Coloring Channel Representations for Visual Tracking

Martin Danelljan, Gustav Häger, Fahad Shahbaz Khan and Michael Felsberg

Original Publication:

Martin Danelljan, Gustav Häger, Fahad Shahbaz Khan and Michael Felsberg, Coloring Channel Representations for Visual Tracking, 2015, 19th Scandinavian Conference, SCIA 2015, Copenhagen, Denmark, June 15-17, 2015. Proceedings, 117-129.
http://dx.doi.org/10.1007/978-3-319-19665-7_10

Coloring Channel Representations for Visual Tracking

Martin Danelljan, Gustav Häger, Fahad Shahbaz Khan, Michael Felsberg

Computer Vision Laboratory, Linköping University

Abstract. Visual object tracking is a classical, but still open research problem in computer vision, with many real world applications. The problem is challenging due to several factors, such as illumination variation, occlusions, camera motion and appearance changes. Such problems can be alleviated by constructing robust, discriminative and computationally efficient visual features. Recently, biologically-inspired channel representations [9] have shown to provide promising results in many applications ranging from autonomous driving to visual tracking.
This paper investigates the problem of coloring channel representations for visual tracking. We evaluate two strategies, channel concatenation and channel product, to construct channel coded color representations. The proposed channel coded color representations are generic and can be used beyond tracking.
Experiments are performed on 41 challenging benchmark videos. Our experiments clearly suggest that a careful selection of color feature together with an optimal fusion strategy, significantly outperforms the standard luminance based channel representation. Finally, we show promising results compared to state-of-the-art tracking methods in the literature.

Keywords: Visual tracking, channel coding, color names

1 Introduction

Visual tracking is the problem of estimating the trajectory of a target in an image sequence. It has a vast number of real world applications, including robotics [5] and surveillance [30]. In generic visual tracking nothing is known about the target except its initial location. It is one of the most challenging computer vision problems due to factors such as illumination variation, occlusions and fast motion. Many of the challenges encountered in visual tracking can be alleviated by the usage of robust and discriminative features. This paper therefore aims at investigating feature representations and fusion strategies for visual tracking.

In recent years, channel coded feature representations [9] have been successfully used in many computer vision applications, including visual tracking [8], autonomous driving [11], image diffusion [16] and real-time object recognition [10]. The biologically inspired channel representation [13], is a technique for representing data. A set of feature values is represented by a number of channel coefficients, which essentially correspond to a soft histogram. In the visual tracking application, the EDFT method [8] employs a channel representation based

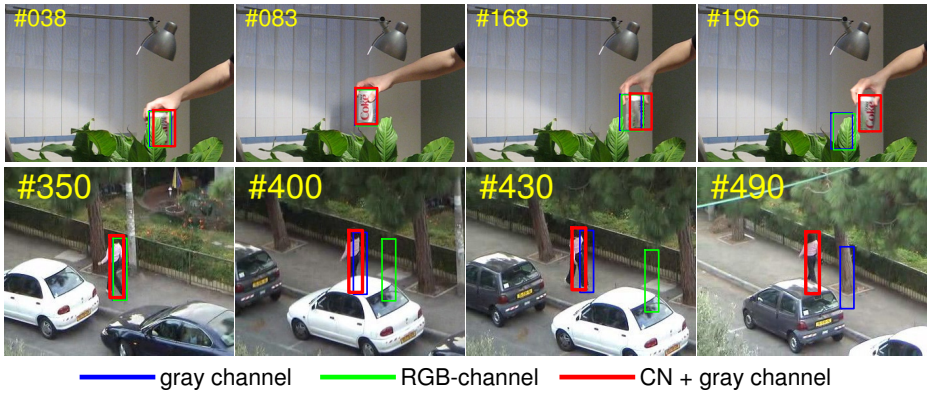

Fig. 1. Comparison of the standard luminance based channel representation (blue) with color augmentation on two sequences (*coke* and *woman*). The straightforward channel representation of RGB (green) fails to improve the performance. The proposed color names extension of the luminance based channel representation (red) significantly improves the performance.

appearance model. However, it only models the luminance distribution over the target template, while ignoring all color information. In this paper, we therefore extend channel representations to incorporate color information.

Channel coded color representations can be constructed using two standard strategies. In both cases, channel coding is first performed on each color space dimension (e.g. R, G and B) independently. In the *channel concatenation* strategy, the channel coefficients for each dimension are concatenated into a final representation. As an alternative, the final representation can be constructed by taking the outer product of the individual channel coefficients, called *channel products*. Typically, a large number of color channels are required to obtain a more discriminative feature representation. However, such high dimensional representations lead to an increased computational cost, and thereby restricting its applicability to real-time tracking.

