TEC2011-25995 EventVideo (2012-2014)

Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring

D5.2 QUALITY MEASURES AND FEEDBACK-DRIVEN ANALYSIS APPROACHES

Video Processing and Understanding Lab

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Supported by GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

AUTHOR LIST

Name	Email
Juan C. SanMiguel	Juancarlos.sanmiguel@uam.es

CHANGE LOG

Versión	**Date**	**Editor**	**Description**
0.1	28/11/2014	*Juan C. SanMiguel*	
1.0	30/11/2014	*José M. Martínez*	

CONTENT:

1. Introduction

1.1. Motivation

Traditionally, the processing scheme of video analysis systems is based on the feed-forward (or open-loop) approach that sequentially analyzes the data. The system output, computed as a function of the input data, is not used as a control variable of the processing. Its simplicity and low cost have motivated the wide spread among the existing video analysis systems. However, it does not consider the uncertainty when dealing with unexpected data and the dependence among processing levels. These limitations lead to low robustness of such systems for different operating conditions.

On the other hand, the feedback approach is proposed as a control method to increase the robustness of the system. It defines a closed-loop control that allows to *fed back* the output to the input of the system. Thus, an iterative analysis is performed until a desired performance level is achieved. Despite its advantages, its application in video analysis is not widely extended as the design of *feedback* control schemes is a complex task.

In this feedback processing loop, a key element is to estimate the quality of the generated data (i.e. processing) so decisions can be made whether to continue operating in the loop or to stop for requesting new data to be analysed. *Quality measures* estimate the uncertainty or the output quality of the generated data by each stage of video processing systems.

Foreground segmentation and tracking are the basic stages for many video applications The evaluation of the output quality of video object segmentation and tracking algorithms is crucial to estimate their accuracy and to tune their parameters for optimal performance. Although analytical approaches exist, this evaluation is typically performed by comparing the obtained results with manual annotations (or ground-truth, GT). However, manual annotation is time consuming and prone to human error. It usually covers a small set of video sequences only representing a small percentage of data variability. This limitation complicates the extrapolation of the performance evaluation results to new (unlabeled) sequences. Moreover, evaluation using ground truth is not feasible for on-line performance analysis. Conversely, the evaluation not-based on ground-truth (NGT) is a desirable option to overcome these limitations.

Few approaches currently exist for performing feedback-based analysis guided by quality measures. Within the context of the EventVideo project, this research line aims to contribute to the state of the art in three areas: video object segmentation, visual tracking and feature extraction.

1.2. Objetives

The objectives are as follows:

- Development of quality measures for generic video tracking approaches

- Development of quality measures for specific video tracking approaches such as Particle Filters.
- Development of quality measures for feature extraction
- Development of feedback control schemes

1.3. Estructure of the document

This document is structured as follows:

- Chapter 1 introduces the research areas covered in the document, its purpose within the scope of the eventVideo project and the document structure.

- Chapter 2 provides an overview of the approaches developed in the area of quality measures where the efforts are driven towards video tracking

- Chapter 3 provides an overview of the developments for analysis based on feedback which applies iterative analysis schemes, thus adapting the analysed content.

- Chapter 4 concludes this document by summarizing the major findings and presenting future lines of research.

2. Quality measures

In this section we present the achievements for quality estimation, focused on video tracking.

2.1. Temporal validation of particle filters for video tracking

2.1.1. Approach overview

We propose an approach for determining the temporal consistency of Particle Filters in video tracking based on model validation of their uncertainty over sliding windows. The filter uncertainty is related to the consistency of the dispersion of the filter hypotheses in the state space. We learn an uncertainty model via a mixture of Gamma distributions whose optimum number is selected by modified information-based criteria. The time-accumulated model is estimated as the sequential convolution of the uncertainty model. Model validation is performed by verifying whether the output of the filter belongs to the convolution model through its approximated cumulative density function. Experimental results and comparisons show that the proposed approach improves both precision and recall of competitive approaches such as Gaussian-based online model extraction, bank of Kalman filters and empirical thresholding. We combine the proposed approach with a state-of-the-art online performance estimator for video tracking and show that it improves accuracy compared to the same estimator with manually tuned thresholds while reducing the overall computational cost. The approach has been published in the Computer Vision and Image Understanding Journal [1]

The following figure shows an example of a Particle-Filter-based tracker where the filter becomes inconsistent as most of the hypotheses are apart from each other and have small weights. The proposed approach *aims to detect such behaviour over time for Particle Filters* and determine that the tracker is not correctly operating.

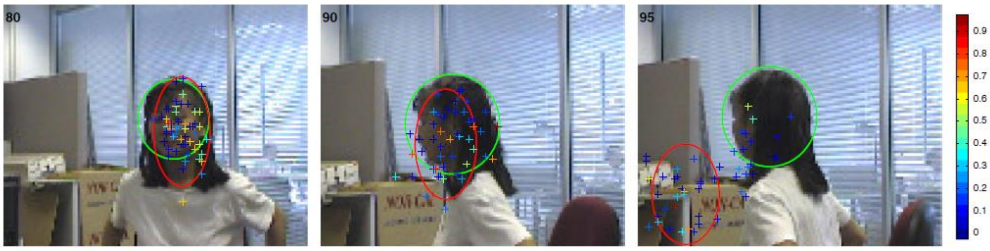


Figure 1. Example of filter consistency for face video tracking using a color-based Particle Filter (with 100 particles). The green ellipse represents the ideal target; the red ellipse represents the estimated target. The left image illustrates a consistent behavior. The central and right image illustrate inconsistent situations. The particles (identified for clarity only by their center) are colored according to their weights: the warmer the color, the higher the weight. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The proposed approach starts from the output generated by a Particle Filter and considers two stages: Uncertainty Estimation and Model Validation. Model validation provides a robust framework for Particle Filter consistency analysis whose performance could be improved by sliding windows. For measuring the consistency of the Particle Filter, we first compute the uncertainty of its posterior and accumulate its change over a temporal window. Then, we validate an uncertainty change model to check consistency (Figure 2). We term the proposed approach as Accumulated Validation of Uncertainty (AVU).

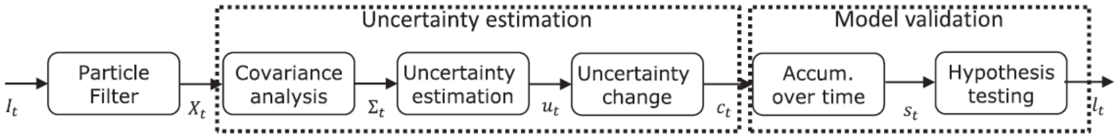


Figure 2. Block diagram of the proposed approach.

We estimate the uncertainty for each time *t* by measuring the spread of the generated hypotheses in the state space (particles) through the covariance matrix of the filter output. The following figure shows and example.

