# Training with Corrupted Labels to Reinforce a Probably Correct Teamsport Player Detector*

Pascaline Parisot, Berk Sevilmiş, and Christophe De Vleeschouwer

Université Catholique de Louvain, ICTEAM-ELEN,
Place du Levant, 2, 1348 Louvain-La-Neuve, Belgique
pascaline.parisot@uclouvain.be

**Abstract.** While the analysis of foreground silhouettes has become a key component of modern approach to multi-view people detection, it remains subject to errors when dealing with a single viewpoint. Besides, several works have demonstrated the benefit of exploiting classifiers to detect objects or people in images, based on local texture statistics. In this paper, we train a classifier to differentiate false and true positives among the detections computed based on a foreground mask analysis. This is done in a sport analysis context where people deformations are important, which makes it important to adapt the classifier to the case at hand, so as to take the teamsport color and the background appearance into account. To circumvent the manual annotation burden incurred by the repetition of the training for each event, we propose to train the classifier based on the foreground detector decisions. Hence, since the detector is not perfect, we face a training set whose labels might be corrupted. We investigate a set of classifier design strategies, and demonstrate the effectiveness of the approach to reliably detect sport players with a single view.

**Keywords:** detection, random ferns, corrupted label.

## 1 Introduction

Detecting people in images is an important question for many computer vision applications including surveillance, automotive safety, or sportmen behavior monitoring. It has motivated a long history of research efforts [8], which have recently converged into two main trends.

On the one hand, background subtraction approaches have gained in popularity since they have been considered in a multi-view framework. In each view, those approaches build on a background model to compute a mask that is supposed to detect the moving foreground objects in the view. The foreground silhouettes computed in each view of a calibrated multi-camera set-up are then merged to mitigate the problems caused by occlusions and illumination changes when inferring people location from a single view. Several strategies have been

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-3-319-02895-8_64

* Part of this work has been funded by the Belgian NSF, and the walloon region project SPORTIC.

J. Blanc-Talon et al. (Eds.): ACIVS 2013, LNCS 8192, pp. 682–694, 2013.

considered to fuse the masks from multiple views [13,10,1,6]. They generally rely on the definition of a ground occupancy probability map, which exploits the verticality of people silhouettes to estimate the likelihood that a particular ground plane position is occupied or not by someone. All of these approaches build on the multiplicity and diversity of viewpoints, and their performances significantly degrade when a single viewpoint is available.

On the other hand, efforts have been carried out to detect people or objects of interest based on their visual appearance. Modern approaches make an extensive use of training samples, to learn how the object is defined in terms of topologically organized components [9,2] and/or in terms of texture statistics [15,4]. The pioneering work of Viola and Jones [19] illustrates the success of those approaches to detect objects in images. It relies on boosting strategies to select and combine a large number of weak binary tests to decide whether the content of a (sub-)image corresponds to the object-of-interest or not. Since the tests are defined in terms of the average luminance observed on small patches defined by their size and location in the image, their statistics intrinsically capture the spatial topological organization of the image textures. Several recent works have been inspired by the same intuition to detect people. Representative examples are the work in [7] and in [20], which analyze the content of an image in terms of a multiplicity of pixel features -like color, gradient, or motion. Those methods appear to be efficient in detecting people, as long as a sufficiently large and representative database is available to train the classifier. The collection of those training samples is however performed manually in most previous works, which prevents to adapt the detector to the appearance specificities encountered in the particular case at hand. Such adaptation capability is especially relevant in a teamsport analysis context, since the background and all the players of each team are characterized by a specific shirt.

Our paper takes advantage of the two trends presented above. It aims at improving the foreground silhouette detector (referred as foreground detector in the following), by using an appearance-based classifier to differentiate false and true positives among the foreground silhouette detections. The main idea of our paper, and its main contribution from a system design point-of-view, consists in training the classifier based on the probably correct decisions taken by the foreground detector. Because it exploits color and gradient visual features, the appearance-based classifier offers a complementary information compared to the one provided by the foreground detector, thereby making the overall detection more reliable. This idea is in-line with co-training approaches [3]. The similarities and differences between our proposal and co-training approaches will be discussed in Section 3. More importantly, because our approach defines the training samples of the classifier based on the foreground detector decisions, no manual annotation is required to generate the training set, which makes it possible to retrain and adapt the classifier to the case at hand.

