Encoding Color Information for Visual Tracking: Algorithms and Benchmark

Pengpeng Liang, Erik Blasch, *Senior Member, IEEE*, Haibin Ling*, *Member, IEEE*

***Abstract*—While color information is known to provide rich discriminative clues for visual inference, most modern visual trackers limit themselves to the grayscale realm. Despite recent efforts to integrate color in tracking, there is a lack of comprehensive understanding of the role color information can play. In this paper, we attack this problem by conducting a systematic study from both the algorithm and benchmark perspectives. On the algorithm side, we comprehensively encode 10 chromatic models into 16 carefully selected state-of-the-art visual trackers. On the benchmark side, we compile a large set of 128 color sequences with ground truth and challenge factor annotations (e.g., occlusion). A thorough evaluation is conducted by running all the color-encoded trackers, together with two recently proposed color trackers. A further validation is conducted on a RGBD tracking benchmark. The results clearly show the benefit of encoding color information for tracking. We also perform detailed analysis on several issues including the behavior of various combinations between color model and visual tracker, the degree of difficulty of each sequence for tracking, and how different challenge factors affect the tracking performance. We expect the study to provide the guidance, motivation and benchmark for future work on encoding color in visual tracking.**

***Index Terms*—Visual tracking, color, benchmark, evaluation.**

I. Introduction

Being an important topic in computer vision, visual tracking has a wide range of applications including human computer interaction, video surveillance, vehicle navigation, robotics, etc. In practice, tracking algorithms are often challenged by various factors such as illumination changes, occlusion, pose change, abrupt motion, and background clutter. Consequently, a great amount of effort has been devoted to extract robust visual cues, such as shape and appearance, to distinguish a tracking target apart from its surrounding. Most modern trackers, however, rely purely on the grayscale version of an input sequence, leaving out the rich chromatic information. There are several possible reasons for this: (1) color information can be corrupted by environmental factors such as change in the illumination; (2) encoding color may increase the computational burden; and (3) grayscale images are sometimes sufficient to produce reasonably good results. In this paper, we show that color information is indeed very helpful to improve visual tracking and the improvement it brings is general for different tracking algorithms. An intuitive example observed in our experiment is shown in Figure 1. The figure shows that, while the target blends into the background in grayscale images (in frame 148), it clearly distinguishes itself when color information is utilized.

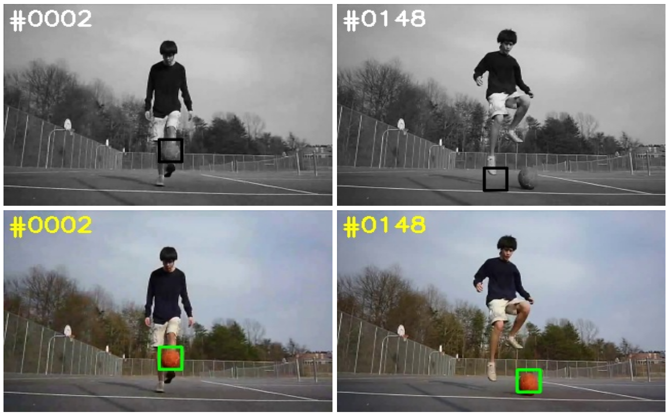


Fig. 1. Color information for visual tracking, results shown in bounding boxes are from Struck [1] (first row) and color-encoded Struck (second row). When color information is discarded, the target can be hardly distinguished from its surrounding (right column), and consequently confuses the tracker. On the contrary, the color information helps the tracker to avoid such drifting. This figure should be viewed in color.

To capture the chromatic information, several visual tracking algorithms have encoded color information (See Sec. II), including recent ones such as [2] that achieves state-of-the-art performance. Despite these efforts, there is a lack of systematic study on the effects of using color for visual tracking, and several questions remain unanswered: Is color useful for visual tracking in general, or just for some specific trackers? How will existing state-of-art trackers behave if color information is encoded? What chromatic representations are the most suitable for visual tracking? What are the main challenges even when color is used? In this paper we seek answers to these questions.

As the first comprehensive investigation of encoding color in visual tracking, our study tackles the problem from two aspects: algorithm and benchmark. On the algorithm side, inspired by recent work on color descriptor evaluation [3], [4], [5], we create a set of 160 trackers by combining various color representations and existing visual trackers. In particular, 10 different color models are chosen to cover different chromatic properties; and 16 state-of-the-art grayscale visual trackers are carefully chosen that have achieved top performances in recent tracking evaluations [6], [7], [8]. On the benchmark

*Corresponding author.