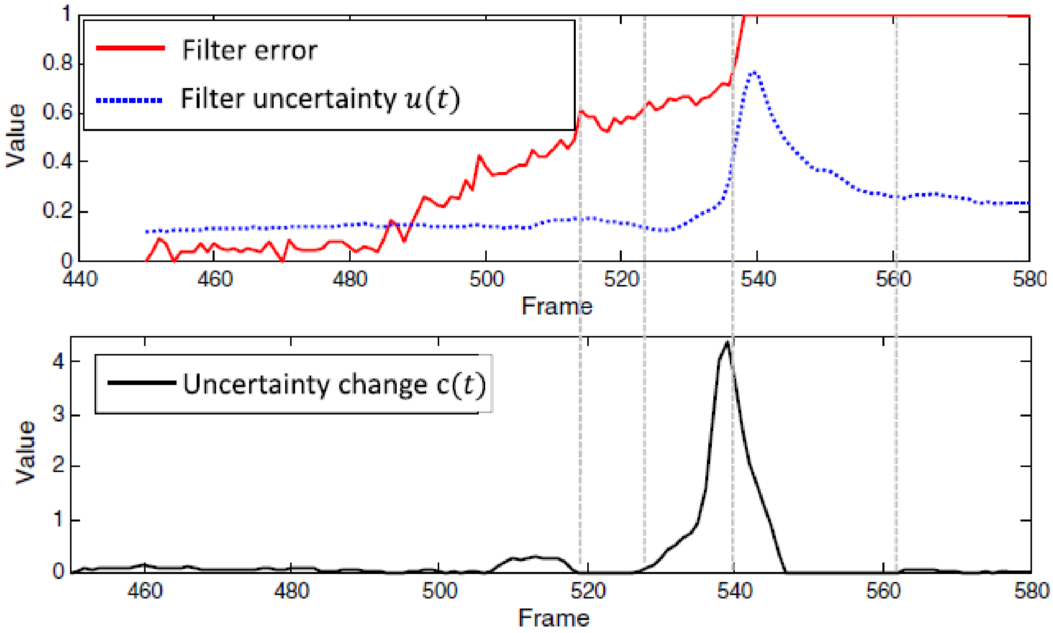


Figure 3. Evolution of the filter uncertainty and its error for color-based Particle Filter video tracking. Green and red ellipses are, respectively, ideal and estimated target locations. Sample frames correspond to vertical dotted lines. The filter error was computed as the Dice coefficient.

We estimate the uncertainty for each time *t* by measuring the spread of the generated hypotheses in the state space (particles) through the covariance matrix of the filter output. The following figure shows and example.

The problem consists of detecting changes in the time series *c(t)* (uncertainty change over time), which is sampled from a random variable Q following a certain probability density function (pdf). For increasing robustness of model validation, we accumulate *c(t)* by using a sliding window of length L, obtaining a test statistic *s(t).* Finally, a null hypothesis test is performed for *s(t)* in order to determine whether *s(t)* belongs to a random variable S following a certain (pdf). The test is as follows:

$$H_0 : s_t \in S$$
$$H_1 : s_t \notin S,$$

After modelling the consistent filter status (i.e. pdfs for *c(t)* and its window-based accumulation *s(t)*), we obtain a decision rule to perform the hypothesis test.

We combine AVU into an online method performance evaluation of Particle Filter-based video tracking, ARTE [6]. ARTE determines whether the Particle Filter is successfully estimating the target state without the use of ground-truth. ARTE analyses the Particle Filter consistency and the time-reversibility property of target motion.

2.1.2. Experimental results

For performing experiments, we use two evaluation sets (D1 and D2) with sequences selected from the following datasets: CAVIAR, PETS2001, PETS2010, CLEMSON, VISOR, AVSS2007, TRECVID and MIT TRAFFIC. D1 is the same set as in [6], which is composed of 18 sequences (>3400 annotated frames). D2 contains 51 sequences (>7500 annotated frames). Samples are shown in the following figure:

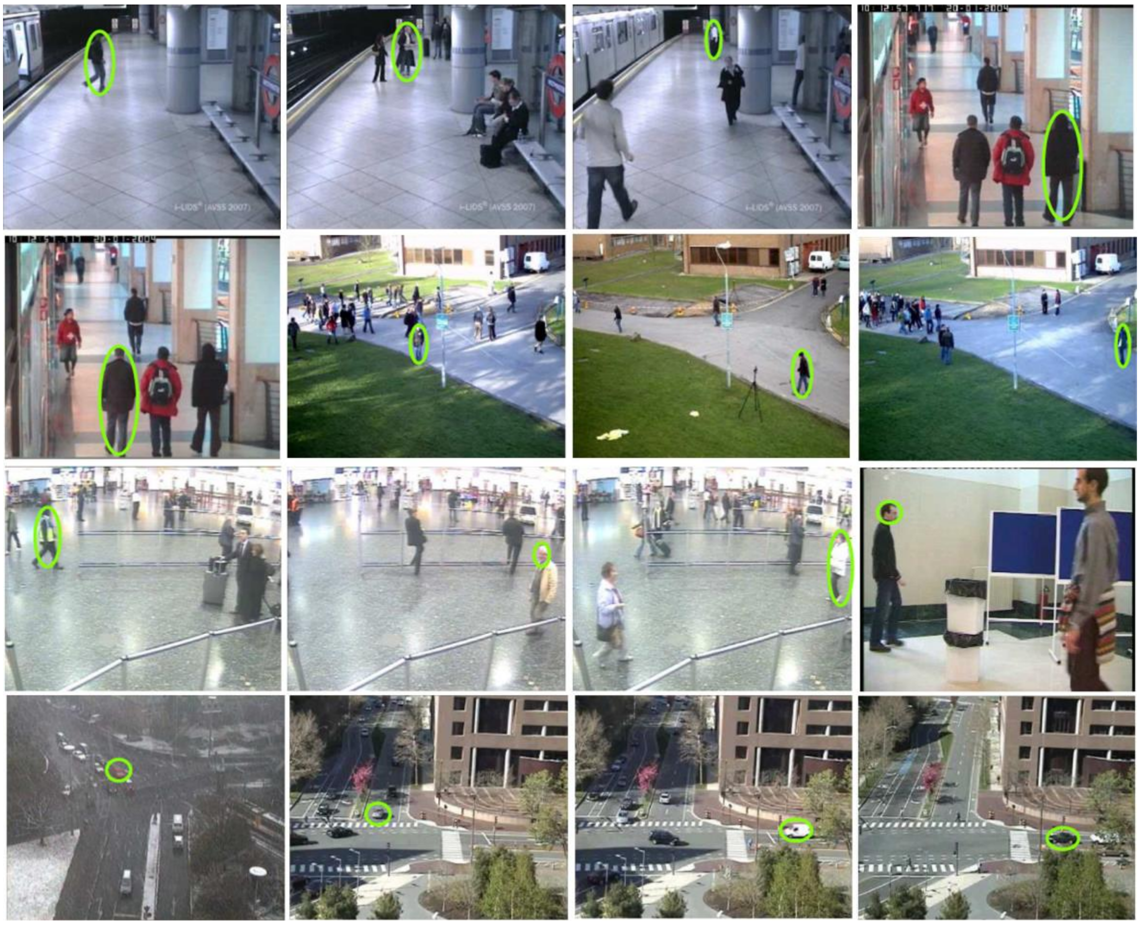


Figure 4. Sample images of the employed dataset

We compare AVU against representative approaches for online change detection without thresholding: the two-model sliding window (Two_MChi) that assumes Gaussian-distributed data, the bank of Kalman filters adapted to various change hypothesis Mmodel) and the empirical thresholding approach (EmpTh) [6], which is tuned using D1. All approaches are applied to the uncertainty change signal *c(t)*. Experiments with different lengths of the sliding window (L) are performed for testing the robustness of AVU and the results are summarized in the following figure.

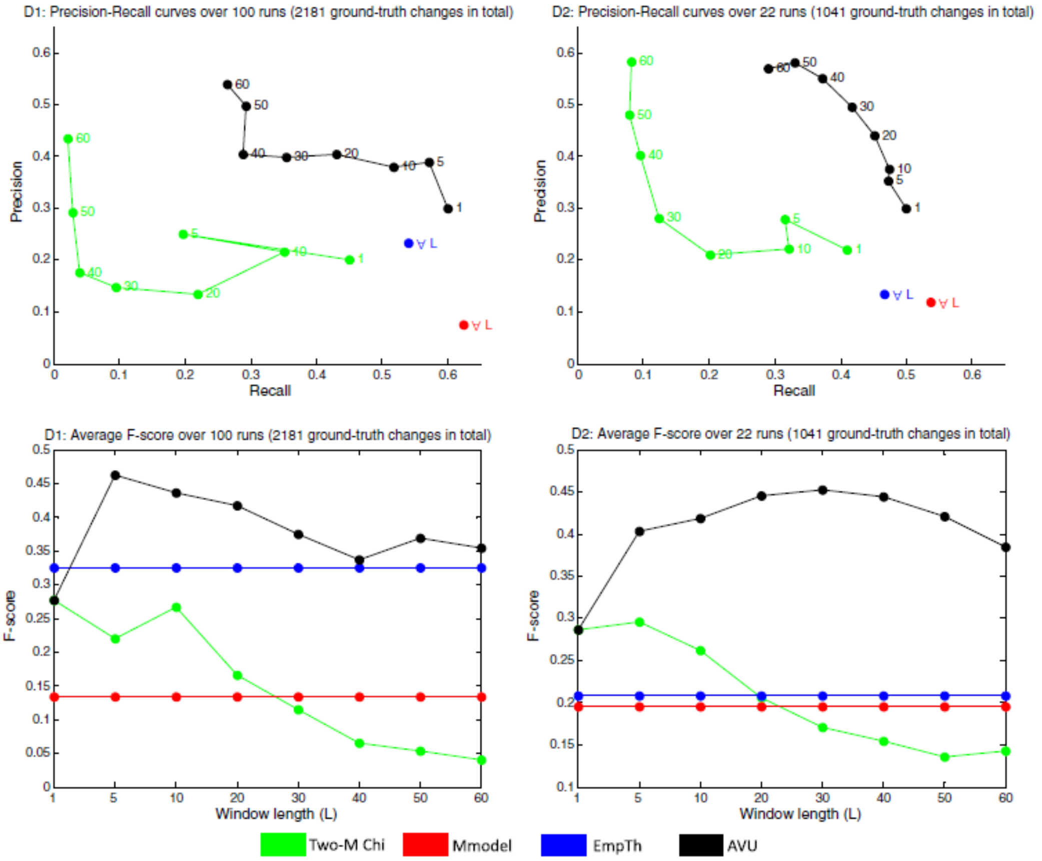
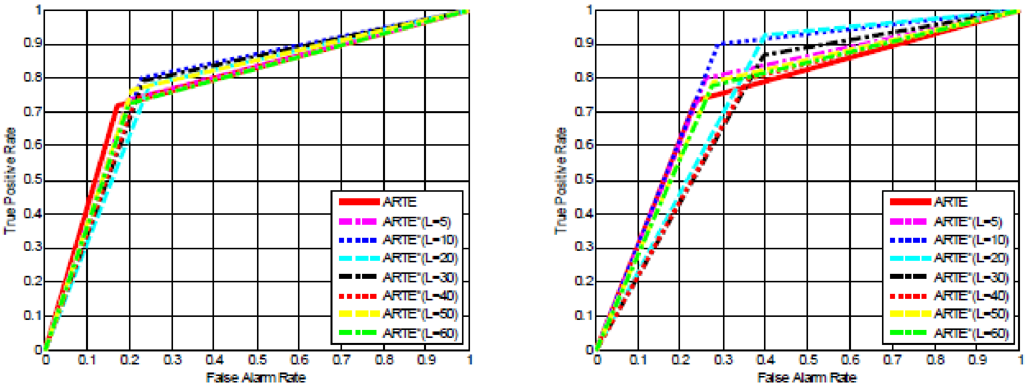


Figure 5. Comparison for selected change detection approaches with different lengths (L) of the sliding windows for evaluation sets D1 (left) and D2 (right).

The results of the proposed approach for online evaluation (hereinafter ARTE*) are presented in the following figure. The left part of the figure shows that ARTE* has similar accuracy to ARTE for D1. A noticeable improvement in TPR is observed for ARTE* with all lengths. However, ARTE* slightly increases the False Positive Rate compared to ARTE because of the use of the sliding window, requiring a higher amount of variation to detect an uncertainty change. This implies in some situations a short delay in the detection of changes. ARTE* reach similar performance to that of the change detector of ARTE whose threshold values were manually tuned on the same dataset (D1). The right part of the figure (results on D2) shows a situation where the thresholds of ARTE are not optimum. As it can be observed, shorter windows got higher results than that of ARTE demonstrating that the proposed approach generalizes better than the optimal thresholding of ARTE. However, a performance decrease is observed as the length of the window increases due to the reduction of the number of detected changes. The main advantage of ARTE* over ARTE is that it does not require to setup any threshold.