When incorporating color information into visual tracking, two main research problems have to be addressed. The first issue is the selection of color representation to be used. Ideally, a color feature should possess a certain degree of photometric invariance while maintaining a high discriminative power. The second problem is how to fuse color and intensity information into a single representation. Recently, Danelljan et al. [6] evaluated several color features for visual tracking. In their evaluation, the color names representation [33] was shown to provide superior performance compared to other color features. However, the work of [6] only investigates what color feature to use, while employing raw pixel gray scale values to represent luminance information. Inspired by the success of channel coded luminance information, we investigate how to augment these representations with color information. Additionally, we extend the evaluation performed by [6] with channel coded color features. We show that our proposed feature representation outperforms the best color-intensity combination of [6].

Contributions: In this paper, we investigate how to incorporate color information into channel representations for visual tracking. Both channel concatenation and channel product coding strategies are evaluated on six different color spaces. Additionally, we investigate combining color names and channel coded luminance representations. The evaluated channel coded color representations are generic and can be used beyond visual tracking.

Experiments are performed on 41 challenging videos including all the color sequences from the online benchmark dataset [34]. Our experiments show that fusion of color names and channel coded luminance information outperforms the combination of color names and raw gray scale values [6]. By selecting the best feature (color names and channel coded luminance) and the optimal fusion strategy (concatenation), we achieve a significant gain of 5.4% in median distance precision compared to the standard channel concatenation using RGB. Finally, our approach is also shown to outperform state-of-the-art trackers in both quantitative and qualitative evaluations. Figure 1 shows the comparison of channel coded color representations with the standard channel coded luminance.

2 Related Work

Generic visual trackers can be categorized into generative [31, 2, 24, 25] and discriminative [14, 35, 7, 17, 6] methods. The generative trackers search for image regions most similar to a generative appearance model. On the other hand, the discriminative approaches use machine learning techniques to differentiate the target from the background. Recently, the discriminative correlation filter [3] based trackers have received much research attention thanks to their accuracy, simplicity and speed. These approches utilize the circulant structure induced by correlation to efficiently train a regularized least squares regressor (ridge regression). Most of the computations required for learning and detection are performed using the Fast Fourier transform (FFT), which is the key for its low computational cost. Henriques et al. [17] further introduced kernels into this framework to allow non-linear classification boundaries. The work of Danelljan et al. [6] proposed a consistent learning approach for increased robustness.

Most of the research effort into generic visual tracking has focused on the learning aspect of appearance modeling, while relatively little work has been done on the problem of constructing robust and discriminative features. Most state-of-the-art methods rely on solely image intensity information [17, 31, 8, 14, 35, 20, 7], while others employ simple color space transformations [29, 27, 28]. On the contrary, feature representations have been thoroughly investigated in the related fields of object recognition and action recognition [22, 21]. Recently, Danelljan et al. [6] introduced the Adaptive Color Tracker (ACT), which learns an adaptive color representation based on Color Names [33]. However, this approach still employs a standard grayscale channel for capturing image intensity information.

Channel representations have been used in a large variety of applications [8, 11, 16, 10]. The Distribution Field Tracker (DFT) [31] utilizes a feature representation similar to channel coding to capture the image intensity statistics

of the target. The Enhanced DFT (EDFT) [8] employs channel coding instead of distribution fields and a more robust metric for computation of the objective function. The work of [12, 19] investigate how to fuse color and channel coded luminance information. However, a comprehensive evaluation of color and channel coded luminance fusion is yet to be investigated for the task of tracking.

3 Tracking Framework

In this work, we use the discriminative correlation filter (DCF) based tracking framework proposed by Danelljan et al. [6], called the Adaptive Color Tracker (ACT). It has been shown to provide superior results on benchmark tracking datasets. The method works by learning a kernelized least squared classifier from several samples of the target appearance. The classifier is then applied to locate the target in the new frame.

To update the classifier at frame n, a template f_n centered around the target is first extracted. The template is of a fixed size of $M \times N$ pixels and contains a D-dimensional feature vector at each pixel location within the template. The features are preprocessed by a normalization step and a windowing operation. The classifier coefficients are updated in each frame through the recursive formula

$$\hat{u}_n = (1-\gamma)\hat{u}_{n-1} + \gamma \hat{y}_n \hat{a}_n \tag{1a}$$

$$\hat{v}_n = (1-\gamma)\hat{v}_{n-1} + \gamma \hat{a}_n (\hat{a}_n + \lambda). \tag{1b}$$

Here, a_n is the kernelized autocorrelation of the template f_n, and y_n is a Gaussian label function. The discrete Fourier transform (DFT) is denoted by a hat. The constants γ and λ are learning and regularization weights respectively. A target template t_n is also updated as: $t_n = (1-\gamma)t_{n-1} + \gamma f_n$.