P. Liang is with the Meitu HiScene Lab, HiScene Information Technologies, and with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: pliang@temple.edu).

E. Blasch is with Air Force Research Lab, 525 Brooks Rd, Rome, NY 13441 (e-mail: erik.blasch@rl.af.mil).

H. Ling is with School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006 China, and with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: hbling@temple.edu).

side, to address the issue of lacking appropriate datasets, we compile a set of 128 color sequences (Fig. 2), named *TColor-128* (Temple Color 128), along with ground truth annotation. Out of the 128 sequences, 78 have never before been used for visual tracking, and as shown in our analysis, they are often more challenging than previously tested ones. The challenge factors (e.g., *occlusion* and *out of plane rotation*) for each sequence is also provided, enabling a more detailed performance analysis. The data set is available for research exploration at http://www.dabi.temple.edu/∼hbling/data/TColor-128.html.

The 160 color-encoded visual trackers, along with recently proposed color trackers, are evaluated together on TColor-128. The results show that, by encoding color information properly, we can consistently improve the baseline trackers where color has not been used. Detailed analysis is also conducted, showing that (1) Some color models (Opponent, HSV and LAB) are in general more effective for improving tracking performance; (2) Color information helps the most when targets are under deformation or rotation (both in plane and out of plane). For further validation, an additional evaluation is conducted on the *Princeton Tracking Benchmark* (PTB) [8], where similar conclusions can be drawn.

To summarize, our contribution is four-fold:

- *Color-encoded visual trackers*. We systematically combine various color models with state-of-the-art grayscale trackers and investigate their performances.
- *Color tracking benchmark*. We create a large color sequence benchmark with annotated groundtruth.
- *Color tracking evaluation*. We thoroughly evaluate different combinations of color models and visual trackers, along with recently proposed color trackers.
- *Color tracking analysis*. We perform comprehensive experimental analysis of the effect of encoding color for tracking over a wide variety of scenarios.

Having observed the increasing popularity of color video in real world applications, we expect our study to provide guidance, motivation and benchmark towards future exploration of color tracking; and, we also expect the data and codes to serve as a basis for future studies.

In the rest of the paper, related work is summarized in Section II. Then, the encoding of color information into state-of-the-art visual trackers is described in Section III. After that, the collected TColor-128 benchmark is introduced in Section IV, followed by the evaluation and analysis in Section V. Finally, Section VI concludes the paper.

II. Related Work

A. Trackers Using Color Information

As a fundamental problem in vision, visual tracking has been drawing research attention for decades. A comprehensive review of the topic can be found in [33]. Since our focus is on integrating color information in tracking, we review only previous color trackers due to space limitation. Table I lists the abbreviations of trackers discussed in this paper.

TABLE I
Abbreviation of Trackers

Tracker	Brief description
ASLA [9]	Adaptive structural local sparse appearance model
BSBT [10]	Beyond semi-supervised tracking
CPF [11]	Color particle filter
CSK [12]	Tracking by detection with circulant structure
CT [13]	Compressive tracking
CXT [14]	Context tracker
DFT [15]	Distribution fields for tracking
FCT [16]	Fast compressive tracking
Frag [17]	Fragments based tracking using the integral histogram
IVT [18]	Incremental learning for tracking
KCF [19]	Tracking with kernelized correlation filters
L1APG [20]	L1 tracker using accelerated proximal gradient
L1T [21]	Tracking via sparse representation
LOT [22]	Locally orderless tracking
LSK [23]	Local sparse appearance model and k-selection
MEEM [24]	Multi-expert restoration using entropy minimization
MTT [25]	Tracking via multi-task sparse learning
OAB [26]	Tracking via on-line boosting
OFS [27]	Online selection of discriminative tracking features
SCM [28]	Tracking via sparsity-based collaborative model
SemiT [29]	Semi-supervised on-line boosting for tracking
Struck [1]	Structured output tracking
TLD [30]	Tracking-learning-detection
VTD [31]	Visual tracking decomposition
VTS [32]	Tracking by sampling trackers

