial relations to the paddler (further details are provided in section 4.4).

3.1. Discriminative Correlation Filters

Discriminative approaches cast tracking as an online learning and classification problem that differentiates the tracked object from the background. Given an image patch containing the object, a classifier is learnt that discriminates the object from the environment, a process akin to tracking-by-detection. Robust object detector classifiers, however, richly characterise not only the object of interest, but importantly its environment through a very large number of negative samples. This computationally expensive and time consuming typically offline process is not feasible for online tracking algorithms, for which speed is a critical performance criterion. For this reason, discriminative tracking approaches use a compromise approach that severely under-samples negatives [2, 17], consequently, acutely hindering the tracker's performance [20].

Current state-of-the-art tracking algorithms use DCF approach, in which a correlation filter is trained from a set of samples and a periodic extension of these samples [3, 27, 20]. Significant improvements in trackers' performance has risen from recent work that formulates the convolution of two patches as an element-wise product in the Fourier domain [3]. Henriques et al [19] demonstrated that translations of an image patch containing the object can be modelled as cyclically shifted signals using circulant matrices. Thus, the classifier training and detection computational cost is significantly reduced when the computations are efficiently performed using FFT.

The periodic extension reduces the contamination of the filter by negative samples. Consequently, a better representation of the object is learned. The approach, however, suffers from boundary contamination effects that result in inferior representation of the object and introduce inaccuracies to the learned object model. Thus, reducing its discriminative power [7, 14]. This problem was partially addressed in Danelljan et al. [7] by utilising spatial regularisation component in the objective function.

3.2. Cascade of Rejectors Classification

The object detection problem involves recognition of the desired object, its location and scale in an image. We note that in recent years deformable parts based methods (DPM) outperform cascade classifiers on standard object detection tasks [13, 38, 5]. DPM, however, strongly relies on a spatial relations of parts model, which cannot handle severe occlusions that are commonly present in our task.

More recently, deep learning of convolutional neural network (CNN) produced the state-of-the-art performance on standard detection tasks [32, 15]. In CNN, high level features replace low and middle level features with improved discriminative power. Notwithstanding, these features are very expensive to compute. Selective search strategies to reduce the computational cost of using CNN [15] resulted in object localisation errors [21]. Importantly, both low computational cost and accurate object localisation are critical to our framework. For these reasons, we opt to use cascade classification for object detection in our framework tasked with tracking initialisation and periodic tracking regularisation.

A cascade detector uses a sequence of node classifiers to distinguish objects from non-objects and simultaneously select weak features to form strong ensemble classifiers using adaptive boosting (AdaBoost). The work of Viola and Jones [34] leverages the scarcity of the object of interest relative to the background to achieve efficient detection by early rejection of most easily classified negative features. They also introduced integral images for fast feature computation and utilising AdaBoost for automatic feature selection. These ideas remain a foundation for modern detectors.

Viola and Jones[35] used low level Haar features due to low computational cost achieved with the aid of integral images. However, the cascade classification approach can be used with other feature descriptors. Significant improvements were obtained by using mid-level features, such as Histogram of Oriented Gradients (HOG) in detection [6] and in speed [39].

3.3. Shot transition

Typical broadcast sport image sequences are characterised by frequent shot transitions. We introduce a robust adaptive shot transition detection algorithm that allows soft initialisation of the tracker. Further details are provided in 4.3

3.4. Our PPRDCF framework

Conceptually, our framework is most similar to Kalal et al. [25] in adopting a unified framework that distinguishes between the detection and tracking tasks. Kalal et al. [25] described a framework of three sub-tasks of Tracking, Learning and Detection (TLD). The TLD uses a naive geometric shape template matching method with median flow for tracking and a cascade classifier with online learning. We argue that while operating independently, the aggregation of the TLD's three sub tasks are equivalent to a modern DCF tracker with an online learning of the tracked patch. Therefore, due to its online learning component it suffers from the error accumulation problem of DCF trackers. Instead, we opt for using a true independent offline learnt detector to complement an online

DCF tracker. Furthermore, to enhance the overall framework performance, our detector uses a different feature descriptor than the tracker. This enhances the cumulative discriminative power, as the two models hold complementary characteristics of the object.

In agreement with Kalal et al. [25], we accept the view that neither tracking nor detection can solve the task independently. We support the view that the two approaches can be complementary. A detector can initialise the tracker, provide tracking validation and failure recovery to a tracker. A tracker accumulates temporal object localisation and can reduce the computational cost and running time of the detection.

We define a spatial discrepancy signal between the prior classifier and the tracker. To overcome the inherent limitation of tracking algorithms, the object evidence accumulated by the tracker and the discrepancy signal are used to prune false positive and overcome false negative detections. Within each shot sequence, our proposed PPRDCF formulation introduces a penalty term on the correlation filter coefficients during online learning. The prior regularisation reduces the corruption of positive samples during online learning of the correlation filters by negative training samples. Consequently the PPRDCF successfully recovers the tracker when the location, view or scale of the object changes or the tracker loses the object.

Alternatively, DCF tracking can be characterised as tracking-by-detection. However, unlike robust object detector classifiers, rich characterisation of the object is not possible due to the high computational cost. Hence, DCF uses a compromise approach that restricts the object region and severely under-samples negatives acutely hindering the trackers performance [20]. Our framework can then be viewed as tracking by weak detection classifier with periodic update by a rich detection classifier. Our observation is that the weak online learnt detector is likely to become contaminated, whilst the rich offline learnt detector will remain immune to contamination and will retain its strong discriminative power.

The basic structure of our framework is as follows; For initialisation and regularisation of the tracker we construct a rejection cascade classifier similar to Viola and Jones [34] and described in section 4.1. For tracking, we construct a DCF following Danelljan et al. [7] as detailed in section 4.2. For shot transition detection we use an adaptive outlier detection method described in section 4.3. The race annotation component is detailed in section 4.4. Figure 4 depicts the outline of our framework.

4. System Implementation

4.1. Paddler Detection

Our object detection framework overview is depicted in figure 5 for learning and in figure 6 for inference. We construct a rejection cascade similar to Viola and Jones [34]. Essentially, a cascade classifier forms a degenerate decision tree, where a negative classification of an image patch results in rejection of the patch. A positive classification is passed on for evaluation at the subsequent classifier. In this manner easily classified patches are rejected early with improved overall efficiency due to the observation that an image consists of mostly negative samples. Only positive samples will be evaluated by a classifier at every stage.

At each stage a classifier is trained on the examples that were evaluated as positives in all preceding stages. Consequently, the classifier's complexity and discriminative power increases as stages increase due to the escalating task difficulty. We use 20 stages in our implementation, predicated on our experiments described in section 5.3.

The computational cost of training a cascade classifier is significant. Inference, however, is fast due to the cascade of rejectors and boosting. This makes the approach suitable for complementary detection in our tracking framework.

In inference, a sliding window approach is employed to evaluate the classifier score function f over rectangular sub-regions of the image I at multiple scales. We select the object's region $\tilde{R}$ to be its maximum as

$$\tilde{R} = arg \max_{R \subseteq I} f(R|\tilde{\mathbf{x}}), \tag{1}$$

where $\tilde{\mathbf{x}}$ is the learnt object's appearance model and R ranges over all rectangular sub-regions of the image I. We incrementally scale the search region as defined by $\lceil (mTS \cdot SF^n) \rceil$, where mTS indicates the median patch size used to train the classification model (117×124), SF a scale factor determined by the ratio between the size of the input image I and mTS and the number of increments N(7 in our implementation), and $n \in \{1, \ldots, N\}$ is an indicator function of the current increment.

