A Generic Feature Selection Method for Background Subtraction Using Global Foreground Models

Marc Braham and Marc Van Droogenbroeck

{m.braham,M.VanDroogenbroeck}@ulg.ac.be

INTELSIG Laboratory, Montefiore Institute, University of Liège, Belgium

Link to the article

1. Introduction

Background subtraction techniques are widely used to extract moving objects from video sequences acquired with a static camera. The vast majority of these techniques manage the spatial distribution of background properties by maintaining a set of local background models. However, the features used to construct each local model are generally the same for all pixels of the image, thus reducing the capability of these models to discriminate between background and foreground objects. To overcome this limitation, we propose a generic feature selection method allowing to select, for each pixel, the most appropriate feature for a given background modeling strategy. Experiments conducted on the ViBe algorithm show that our feature selection framework improves the detection results.

2. Proposed local feature selection strategy

Our generic selection method is based on three main ideas:

1. Features should be selected according to their capability to discriminate between background and foreground values, which indirectly implies that we are able to estimate both background and foreground statistical distributions for each candidate feature.
2. The background modeling strategy impacts on the discriminating power of features.
3. Each feature choice tends to favor particular performance metrics. Therefore, the application performance metric should be provided to the local selection process.

Fig. 1 illustrates how these three main ideas are incorporated into our selection framework.

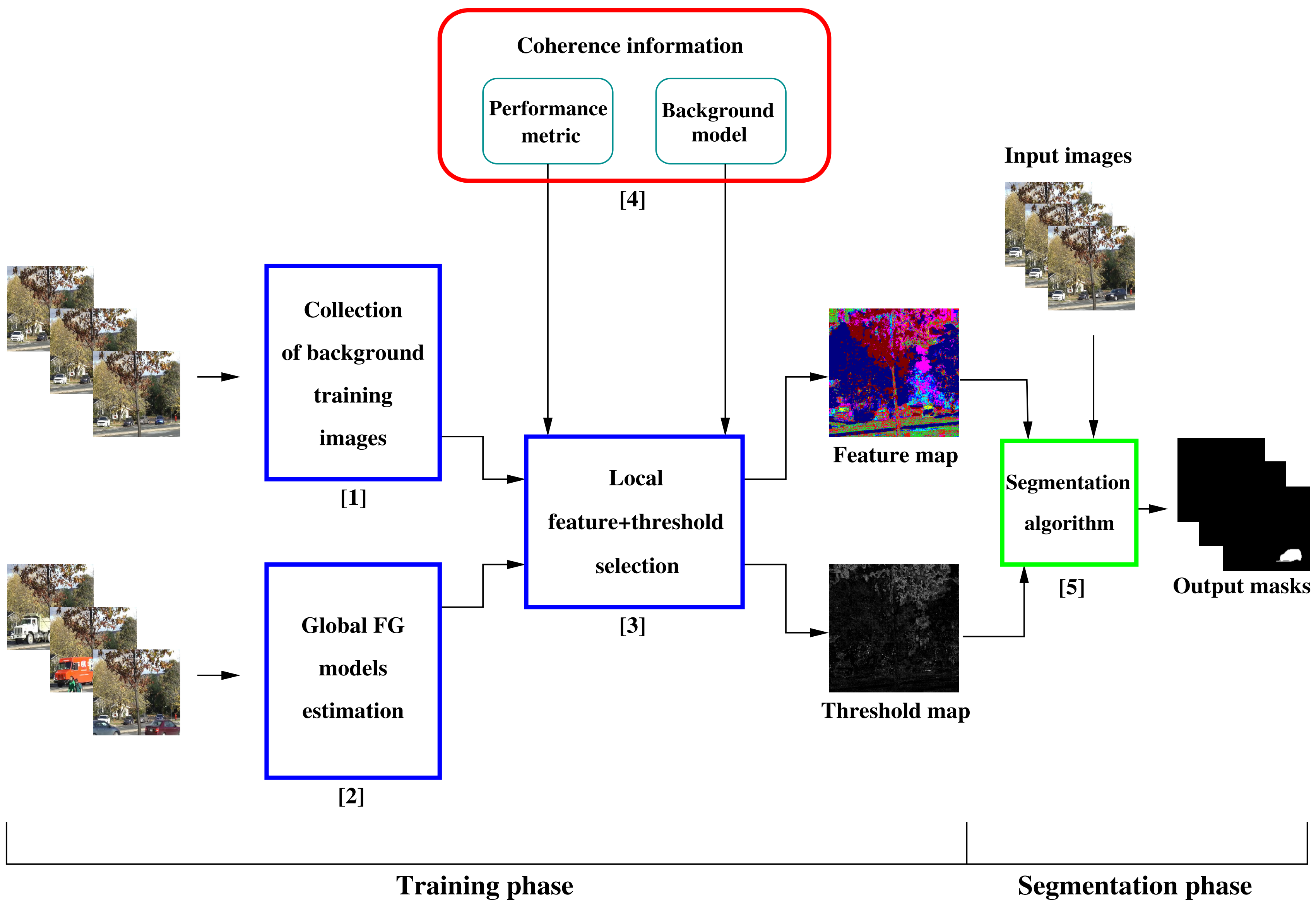

Figure 1 : Schematic representation of the proposed feature selection framework.

The selection process occurs during a training phase, which allows to avoid extra computations during normal background subtraction operations. This training phase is divided into three parts (boxes [1], [2], and [3] in Fig. 1). The first part consists to accumulate a few hundreds of images without foreground objects which are further processed to build local background statistical models. The second part requires another sequence of images, this time including moving objects in order to estimate a global foreground statistical distribution for each candidate feature. Our motivation behind a global foreground estimation is twofold:

- Compared to a local foreground estimation, it reduces drastically the required estimation time. Indeed, to the contrary of background features, which are frequent and stable, foreground features have generally wider probability density functions and lower priors. Assuming that local foreground distributions can be approximated by global foreground distributions makes a practical estimation feasible.
- If moving objects are very different from the background for some areas of the image but close for other areas, a global estimation will help to learn from the foreground values detected in the easiest areas of the scene.

The global foreground distributions are estimated by feeding an arbitrary background subtraction algorithm (named "estimator") with the second sequence of images. The segmentation masks of the estimator are processed to build $\#_f$ global foreground histograms of observed values, $\#_f$ denoting the cardinality of the set of candidate features.

The third part of the training phase consists to select locally the most appropriate combination feature/threshold. For each pixel, the collected background samples are used to build $\#_f$ background models (one per feature), according to the considered background modeling strategy. Then, each feature model is assessed for its capability to predict the correct class of input samples, these one being both (1) all local samples of the background training images, and (2) all global samples collected in the corresponding global foreground histogram. This capability is assessed for all candidate thresholds of a given feature, and described by a confusion matrix. The application performance metric is computed from the confusion matrix and the best feature/threshold combination is selected.

3. Experimental results

We particularized our generic method for the ViBe algorithm and evaluated it for several categories of the CDnet 2012 dataset. In our experiments, the feature set contains 9 individual color components taken from the 3 common color spaces RGB, HSV and YCbCr:

$$CF = \{R, G, B, H, S, V, Y, Cb, Cr\},\ \#_f = 9.$$

The Euclidean distance in the ROC space defined by $d_{euc} = \sqrt{FNR^2 + FPR^2}$ is used as application performance metric. This prior-independent metric measures the distance between the position of the assessed classifier and the position of the best theoretical classifier in the ROC space (top left corner, also called *oracle*).

In Fig. 2, we compare the results of ViBe applied on the 9 uniform feature maps (one per color component) to those obtained when the algorithm is fed by our local feature selection method (denoted by "ViBe-FT").

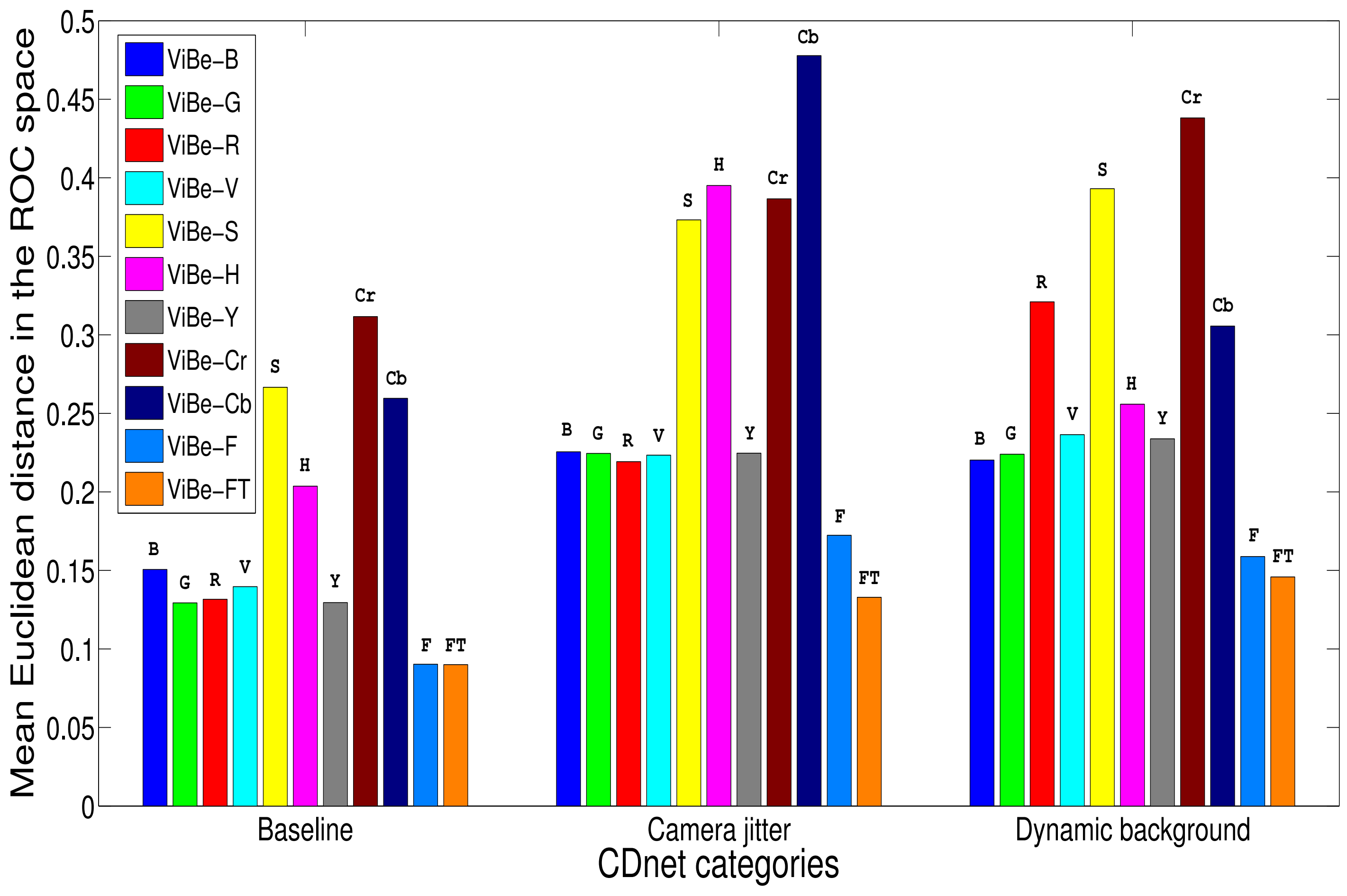

Figure 2 : Mean Euclidean distances in the ROC space for relevant CDnet 2012 categories.

Results displayed in Fig. 2 show that the proposed local feature selection framework significantly improves the detection results. For each category, our feature and threshold maps reduce the mean Euclidean distance of ViBe. This means that our framework pushes the background subtraction algorithm towards the oracle of the ROC space , and thus improves the detection. The conclusion is similar when the threshold selection mechanism is deactivated (denoted by "ViBe-F" on the graphic), which proves the robustness of the local feature selection process with respect to the threshold values. Fig. 3 shows several feature maps obtained after the local feature selection process, as well as the resulting improved masks.