A notable early work on color tracking is the color particle filter introduced in [11], which calculates the likelihood of each particle by comparing its color histogram from the HSV color space with the reference color model. In [34], the target model and target candidates are represented by smoothed color histograms quantized from the RGB color space, and mean shift is used to minimize the distance between the discrete distributions of the target model and target candidates. In [35], RGB color distribution was used to describe the target model and candidates, and the target object was located by minimizing the Kullback Leibler distance between the color distributions of the target model and candidates with the help of a trust-region method. VTD [31] integrates basic trackers derived from the combination of different basic observation and motion models, and four basic observation models, which use hue, saturation, intensity and edge templates as features respectively, are adopted. LOT [22] measures the similarity between a candidate and the target using locally orderless matching, and HSV color space is used to describe the appearance of each pixel. MEEM [24] uses features extracted in the LAB color space. In the most recent work [2], CSK [12] is extended with color names [36], and to speed up, the dimension of the original color names is reduced with an adaptive dimensionality reduction technique. There are also trackers that take color input (e.g., [37]), but do not explicitly exploit the use of color information.

Despite previous arts, there is a lack of a systematic study and understanding of how color information can be used to improve visual tracking. Our work aims to fill the gap by thoroughly investigating the behavior of numerous state-of-the-art visual trackers with various color representations.

B. Visual Tracking Benchmark and Evaluation

The advance in visual tracking algorithms makes it imperative to have large scale benchmarks for evaluation purposes. Recently, there are several remarkable studies along this line [6], [7], [8], [38], [39]. In [6], the authors collected 50

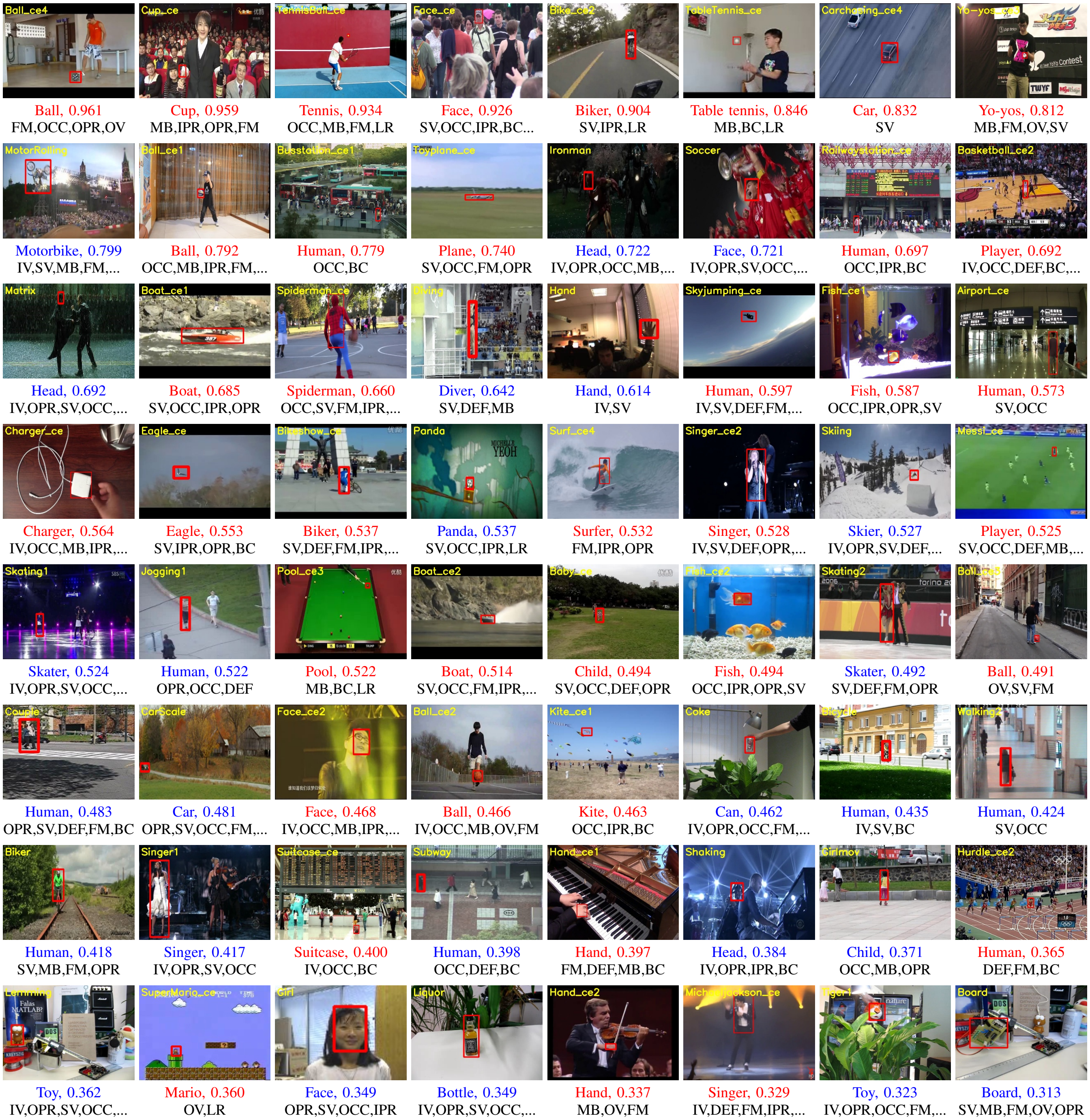


Fig. 2. Selected sequences from TColor-128. The first frame with the ground truth bounding box is shown. The sequences are ordered from hardest to easiest based on the "degree of difficulty" estimated in our evaluation (Sec. V-A4). The tracking target and its degree of difficulty are listed below each frame. We use red font for newly collected sequences and blue for those used in previous studies. The challenge factors are also listed, and for the meanings of the acronyms, please refer to Sec. IV.

sequences which were most commonly used in previous literatures and evaluated the overall performances of 29 tracking algorithms on the benchmark and their performance on the different subsets of the benchmark having different challenging factors. To rank the state-of-the-art trackers with as little subjective bias as possible, four different ranking algorithms were adopted in [7] to analyze the comparison results extracted from selected published papers. In [8], in order to study the role of depth in helping visual tracking, a benchmark with 100 RGBD sequences was constructed, and several RGBD baseline algorithms were proposed and evaluated. In the Visual Object Tracking (VOT) workshop 2013 [38], a dataset containing 16 sequences was provided along with an evaluation protocol, and the competition results of 27 trackers were reported. In [39], 19 trackers were evaluated on a large benchmark and the overall performance and the results for different challenge factors were investigated. Aside from these works, there are some classic efforts for tracking evaluation, such as the VIVID

TABLE II
INVARIANCE PROPERTIES OF DIFFERENT PHOTOMETRIC REPRESENTATIONS [3], [5].

	RGB	HSV	LAB	rg	TRGB	OPP	C-OPP	N-OPP	Hue
Highlight	×	×	×	×	√	×	√	×	√
Shadow	×	×	×	√	√	×	×	√	√

Tracking Evaluation Website [40].

Compared with these studies, our work is the first large scale evaluation for studying color factors in visual tracking. Note that VOT 2013 [38] also paid attention to color and conducted separated experiments using grayscale information and color information. However, VOT 2013 does not explore the effects of enhancing existing grayscale visual trackers with color information. By contrast, we extend such grayscale trackers to color versions by integrating different color models.

C. Color Information in Other Vision Tasks

Not surprisingly, the discriminative power of color information has been systematically investigated for various vision topics, such as object recognition [3], [4], [41], [42], human action recognition [5], [43], object detection [44], etc. While being highly motivated by these pioneering works and borrowing some ideas from them, our work however focuses on visual tracking. To the best of our knowledge, this is the first comprehensive study on encoding color information for visual tracking. In fact, as shown in our experiments, many modern grayscale trackers, when augmented with color information, outperform previously proposed color trackers.

III. ENCODING COLOR INFORMATION IN VISUAL TRACKING

A. Photometric Representations

There are various color models used in computer vision, each with different photometric invariance. Thorough evaluations of these models have been conducted for visual recognition tasks [3], [5]. Motivated by these works, we inherit the photometric representations in these studies for visual tracking. In this subsection, we review these photometric representations and their photometric invariance properties.

In [3] and [5], the diagonal model [45] and the dichromatic reflection model [46] are used respectively to analyze the invariance properties of different photometric representations. In visual tracking, we consider invariance against light intensity changes due to highlights, shadows or shading. In the following, we list the photometric representations explored in our work, and their invariance properties are summarized in Table II.

RGB: The standard RGB color model with three color channels; namely *red*, *blue* and *green*.

HSV: The standard HSV color model, where H (hue) and S (saturation) are invariant to light intensity change, and H is also invariant to highlights. Such invariance, however, does not hold for the V (value) channel.

LAB: The LAB is a perceptually uniform color space. More specifically, the same amount of change in the LAB color value produces the same amount of change in perception. The LAB color space also has no invariance properties.

rg [3]: The rg color model refers to the r and g channels of the normalized RGB color model:

$$\begin{pmatrix} r \\ g \end{pmatrix} = \begin{pmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \end{pmatrix} \tag{1}$$

The normalization makes r and g invariant to shadows and shading.

TRGB [3]: Transformed RGB (TRGB) is obtained by normalizing each channel of the RGB color space to a zero-mean and unit-variance distribution, i.e.,

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \frac{R-\mu_R}{\sigma_R} \\ \frac{G-\mu_G}{\sigma_G} \\ \frac{B-\mu_B}{\sigma_B} \end{pmatrix} \tag{2}$$

where μ_C and σ_C are the mean and standard deviation of the distribution in channel C. The normalization makes TRGB invariant to highlights and light intensity change.

OPP [3], [5]: The opponent color space (OPP) is transformed from the RGB color space,

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix} \tag{3}$$

The intensity information is contained in O_3, and the chromatic information in O_1 and O_2. While O_1 and O_2 are invariant to highlights due to the intensity cancelation, the combination of all three channels is not.

C-OPP [5]: The C-OPP $[O_1, O_2]$ refers to the chromatic components of the opponent space, which is invariant to highlights.

N-OPP [5]: The N-OPP $\left[\frac{O_1}{O_3}, \frac{O_2}{O_3}\right]$ is the normalized C-Oppo. The normalization makes N-Oppo invariant to shadows and shading.

Hue [5]: Hue is defined as the ratio between the two chromatic opponents: $\left[\frac{O_1}{O_2}\right]$. It is invariant to shadows, shading and highlights. Hue used here is a little different from the H channel of HSV color space which contains an extra hexagon-to-circle transformation.

B. Selection of Trackers

It is unrealistic to test the effect of using color information in all existing visual trackers. Instead, we include in our list representative trackers that have ranked high in recent benchmark evaluations. The selection of visual trackers includes two stages: initial selection and fine adjustment. Short descriptions of the selected trackers can be found in Table I.

Initial selection. For initial selection, the basic idea is to construct a set of trackers, such that each selected tracker has been ranked in the top 10 in at least one of the three recent evaluations ([6], [7], [8]) according to at least one of the criteria (whenever multiple criteria are used). The evaluation in [6] involves 29 tracking algorithms on 50 sequences using two evaluation criteria and three robustness evaluation strategies: one-pass evaluation (OPE), temporal robustness evaluation

(TRE), and spatial robustness evaluation (SRE). The initial selection derived from [6] is denoted by

$$\begin{aligned}\mathcal{T}_1 = \{&\mathrm{ASLA}, \mathrm{CPF}, \mathrm{CSK}, \mathrm{CXT}, \mathrm{DFT}, \mathrm{LOT}, \mathrm{LSK}, \mathrm{MTT},\\ &\mathrm{OAB}, \mathrm{SCM}, \mathrm{Struck}, \mathrm{TLD}, \mathrm{VTD}, \mathrm{VTS}\}.\end{aligned}$$

Similarly, the set of selected trackers derived from [7] is

$$\begin{aligned}\mathcal{T}_2 = \{&\mathrm{BSBT}, \mathrm{CPF}, \mathrm{Frag}, \mathrm{IVT}, \mathrm{L1T}, \mathrm{MIL}, \mathrm{MTT}, \mathrm{OAB},\\ &\mathrm{OFS}, \mathrm{SemiT}, \mathrm{Struck}, \mathrm{TLD}, \mathrm{VTD}\}.\end{aligned}$$

In [8], six state-of-the-art 2D trackers are evaluated and they form the selected tracker set

$$\mathcal{T}_3 = \{\mathrm{CT}, \mathrm{MIL}, \mathrm{SemiT}, \mathrm{Struck}, \mathrm{TLD}, \mathrm{VTD}\}.$$

Combining $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}_3$, we get the initial selection as

$$\begin{aligned}\mathcal{T}_{\mathrm{init}} =& \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3\\ =& \{\mathrm{ASLA}, \mathrm{BSBT}, \mathrm{CPF}, \mathrm{CSK}, \mathrm{CT}, \mathrm{CXT}, \mathrm{DFT}, \mathrm{Frag},\\ &\mathrm{IVT}, \mathrm{L1T}, \mathrm{LSK}, \mathrm{LOT}, \mathrm{MIL}, \mathrm{MTT}, \mathrm{OAB}, \mathrm{OFS},\\ &\mathrm{SCM}, \mathrm{SemiT}, \mathrm{Struck}, \mathrm{TLD}, \mathrm{VTD}, \mathrm{VTS}\}.\end{aligned}$$