In our experiments (see section 5.3), using rich mid-level feature descriptors (HOG and LBP) in our detection classifier results in rare false positives, but high levels of false negatives. In contrast, using Haar features produced fewer false negatives, but significantly more false positives. In our framework, detection of false positives corresponds to a high discrepancy signal between detector and tracker. Hence, they are naturally handled by a severe penalty imposed by a penalty function (eq. 5) that changes the tracker's online learning parameters.

Furthermore, Wojek and Schiele [36] showed that a combination of several feature descriptors outperforms any single feature descriptor. Considering our tracker already uses HOG features, a complementary detector using a different feature descriptor is preferred, as it is likely to capture ancillary aspects of the object. For these reasons, we opt for using Haar features in our detector model.

4.2. Paddler Tracking

The standard DCF tracker [19, 20] is essentially a regressor $g(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle$, where ϕ represents the mapping to the Hilbert space induced by a kernel function and $\mathbf{w}$ is the discriminative model. Considering all the previous image patches $\{\mathbf{x}^k,\ k = 1, ..., t-1\}$ of size $M \times N$ centred

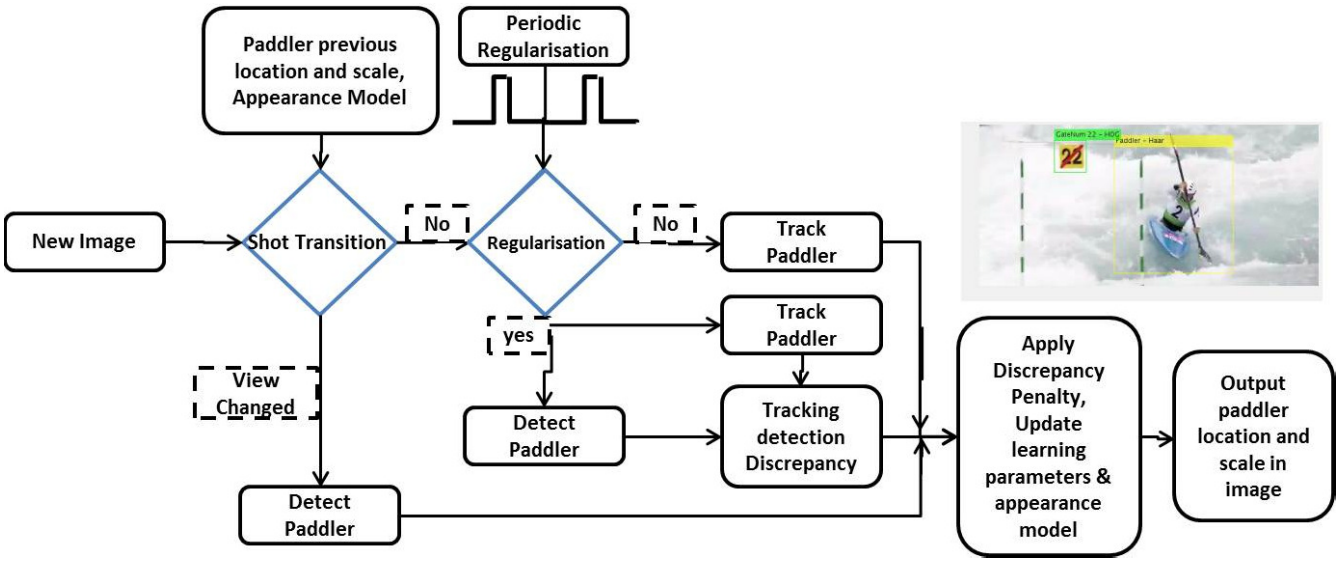


Figure 4: Overview of our PPRDCF algorithm overview (see details in text)

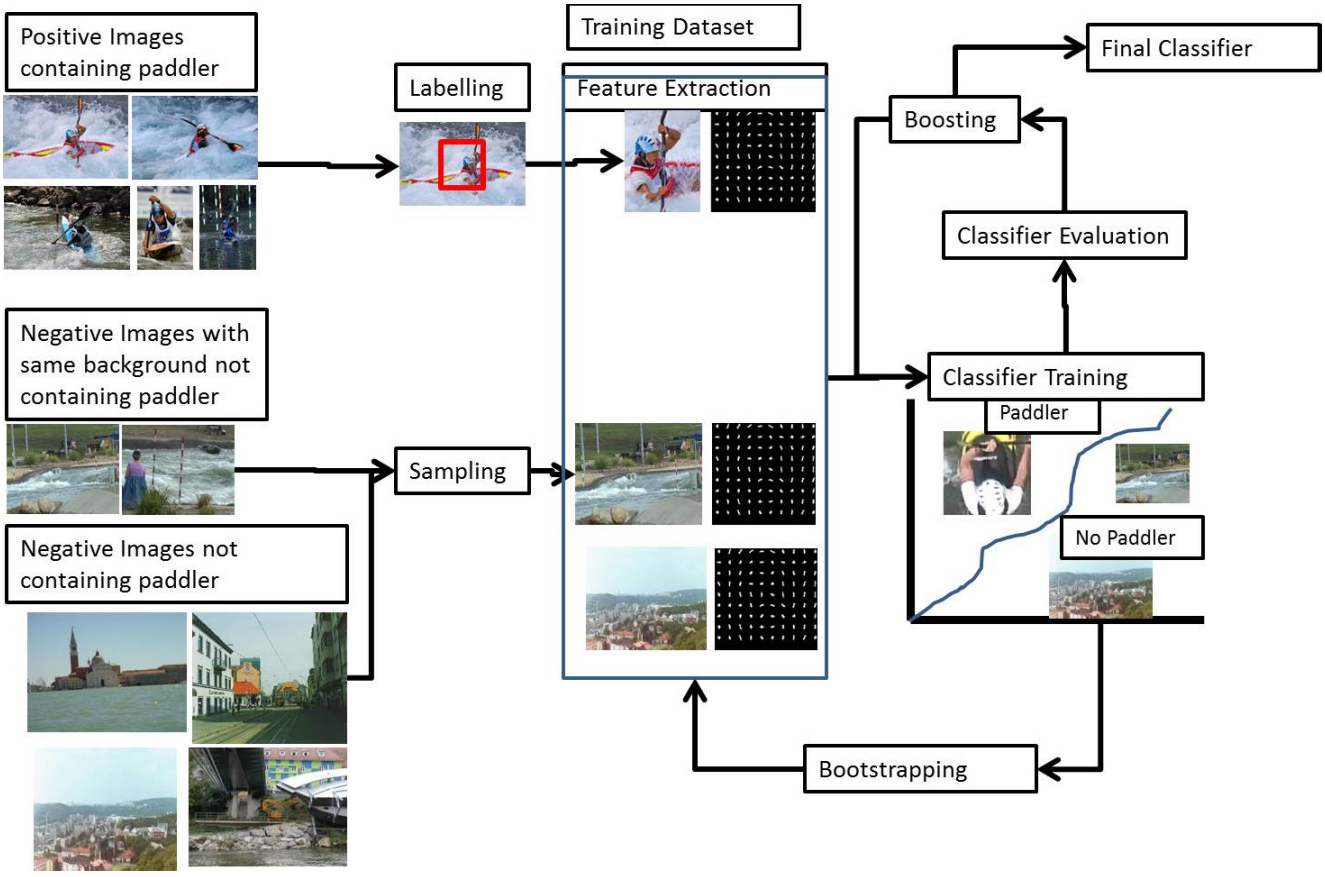


Figure 5: Training a cascade classifier of a Slalom paddler. See text for details.