The classifier is applied to an image template g_n by first computing its kernelized cross-correlation b_n with the learned target template t_{n-1} from the previous frame. The classification scores s_n are obtained by evaluating

$$s_n = \mathscr{F}^{-1}\left\{ \frac{\hat{u}_{n-1}\hat{b}_n}{\hat{v}_{n-1}} \right\}. \tag{2}$$

Here, $\mathscr{F}^{-1}$ denotes the inverse DFT. Henriques et al. [17] showed that the kernelized correlations a_n and b_n can be computed efficiently for radial basis function kernels, using the FFT. For more details, we refer to [6].

4 Channel Representation

Channel representation [9] is a biologically inspired approach for representing data [13]. It is closely related to soft histograms and is used in many computer vision applications, including visual tracking [8]. A scalar value x is represented in terms of its channel coefficients $\{c_k\}_1^n$. The k:th coefficient c_k is computed by evaluating a kernel function K located at the k:th channel center $\tilde{x}_k$ using:

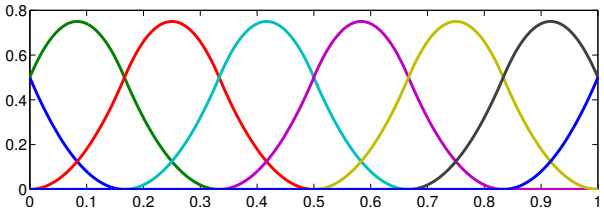

Fig. 2. A graphichal visualization of the binning functions employed in our channel representations. Here the configuration is shown for $n = 8$ channels.

$c_k = K(x - \tilde{x}_k)$. Common choices for the kernel function K include Gaussian, $\cos^2$ and B-spline functions. The coefficients $\{c_k\}_1^n$ can be interpreted as a soft histogram of the data, where the bins are centered at $\{\tilde{x}_k\}_1^n$ and the binning function K weights the contribution of x to each bin.

In this paper, we construct the channel representation using regularly spaced second order B-spline kernel functions. We set $K_w(x) = B(x/w)$, where w is the spacing between the channel centers and the second order B-spline is given by:

$$B(x) = \begin{cases} \frac{3}{4} - x^2 & , \ |x| \leq \frac{1}{2} \\ \frac{1}{2}\left(|x| - \frac{3}{2}\right)^2 & , \ \frac{1}{2} \leq |x| \leq \frac{3}{2} \\ 0 & , \ |x| \geq \frac{3}{2} \end{cases}. \tag{3}$$

All color and luminance features used in this work only take values within a bounded interval, e.g. the red component of an RGB image pixel. It can therefore be assumed that $x \in [0, 1]$ by simply translating and re-scaling the feature appropriately. The range $[0, 1]$ is covered by n channels, which are centered at

$$\tilde{x}_k = wk - \frac{3w}{2} \quad , \ k = 1, \ldots, n. \tag{4}$$

The spacing is set to $w = \frac{1}{n-2}$. With this configuration the channel coefficients always sum up to one, and thus have a direct probabilistic interpretation. The used channel configuration is visualized in figure 2 for $n = 8$ channels.

Channel representations can be extended to multi-dimensional features $\mathbf{x} = (x_1, \ldots, x_m)$ (e.g. the RGB value of a pixel) using either *channel concatenation* or *channel products*. In the former case, the final representation is obtained as the collection of channel coefficients for each scalar component x_j. The number of coefficients in the channel concatenation is $n = n_1 + \ldots + n_m$, where n_j denotes the number of channels used for representing x_j.

The channel product representation considers m-dimensional channels $c_{\mathbf{k}} = \hat{K}(\mathbf{x} - \tilde{\mathbf{x}}_{\mathbf{k}})$. For a separable kernel $\hat{K}(\mathbf{x}) = K_1(x_1) \cdots K_m(x_m)$, the final representation is obtained as the outer product of the individual channel coefficients

$$c_{\mathbf{k}} = c_{k_1, \ldots, k_m} = \prod_{j=1}^{m} c_{k_j}^{(j)}. \tag{5}$$

Here, $\{c_1^{(j)}, \ldots, c_{n_j}^{(j)}\}$ is the channel representation of x_j. The number of coefficients in the channel product representation is hence $n = n_1 \cdot \ldots \cdot n_m$.