In addition to the original integration of two families of people detection algorithms, our paper also brings significant contributions related to the design of the classifier itself. Indeed, primarily, our paper introduces an original people

detection method that relies on an ensemble of random sets of binary tests to characterize the texture describing the visual appearance of the target. The binary tests consist of comparison of pixel values within a block. Specifically, we extend the approaches in [15] and [17] to the description of large image patterns (see Section 4.2). Our experimental results demonstrate that the use of simple binary tests on raw pixel color or gradients of image blocks is more effective in characterizing sport player patterns than the integral image features recommended in [7] for pedestrians.

As a second contribution related to the design of appearance-based classifiers, our paper shows that, in the particular case of large deformations of the objects of interest as encountered in a sport context, ensembles of random classifiers outperform the boosted classification methods, traditionally adopted for pedestrian detection [7]. Ensembles of random classifiers have gained popularity in recent years, mainly because they reduce the risk of overfitting and offer good generalization properties in case of training samples scarcity [12]. Our work reveals that those random classifiers are also more robust to labels corruption than AdaBoost solutions.

The rest of the paper is organized as follows. Section 2 presents the overview of our system. Section 3 discusses the similarities and differences between the co-training framework and our approach. Section 4 then defines our proposed classifier. It is supposed to differentiate human player patterns from arbitrary background patterns, and consists of an ensemble of random ferns, each fern characterizing a block of the image in terms of the stochastic distribution of its visual features. Section 5 validates our approach.

## 2 System Overview: Training with Corrupted Labels

The proposed detection scheme is depicted in Figure 1.

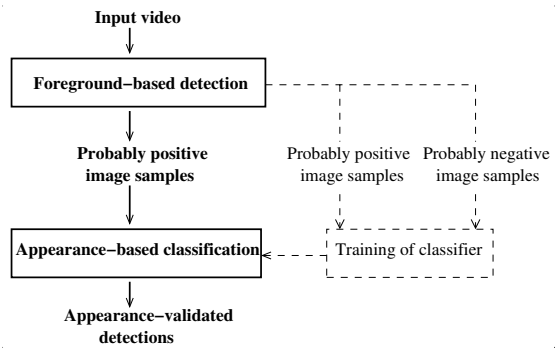



**Fig. 1.** Solid lines depict the proposed people detection scheme. The foreground-based detections are validated or rejected based on their appearance. Dashed lines depict the training phase. The appearance-based classifier is trained with image samples that are labelled with a good rate of success by the foreground detector.

People/player image samples are continuously detected with a high detection rate and a reasonable false alarms rate, based on the foreground mask approach described in [6]. The resulting probably positive samples are then processed by a classifier, which further investigates the visual features of each foreground detected object to decide whether it corresponds to a human/player or not. Optionally, the foreground detected samples can feed the training of the classifier. Specifically, two classes of training samples are defined based on the ground occupancy map computed in [6]. The first class of training samples corresponds to the probably positive samples. Those samples are defined by cropping a rectangular sub-image in the camera view, around the backprojection of a probably occupied ground position. The training samples of the second class correspond to probably negative samples, which are randomly cropped around backprojected ground positions that are considered to be unoccupied by the detector. Examples of image samples from both classes are presented in Figure 2.


**Fig. 2.** Examples of samples labelled as probably positive (a) or probably negative (b) by the foreground detector

We observe a significant variability among the samples of each class, which makes the learning of a classifier challenging, and motivates the careful investigation carried out in Section 5. Regarding the people/no-people decision expected from the classifier, those samples are subject to label corruption. This is because those labels are defined based on the error-prone decisions of the foreground detector. As a consequence, the appearance-based classifier should be designed so that its learning is robust to label corruption. In the next section, we motivate the need of online training and position our work with regard to the co-training framework, which shares similarities with our approach, whilst being different.

## 3 Specificities of Our Applicative Context vs. Related Work

To motivate both the need for online training, and the development of an original solution to this problem, it is worth presenting the specificities of our application context.

In short, we are interested in the detection of teamsport players to control the autonomous production of images to render a sport game action [5]. In other

words, the information about players positions is used to select the view point to adopt to render the action, typically by cropping within a fixed view. Hence, we are not interested in the accurate segmentation of each individual, but we are eager to determine whether a given foreground activity either results from (one or several) players, or is caused by some other reason like, for example, dynamic advertisement panels or spot lightings. As another consequence of our application context, our system has to deal with severe deformations of the object-of-interest (players are running, jumping, falling down, connecting to each others, etc). Hence, to be effective, it can not only rely on the characterization of the standard appearance of a standing human, like it is done for pedestrian detection for example, but it has to exploit as much of the *a priori* information that is available about the appearance of the object (e.g. players'jerseys have a known color) and of the scene (sport hall, known background advertisements). Since this *a priori* information changes from one game to another, the classifier has to be trained online, so as to adapt to the game at hand.

Besides motivating the online training, the wide range of deformations encountered by our application also prevents the use of most of the solutions that have been proposed in the past to learn online, without manual labelling of the training samples. Specifically, in the late nineties, Blum and Mitchell [3] have introduced the co-training framework to reduce the amount of labelled samples required to train a classifier. Their purpose was to exploit unlabelled samples to jointly reinforce two complementary classifiers, i.e. that look at the data from different points of view, using independent features. In a straightforward implementation of their framework, the two classifiers are initially trained based on a small set of manually labelled samples, and are then jointly improved by increasing the training set of one classifier based on the *reliable* labels assigned by the other classifier [14]. In more recent works, motion detection has been considered to initialize the learning process, so that manual labelling is not required anymore [18,16]. In both kinds of approaches, however, a key issue lies in the selection of *reliably labelled* samples. To identify those reliable samples, earlier works make the explicit or implicit assumption that the appearances of the objects-of-interest are sufficiently similar to be accurately described by some fixed discriminative (appearance) model. They then propose to learn such discriminative model from the dominant statistics observed among the positively labelled samples of each classifier, either in terms of PCA [18] or simply in terms of motion blobs aspect ratio [16]. In our sport analysis context, however, the assumption about the existence of a stable appearance model does not hold anymore. Players are very active, and their silhouettes change a lot depending on the action at hand (see variability in Fig. 2-(a)). Bottom line, we can not rely on some simplistic appearance model to select reliable samples among the ones detected based on motion analysis. For this reason, we have to deal with erroneous labels during the training. We show in the rest of the paper that ensembles of random classifiers better support such errors in labelling than AdaBoost solutions.

## 4 Classification of Human Patterns Based on Weak Binary Tests Combination

Many recent works have demonstrated the advantages of combining (weak) binary tests to solve image classification problems [15,4,19,7,17]. We follow this paradigm. Section 4.1 defines the binary tests either in terms of pixel values or integral images comparisons. Section 4.2 presents two approaches to combine the binary tests. The first one follows the well-known AdaBoost method [11], as used in [19,7]. The second one adopts a more recent Semi-Naive Bayesian formulation, and classifies samples based on the joint probability distributions associated to random ferns, i.e. to small sets of randomly selected binary tests [17]. In contrast to previous usages of ferns, which have focused on the description of small texture patches around keypoints, our paper proposes to exploit ferns to classify entire and semantically meaningful image patterns.

### 4.1 Definition of Binary Tests

In our work, the tests are carried out on so-called image channels, defined in [7] as the $R$, $G$, and $B$ components, the gradient magnitude $GM$, and the magnitude of oriented gradients $OG_j, 0 \leq j \leq 5$.

For a given channel, a binary test is then defined to compare either the intensities of two pixel locations, or the integrals of pixel intensities over two rectangular supports. Comparisons of pixel intensities are performed within a small block, e.g. limited to $16 \times 16$ pixels, because they aim at describing local textures through the combination of many local comparisons of pixel intensities. In contrast, integral supports are defined on the entire image since those integral values are supposed to capture discriminating behavior of the image signal on some spatial area. The first kind of test follows the approaches in [15] and [17], while the second one follows [19] and [7].

In a more formal way, a binary test $b_i$ is defined by (i) the test image channel $I_i \in \{R, G, B, GM, \{OG_k\}_{0 \leq k \leq 5}\}$, (ii) the test type $t_i \in \{pixel, integral\}$, and (iii) a pair of pixel locations $(m_{i,1}, m_{i,2})$ (defined in a 16x16 block) or a rectangular support $(r_{i,1}, r_{i,2})$ (defined over the entire image). Letting $w_{i,1}$ and $w_{i,2}$ denote two intermediate values defined as follows:

$$\forall j \in \{1,2\}, \quad w_{i,j} = \begin{cases} I_i(m_{i,j}), & \text{if } t_i = pixel \\ \frac{1}{|r_{i,j}|} \sum_{m \in r_{i,j}} I_i(m), & \text{if } t_i = integral \end{cases} \tag{1}$$

where $|r_{i,j}|$ is the number of pixels in the rectangle, we simply write:

$$b_i = \begin{cases} 1, & \text{if } w_{i,1} > w_{i,2} \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

### 4.2 Combination of Binary Tests

Two approaches are considered to combine the weak binary classifiers.

The first one follows the AdaBoost algorithm [11]. It will be used as a baseline reference since its effectiveness and efficiency in solving object detection problems in images have already been extensively demonstrated [19,7].

The second combination approach is an original contribution of our paper. As told above, it is inspired by a number of earlier works dealing with image texture classification [15] and keypoint identification [17]. It differs from those previous works by the fact that it is designed to describe the semantically meaningful pattern corresponding to the projection of an object or a human-being. Therefore, the binary tests are selected over the entire image support, and have to be defined in terms of their relative position compared to the image support. This is simply done by normalizing the image sizes, typically to $128 \times 64$ pixels in our work. To explain the other specificities of our approach compared to [17], it is worth reminding the principle underlying the classification with ensemble of random sets of binary tests, also named random ferns (RF) classification.

Let $D$ denote the random variable that represents the class of an image sample. In our problem, $D = 1$ if the sample corresponds to a player, and $D = 0$ otherwise. Given a set of $N$ binary tests $b_i, i = 1, ..., N$, the sample class MAP estimate $\hat{d}$ is defined by:

$$\hat{d} = \underset{d \in \{0,1\}}{\operatorname{argmax}} P(D = d | b_1, ..., b_N). \tag{3}$$

Bayes' formula yields:

$$\hat{d} = \underset{d \in \{0,1\}}{\operatorname{argmax}} P(b_1, ..., b_N | D = d), \tag{4}$$

if we admit a uniform prior $P(D)$.

Learning and maintaining the joint probability in Equation (4) is not feasible for large $N$ since it would require to compute and store $2^N$ entries for each class. A naive approximation would assume independence between binary tests, which would reduce the number of entries per class to $N$. However, such representation completely ignores the correlation between the tests. The semi-naive bayesian approach proposed in [17] accounts for dependencies between tests while keeping the problem tractable, by grouping the $N$ binary tests into $M$ sets of size $S = N/M$. These groups are named ferns, and the joint conditional probability is approximated by:

$$P(b_1, ..., b_N | D = d) = \prod_{k=1}^{M} P(F_k | D = d), \tag{5}$$

where $F_k$ denotes the $k^{th}$ fern.

The training phase estimates the class conditional probability distribution of each fern independently, and is detailed in [17]. Compared to AdaBoost, random ferns have the advantage to support incremental training. This is especially interesting in our teamsport analysis context, since it allows to initialize the process with default ferns distributions (e.g. averaged on several games), and to progressively update the distributions along the game, as new samples are collected.

Now that the random ferns classification principles have been reminded, we explain how our approach differs from earlier works in terms of tests assignment to ferns. This subtle change is required to characterize large image patterns, and not just small texture patches as in [17]. In [17], to split the $N$ tests into ferns of $S$ tests, they use a random permutation function with range $1...N$. This is motivated by the fact that all tests have *a priori* the same chance to be (in)dependent. In our case, this assumption reasonably holds for integral image tests, since their supports cover large fractions of the image, which gives all pairs of tests a similar chance to be (in)dependent. In contrast, the assumption does not hold anymore for the tests dealing with pixel intensities. Those tests are local by definition, since they compare the intensities of two locations that are close to each other. Hence, two tests dealing with the same image area are more likely to depend on each other than two tests dealing with far apart pixels. Since dependencies are only handled within a fern, it becomes relevant to assign to each fern a set of tests that correspond to the same spatial area. In final, when using pixels intensities comparisons, our proposed approach can be summarized as follows. The image support is split into a grid of non-overlapping blocks of $16 \times 16$ pixels, and all tests of a given fern are defined based on a pair of pixels that are selected within the same block.

## 5 Experimental Validation

This section considers a typical real life basket-ball player detection scenario. The training sets are defined automatically, as explained in Section 2. The set of probably positive samples detected by the foreground detector [6] is referred to as the detector set in the following, while the set of probably negative samples is named random set. In addition, a reference ground truth label has been assigned manually to each detector sample, so as to split the detector set into a positive and a negative set. The positive set includes the valid detections, while the negative set contains the false detections, resulting from a foreground detector error.

In our experiments, we train the classifiers based on detector and random training sets, and measure how well those classifiers discriminate between positive and negative test sets.

The detector sets considered in our experiments are derived from a game that happened in the Spiroudome sporthall (http://www.spiroudome.com), on a period of time during which 2723 positive samples have been manually annotated. Several detector (and random) sets have been defined on the same period of time. Each detector set is composed of 1000 samples randomly picked among the foreground detected samples, but is affected by a different rate $n$ of false detections, ranging from $n = 2\%$ to $n = 10\%$, as a function of the foreground detector operating point. In addition, an Oracle defines uncorrupted detector sets ($n = 0\%$), based on the manual groundtruth. Five pairs of detector and random sets have been picked up randomly at each corruption rate, so as to repeat the experiments and compute average and standard deviation performance metrics.

The test set is defined by 1000 manually labelled samples (900 positive samples, and 100 negative ones), extracted in a different period of time of the same Spiroudome game.

For all those sets, each image sample is characterized by 10 image channels, and the classifiers parameters are set as follows. There are 5 tests per fern, and 200 ferns per 16x16 image block. For each fern, the common channel of the 5 tests is randomly selected among the 10 image channels. Hence, there are 32000 tests for a normalized image of size 128x64. The same number of tests is considered for AdaBoost classifiers.

In the first experiment, we compare different kinds of binary tests. Therefore, we train the random ferns classifier on uncorrupted labels, and consider three kinds of binary tests: pixel comparisons within a block, pixel comparisons within the whole image and integral image comparisons within the whole image. Figure 3-(a) plots the obtained ROC curves, that is the detection rate on the positive set versus the detection rate on the negative set (which corresponds to the false alarm rate on the detector set). The number of binary tests are the same in the three cases. We observe that significantly better performances are obtained with tests comparing two pixels in a block. Tests comparing integral images or pixels on the whole image are not able to discriminate player activity patterns from background activity. In the following, we only consider comparison of pixels within a block.

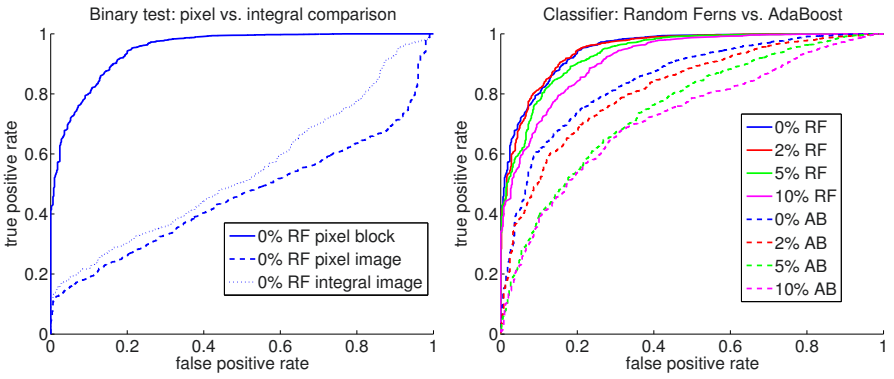



**Fig. 3.** (a) Receiver Operating Characteristic (ROC) curves resulting from the random ferns (RF) classifier trained on uncorrupted labelled samples (0%) for three kinds of binary test: "pixel block", i.e. pixel comparison within a block, "pixel image", i.e. pixel comparison within the whole image and "integral image", i.e. integral comparison within the whole image. Binary test based on the comparison of pixels within a block outperforms the two other kinds of binary test ; (b) ROC curves resulting from training sets with uncorrupted labels (0%) and corrupted labels (rate of 2, 5 and 10%), for both kind of classifiers: random ferns (RF) and AdaBoost (AB). Random ferns classifier outperforms AdaBoost ones and is less sensitive to uncorrupted labels.

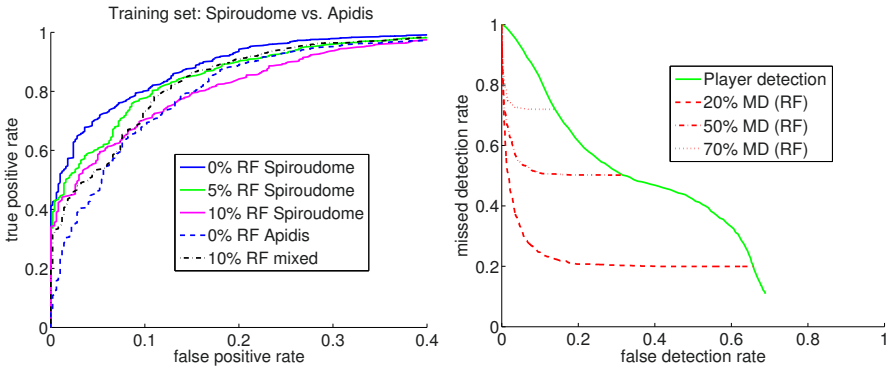
In the second experiment, we analyze the impact of the label corruption rate when training AdaBoost and random ferns classifiers. Figure 3-(b) plots the ROC curves for both kinds of classifiers, and for different corruption rates. It reveals that our proposed random ferns approach outperforms the AdaBoost classifier, and is more robust to label corruption than AdaBoost (see the decrease of "area under curve" in Table 1). Additional experiments that are not reported here have shown that increasing the number of weak classifiers in the case of AdaBoost has no significant impact on the obtained performances.

**Table 1.** Area under curve measured in Fig. 3-(b) (mean ± standard deviation)

| | Uncorrupted labels (0%) | Corrupted labels 2% | 5% | 10% |
|---|---|---|---|---|
| Random Ferns | 0.949 ± 0.006 | 0.947 ± 0.009 | 0.934 ± 0.007 | 0.915 ± 0.006 |
| AdaBoost | 0.848 ± 0.042 | 0.822 ± 0.035 | 0.757 ± 0.047 | 0.730 ± 0.040 |

In the third experiment, we investigate how online training improves performances compared to offline training. We train the random ferns classifier on different training sets. The first training set is based on the Spiroudome game with different rates of corrupted labels. The second training set is based on the Apidis dataset (http://www.apidis.org/Dataset) with uncorrupted labels. Finally, the last training set is composed by a mixture of samples from the Spiroudome and Apidis datasets: 500 positive samples from the Apidis dataset, 500 probably positive samples from the Spiroudome dataset (with 10% of corrupted labels) and 1000 probably negative samples from the Spiroudome dataset. The obtained ROC curves are plotted in Figure 4-(a). The best curves are obtained from the Spiroudome training set with small corruption rates ($\leq 5\%$), or from a mixed training set when the corruption rate increases. We also observe that keeping the false alarm rate below 5% (which is reasonable to avoid video production inconsistencies) results in a selection rate lower than 40% for offline training, but higher than 60% with mixed online training.

As a last experiment, we have measured the impact of our proposed system on the operating points of a single-view player detector integrating the foreground detector and the random ferns classifier. For this purpose, we have defined manually a detection ground truth over 280 regularly spaced frames, in an interval of 4min40s of a Spiroudome basketball game. This ground truth information consists of the bounding boxes of the players and referees in the frame view coordinate system. We have then compared this ground truth to the detections computed by the foreground detector in [6], and to the subsets of those detections that are considered to be positive by the classifier trained with the 10% corrupted training set in the second experiment (see Fig. 3-(b)). For this comparison, we consider that two objects cannot be matched if the overlapping of the detected bounding boxes on the frame is smaller than 50%. Figure 4-(b) presents, in solid line, the ROC curve of [6], i.e. using the foreground detector only. The dotted, dashed-dotted and dashed lines correspond to the ROC curves




(a) (b)

**Fig. 4.** (a) ROC curves for three kinds of training sets: Spiroudome set with different rates of label corruption, Apidis set and mixture of them. We conclude that online training helps even in case of corrupted labels ; (b) Improvement of ROC curve resulting from our proposed random ferns (RF) classifier, trained on corrupted labels: The solid green line depicts the initial foreground detector ROC. Dotted, dot-dashed, and dashed lines plot the ROC curves obtained after classification of the samples detected by the foreground detector, respectively working with 70%, 50% or 20% of missed detections (MD).

obtained when using the random ferns classifier to sort the foreground detections into false and true positives. Each of these 3 curves is derived from a particular foreground detector operating point, respectively corresponding to 20%, 50% and 70% of missed detections. We conclude from Figure 4-(b) that the classifier significantly improves the operating trades-off compared to the ones obtained based on foreground detection only, which definitely demonstrates the relevance of the scheme proposed in Fig. 1.

## 6 Conclusion

As a first and primary contribution, the paper has proposed an original framework to reinforce a (visual object) detector. The framework assumes that a reasonably correct detector is available, but that it fails to use some available (visual) features that are actually discriminating with respect to the detection task. Based on those assumptions, our framework proposes to train a classifier to discriminate between detected samples, which are probably positive regarding the detection goal, and randomly selected samples, which are probably negative. Our experimental results demonstrate that the resulting classifier offers good generalization properties and captures the essence of the knowledge needed to differentiate false and true positives among the samples detected by the initial foreground detector, thereby shifting the receiver operating characteristics of the reinforced detector towards smaller false alarm rates for a given detection rate.

As a second contribution, our paper has shown that an ensemble of random classifiers achieves better performances than conventional boosted solutions for large intra class variability, and when the labels of the training samples are corrupted. Regarding the definition of the binary tests that are combined through boosting or random strategies, it appears that the comparisons of neighbouring pixel values offer better performances than comparison of pixels or integral images on the whole image.

## References


1. Alahi, A., Jacques, L., Boursier, Y., Vandergheynst, P.: Sparsity driven people localization with a heterogeneous network of cameras. Jour. of MIV 41(1-2), 39–58 (2011)
2. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. Neural Computation 9(12), 1545–1588 (1997)
3. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proc. of COLT, pp. 92–100 (1998)
4. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: Proc. of ICCV (2007)
5. Chen, F., Delannay, D., De Vleeschouwer, C.: An autonomous framework to produce and distribute personalized team-sport video summaries: a basket-ball case study. IEEE Trans. on Multimedia 13(6), 1381–1394 (2011)
6. Delannay, D., Danhier, N., De Vleeschouwer, C.: Detection and recognition of sports (wo)men from multiple views. In: Proc. of ACM/IEEE ICDSC (2009)
7. Dollar, P., Tu, Z., Perona, P., Belongie, S.: Integral channel features. In: Proc. of BMVC (2009)
8. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: a benchmark. In: Proc. of IEEE CVPR (2009)
9. Felzenszwalb, P., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Trans. on PAMI 32(9), 1627–1645 (2010)
10. Fleuret, F., Berclaz, J., Lengagne, R., Fua, P.: Multi-camera people tracking with a probabilistic occupancy map. IEEE Trans. on PAMI 30(2), 267–282 (2008)
11. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Jour. of CSS 55(1), 119–139 (1997)
12. Geurts, P., Ernst, D., Wehenkel, L.: Extremely Randomized Trees. Machine Learning 36(1), 3–42 (2006)
13. Khan, S.M., Shah, M.: A multiview approach to tracking people in crowded scenes using a planar homography constraint. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 133–146. Springer, Heidelberg (2006)
14. Levin, A., Viola, P., Freund, Y.: Unsupervised improvement of visual detectors using co-training. In: ICCV, pp. 626–633 (2003)
15. Marée, R., Geurts, P., Piater, J., Wehenkel, L.: Random subwindows for robust image classification. In: Proc. of IEEE CVPR, pp. 34–40 (2005)
16. Nair, V., Clark, J.J.: An unsupervised, online learning framework for moving object detection. In: Proc. of IEEE CVPR, vol. 2, pp. 317–324 (2004)

17. Ozuysal, M., Calonder, M., Lepetit, V., Fua, P.: Fast keypoint recognition using random ferns. IEEE Trans. on PAMI 32(3), 448–461 (2010)
18. Roth, P., Grabner, H., Skočaj, D., Bischof, H., Leonardis, A.: Conservative visual learning for object detection with minimal hand labeling effort. In: Kropatsch, W.G., Sablatnig, R., Hanbury, A. (eds.) DAGM 2005. LNCS, vol. 3663, pp. 293–300. Springer, Heidelberg (2005)
19. Viola, P., Jones, M.: Robust real-time object detection. In: Proc. of the Int. Workshop on SCTV (2001)
20. Xing, J., Ai, H., Liu, L., Lao, S.: Multiple player tracking in sports video: A dual-mode two-way bayesian inference approach with progressive observation modeling. IEEE Trans. on Image Processing 20(6), 1652–1667 (2011)