Approach	AUC	TPR	FPR
ARTE [13]	.772 ± .06	.717 ± .05	.172 ± .02
ARTE* (*L*=*5*)	.770 ± .07	.737 ± .05	.197 ± .03
ARTE* (*L*=10)	.785 ± .05	.799 ± .03	.228 ± .01
ARTE* (*L*=20)	.766 ± .04	.779 ± .04	.247 ± .02
ARTE* (*L*=30)	.750 ± .07	.760 ± .05	.235 ± .01
ARTE* (*L*=40)	.743 ± .06	.739 ± .03	.224 ± .01
ARTE* (*L*=50)	.700 ± .02	.765 ± .03	.236 ± .03
ARTE* (*L*=*60*)	.698 ± .04	.723 ± .06	.300 ± .04

Approach	AUC	TPR	FPR
ARTE [13]	.747 ± .06	.732 ± .14	.237 ± .05
ARTE* (*L*=*5*)	.770 ± .07	.800 ± .14	.261 ± .04
ARTE* (*L*=10)	.806 ± .05	.900 ± .10	.289 ± .03
ARTE* (*L*=20)	.763 ± .04	.926 ± .08	.401 ± .03
ARTE* (*L*=30)	.735 ± .07	.868 ± .14	.398 ± .03
ARTE* (*L*=40)	.717 ± .06	.796 ± .12	.361 ± .03
ARTE* (*L*=50)	.722 ± .06	.767 ± .14	.279 ± .05
ARTE* (*L*=*60*)	.710 ± .06	.776 ± .14	.276 ± .04

Figure 6. ROC analysis for successful-unsuccessful segmentation of video tracking for sets D1 (left) and D2 (right). Data are presented as mean _ standard deviation. (Key. ARTE: Adaptive Reverse Tracking Evaluation [6]; ARTE*: threshold-automatic ARTE; AUC: area under the curve, FPR: false positive rate, TPR: true positive rate).

We demonstrate the generality of the proposed approach by evaluating two state-of-the-art trackers [2][3]. The first tracker models targets as fragments adaptively selected over time which are embedded in the PF framework [2]. The second tracker performs multi-hypothesis estimation based on sparse appearance models, presenting a PF-like structure [3]. We employ the code provided by the authors with the default parameter settings. For the proposed approach, we learn the pdfs for each tracker using D1 dataset and we use L = 20 as a compromise between the previously described results for D1 and D2 datasets. The *EmpTh* approach is tuned to get best results for D1. The presented results are the mean of 10 runs.

(a) Dataset D1

Approach	Color-tracker [17]			Frag-tracker [34]			Sparse-tracker[35]		
	P	R	F	P	R	F	P	R	F
Two-M Chi	.133 ± .01	.220 ± .02	.166 ± .01	.186 ± .03	.251 ± .02	.214 ± .01	.080 ± .01	.788 ± .04	.145 ± .03
Mmodel	.074 ± .03	**.624 ± .01**	.133 ± .02	.134 ± .02	.541 ± .01	.215 ± .01	.104 ± .02	.718 ± .03	.182 ± .02
EmpTh	.233 ± .06	.539 ± .03	.326 ± .04	.142 ± .01	.530 ± .03	.224 ± .02	.102 ± .06	.410 ± .04	.163 ± .04
AVU (L=20)	**.404 ± .03**	.430 ± .04	**.417 ± .02**	**.264 ± .04**	**.587 ± .02**	**.364 ± .03**	**.264 ± .09**	**.503 ± .03**	**.346 ± .05**

(b) Dataset D2

Approach	Color-tracker [17]			Frag-tracker [34]			Sparse-tracker[35]		
	P	R	F	P	R	F	P	R	F
Two-M Chi	.210 ± .00	.202 ± .00	.206 ± .00	.139 ± .00	.071 ± .00	.094 ± .00	.080 ± .00	.525 ± .00	.138 ± .00
Mmodel	.119 ± .00	**.537 ± .00**	.195 ± .00	.119 ± .00	.245 ± .00	.160 ± .00	.099 ± .00	.475 ± .00	.164 ± .00
EmpTh	.134 ± .00	.466 ± .00	.208 ± .00	.134 ± .00	.245 ± .00	.173 ± .00	.102 ± .00	.468 ± .00	.167 ± .00
AVU (L=20)	**.440 ± .00**	.451 ± .00	**.446 ± .00**	**.328 ± .00**	**.263 ± .00**	**.292 ± .00**	**.253 ± .00**	**.728 ± .00**	**.375 ± .00**

Table 1. Comparison of change detection approaches for the selected PF-based trackers. Best results are indicated in bold.

2.2. Feature-based online validation of video tracking

2.2.1. Approach overview

To overcome the problems related to evaluation, we propose an alternative approach for online evaluation of single-object trackers without ground-truth data. It is based on the temporal evolution of covariance features only requiring a bounding box as tracker output. Unlike previous work focused on the unsuccessful tracker case, the proposed approach models the successful case and identify model deviations via a validation strategy. Then, a two-state machine determines the successful tracker results. This work has been published as a letter [4] and final degree project [5].

An overview of the proposed approach is shown in the following figure. It starts from the target location estimated by the tracker at time *t*, $\boldsymbol{x_t}$ = *[x_t; y_t;w_t; y_t; o_t]*, where (x_t; y_t), w_t;, h_t and o_t are the center, width, height and orientation of the target, respectively. The proposed approach can used most of existing trackers as they output $\boldsymbol{x_t}$. Then, we measure the target appearance structure in $\boldsymbol{x_t}$ via the covariance feature Σ.

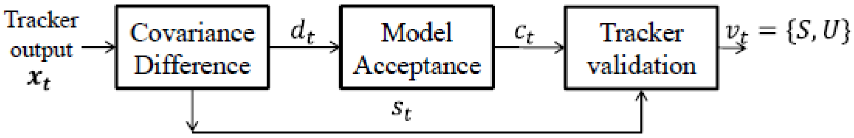


Figure 7. Block diagram of the proposed approach.

We detect dissimilar covariance features over time via a model acceptance strategy. We consider a model D to define the variability of the distance between consecutive convariance matrices during successful tracker operation, which follows a probability density function. We perfom hypothesis testing for model acceptation where the null hypothesis H0 indicates that the covariance change is consistent with the model D. Let H1 be the hypothesis that an unknown change has occurred. Model acceptance is formulated via simple hypothesis testing.

Once we determine the consistency of the filter, we employ a finite state machine to validate the tracker operation (see the following figure) where two states are defined for the successful (S) and unsuccessful (U) cases. Starting from the S state, the S->U transition is triggered when the H1 hypothesis is detected due to tracker failures (target loss). The U->S transition is when the tracker recovers to the correct target after a failure. It is activated when H1 hypothesis is accepted and the new tracker output is similar to the previously tracked target. We compute the similarity between the last successful and the new tracker outputs to determine if we are tracking an old target.

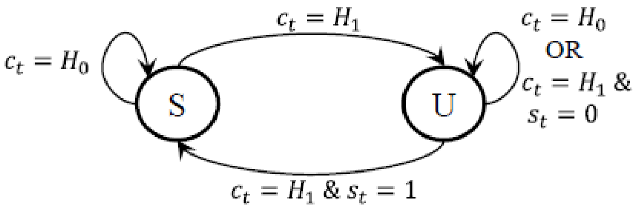
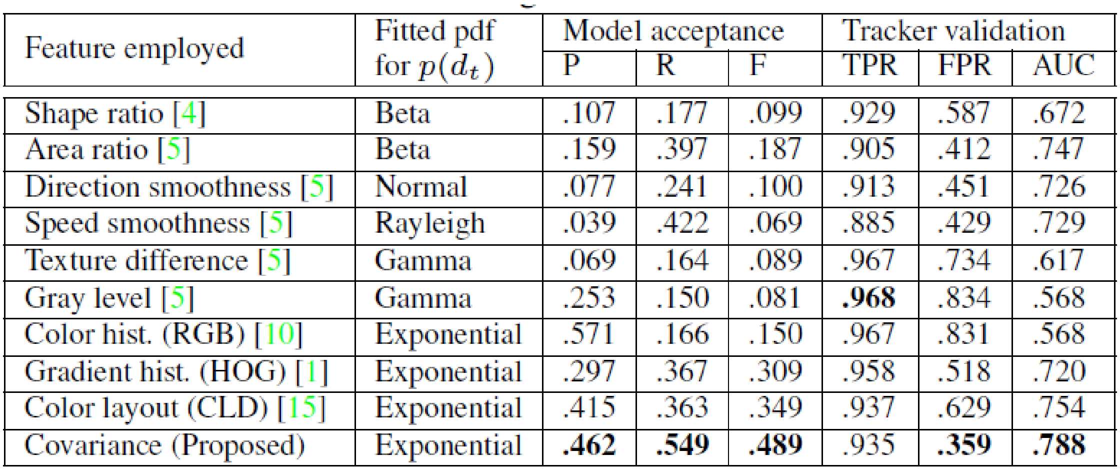


Figure 8. Finite state machine to validate the tracker output using two states: successful (S) and unsuccessful (U). *c(t)* and *s(t)* are variables for model acceptance and tracker recovery checks.

2.2.2. Experimental results

We use the SOVTds dataset [8] for evaluation and we validate the results of six well known trackers based on Mean-Shift, Color-based Particle Filter, Incremental Visual Tracking, Tracking- Learning-Detection, SuperPixel Tracking and Locally Orderless Tracking. The code of the original authors is used to analyse the dataset and get the tracker results for validation (~138000 in total).

The following table compares common features in video tracking against the covariance feature, all applied within the proposed approach. For each feature, the pdf is modeled as the best fitting of popular distributions using the Kolmogorov-Smirnov statistic over the training set. The results show low performance for features based on contour (shape and area), motion (speed and direction) and color (gray, RGB histograms and texture) information, demonstrating their low discriminative power between the successful and unsuccessful cases. Structure-based features (HOG, CLD and Covariance) present the best results showing that the target appearance structure exhibits short-term stability. Figure 9 shows an example of the proposed approach where the three tracker errors (frames 90, 131-164 and 195-214) are correctly identified.



Feature employed	Fitted pdf for $p(d_t)$	Model acceptance			Tracker validation		
		P	R	F	TPR	FPR	AUC
Shape ratio [4]	Beta	.107	.177	.099	.929	.587	.672
Area ratio [5]	Beta	.159	.397	.187	.905	.412	.747
Direction smoothness [5]	Normal	.077	.241	.100	.913	.451	.726
Speed smoothness [5]	Rayleigh	.039	.422	.069	.885	.429	.729
Texture difference [5]	Gamma	.069	.164	.089	.967	.734	.617
Gray level [5]	Gamma	.253	.150	.081	**.968**	.834	.568
Color hist. (RGB) [10]	Exponential	.571	.166	.150	.967	.831	.568
Gradient hist. (HOG) [1]	Exponential	.297	.367	.309	.958	.518	.720
Color layout (CLD) [15]	Exponential	.415	.363	.349	.937	.629	.754
Covariance (Proposed)	Exponential	**.462**	**.549**	**.489**	.935	**.359**	**.788**

Table 2. Performance (mean results) of the proposed approach using common features for video tracking. Bold indicates best results.

[1] Smeulders, A., Chu, D., Cucchiara, R., Calderara, S., Dehghan, A., and Shah, M.: 'Visual Tracking: An Experimental Survey', IEEE Trans. Pattern Anal. Mach. Intell., 2014, 36, (7), pp.1442-1468

[4] Chau, D., Thonnat, M., Brémond, F., and Corvée, E.: 'Online parameter tuning for object tracking algorithms', Image Vis. Comput., 2014, 32, (4), pp. 287-302

[5] Spampinato, C., Palazzo, S., and Giordano, D.: 'Evaluation of tracking algorithm performance without ground-truth data', Proc. of IEEE Conf. on Image Process., Orlando (USA), Oct. 2012, pp.1345-1348

[10] Nummiaro, K., Koller-Meier, E., and Gool, L.V.: 'An adaptive colour-based particle filter', Image Vis. Comput., 2002, 21, (1), pp. 99-110

[15]Manjunath, B., Ohm, J., Vasudevan, V., and Yamada, A.: 'Color and texture descriptors', IEEE Trans. Circ. Syst. Video Technol., 2001, 11, 6, pp.703-715

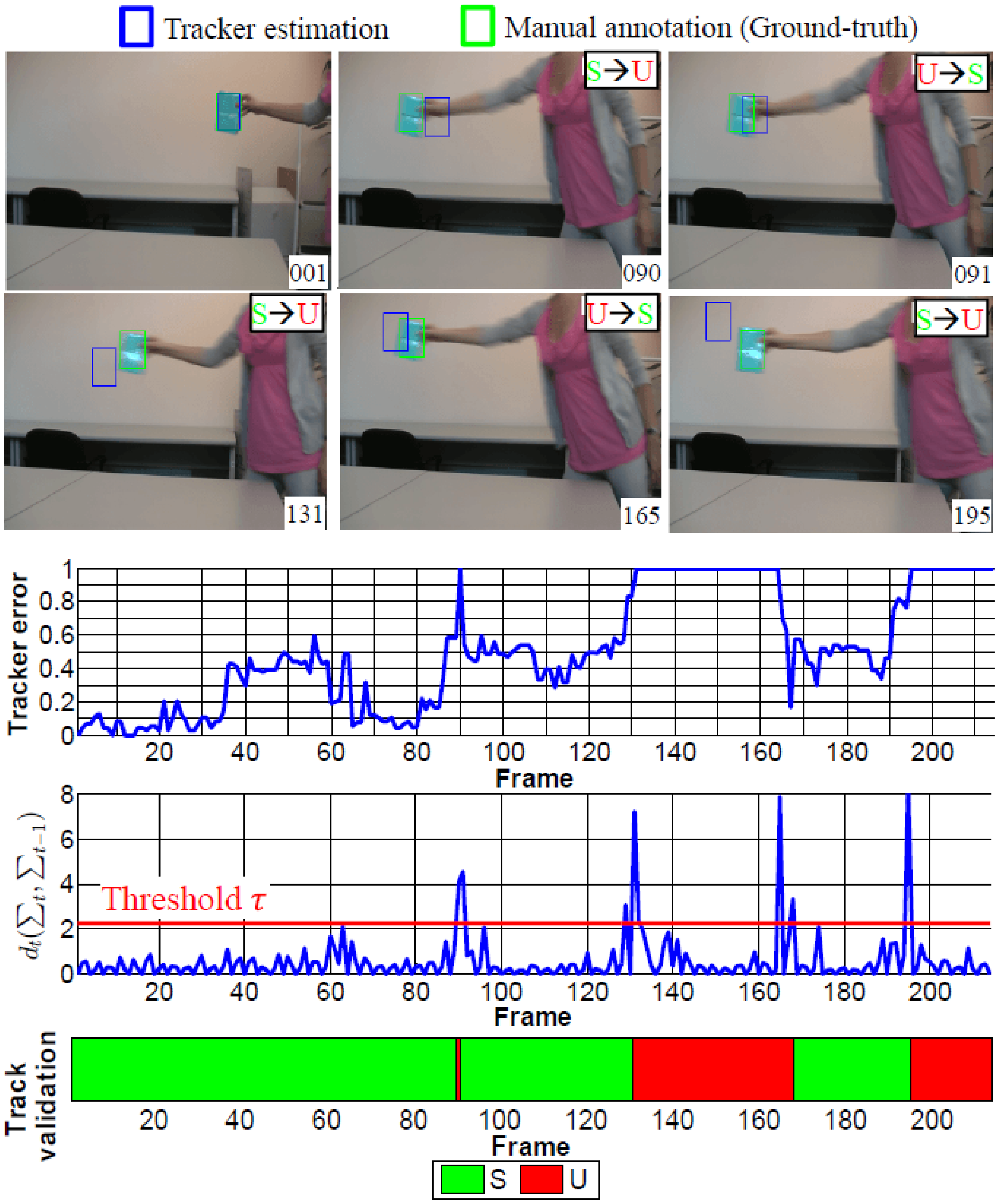
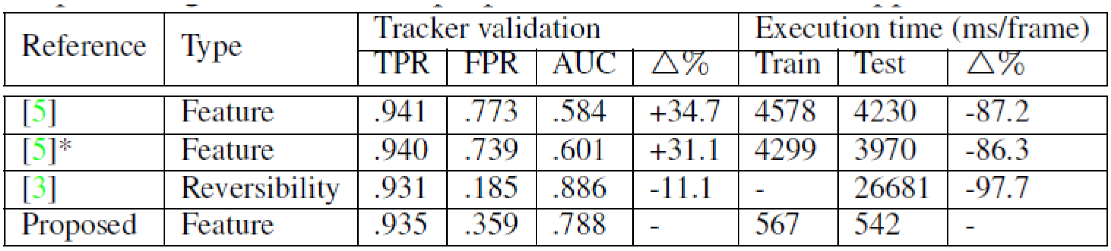


Figure 9. Example for online validation of tracking results between successful (S) and unsuccessful (U) for the Mean-Shift (MS) tracker. From top to bottom graphs: error as the spatial overlap between the estimation and ground-truth data, covariance difference d_t and final tracker validation.

Table 3 compares the results of the proposed approach against the related state-of-the-art in terms of accuracy and computational cost. For feature-based approaches, the proposal clearly improves the accuracy of Hao et al (and its modification using the best feature), showing the benefits of model validation over a two-model Bayesian classifier for successful and unsuccessful cases. Moreover, the computational cost is reduced as only covariance feature is employed instead of multiple features in Spampinato et al. Compared to reverse validation of Hao et al, the proposed approach reduces the computation cost around 50x as compared to Hao et al. Moreover, the computations of Hao et al depend on the sequence length whereas the proposed approach has a bounded computation. This limitation of Hao et al prevents its use for many applications where execution time is critical and for long sequences as the computational cost is not affordable. Therefore, the proposed approach allows a broader application of online validation as compared to Hao et al, offering a trade-off between accuracy and cost. Note that we do not compare with PF-based approaches and approaches with low-performing features (motion speed and smoothness, see Table 2).



Reference	Type	Tracker validation				Execution time (ms/frame)		
		TPR	FPR	AUC	△%	Train	Test	△%
[5]	Feature	.941	.773	.584	+34.7	4578	4230	-87.2
[5]*	Feature	.940	.739	.601	+31.1	4299	3970	-86.3
[3]	Reversibility	.931	.185	.886	-11.1	-	26681	-97.7
Proposed	Feature	.935	.359	.788	-	567	542	-

Table 3. Comparative results (mean) for online tracker validation. The symbol '*' is for [5] using only the best feature. Δ% shows the difference (in percentage) between the proposed and each selected approach.

[3] Hao, W., Sankaranarayanan, A., and Chellappa, R.: 'Online Empirical Evaluation of Tracking Algorithms', IEEE Trans. Pattern Anal. Mach. Intell., 2010, 32, (8), pp.1443-1458

[5] Spampinato, C., Palazzo, S., and Giordano, D.: 'Evaluation of tracking algorithm performance without ground-truth data', Proc. of IEEE Conf. on Image Process., Orlando (USA), Oct. 2012, pp.1345-1348

3. Feedback-driven analysis

In this section, we present a feedback-based approach to extract features (skin) in images, that correspond to human body parts is an important task in many areas such as human– computer interaction, gesture analysis and content-based image retrieval.

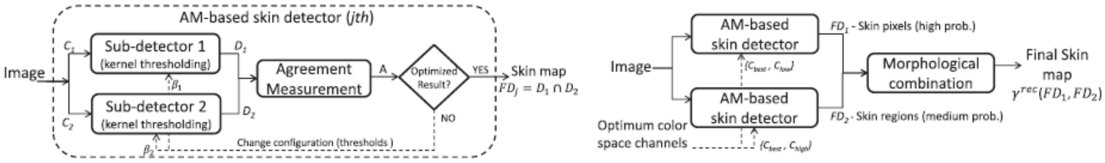
3.1. Skin detection by dual maximization of detectors agreement for video monitoring

3.1.1. Approach overview

We propose an approach to detect skin in single images of human activity recognition videos where, for each image, it dynamically selects the best configuration starting from a predefined one. The approach has been published in the Pattern Recognition Letters Journal [7].

First, we introduce such adaptation using the AM framework. It selects the best detectors's configuration based on agreement maximization (AM) and consists of three basic elements (detectors applied, agreement measure and optimization process). However, this framework has no constraints in the parameter optimization process which makes the thresholds tend to increase the number of false positives or negatives as agreement is high in certain non-desirable situations. Moreover, there is no indication of which channels of colour spaces are better for increasing the agreement and complex combination schemes can be designed considering the properties of the employed detectors. These detectors are improved by improved by learning parameter relations through kernel thresholding and including a new agreement measure (see Figure 10(a)).

Then, two AM-based detectors are composed to detect skin-like regions and high-probable skin pixels (via optimal selection of color space channels), which are later combined using binary morphology (see Figure 10(b)) for maximizing performance.

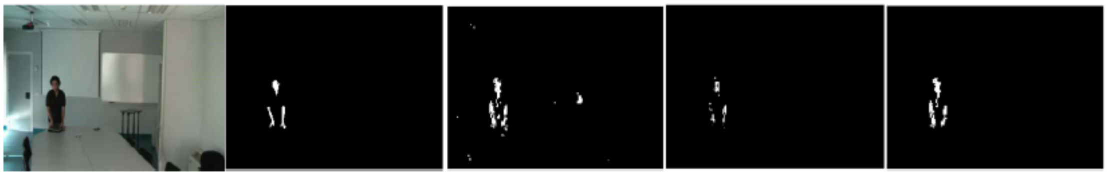


(a) (b)

Figure 10. Block diagrams of the proposed (a) detector based on agreement maximization (AM) and (b) framework for skin detection in images.

For each type of scenario, we obtain the best color space channels among the most popular ones (RGB, HSV, YCbCr and Lab) to detect skin pixels by determining their discriminative capabilities over the training data. We conform the two detectors by using the color channels {*Cbest,Clow*} for *FD*1 and {*Cbest,Chigh*} for *FD*2.

After selecting the optimum channels of the AM-based skin detectors and optimizing their parameters, they are combined to improve the final result of the skin detection. Considering that one of the detectors obtains highly probable pixels whereas the other gets compact skin-like regions (that might correspond to skin or similar objects), we propose to use a morphological reconstruction filter to retain only the skin regions marked by the highly probable pixels of FD2(among all the regions of FD1). The following figure shows an example of the proposed approach



Input image (a) Ground-truth (b) FD_1 (c) FD_2 (d) $\gamma^{rec}(FD_1, FD_2)$ (e)

Figure 11. Sample results for an image of the EDds dataset showing the output of the skin detectors FD1 (H-a colour channels) and FD2 (H-b colour channels) after optimum channel selection and their combination through mathematical morphology.

3.1.2. Experimental results

As evaluation set, we have selected images from public datasets for human activity recognition: AMI (http://corpus.amiproject.org/), EDds (http://www-vpu.eps.uam.es/DS/EDds/), SSG (http://www-vpu.eps.uam.es/publications/SkinDetDM) LIRIS (http://liris.cnrs.fr/harl2012) and UT (http://cvrc.ece.utexas.edu/SDHA2010). This set covers a wide variety of situations, viewing distances and resolutions (ranging from 320x240 to 720x576) where skin detection has many challenges due to, among others, illumination changes or poor visibility. For each dataset, around 50 images have been selected and the corresponding ground truth has been manually generated at pixel level. In total, 290 images compose the evaluation set containing more than 870000 skin pixels, which have been equally divided into two sets for training (~450000) and testing (~420000). This subsection presents selected results of the proposed approach.

The following figure depicts the mean detection ratio of the histogram-models computed for each channel over the training set. As it can be observed, the minimum value corresponded to the H channel (of HSV) for all the datasets except for UT, where a channel (of Lab) obtained best results, closely followed by H. This indicates that H channel had stable results and high discriminative power (for skin and non-skin regions) in the considered scenarios.

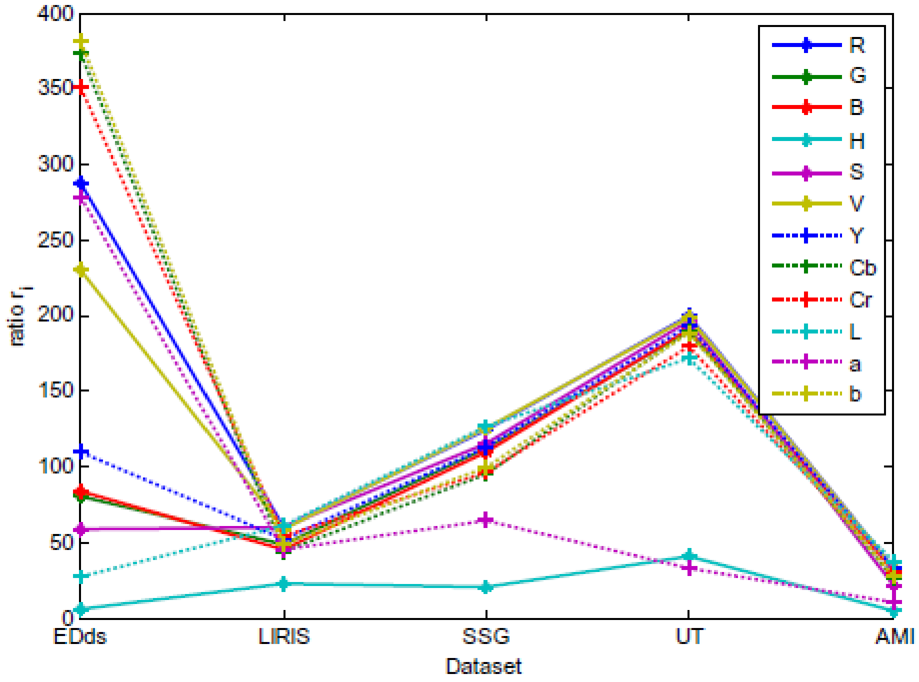
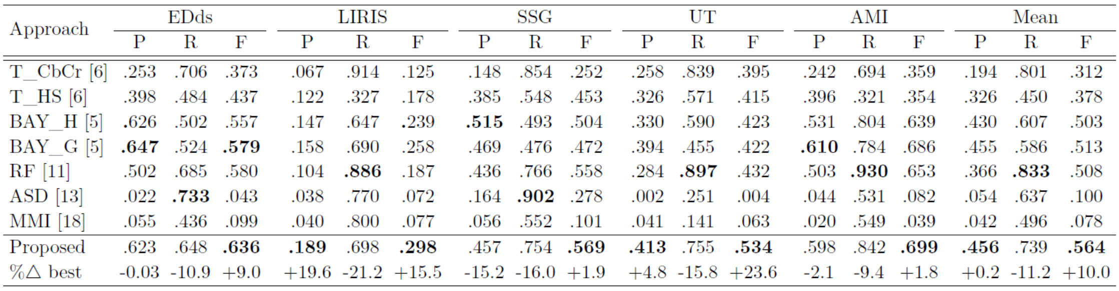


Figure 12. Results for base optimum channel selection over the selected datasets using mean detection ratio.

Skin detection results of the proposed and compared approaches are presented in the following Table. As it can be observed, the results of the LIRIS dataset exhibited a clear decrease in performance for all the approaches as compared to the results of the other datasets. This is due the office furniture used in the dataset, which contains several skin-like surfaces affecting the precision results. In general, fixed-thresholding approaches (T_HS and T_CbCr) got medium performance showing that, albeit effective, the use of parameters with fixed values does not generalize well for a variety of heterogeneous scenarios. BAY versions (BAY_G and BAY_H) obtained good performance demonstrating that non-skin data can be efficiently used to improve final skin detection. Adaptive approaches (ASD and MMI) presented very low performance indicating that introducing adaptive capabilities into skin detection is not an easy task.



Approach	EDds			LIRIS			SSG			UT			AMI			Mean		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
T_CbCr [6]	.253	.706	.373	.067	.914	.125	.148	.854	.252	.258	.839	.395	.242	.694	.359	.194	.801	.312
T_HS [6]	.398	.484	.437	.122	.327	.178	.385	.548	.453	.326	.571	.415	.396	.321	.354	.326	.450	.378
BAY_H [5]	.626	.502	.557	.147	.647	**.239**	**.515**	.493	.504	.330	.590	.423	.531	.804	.639	.430	.607	.503
BAY_G [5]	**.647**	.524	**.579**	.158	.690	.258	.469	.476	.472	.394	.455	.422	**.610**	.784	.686	.455	.586	.513
RF [11]	.502	.685	.580	.104	**.886**	.187	.436	.766	.558	.284	**.897**	.432	.503	**.930**	.653	.366	**.833**	.508
ASD [13]	.022	**.733**	.043	.038	.770	.072	.164	**.902**	.278	.002	.251	.004	.044	.531	.082	.054	.637	.100
MMI [18]	.055	.436	.099	.040	.800	.077	.056	.552	.101	.041	.141	.063	.020	.549	.039	.042	.496	.078
Proposed	.623	.648	**.636**	**.189**	.698	**.298**	.457	.754	**.569**	**.413**	.755	**.534**	.598	.842	**.699**	**.456**	.739	**.564**
%△ best	-0.03	-10.9	+9.0	+19.6	-21.2	+15.5	-15.2	-16.0	+1.9	+4.8	-15.8	+23.6	-2.1	-9.4	+1.8	+0.2	-11.2	+10.0

Table 4. Comparison for selected skin detection approaches. Best results are bold marked. Last row indicates the Percentage increase (%) of each measure with respect to the best performance.

[5] M. Jones, J. Rehg, Statistical color models with application to skin detection, Int. Journal of Computer Vision 46 (1) (2002) 81–96.

[6] Y. Wang, B. Yuan, A novel approach for human face detection from color images under complex background, Pattern Recognition 34 (10) (2001) 1983–1992.

[11] R. Khan, A. Hanbury, J. Stoettinger, Skin detection: A random forest approach, in: IEEE Int. Conf. on Image Processing (ICIP), 2010, pp. 4613–4616.

[12] B. Jedynak, H. Zheng, M. Daoudi, Skin detection using pairwise models, Image and Vision Computing 23 (13) (2005) 1122–1130.

[13] F. Dadgostar, A. Sarrafzadeh, An adaptive real-time skin detector based on hue thresholding: A comparison on two motion tracking methods, Pattern Recognition Letters 27 (12) (2006) 1342–1352.

[18] C. Conaire, N. O'Connor, A. Smeaton, Detector adaptation by maximising agreement between independent data sources, in: IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–6.

4. Conclusions and future work

4.1. Quality measures

Two approaches have been presented for online validation of video tracking. The first is based on accumulation of spatial uncertainty of filter hypotheses (as convolutions of mixtures of Gamma distributions) whereas the second is based on short-term evolution of covariance features.

The results for both approaches show that focusing on temporal consistency of features is more effective than the traditional approaches. Moreover, the structure of target appearance (covariance) performs better than common features to determine tracker errors. Finally, the two proposed approaches outperform competitive feature-based approaches and provide a generic cost-bounded validation that can be applied for long-term and time-critical applications.

As future work, we will explore the application of both approaches to deterministic filters through appropriate adaptations, model validation based on multiple detectors and the selection of the optimum window length for time-based analysis.

4.2. Feedback-driven analysis

For the task of skin-detection, we have presented the benefits of the feedback approach for video analysis compared with the traditional feed-forward one. Its main advantages are the possibility of achieving a desired performance level and adapting to unknown conditions. However, its complex design has limited its use by the video processing community. Conversely, the feed-forward approach is easy to develop and the control over the process performance is not feasible.

Experimental results demonstrate the adaptive capabilities of the proposed approach to improve performance of parameter-fixed, adaptive and learning-based state-of-the-art approaches.

As future work, we will explore the adaptive estimation of skin proportions for the agreement function and the application of the proposed approach to video sequences exploiting temporal relations between the frames.

Bibliography

[1] Juan C. SanMiguel and Andrea Cavallaro, “Temporal validation of particle filters for video tracking”, Computer Vision and Image Understanding, Elsevier Science Inc, (aceptado), ISSN 1077-3142, (Digital Object Identifier: 10.1016/j.cviu.2014.06.016)

[2] E. Erdem, S. Dubuisson, I. Bloch, Fragments based tracking with adaptive cue integration, Computer Vision and Image Understanding 116(7):827-841, 2012.

[3] D.Wang, H. Lu, M.-H. Yang, Online object tracking with sparse prototypes, IEEE Trans. Image Process. 22 (1):314-325, 2013.

[4] Juan C. SanMiguel and Álvaro Calvo, Covariance-based online validation of video tracking, IET Electronic Letters, (aceptado), IET, ISSN 0013-5194 (Digital Object Identifier: 10.1049/el.2014.3405)

[5] Alvaro Calvo, “Estimación de fiabilidad del seguimiento de objetos en video”, Master Thesis, 22 July 2014.

[6] Juan C. SanMiguel, Andrea Cavallaro and José M. Martínez, “Adaptive online performance evaluation of video trackers”, IEEE Transactions on Image Processing, 21(5): 2812-2823, Mayo 2012, IEEE Computer Society, ISSN 1057-7149 (Digital Object Identifier: 10.1109/TIP.2011.2182520)

[7] Juan C. SanMiguel and Sergio Suja, “Skin detection by dual maximization of detectors agreement for video monitoring”, Pattern Recognition Letters, Elsevier Science Inc., 34(16):2102-2109, 2013 (Digital Object Identifier: 10.1016/j.patrec.2013.07.016).