5 Channel Coded Color Representations

In this paper, we investigate different channel coded color representations for visual tracking. We evaluate the two strategies mentioned in section 4 to construct channel coded color representations. Six color spaces are used for our evaluation: **RGB**, **Opp**, **C**, **HSV**, **YCbCr** and **LAB**. The opponent (Opp) color space is an orthonormal transformation of the RGB cube aligning the third dimension with the diagonal $O_3 = 3^{-\frac{1}{2}} \cdot (R+G+B)$. The image intensity is thus captured by O_3 while O_1 and O_2 are its color opponent dimensions. The C space further adds photometric invariance to Opp space by dividing O_1 and O_2 with the intensity dimension O_3. The HSV representation instead maps the RGB cube to a cylinder, providing the hue angle H, saturation S and value V components. The YCbCr space contains a luminance component Y and the two chroma components Cb and Cr. LAB is a perceptually uniform color space, which contains the lightness dimension L and the two color opponent dimensions A and B.

We evaluate the channel concatenation and product representations for each of the six aforementioned color spaces. In both cases, we use the channel configuration described in section 4 to code the individual color space dimensions.

The channel coded color representations are compared with Color Names (CN) [33], which achieved the best results among the evaluated color features in [6]. The CN representation is inspired by linguistics. An RGB value is mapped to probabilities for the 11 basic color names in the English language: black, blue, brown, grey, green, orange, pink, purple, red, white and yellow. This mapping was automatically learned from images retrieved by Google image search.

Color names have successfully been applied in object recognition [22], action recognition [21] and image stitching [26] to capture color information. On the other hand, channel coded intensity features have been used in visual tracking [31, 8] to capture the image intensity statistics. The EDFT tracker [8] employs channel coded grayscale values with the same channel configuration as described in section 4. Inspired by their success, we propose to combine the two features into a single representation. Given an image template, color names and channel coded luminance are computed at each pixel. The two representations are then concatenated into a single feature vector.

6 Experiments

We perform a comprehensive evaluation of the color representations for visual tracking described in section 5. The best performing representation is then compared to several state-of-the-art tracking methods.

6.1 Evaluation Methodology and Dataset

The results are evaluated using the standard benchmark protocol suggested by Wu et al. [34]. We present the results using three standard evaluation metrics,

	IC	CN [6]	RGB-c	RGB-p	LAB-c	LAB-p	YCbCr-c	YCbCr-p	HSV-c	HSV-p	Opp-c	Opp-p	C-c	C-p	IC+CN
DP	77.1	81.4	71.7	71.2	71.1	62.1	79.8	57.3	75.6	73.4	74.1	56.7	*81.5*	60.1	**83.1**
OP	53.3	51	52.3	49.2	46.1	42	47.5	40.5	53.3	52.2	43.9	44.4	*58.6*	41.5	**59**
CLE	19.1	*13.8*	17.8	17.2	19.1	24.3	17.3	26.1	15.5	20.4	16.6	31.2	14.9	25.2	**13.7**

Table 1. The median Distance Precision (DP) (%), Overlap Precision (OP) (%) and Center Location Error (CLE) (in pixels) results using different features on 41 videos. The best two results shown in red and blue fonts. In all cases, the channel concatenation using color names significantly improves the performance compared to luminance based channels and color names alone.

namely center location error (CLE), distance precision (DP) and overlap precision (OP). CLE is computed as the euclidean distance between the ground truth bounding box and the tracked bounding box centers. The average CLE value is then used for each sequence. Distance precision is the percentage of frames where the CLE is below a threshold. We present the DP value at 20 pixels, following [34], [17]. Overlap precision is the percentage of frames where the intersection-over-union overlap between the ground truth and tracked bounding boxes is greater than a threshold. We present numeric values of OP at the threshold 0.5, which corresponds to the PASCAL evaluation metric.

The results are also presented as precision plot and success plots [34]. In the plots, the average DP and OP is plotted over a range of thresholds. The mean DP value over all sequences is included in the legend of the precision plot, while the area under curve (AUC) is shown in the legend of the success plot.

We use the same dataset as employed in the evaluation performed by Danelljan et al. [6]. It consists of all the 35 color sequences from the benchmark dataset [34] and 6 additional color videos. All methods are thus evaluated on 41 videos.