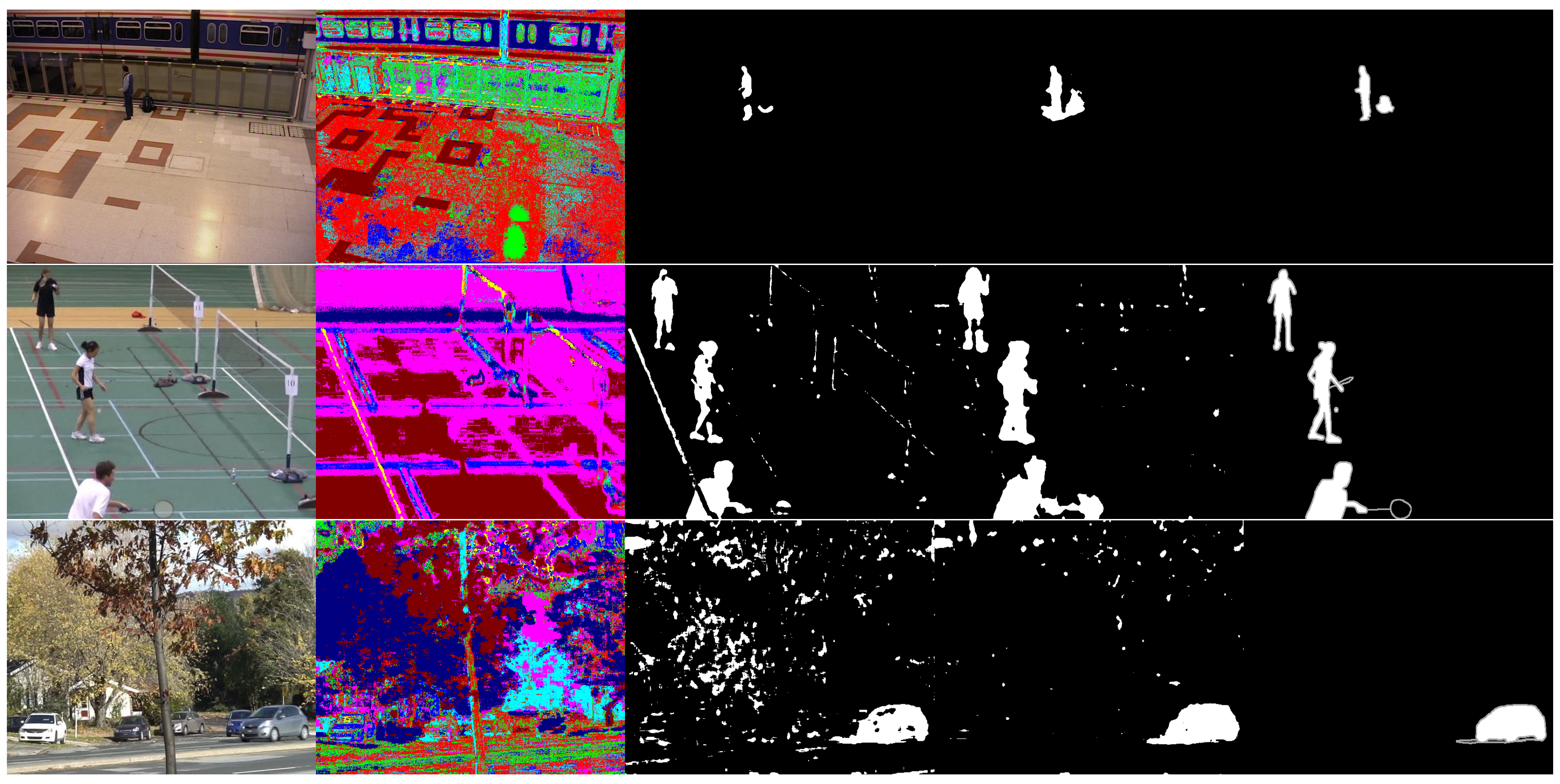

Figure 3 : Improvements obtained with our selection framework for three videos of CDnet. Columns from left to right show: the input images, the computed feature maps (see legend of Fig. 2 for the color-feature matching), the results of ViBe-G (lowest mean d_{euc} among the uniform feature maps), the results of ViBe-FT, and the ground-truths. Note that both ViBe-G and ViBe-FT are post-processed by a 9x9 median filter.

4. Conclusion

In this paper, we present a generic feature selection framework for background subtraction in video sequences. Our approach consists to select features locally depending on their capability to discriminate between local background samples and global foreground samples, according to the considered background modeling strategy and application performance metric. The selection process does not affect the computational complexity of the background subtraction step. Experiments led for the ViBe algorithm show that our method significantly improves the performance in the ROC space.