Fine adjustment. Since we need to modify the original code of the trackers to encode color information, we remove from $\mathcal{T}_{\mathrm{init}}$ the trackers without available source code. In addition, we adjust the selection according to the following rules: (1) We exclude some trackers that are very slow even before encoding color information. (2) Some trackers have several components and it is hard to determine which part is more critical for encoding color information. (3) For L1T, we replace it by its new version L1APG, which speeds up the L1 minimization using an accelerated proximal gradient approach. (4) For CPF, although there is no available source code for this tracker, we implement it ourselves since the particle filter tracking framework is fairly standard. (5) We also incorporate FCT [16], which accelerates CT with a coarse-to-fine search search strategy. (6) There are several recently proposed trackers achieving promising performances on the CVPR2013 benchmark, such as [19], [24], [47], [48], [49], [50], [51], [52]. Among them, we sample KCF [19] and MEEM [24] in our study due to the availability of source code. With the above adjustments, our final selection contains the following 16 state-of-the-art trackers:

$$\begin{aligned}\mathcal{T}_{\mathrm{base}} = \{&\mathrm{ASLA}, \mathrm{CPF}, \mathrm{CSK}, \mathrm{CT}, \mathrm{DFT}, \mathrm{FCT}, \mathrm{Frag}, \mathrm{IVT}, \mathrm{KCF},\\ &\mathrm{L1APG}, \mathrm{LOT}, \mathrm{MEEM}, \mathrm{MIL}, \mathrm{OAB}, \mathrm{SemiT}, \mathrm{Struck}\}.\end{aligned}$$

In the rest of the paper, we call these trackers *base trackers* to distinguish from the versions using color information.

C. Encoding Color in Visual Tracking

To maximize the benefit of using color information, one may need to design a tracker-specific strategy for encoding color for each individual tracker. It is however far from trivial to come up with “best” strategies for all base trackers in $\mathcal{T}_{\mathrm{base}}$. Consequently, we rely on straightforward solutions, which, besides being relatively fair, have demonstrated clear improvement over the original grayscale base trackers. We mainly use two strategies: First, for trackers that learn their model via feature selection mechanisms, we enrich the feature pool by including features from multiple color channels. Second, for trackers using fixed grayscale-based representation, we extend such representation to multiple color channels and concatenate the results as a new representation. In the rest of the paper, we call a color-enhanced version of a base tracker a *color enhanced tracker*.

ASLA. In [9], a robust tracking algorithm using adaptive structural local sparse appearance model is proposed to handle occlusion and locate the target. For appearance modeling, both holistic templates and local patches are represented by vectorized ℓ_2 normalized pixel intensities. To encode color information, we concatenate the ℓ_2 normalized vectors from all color channels of a given color model, for representing both local image patches and holistic templates.

CPF. In [11], color histogram in the HSV color space is used to model an image patch in the particle filter tracking framework [53]. Since CPF already uses color information from the HSV space, we directly generalize it to other photometric representations for comparison.

CSK. CSK [12] uses a kernel regularized least squares classifier trained on all subwindows around the target, by exploiting the circulant structure in the Fourier domain. To encode color information, we concatenate the vectorized pixel values from all color channel to represent an image patch (i.e., a sample), and extend the RBF kernel in the original CSK by summing over the RBF kernels of individual color channels.

CT. In [13], a compressive tracker is presented that maps a high-dimensional feature vector, extracted from a target sample, to a low dimensional space, through a sparse random measurement matrix. The resulting low-dimensional representation is then fed into a naive Bayes classifier for target localization. To encode color information given a specific color model, we extend the feature projection to all color channels, and concatenate the resulting low-dimensional feature vectors as a new target representation.

DFT. In [15], distribution fields (DFs) is proposed for building image descriptors in visual tracking. The tracking inference is based on the L1 distance between the smoothed DFs of a candidate region and that of the target model. To encode color information, given a color model, we concatenate the smoothed DFs of the target or candidates from each color channel to form the representation.

FCT. FCT [16] extends CT mainly in using a coarse-to-fine sliding window for acceleration. Similar to CT, we concatenate the low-dimensional feature vectors from all the channels to encode color information.

Frag. The Frag tracker [17] represents a tracking object (template or candidate patch) by multiple image fragments to handle partial occlusion and capture spatial structures. Each fragment patch is represented by a histogram that is efficiently calculated by the integral histogram [54] algorithm. To convert Frag to a color version, we represent each fragment using the concatenation of the histograms from each color channel of a given color model.

IVT. IVT [18] uses incremental subspace learning for robust visual tracking. The idea is to dynamically and incrementally maintain and update a subspace for target appearance representation. In the original IVT algorithm, an object patch is first represented by vectorized pixel intensities. To include color information in IVT, we concatenate the pixel values from all channels of a color model, and then follow the same subspace learning and inference procedure as the original IVT.

KCF. KCF [19] is the extension of CSK in that KCF uses multiple channels by summing over the results from all the channels in the Fourier domain. In addition, HOG is used to further boost the performance. To incorporate color information, we concatenate HOG from each channel of a given color model.

L1APG. L1APG [20] is an extension of the sparse representation-based visual tracker [21]. The key idea is to use an accelerated proximal gradient algorithm to speed up the L1 minimization used to compare a candidate to a set of templates. In L1APG, a candidate or template is represented by a normalized vector of pixel intensities. To incorporate color information, we concatenate such vectors from all channels of a given color model before seeking solutions of the sparse representation.

LOT. Capturing the degree of local disorder of the tracking target, LOT [22] deals with both rigid and deformable targets with no prior assumption. In LOT, each pixel in an image patch is represented by its location and appearance, and superpixels are used in practice for efficiency. Since LOT itself has already used the HSV color model, we modify the original LOT by replacing HSV with other color models for comparison.

MEEM. In [24], historical tracking results are used to build an expert ensemble. To avoid the contamination of the target model, the best expert is selected to restore the tracking result when needed based on a minimum entropy criterion. The original MEEM already uses LAB color space to extract features, so we just replace LAB with other color models for evaluation.

MIL. The key idea of MIL [55] is to handle the unavoidable label noise in a multiple instance learning framework. MIL uses Haar-like features [56] to model the appearance of the target. In particular, a feature pool containing Haar-like features is generated and the learning process selects effective features from the pool to construct the classifier. To encode color information using a specific color model, we enlarge the feature pool by including Haar-like features from all channels of the color model. Then, the learning process is the same as the original MIL.

OAB. OAB [26] employs an on-line Adaboost algorithm to address the target appearance variation during the course of tracking. According to the implementation from the authors' website[1], Haar-like features are used to form the feature pool for Adaboost. Then, similar to MIL, we encode color information into OAB by enlarging the feature pool with features extracted from involved color channels of a given color model.

SemiT. SemiT [29] uses a semi-supervised on-line boosting algorithm for visual tracking. The implementation of SemiT is based on OAB and the Haar-like features are again used for constructing the feature pool. As a result, we encode color information for SemiT in the same way as for OAB.

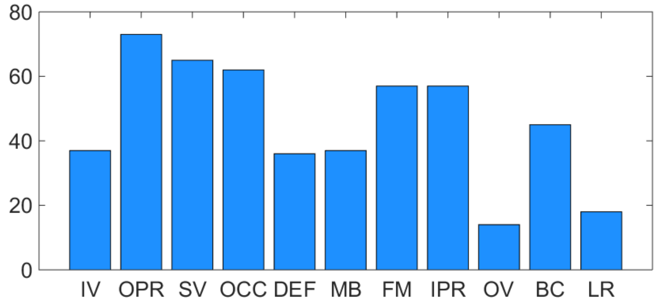


Fig. 3. The distribution of challenge factors in TColor-128.

Struck. Struck [1] uses a kernelized structured output support vector machine to distinguish the target from the background. For modeling target appearance, Struck uses a 192 dimensional Haar-like [56] feature vector obtained from a 4×4 grid at two scales with six different filters. To encode color information, we concatenate the features from different color channels into a larger feature vector for both training and updating the classifier. In particular, if the photometric representation has n channels, the final feature vector is $n \times 192$ dimensional.

IV. Color Tracking Benchmark

As summarized in Section II, there is a lack of benchmark datasets devoted to color visual tracking. Addressing this issue, we construct a large dataset with 128 color sequences, named TColor-128, as a color tracking benchmark. The sequences in TColor-128 come from two main sources: previous studies and new collections.

For the first part, we have collected 50 frequently tested color sequences used in previous studies, such as [6] and [38]. These sequences are however insufficient for thoroughly evaluating color trackers. On the one hand, due to the large number of factors involved in visual tracking and many tunable parameters in visual trackers, experiments on 50 sequences may not be enough to reach a significant conclusion. On the other hand, due to the popularity of these sequences in previous studies, the performance on them are often close to saturation and many of them are not as difficult as they originally appear to be (see Fig. 2). The two observations motivate us to collect more color sequences, which form the second part of TColor-128.

The second part of TColor-128 contains 78 color sequences newly collected from the Internet. The 78 sequences, by design, largely increase the diversity and difficulty over the first 50 sequences: various circumstances are involved such as highway, airport terminal, railway station, concert, etc.; none of them were taped purposely for evaluating visual tracking algorithms; these sequences have many challenge factors such as full target occlusion, large illumination change, significant target deformation and low resolution.

Fig. 2 shows the first frame of 80 selected sequences with the bounding box of the tracking target. These sequences are ordered according to their degree of challenges in visual

[1] http://www.vision.ee.ethz.ch/boostingTrackers/

tracking, measured by the average performances of the color trackers experimented in our study (details in Sec. V). Fig. 2 shows that newly collected sequences are often much harder than classically used ones, justifying the inclusion of these new sequences.

In addition to tracking ground truth, each sequence in TColor-128 is also annotated by its challenge factors. Same as in [6], 11 factors are used for TColor-128, including illumination variation (IV), scale variation (SV), occlusion (OCC), deformation (DEF), motion blur (MB), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), out-of-view (OV), background clutters (BC), and low resolution (LR). In particular, scale variation is decided when the ratio of the size of the bounding box in the current frame to that in the first frame falls out of the range $[0.5, 2]$; fast motion is decided when the target motion is larger than 20 pixels; low resolution is decided when the number of pixels inside the groundtruth bounding box is fewer than 400 pixels. Fig. 3 gives the distribution of challenge factors in TColor-128. Although we try to make the dataset balanced in terms of challenging factors, trackers that handle OPR, SV, OCC, FM and IPR better may have some advantages over those who handle OV and LR better.

V. Evaluation

We first evaluate color tracking using the proposed TColor-128 benchmark, with detailed analysis on the effects of different combinations of color-representations and visual trackers, as well as comparison with recently proposed color trackers. Then, for further validation, we run the color-encoded trackers on the Princeton Tracking Benchmark (PTB) where the results are consistent with the results on TColor-128.

A. Evaluation Color Trackers on TColor-128

Our main goal is to study the effectiveness of encoding color information for visual tracking. Toward this goal, we conduct thorough experiments on TColor-128 to explore the following issues: the gain by encoding color information, effects of different color representations on base trackers, and degree of difficulty of the sequences in TColor-128.

1) Evaluation Metrics: Following the protocol in [6], we use two widely used metrics for tracking evaluation. The main metric we use is the *Area Under Curve* (AUC) derived from the *success plot* of tracking algorithms. More specifically, for each frame, given the tracking output bounding box (r_t) and ground truth bounding box (r_g), the *overlap ratio* (S) is used as the basic measure for tracking success. The overlap ratio is defined by $S = \frac{|r_t \bigcap r_g|}{|r_t \bigcup r_g|}$, where $|\cdot|$ denotes the area. Then, the success rate of a tracker on a sequence is the percentage of frames whose overlap score S is larger than a given threshold. By varying the threshold from 0 to 1, one can generate the success plot, and the AUC can be derived afterwards.

Another metric used in our evaluation is the *precision plot* as in [55], [6]. It is based on the *center location error* (CLE), defined as the Euclidean distance between the centers of the tracking result and groundtruth. Traditionally, the average CLE over all the frames have been used to measure the tracking performance. Such measurement, however, can be meaninglessly large when a tracker completely loses the target. The precision plot addresses this issue by showing the *precision*, defined as the percentage of frames whose CLEs are smaller than a threshold, against the CLE threshold. Same as in [6], we use the precision at the threshold 20 to rank the performance. It is worth mentioning that our goal is to comprehensively study the performance of different color models rather than the robustness of the state-of-the-art trackers, we use one-pass evaluation (OPE) in the following evaluation.

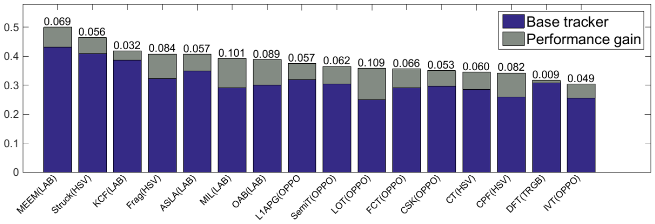


Fig. 6. