Classifying Global Scene Context for On-line Multiple Tracker Selection

Salma Moujtahid
salma.moujtahid@liris.cnrs.fr
Stefan Duffner
stefan.duffner@liris.cnrs.fr
Atilla Baskurt
atilla.baskurt@liris.cnrs.fr

Université de Lyon, CNRS, INSA-Lyon,
LIRIS, UMR5205,
F-69621, France

In this paper, we present a novel framework for combining independent on-line trackers using visual scene context. The aim of our method is to decide automatically at each point in time which specific tracking algorithm works best under the given scene or acquisition conditions.

In the literature, many ways of combining, fusing or selecting visual features have been presented. For example, low-level fusion of features (like motion or shape) is applied to improve the foreground-background discrimination (*e.g.* [2, 8]). Fusion is also possible at a higher level, where several trackers are run in parallel in order to select or combine their respective results (*e.g.* [1, 5]). In terms of model or feature fusion, our previous work Moujtahid *et al.* [6] concentrated on using confidence values of several individual trackers with different visual features coupled with a spatial-temporal coherence criteria to select the most suitable tracker at a given instant and enforce the continuity of tracking.

The main idea behind our framework is to use the strengths of different tracking algorithms as well as scene context information in order to improve the tracking performance. To this end, we introduce a framework that combines several independent and complementary trackers, each specialised on different scene conditions. The decision on which tracker to select is proposed by an off-line trained classifier which, in turn, is based on general scene context features that are independent from the trackers.

The general procedure of the proposed tracking framework is illustrated in Fig. 1. On a given video, N independent trackers $T_n, (n \in 1..N)$ run in parallel and, at every frame t, produce each an estimate of the object's state. This is usually a bounding box B_t^n with an associated confidence value $c_{t,n}$. The objective is to select at each frame the best tracker, *i.e.* the one that outputs the bounding box that fits best the object to track.

At the same time, the scene context features $\mathbf{f}_t$ are extracted. These features correspond to first and second order statistics of a given image-related variable such as intensity, colour and motion. They are computed on different image regions giving local, global and differential values.

The scene features $\mathbf{f}_t$ are concatenated with additional measures like the trackers confidences values $\mathbf{c}_t$ and the identifier of the last selected tracker s_{t-1} to form a large feature vector $\mathbf{i}_t$. An N-class classifier, that has been trained off-line on annotated data, is then applied on these features to estimate the best tracker for the given scene context. The classifier responds with $\mathbf{y}_t$, a probability for each class which is subsequently filtered by a Hidden Markov Model to ensure the temporal continuity of the tracker selection and reject outliers. The HMM estimates the posterior probability distribution $\mathbf{x}_t$, which is used to select the best tracker.

Finally, a Kalman Filter is applied as a post-processing step to temporally smooth the resulting object bounding box B_t^s from the selected trackers T_s. The result of the Kalman filter represents the final output of our tracking algorithm, and is further used to update the models of the individual trackers T_n.

Apart from this last update step, all the trackers are completely independent and do not cooperate or interact with each other. It is also important to mention that this approach is very generic, and in theory *any* on-line tracking algorithm can be integrated in this framework.

For our experiments, we used 3 on-line AdaBoost trackers [3] with different visual cues : Haar like features (HAAR), Histograms of Oriented Gradients (HOG) and Histograms of Colour (HOC). We also chose a fully connected Multi-Layer Perceptron (MLP) as scene context classifier.

We evaluated the performance of our framework on the Visual Object Tracking (VOT2013) benchmark [4]. First, we measured the classification rate of the proposed scene context classifier trained on the Princeton Dataset [7] and achieved a 81.80% rate of prediction. Secondly, we evaluated the overall tracking algorithm on the Visual Object Tracking (VOT2013) analysing the contribution of the different components, *i.e.* scene context classifier, HMM, and Kalman filter and comparing the proposed framework with other state-of-the-art tracking algorithms.

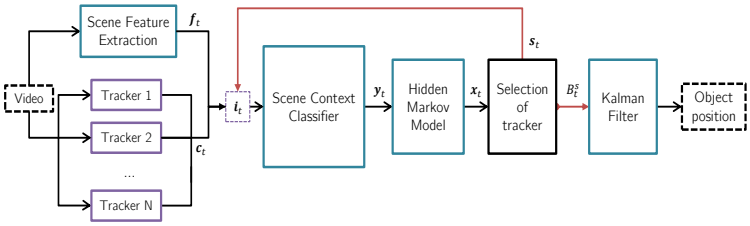

Figure 1: Overall framework of the proposed scene context-based tracking algorithm.

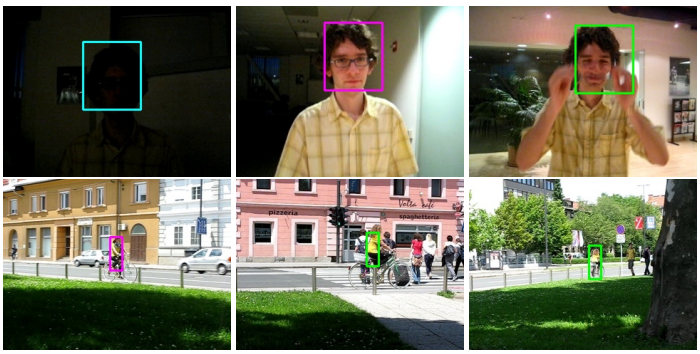

Figure 2: Illustration of our proposed framework's tracking results on the "David"(1^{st} row) and "Bicycle"(2^{nd} row) videos. Different scene context variations in lighting, texture or background are present throughout the videos. In the first part of the "David" video, the lighting is very poor and texture-based trackers usually work better, whereas in the second part the texture changes due to varying pose so colour-based trackers are more suitable. Our framework selects the most suitable tracker in each scenario (*pink*: HAAR, *blue*: HOG, *green*: HOC).

The proposed algorithm proved to increase the performance of our individual trackers, ranking among the top trackers of the VOT2013 challenge in terms of robustness.

[1] Christian Bailer, Alain Pagani, and Didier Stricker. A superior tracking approach: Building a strong tracker through fusion. In *Proc. of ECCV*, pages 170–185, 2014.

[2] Robert T. Collins and Yanxi Liu. On-line selection of discriminative tracking features. *IEEE Trans. on PAMI*, 27(10):1631–1643, 2005.

[3] Helmut Grabner, Michael Grabner, and Horst Bischof. Real-time tracking via on-line boosting. In *Proc. of BMVC*, pages 47–56, 2006.

[4] Matej Kristan, Luka Cehovin, Roman Pflugfelder, Georg Nebehay, Gustavo Fernandez, Jiri Matas, and et al. The Visual Object Tracking VOT2013 challenge results. In *Proc. of ICCV (Workshops)*, 2013.

[5] Ido Leichter, Michael Lindenbaum, and Ehud Rivlin. A general framework for combining visual trackers – "black boxes" approach. *IJCV*, 67(3):343–363, March 2006.

[6] Salma Moujtahid, Stefan Duffner, and Atilla Baskurt. Coherent selection of independent trackers for real-time object tracking. In *VISAPP*, pages 584–592, 2015.

[7] Shuran Song and Jianxiong Xiao. Tracking revisited using rgbd camera: Unified benchmark and baselines. In *Proc. of ICCV*, 2013.

[8] Alper Yilmaz, Xin Li, and Mubarak Shah. Object contour tracking using level sets. In *Proc. of ACCV*, 2004.