6.2 Experiment 1: Channel Coded Color Representations

Here we present results of augmenting channel representations with color information. All features are evaluated in the DCF-based tracker proposed by [6] with the suggested parameters. However, no adaptive dimensionality reduction is performed. For the six color spaces mentioned in section 5, we use 16 channels per dimension for the channel concatenation and 4 channels per dimension for the channel product representation. The feature dimensionality is thus 48 and 64 for the concatenation and product representations respectively. We denote the concatenation representation by adding a "c" to the color space name, e.g. RGB-c. Similarly, RGB-p denotes the channel product representation of RGB.

The channel coded color spaces are compared with the color names (CN) feature and the standard luminance based channel representation (IC). As in [8], we use 16 channels for the IC representation. For the IC+CN combination approach presented in section 5, we use 16 intensity channels combined with the 11 color names. For a fair comparison, we append the usual grayscale component (obtained by Matlab's `rgb2gray`) to all evaluated feature representations, including the channel representations. We further perform the appropriate normalization steps [6] to reduce the windowing effect within the DCF framework.

	CT	TLD	DFT	EDFT	ASLA	L1APG	CSK	SCM	LOT	CPF	CXT	Frag	Struck	LSST	LSHT	**Ours**
DP	20.8	45.4	41.4	49	42.2	28.9	54.5	34.1	37.1	37.1	39.5	38.7	*71.3*	23.4	55.9	**83.1**
OP	13.3	36.7	34.3	44.8	42.2	26.3	37.7	33.6	31.1	33.1	33.2	36.8	*53.8*	19.5	40.4	**59**
CLE	78.4	54.4	47.9	53.5	56.8	62.9	50.3	54.3	60.9	41.1	43.8	70.8	*19.6*	78.4	32.3	**13.7**
FPS	*68.9*	20.7	9.11	19.7	0.946	1.03	**151**	0.0862	0.467	55.5	11.3	3.34	10.4	3.57	12.5	36.6

Table 2. Quantitative comparison of our approach with 15 state-of-the-art trackers on 41 videos. The results are presented in median Distance Precision (DP) (%), Overlap Precision (OP) (%), Center Location Error (CLE) (in pixels) and frames per second (FPS). The two best results are shown in red and blue. In all cases, our approach significantly outperforms the best reported tracker (Struck) in the literature.

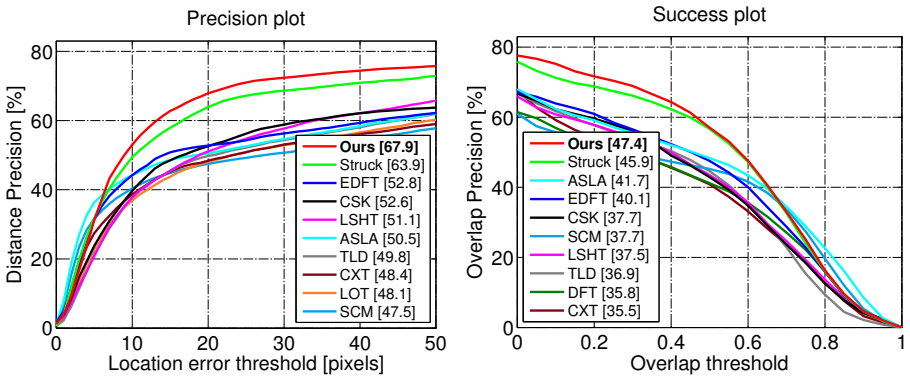

Fig. 3. Precision and success plots for comparison with state-of-the-art trackers. Only the top 10 trackers are displayed for clarity. Our approach outperforms the second best tracker (Struck) in both mean distance and overlap precision.

Table 1 shows a comparison of the evaluated feature representations. The standard luminance based channel representation achieves a median DP of 77.1%. The channel concatenation and product representations using the C color space achieve a median DP of 81.5% and 60.1% respectively. In all cases, the concatenation strategy provides improved results compared to the channel product representation. The CN approach of [6], employing color names and an intensity component, provides a median DP of 81.4%. Our channel concatenation using color names provides an improvement of 1.7% in median DP compared to [6]. Similarly, our channel concatenation using color names also provides the best results in median OP and CLE. Based on this analysis, we select the channel-color names combination (IC+CN) as our proposed representation for tracking.

6.3 Experiment 2: State of the art comparison

We compare our proposed feature representation with 15 state of the art trackers: CT [35], TLD [20], DFT [31], EDFT [8], ASLA [18], L1APG [2], CSK [17], SCM [36], LOT [28], CPF [29], CXT [7], Frag [1], Struck [14], LSHT [15] and LSST [32]. Table 2 shows the comparison of our tracker with the state-of-the-art tracking methods using median DP, OP and CLE. The two best results are presented in red and blue fonts. The CSK tracker [17] achieves a median DP of 54.5%. The EDFT method [8] based on channel coded luminance provides a median DP of

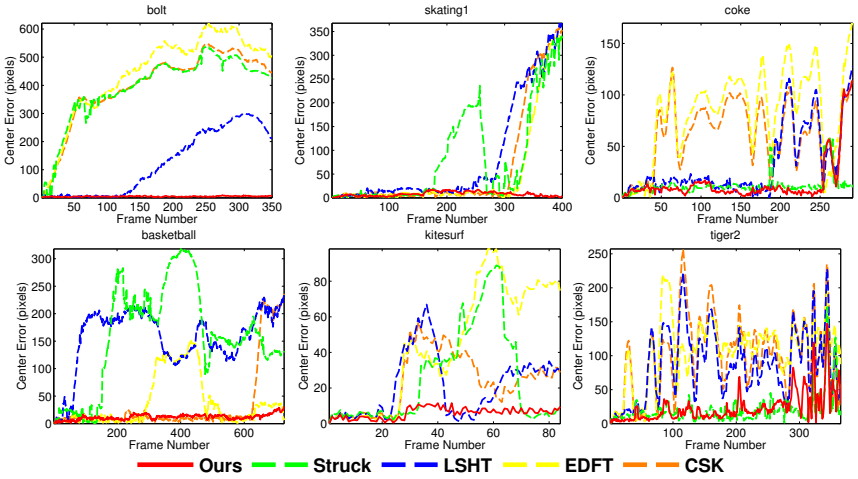

Fig. 4. A frame-by-frame comparison with four state-of-the-art trackers on six example videos. The results are shown in terms of center location error in pixels for each frame. Our approach provides favorable performance compared to the existing trackers.

49.0%. Among the existing tracking approaches, Struck [14] provides a median DP of 71.3%. Our approach significantly outperforms Struck by 11.8% in median DP. Similarly, our tracker achieves the best performance by providing a gain of 5.2% and 5.9 pixels in median OP and CLE respectively compared to Struck.

Figure 3 shows the results using precision and success plots, containing mean distance and overlap precision. The mean results are calculated over all the 41 videos. The values in the legends of precision and success plots are the mean DP at 20 pixels and the AUC respectively. Among the existing trackers, Struck provides the best results with mean DP of 63.9% in the precision plot. Our approach outperforms Struck by 5.6% in mean DP. Similarly, our approach also provides superior performance compared to existing methods in success plot.

Figure 4 shows a frame-by-frame comparison of our tracker with existing tracking methods in terms of center-pixel error. Our tracking method provides favorable performance compared to existing trackers on the six example videos.

Robustness to Initialization: We follow the protocol suggested by Wu et al. [34] to validate the robustness of our approach with respect to initialization. The performance is evaluated using two different strategies: temporal robustness (TRE) and spatial robustness (SRE). In the case of TRE, the trackers are initialized at different frames. In the case of SRE, the trackers are instead initialized at different locations in the first frame of the sequence. As in [34], twelve different initializations are performed for SRE whereas each video is segmented into 20 partitions for TRE. Figure 5 shows the results for both TRE and SRE. For clarity, we only compare with the top six trackers in our evaluation. In both cases, our approach provides promising results compared to existing methods.

Attribute-based comparisons: Here, we investigate the factors that can affect the performance of a visual tracker. The videos in the benchmark dataset [34] are annotated with 11 attributes: illumination variation, occlusion, deformation,

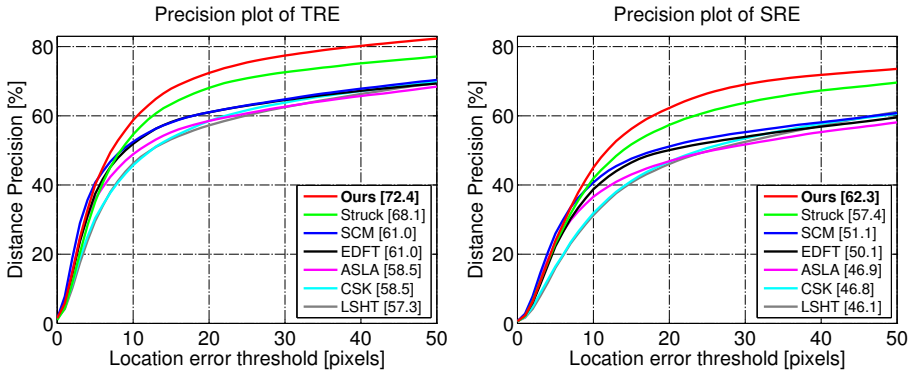

Fig. 5. Precision plots to compare the robustness of our approach with respect to initialization. The performance is validated using temporal and spatial robustness (TRE and SRE). Our method achieves superior performance compared to existing trackers.

scale variation, motion blur, fast motion, in-plane rotation, out-of-plane rotation, out-of-view, low resolution and background clutter. We show a comparison of our tracker with existing methods on 35 videos annotated with these 11 attributes.