on the object, this regressor can be effectively trained by optimizing

$$\min_{\mathbf{w}} \sum_{k=1}^{t-1} \alpha_k \sum_i (\langle \mathbf{w}, \phi(\mathbf{x}_i^k) \rangle - y_i)^2 + \lambda \|\mathbf{w}\|^2, \qquad (2)$$

where k denotes the frame index and α_k is the frame weight. The matrix $\mathbf{x}_i^k$, $i \in \{0, \ldots, M-1\} \times \{0, \ldots, N-1\}$ is a cyclic shift version of the image patch $\mathbf{x}^k$. The scalar y_i is the Gaussian-shaped regression target based on the periodic shift of patch $\mathbf{x}^k$.

The power of the DCF tracker lays in the fact that

all possible cyclic shifts of the object image patches are taken into account to train the model while the solution to the optimisation problem (eq. 2) can be efficiently computed using Discrete Fourier Transform (DFT). To track an object at frame t, the responses of all cyclic shifts of a test image sample can be obtained efficiently in the same way. The location corresponding to the cyclic shift with the maximal response is treated as the final result.

SRDCF [7] extended the standard formulation of DCF to address issues caused by the underlying periodic assumption. It also adapts to object scale change by applying the regressor at multiple resolutions similar to Li and Zhu [28]. We therefore employ SRDCF as the tracking component and denote its output as the object's region $\hat{R}^t$.

The discriminative model $\mathbf{w}$ can be updated via incremental learning as the new training sample $\mathbf{x}^t$ becomes available. The weight α_k is a key factor associated with each training sample from frame k. In the original DCF formulation, it is updated by

$$\alpha_k^t = (1-\gamma)\alpha_k^{t-1} \tag{3}$$

where $\gamma = \gamma_0 = 0.01$ is a fixed learning rate used to control the speed of adapting to the new object appearance.

To properly incorporate the regularization information from the detector, we choose to adapt the learning rate γ according to the discrepancy signal between the detector and the tracker. Specifically, we enforce a strong impact upon the tracker's update when the discrepancy signal is large by setting γ to a higher value. This implies a detection bounding box that is far from the estimated tracking result, and indicates that the tracker has most likely experienced significant failure or drift. Thus,

$$\gamma^t = c\exp(-\frac{1}{\sigma_l^2}\|d(\hat{R}^t,\tilde{R}^t)-1\|^2), \tag{4}$$

where c and σ_l are constants and $d(\hat{R}^t,\tilde{R}^t) = 1-(\hat{R}^t \cap \tilde{R}^t)/(\hat{R}^t \cup \tilde{R}^t)$ is the discrepancy metric based on overlap between two bounding boxes $\hat{R}^t$ and $\tilde{R}^t$, estimated from the paddler tracker and detector respectively.

We update the new region of the object from $\hat{R}^t$ and $\tilde{R}^t$ by linear interpolation using the discrepancy signal $d(\hat{R}^t,\tilde{R}^t)$, such that

$$R^t = (1-d(\hat{R}^t,\tilde{R}^t))\hat{R}^t + d(\hat{R}^t,\tilde{R}^t)\tilde{R}^t. \tag{5}$$

This is to say, when the discrepancy signal is large, we enhance the impact of the detection result when updating the new position and scale of the object. Essentially, we impose a penalty on the tracker's proposed position based on the discrepancy signal with respect to the detector.

4.3. Shot Transition Detection

Frequent shot transitions are characteristic of broadcast image sequences. This severely violates the spatial continuity assumption of tracking algorithms. It is therefore necessary to re-initialise the tracker when a change of view takes place.

Algorithm 1 PPRDCF Paddler Tracking

Input: Image I^t, Image I^{t-1}
Previous target region R^{t-1}
The paddler tracker model $\mathbf{w}^{t-1}$, a constant learning rate $\gamma_0 = 0.01$ and the global static detector's object appearance model $\tilde{\mathbf{x}}$
Output: target region R^t and updated tracker model $\mathbf{w}^t$

if detectShotTransition$(I^t, I^{t-1}) = 0$ **then**
  if $period$ mod $frameNum \neq 0$ **then** ▷ Standard SPRDCF
    $R^t \leftarrow trackPaddler(I^t, R^{t-1}, \mathbf{w}^{t-1})$
    $\mathbf{w}^t \leftarrow updateTrackerModel(R^t, \gamma_0)$
  else ▷ Periodic regularisation - detector's model
    $\tilde{R}^t \leftarrow detectPaddler(I^t, \tilde{\mathbf{x}})$ ▷ eq 1
    $\hat{R}^t \leftarrow trackPaddler(I^t, R^{t-1}, \mathbf{w}^{t-1})$
    update γ^t using $d(\hat{R}^t, \tilde{R}^t)$ ▷ eq 4
    update R^t using $d(\hat{R}^t, \tilde{R}^t), \hat{R}^t, \tilde{R}^t$ ▷ eq 5
    $\mathbf{w}^t \leftarrow updateTrackerModel(R^t, \gamma^t)$
else ▷ Shot transition detected
  $R^t \leftarrow detectPaddler(I^t, \tilde{\mathbf{x}})$ ▷ eq 1
  $\gamma^t = 1$
  $\mathbf{w}^t \leftarrow updateTrackerModel(R^t, \gamma^t)$

To detect shot transition we employ a simple yet effective method. We let the distance metric $d_{t-1,t}$ between two consecutive frames be the Mean Square Error (MSE) of pixel intensities between the frames. We fit a Gaussian distribution $\{\mu^{t-1}, \sigma^{t-1}\}$ to the accumulated distance metric of all preceding images. A new frame is considered to be a shot transition, if the likelihood of its distance metric from its predecessor is outside 3 standard deviations from the expectation, i.e., $|d_{t-1,t} - \mu^{t-1}| > 3\sigma^{t-1}$, where $d_{t-1,t}$ is the distance metric between two consecutive frames. The Gaussian model is then incrementally updated with the new data if no shot transition is detected.

Upon detection of shot transition, we employ a 'soft' re-initialisation scheme. The scheme involves re-localisation of the object's position using the detector, and short-term enhanced learning parameters through boosted γ in eq. 5. This reflects a higher level of trust in the rich detector's model.

4.4. Race Annotation

For kinematic analysis, in contrast to enhancing spectator experience and entertainment [26], athlete tracking is only useful, when the context of the motion is understood. Unlike sports like Football [23] or Athletics [12], where the field of play is known, constrained and can be modelled, in CK Slalom no two venues are the same, the water flow is rapidly changing as are the obstacles and navigation gates' positions. Hence, in order for athlete tracking to be

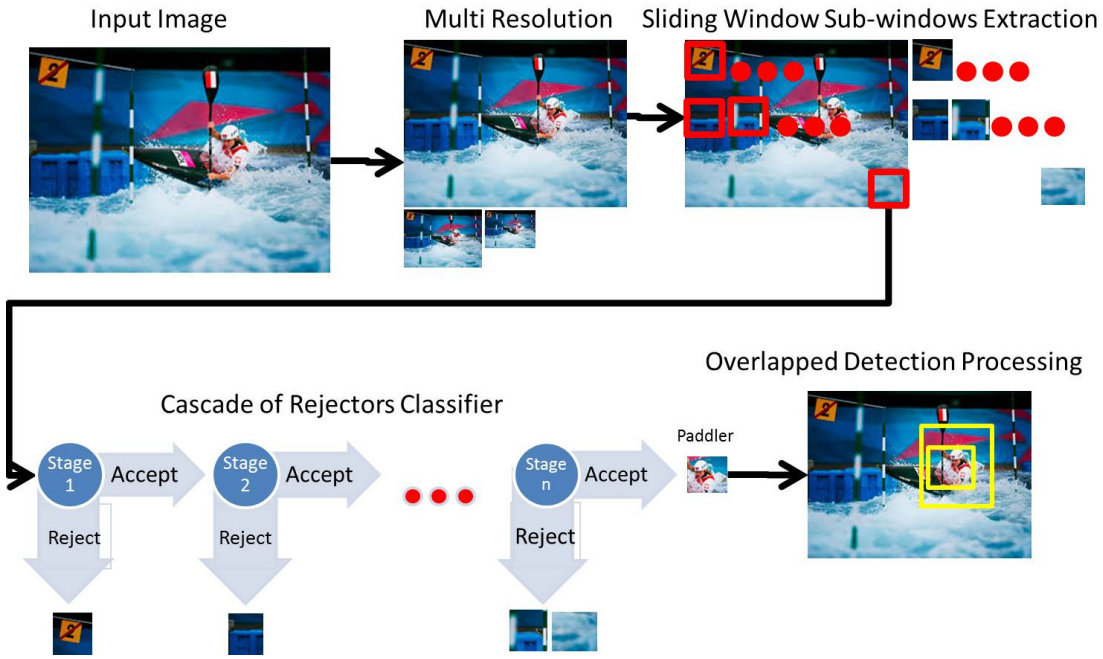


Figure 6: Inference detection of a slalom paddler in a new sample using a cascade classifier. See text for details

relevant for further kinematic analysis, the context of the motion in relation to the environment is necessary.

For CK Slalom, our framework detects the location of the gates, through which the athletes need to navigate, and their assigned order (see figure 7). This is performed via a discriminative cascade classifier similar to our paddler detector in 4.1 that is trained offline, with the exception of the feature descriptor used (HOG). The classifier outputs the location of the gate poles and number. This enables further association between the athlete tracking and the race context. This analysis is outside the scope of this paper.

Figure 7: Gate poles and number detection using discriminative cascade classifier

Gate Number Identification. Once a gate number object is detected using a cascade classifier, the gate number is identified using a trained multi-class linear Support Vector Machine (SVM) classification with a 'one-vs-one' scheme over HOG features (see figure 9). Since a finite maximal number of 24 gates may be used in slalom competition and to avoid aggregation of single digit models for two-digit numbers, our framework learns 24 distinct number classes. The model is learnt from 4×4 cell HOG features extracted for each training image in our gate number dataset. This dataset contains 201 32×32 image patches per number class, which were extracted from the SlalomImRV dataset (described below) and scaled. We note that a diagonal red line may be present on gate numbers in slalom to indicate illegal direction of gate negotiation. The gate number dataset contains both appearance types. The number identification training framework is depicted in figure 8.

Multi-class linear SVM classification is a mature technique. Its goal is to construct a function that will correctly predict the class of a new sample to one of K different classes. In its basic form the problem can be trivially decomposed into mutually exclusive binary classification problems. The one-vs-one method constructs $k(k-1)/2$ classifiers, where each classifier is trained as a binary classifier on two classes. Thus, given a set of m training samples $(x_1, y_1), \ldots, (x_m, y_m)$, where $x_i \in R^K, i = 1, \ldots, m$ and $y_i \in \{1, \ldots, k\}$ is the class of x_i, the i, j SVM classifier solves the binary classification problem

$$D_{ij}(\mathbf{x}) = \mathbf{w}_{ij}^t x + b_{ij}, \tag{6}$$

where each $\mathbf{w}_{ij}$ is a m-dimensional vector, b_{ij} is a scalar and $D_{ij}(\mathbf{x}) = -D_{ji}(\mathbf{x})$. For the input vector $\mathbf{x} \in R^K$ we

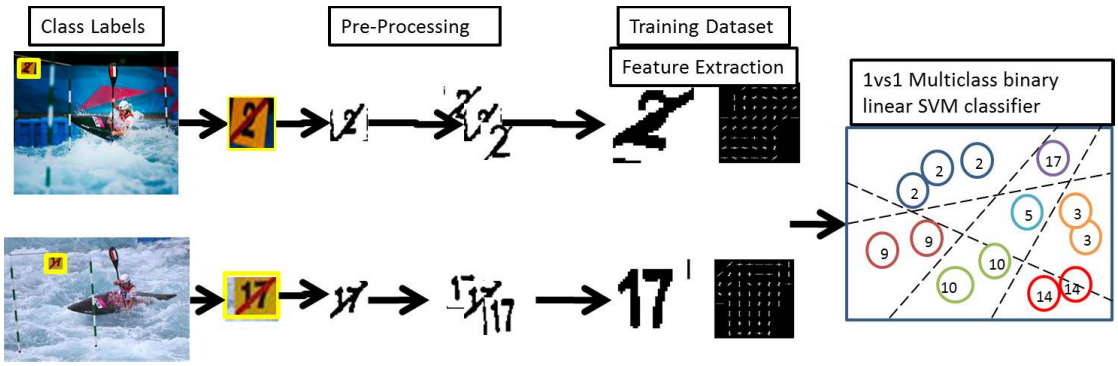


Figure 8: Gate number identification using a multi-class linear SVM classification

calculate

$$D_i(\mathbf{x}) = \sum_{j \neq i, j=1}^{k} \text{sign}(D_{ij}(\mathbf{x})) \quad (7)$$

and classify $\mathbf{x}$ into the class

$$\arg \max_{i=1,\ldots,k} D_i(\mathbf{x}). \quad (8)$$

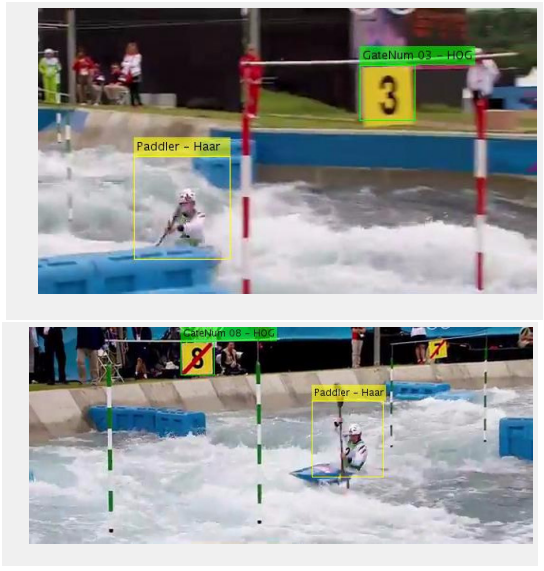


Figure 9: Gate Number identification results using multiclass linear SVM classifier.

5. Quantitative Evaluation

This section reports on a set of quantitative experiments to evaluate the performance of PPRDCF. The experiments are conducted on a new challenging task-specific dataset containing image sequences of slalom paddling with rapid shot transition, and frequent changes in object appearance and pose in unconstrained environment.

5.1. Datasets

The SlalomImRV dataset contains 404 images of slalom paddlers, typically in training or competition *in natura*, whose location in each image was manually annotated by a bounding box. The dataset contains images that have been either captured by the authors, or obtained from publicly available online repositories with a licence search criteria set to creative commons (Flickr, Vimeo), or labelled for reuse (Google Images). The dataset contains images of paddlers with a wide variety of appearance, pose, garments, and in varied lighting and illumination conditions. Moreover, this dataset contains images with many instances of occlusions and self-occlusion conditions including severe cases caused by partial submersion in the water environment, which present a significant challenge to the classifier's model.

To train our *detection* models, we have split the SlalomImRV dataset into a standard 80% and 20% for model learning and cross-validation sets respectively. Our negative set is comprised of 1694 images from the positive and negative training images of the INRIA Person [6] and Parse [31] datasets. In both datasets, the positive training images contain images of people, whilst the negative sets contain mostly background scenery images. Using images that contain people in our negative set ensures that our paddler detection model discriminates well between people displaying a variety of activities and paddlers for our specific task.

The SlalomVidRV dataset contains 30 broadcast image sequences of slalom competition races of 70 to 110 seconds in duration captured at 25 to 30Hz. These sequences have been either captured by the authors, or obtained from publicly available online repositories with a licence search criteria set to creative commons (Flickr, Vimeo) from three competition venue locations that distinctively differ from the images in the SlalomImRV dataset. Specifically, the datasets differ in venue locations where the images and sequences were captured. This is reflected in distinct appearance of the scene, obstacles, slalom gates, and gate numbering, as well as the environmental and lighting conditions in the SlalomVidRV dataset compared to the SlalomImRV dataset. Importantly, the majority of

camera views of the capture differed. In generating our detector model from a dataset distinct from the dataset used for testing the tracking algorithm, we ensure that our framework generalises well within the activity-specific target application. Using this dataset, section 5.2 reports on the performance of our shot transition detector. We report on the performance of our paddler detection module and the empirical selection of feature type and number of cascade stages, which were performed on a subset of this dataset as described in section 5.3. Section 5.4 evaluates the performance of the unified tracking framework.

5.2. Shot Transition Results

For all image sequences of the SlalomVidRV dataset, we manually annotated all frames with a latent random binary variable indicating the ground truth for shot transition. To evaluate the performance of our shot transition detection we calculate precision and recall using $p_t/(p_t + p_f)$ and $p_t/(p_t + n_f)$ respectively, where p_t is true positive, p_f is false positive and n_f is false negative with respect to the ground truth. The high precision and recall and low p_f achieved by our results (see table 1) indicate that this algorithm is very effective in detecting shot transition.

Table 1: Shot Transition Detection Results

#sequences	#frames evaluated	$\#p_t$	$\#p_f$	Precision	Recall
30	60,669	233	14	0.94	1

5.3. Paddler Detection Results

For selection of detector feature type and number of cascade stages, and to separately evaluate the performance of our paddler detection module, we empirically tested the detector on 3 levels of features (Haar, LBP and HOG) and 3 levels of number of cascade stages (15, 20 and 30) experimental conditions. The image test set that was used for these experiments consisted of 1500 images randomly extracted from 3 image sequences in our SlalomVidRV dataset (500 random images per sequence) that were manually annotated for the paddler's ground truth location with a bounding box. The images in this set distinctively differ from the images in the SlalomImRV dataset that was used to train the detector. The difference in venue location is reflected in distinct appearance of the scene and the environmental and lighting conditions.

In addition to precision and recall metrics, we considered the *precision score* defined as the centre of the paddler's bounding box relative to the ground truth, and the *success ratio* defined in 5.4. The scores on each comparison metric were then averaged across the image sequences and are presented for all experimental conditions in table 2 (best score for each metric is indicated in bold).

These results indicate that using Haar features resulted in slightly lower precision, and significantly inferior p_f, precision and success ratio scores compared to using rich mid-level feature descriptors (HOG and LBP). However, since the role of the detector in our framework is to initialise and recover the tracker, we consider the Haar feature's superiority in n_f, number of detections and recall more critical to the overall performance of the framework. Hence, the paddler model trained on Haar feature descriptor was selected for our detection module. In addition, since p_f corresponds to a high discrepancy signal between detector and tracker, they are naturally handled by the penalty imposed by the penalty function (eq. 5).

Intuitively, a feature descriptor that aggregates a number of existing descriptors may result in superior detection performance. However, due to the computational efficiency of using Haar features, we decided against it.

Likewise, the empirical results in table 2 show the superior performance of using 20 cascade stages over the alternative experimental conditions in the number of detections, n_f, and recall. Hence, the corresponding paddler model was selected for use in our framework.

Table 2: Paddler Detection Results (mean per 500 images)

	Feature Type			**# Cascade Stages**		
	Haar	**LBP**	**HOG**	**15**	**20**	**30**
# Detections	**367.67**	203.50	268.00	257.50	**297.33**	284.33
$\# p_t$	**298.67**	180.33	231.67	214.33	**259.67**	236.67
$\# p_f$	69.00	**23.17**	36.33	43.17	**37.67**	47.67
$\# n_f$	**132.33**	296.50	232.00	242.50	**202.67**	215.67
Precision	0.81	**0.89**	0.86	0.83	**0.89**	0.84
Recall	**0.69**	0.38	0.50	0.47	**0.57**	0.53
Precision Score	103.24	22.67	**16.74**	**31.60**	53.24	57.81
Success Ratio	0.25	**0.44**	0.36	**0.36**	0.34	0.35

5.4. Paddler Tracking Results

To evaluate the PPRDCF, we compare its performance with two tracking algorithms; SPRDCF [7] and TLD [25]. The SPRDCF is currently the state-of-the-art tracker, and TLD is a tracker that conceptually is comparable to our framework in using detection to initialise the tracker and assist in tracker failure recovery.

The experiments in this section adopt the following evaluation protocol; We employ the one-pass evaluation that takes the ground truth at the first frame of a sequence as the initialization bounding box then run each tracker until the last frame. The produced trajectory is then compared to manually labelled ground truth using the standard *precision score* and *success ratio* metrics[37]. For each tracker, we calculate a discrepancy signal for the detected objects' location error and overlap ratio with respect to the ground truth. The *precision score* calculates the rate of frames whose centre location is within a certain threshold distance with the ground truth. Here, we use a commonly used threshold of 20 pixels following Wu et al. [37]. This metric emphasizes how well a tracker is able to clasp the target. The *success ratio* calculates the same ratio based on bounding box overlap threshold $(B^* \cap B_{gt})/$

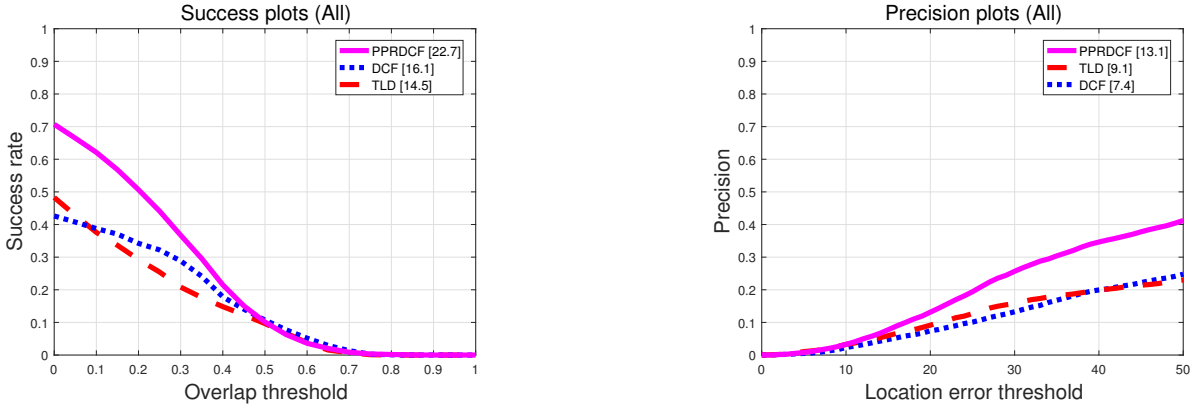


Figure 10: *Success* and *precision* plots comparing the performance of our PPRDCF with state of the art trackers initialised by detection. Algorithms are ranked by the area under the curve and the precision score (20 pixels threshold [37]). Our method (magenta) consistently achieves superior performance.

$(B^* \cup B_{gt})$, where B^* and B_{gt} are the estimated and ground truth bounding boxes' areas, respectively. This metric indicates how well a tracker adapts and covers the target. A typical value is 0.5 as used in object detection evaluation [11]. Thus, for both metrics, higher performance is represented by a greater area under the graph. The results are summarised in figure 10. We also present results on 3 sample image sequences of two different slalom disciplines (C1 - single blade canoe, and K1 - double blade kayak) that were captured in different venues and present distinct variation in the appearance of the scene and environmental conditions in figure 11. Qualitative results from these image sequences are presented in figure 12. We provide the Area Under Curve (AUC) in the figures, which represents the average of all success ratios at different thresholds when the thresholds are evenly distributed.

Further, we performed initialisation experiments with two additional experimental conditions; For the TLD and SPRDCF trackers, we performed separate experiments with initialisation by our paddler detector result and with ground truth bounding box input *at each shot transition*. The former represents an equal opportunity for the three trackers tested, having an identical initialisation. The latter is a standard tracker testing procedure where initialisation is provided by the ground truth, but is only applied to the TLD and SPRDCF trackers. This places our PPRDCF at a disadvantaged starting point. Nevertheless in both experimental conditions the PPRDCF outperforms the TLD and SPRDCF trackers for both precision score and success ratio. We note that the first experimental condition represents a more realistic scenario for automatic systems, where tracker initialisation requires detection.

6. Discussion

In this paper, we investigated the challenging problem of simultaneous human detection and long-term tracking from image sequences comprised of persistent transitioned shots obtained from multiple moving cameras typical of broadcast image sequences, where the object changes appearance frequently as it moves in and out of the camera view.

We introduced Periodically Prior Regularised DCF framework, which uses complementary detection and tracking models. We introduced a robust shot transition detection algorithm for tracking re-initialisation. For tracking our framework uses spatially regularised discriminative correlation filters. For detection, we use offline trained cascade of rejectors classifier. We demonstrated that exploiting the periodic regularisation and camera shot transition detection results in tracker failure and drift recovery.

Our experiments demonstrated that our framework outperforms the state-of-the-art trackers on a new task-specific dataset. The output of our framework forms a critical component and a crucial evidence base pre-requisite to kinematic motion analysis of athletes aimed to optimise technique and performance in sport.

One interesting consequence of our approach is the case where no paddlers are present in an image sequence. As afore stated, initialisation is a pre-requisite to our tracking module. This means that no tracking takes place unless the detector finds a likely object candidate. It is possible that in the absence of a paddler, a false detection initialises the tracker, but then the tracker's performance deteriorates rapidly. However, when a true detection becomes available, the framework naturally overrides the tracker and recovers the position to the paddler's location in the image. In fact, the periodic regularisation of the online tracker's object model by a fixed rich object descriptor highlights a major advantage of our approach. This is further enhanced by our use of different feature descriptors in the detector module (Haar) and the tracker (HOG), as it allows a capture of ancillary characteristics of the object's appearance.

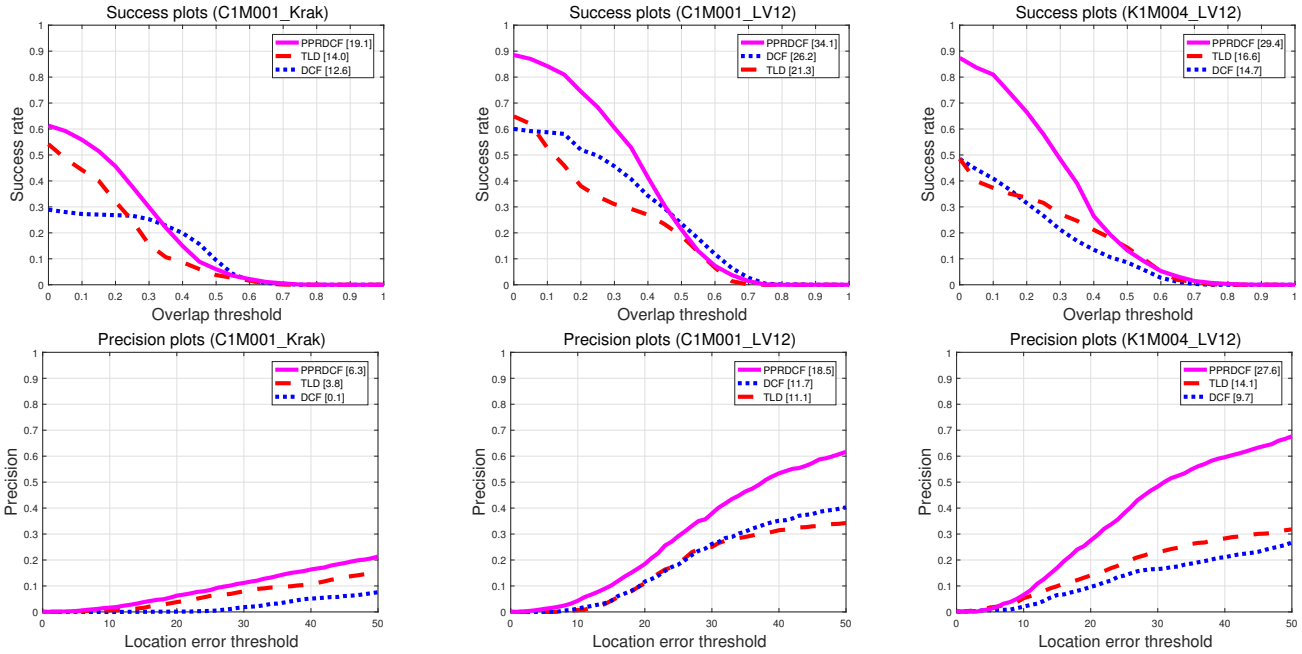


Figure 11: *Success* and *precision* plots comparing the performance of our PPRDCF (magenta) with state of the art trackers on three specific image sequences from two distinct venues (_krak and _LV) and two slalom paddling types (C1 - single blade canoe; and K1 - double blade kayak) reflecting varied appearance and environmental conditions. Our method consistently achieves superior performance.

Limitations and Future Work. A number of challenges remain that need addressing for enhanced robustness of the PPRDCF. For instance, despite the periodic update by the detector, which reduces the contamination of the tracker model, the DCF tracker still suffers from drift [1, 18, 30]. This fundamental challenge of all tracking algorithms is exemplified by the performance of the DCF and and TLD on our datasets, which is particularly challenging since the object rapidly deforms or undergoes severe appearance and occlusion changes. Whilst our approach yields superior tracking results, it does not directly solve the drift problem. Instead it provides a regulating detector to compensate for, and recover from the drift. At high frequency of regularisation, the approach is conceptually akin to tracking-by-detection.

Drift compensation regimes are widely used in vision tracking tasks, for instance, maintenance of a probability distribution function over the object's state spaceAvidan [1] or enforcing structural constraints [30]. Most discriminative approaches, however, adapt the appearance model. To directly address the drift problem, stronger discriminative trackers are needed, albeit, at a high computational cost. This strengthens the argument for enhanced negative sampling of the background used in the online training of the tracker. In the current implementation, the detection method was selected for its low computational cost in inference with disregard to its cost in offline training. It would be interesting to test the performance of recent state-of-the-art CNN detection techniques as an alternative in the framework. Furthermore, inherent to the common use of rectangular image patch to represent the object, is a degree of background contamination in the object's model. The extent to which this degree of contamination should be controlled by non-rectangular patches or advantageous is still debated by the community.

For our task, we achieved robust results with a task-specific trained classifier. It is, however, impossible to design a discriminative classifier for the general case because of the high variability in the appearance of human ambulation. The performance of the approach critically relies on time consuming and costly process and the availability of a large quantity of training samples. Further, the generalisation of the approach can only be achieved through the addition of adequate training samples, as its adaptability to unseen body postures is low and typically manifested by poor performance were occlusions exist.

An interesting extension to our framework will exploit recent advances in simultaneous detection and human pose estimation. These methods exploit in addition to the appearance of the object's parts, the spatial [38] and temporal [5] relations of the parts. This, however, requires pose estimation algorithms to better handle occlusions and self-occlusions than has so far been achieved.

Detailed performance and skill execution analysis of a paddler negotiating a slalom course requires the construction of a 3D model of the scene (slalom course) that is outside the scope of this paper. Nevertheless, whilst 3D reconstruction of *dynamic* scenes from images remains an open problem, the information extracted by our race annotation framework naturally provides additional scene

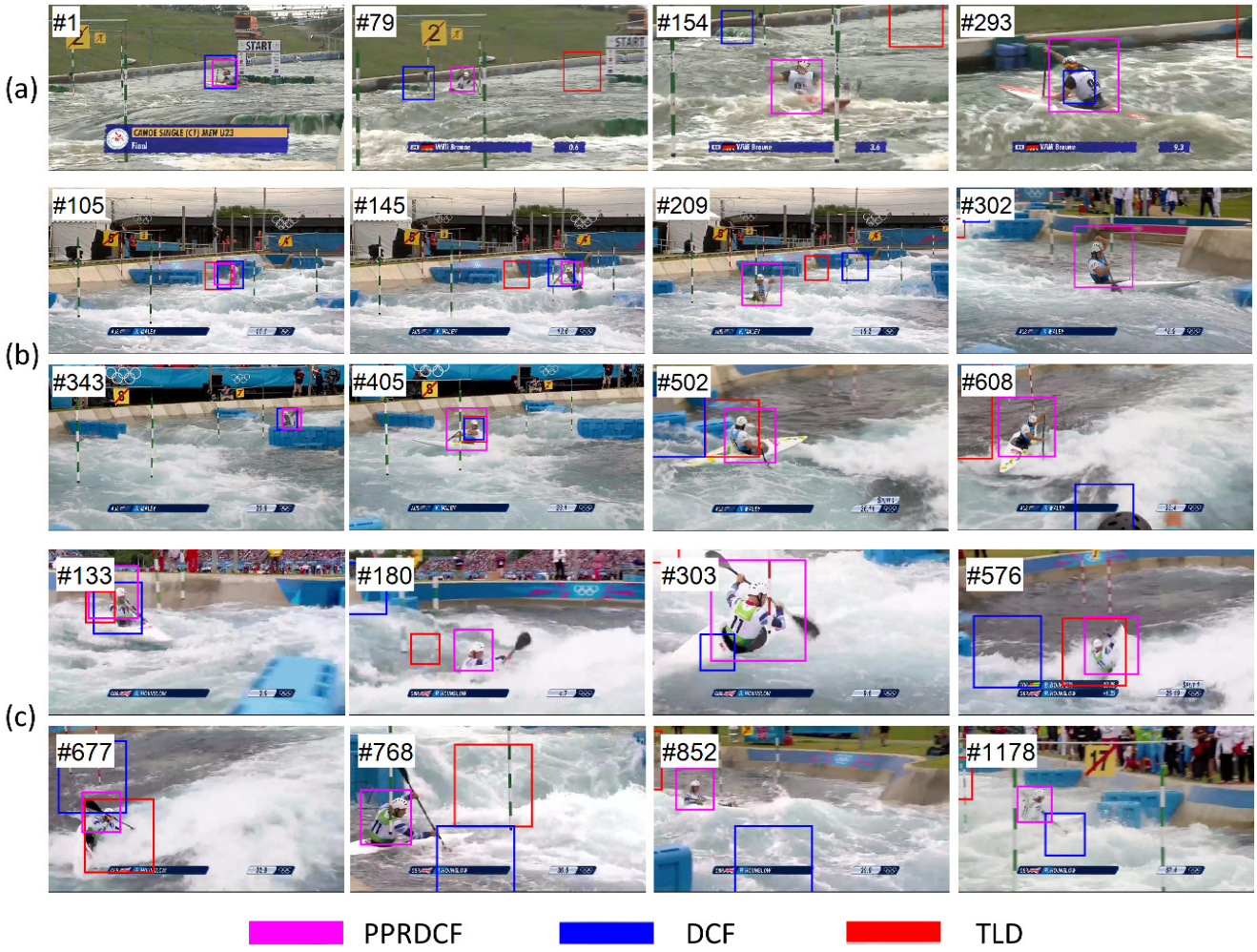


Figure 12: Qualitative comparisons of our PPRDCF (magenta) with state of the art trackers on sample image sequences; (a) C1M001_Krak depicts a sequence of a C1 (single blade canoe) at venue #1. (b) C1M001_LV12 depicts a sequence of a C1 at venue #2. (c) K1M004_LV12 depicts a sequence of a K1 (double blade kayak) at venue #2. Our method attains robust tracking performance and detection recovery in challenging scenarios including fast motion, motion blur, deformation, and severe occlusion resulting from paddler submersion (#180 and #852).

depth information and thus can serve as a strong cue for computation of partial sparse reconstruction using known multi-view relations. In particular, the slalom gate design is standardised to include known sized poles with a known within-gate gap, as well as height-fixed alternating pole colour segments (green/white for downstream gates and red/white for upstream gates, see fig. 7). Intensity and feature-based tracking of these structures should simplify the correspondence problem for stereo matching techniques from which globally consistent slalom course mosaics can be generated and is intended for future work.

References

[1] Avidan, S., 2007. Ensemble tracking. IEEE transactions on pattern analysis and machine intelligence 29 (2), 261–271.

[2] Babenko, B., Yang, M.-H., Belongie, S., Aug 2011. Robust object tracking with online multiple instance learning. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (8), 1619–1632.

[3] Bolme, D., Beveridge, J., Draper, B., Lui, Y. M., June 2010. Visual object tracking using adaptive correlation filters. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. pp. 2544–2550.

[4] Brox, T., Bruhn, A., Papenberg, N., Weickert, J., May 2004. High accuracy optical flow estimation based on a theory for warping. In: European Conference on Computer Vision (ECCV). Vol. 3024 of Lecture Notes in Computer Science. Springer, pp. 25–36.

[5] Cherian, A., Mairal, J., Alahari, K., Schmid, C., 2014. Mixing body-part sequences for human pose estimation. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition.

[6] Dalal, N., Triggs, B., June 2005. Histograms of oriented gradients for human detection. In: Schmid, C., Soatto, S., Tomasi, C. (Eds.), International Conference on Computer Vision & Pattern Recognition. Vol. 2. INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, pp. 886–893.

[7] Danelljan, M., Hager, G., Shahbaz Khan, F., Felsberg, M., 2015. Learning spatially regularized correlation filters for visual tracking. In: Proceedings of the IEEE International Conference on

Computer Vision. pp. 4310–4318.

[8] Danelljan, M., Khan, F. S., Felsberg, M., van de Weijer, J., 2014. Adaptive color attributes for real-time visual tracking. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, pp. 1090–1097.

[9] Dollár, P., Belongie, S., Perona, P., 2010. The fastest pedestrian detector in the west. In: BMVC.

[10] Doyle, D. D., Jennings, A. L., Black, J. T., 2014. Optical flow background estimation for real-time pan/tilt camera object tracking. Measurement 48, 195 – 207.

[11] Everingham, M., Eslami, S. M. A., Gool, L. V., Williams, C. K. I., Winn, J. M., Zisserman, A., 2015. The pascal visual object classes challenge: A retrospective. IJCV 111 (1), 98–136.

[12] Fastovets, M., Guillemaut, J.-Y., Hilton, A., June 2013. Athlete pose estimation from monocular tv sports footage. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on. pp. 1048–1054.

[13] Felzenszwalb, P. F., Huttenlocher, D. P., Jan. 2005. Pictorial structures for object recognition. Int. J. Comput. Vision 61 (1), 55–79.

[14] Galoogahi, H., Sim, T., Lucey, S., June 2015. Correlation filters with limited boundaries. In: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on. pp. 4630–4638.

[15] Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[16] Hare, S., Golodetz, S., Saffari, A., Vineet, V., Cheng, M.-M., Hicks, S., Torr, P., 2015. Struck: Structured output tracking with kernels. Pattern Analysis and Machine Intelligence, IEEE Transactions on PP (99), 1–1.

[17] Hare, S., Saffari, A., Torr, P., Nov 2011. Struck: Structured output tracking with kernels. In: Computer Vision (ICCV), 2011 IEEE International Conference on. pp. 263–270.

[18] He, S., Yang, Q., Lau, R. W., Wang, J., Yang, M.-H., 2013. Visual tracking via locality sensitive histograms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2427–2434.

[19] Henriques, J. F., Caseiro, R., Martins, P., Batista, J., 2012. Exploiting the circulant structure of tracking-by-detection with kernels. In: Computer Vision–ECCV 2012. Springer, pp. 702–715.

[20] Henriques, J. F., Caseiro, R., Martins, P., Batista, J., 2015. High-speed tracking with kernelized correlation filters. Pattern Analysis and Machine Intelligence, IEEE Transactions on 37 (3), 583–596.

[21] Hoiem, D., Chodpathumwan, Y., Dai, Q., 2012. Diagnosing error in object detectors. In: Proceedings of the 12th European Conference on Computer Vision - Volume Part III. ECCV'12. Springer-Verlag, Berlin, Heidelberg, pp. 340–353.

[22] Horn, B. K. P., Schunck, B. G., 1981. Determining optical flow. Artificial Intelligence 17, 185–203.

[23] Huang, P., Hilton, A., 2006. Football player tracking for video annotation. IET European Conference on Visual Media Production, 175–175.

[24] Hunter, A., 2009. Canoe slalom boat trajectory while negotiating an upstream gate. Sports Biomechanics 8 (2), 105–113.

[25] Kalal, Z., Mikolajczyk, K., Matas, J., Jul. 2012. Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. 34 (7), 1409–1422.

[26] Kilner, J., Starck, J., Hilton, A., 2006. A comparative study of free viewpoint video techniques for sports events. IET European Conference on Visual Media Production, 87–96.

[27] Li, Y., Zhu, J., 2014. A scale adaptive kernel correlation filter tracker with feature integration. In: Computer Vision-ECCV 2014 Workshops. Springer, pp. 254–265.

[28] Li, Y., Zhu, J., 2014. A scale adaptive kernel correlation filter tracker with feature integration. In: Computer Vision - ECCV 2014 Workshops - Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part II. pp. 254–265.

[29] Lisani, J., Nov 2014. Adaptive thresholds for robust face detection with a short cascade of classifiers. In: Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on. pp. 27–31.

[30] Oron, S., Bar-Hille, A., Avidan, S., 2014. Extended lucas-kanade tracking. In: European Conference on Computer Vision. Springer, pp. 142–156.

[31] Ramanan, D., 2007. Learning to parse images of articulated bodies. Advances in Neural Information Processing Systems 19, 1129.

[32] Ranzato, M. A., lan Boureau, Y., Cun, Y. L., 2008. Sparse feature learning for deep belief networks. In: Platt, J., Koller, D., Singer, Y., Roweis, S. (Eds.), Advances in Neural Information Processing Systems 20. Curran Associates, Inc., pp. 1185–1192.

[33] Stauffer, C., Grimson, W. E. L., 1999. Adaptive background mixture models for real-time tracking. In: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on. Vol. 2. IEEE Computer Society, Fort Collins, CO, USA, pp. 246–252.

[34] Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. pp. I–511–I–518 vol.1.

[35] Viola, P., Jones, M., 2001. Robust real-time face detection. In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. Vol. 2. pp. 747–747.

[36] Wojek, C., Schiele, B., 2008. A performance evaluation of single and multi-feature people detection. In: Proceedings of the 30th DAGM Symposium on Pattern Recognition. Springer-Verlag, Berlin, Heidelberg, pp. 82–91.

[37] Wu, Y., Lim, J., Yang, M.-H., 2013. Online object tracking: A benchmark. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2411–2418.

[38] Yang, Y., Ramanan, D., Dec 2013. Articulated human detection with flexible mixtures of parts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35 (12), 2878–2890.

[39] Zhu, Q., Yeh, M.-C., Cheng, K.-T., Avidan, S., 2006. Fast human detection using a cascade of histograms of oriented gradients. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. CVPR '06. IEEE Computer Society, Washington, DC, USA, pp. 1491–1498.