Figure 6 shows the precision plots of six different attributes: in-plane rotation, motion blur, fast motion, illumination variation, occlusion and out-of-plane rotation. For clarity, we only show the results for the top 10 trackers in each attribute plot. Our approach performs favorably compared to existing methods. A significant improvement in performance is achieved in case of in-plane rotation, motion blur, illumination variation and out-of-plane rotation. This is due to the robustness and complementary properties of our feature representation. In the presence of fast motion, Struck provides the best results. This is attributed to the local search strategy employed in the baseline DCF-based tracking algorithm.

7 Conclusions

In recent years, luminance based channel representations have shown to provide promising results in many vision applications. In this work, we investigate the problem of augmenting channel representations with color information. Our results clearly suggest that channel conatenation using color names significantly improves the performance of conventional channel coded luminance features. Our quantitative and attribute-based qualitative evaluations demonstrate promising results compared to existing methods.

Recently, efficient tracking methods with scale estimation capability have shown promising results in the VOT 2014 challenge [23]. Currently, our approach has no explicit scale estimation component. Future work involves investigating how to incorporate our feature representation into efficient scale adaptive trackers, e.g. the Discriminative Scale Space Tracker [4].

Acknowledgments: This work has been supported by SSF through a grant for the project CUAS, by VR through a grant for the projects ETT and EMC[2],

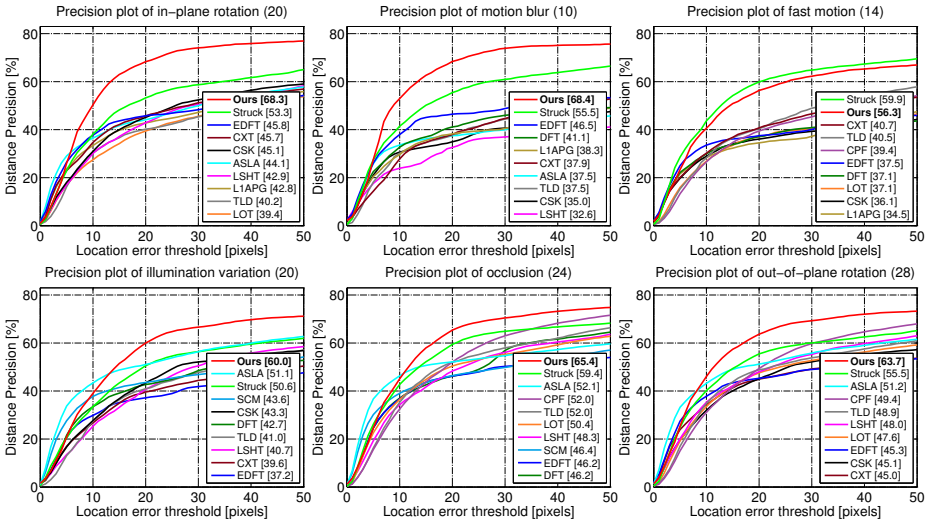

Fig. 6. Attribute-based comparison with the state-of-the-art trackers. The results are shown for in-plane rotation, motion blur, fast motion, illumination variation, occlusion and out-of-plane rotation. The number of videos for an attribute is mentioned in each title. Our approach provides favorable results compared to the existing methods.

by EU's Horizon 2020 Program through a grant for the project CENTAURO, through the Strategic Area for ICT research ELLIIT, and CADICS.

References

1. Adam, A., Rivlin, E., Shimshoni: Robust fragments-based tracking using the integral histogram. In: CVPR (2006)
2. Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust l1 tracker using accelerated proximal gradient approach. In: CVPR (2012)
3. Bolme, D.S., Beveridge, J.R., Draper, B.A., Lui, Y.M.: Visual object tracking using adaptive correlation filters. In: CVPR (2010)
4. Danelljan, M., Häger, G., Khan, F.S., Felsberg, M.: Accurate scale estimation for robust visual tracking. In: BMVC (2014)
5. Danelljan, M., Khan, F.S., Felsberg, M., Granström, K., Heintz, F., Rudol, P., Wzorek, M., Kvarnström, J., Doherty, P.: A low-level active vision framework for collaborative unmanned aircraft systems. In: ECCVW (2014)
6. Danelljan, M., Khan, F.S., Felsberg, M., van de Weijer, J.: Adaptive color attributes for real-time visual tracking. In: CVPR (2014)
7. Dinh, T.B., Vo, N., Medioni, G.: Context tracker: Exploring supporters and distracters in unconstrained environments. In: CVPR (2011)
8. Felsberg, M.: Enhanced distribution field tracking using channel representations. In: ICCV Workshop (2013)
9. Felsberg, M., Forssen, P.E., Scharr, H.: Channel smoothing: Efficient robust smoothing of low-level signal features. PAMI 28(2), 209–222 (2006)
10. Felsberg, M., Hedborg, J.: Real-time visual recognition of objects and scenes using p-channel matching. In: SCIA (2007)

11. Öfjäll, K., Felsberg, M.: Biologically inspired online learning of visual autonomous driving. In: BMVC (2014)
12. Forssen, P.E., Granlund, G., Wiklund, J.: Channel representation of colour images. Tech. rep., Linköping University (2002)
13. Granlund, G.H.: An associative perception-action structure using a localized space variant information representation. In: AFPAC, pp. 48–68. Springer (2000)
14. Hare, S., Saffari, A., Torr, P.: Struck: Structured output tracking with kernels. In: ICCV (2011)
15. He, S., Yang, Q., Lau, R., Wang, J., Yang, M.H.: Visual tracking via locality sensitive histograms. In: CVPR (2013)
16. Heinemann, C., Åström, F., Baravdish, G., Krajsek, K., Felsberg, M., Scharr, H.: Using channel representations in regularization terms: A case study on image diffusion. In: VISAPP (2014)
17. Henriques, J., Caseiro, R., Martins, P., Batista, J.: Exploiting the circulant structure of tracking-by-detection with kernels. In: ECCV (2012)
18. Jia, X., Lu, H., Yang, M.H.: Visual tracking via adaptive structural local sparse appearance model. In: CVPR (2012)
19. Jonsson, E.: Channel-Coded Feature Maps for Computer Vision and Machine Learning. Linköping Studies in Science and Technology. Dissertations No. 1160, Linköping University, Sweden (2008)
20. Kalal, Z., Matas, J., Mikolajczyk, K.: P-n learning: Bootstrapping binary classifiers by structural constraints. In: CVPR (2010)
21. Khan, F.S., Anwer, R.M., van de Weijer, J., Bagdanov, A., Lopez, A., Felsberg, M.: Coloring action recognition in still images. IJCV 105(3), 205–221 (2013)
22. Khan, F.S., van de Weijer, J., Vanrell, M.: Modulating shape features by color attention for object recognition. IJCV 98(1), 49–64 (2012)
23. Kristan, M., Pflugfelder, R., Leonardis, A., Matas, J., et al.: The visual object tracking VOT 2014 challenge results. In: ECCVW (2014)
24. Kwon, J., Lee, K.M.: Tracking by sampling trackers. In: ICCV (2011)
25. Liu, B., Huang, J., Yang, L., Kulikowski, C.: Robust tracking using local sparse appearance model and k-selection. In: CVPR (2011)
26. Meneghetti, G., Danelljan, M., Felsberg, M., Nordberg, K.: Image alignment for panorama stitching in sparsely structured environments. In: SCIA (2015)
27. Nummiaro, K., Koller-Meier, E., Gool, L.J.V.: An adaptive color-based particle filter. IVC 21(1), 99–110 (2003)
28. Oron, S., Hillel, A., Levi, Avidan, S.: Locally orderless tracking. In: CVPR (2012)
29. Perez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-based probabilistic tracking. In: ECCV (2002)
30. Prokaj, J., Medioni, G.: Persistent tracking for wide area aerial surveillance. In: CVPR (2014)
31. Sevilla-Lara, L., Miller, E.: Distribution fields for tracking. In: CVPR (2012)
32. Wang, D., Lu, H., Yang, M.H.: Least soft-threshold squares tracking. In: CVPR (2013)
33. van de Weijer, J., Schmid, C., Verbeek, J.J., Larlus, D.: Learning color names for real-world applications. TIP 18(7), 1512–1524 (2009)
34. Wu, Y., Lim, J., Yang, M.H.: Online object tracking: A benchmark. In: CVPR (2013)
35. Zhang, K., Zhang, L., Yang, M.: Real-time compressive tracking. In: ECCV (2012)
36. Zhong, W., Lu, H., Yang, M.H.: Robust object tracking via sparsity-based collaborative model. In: CVPR (2012)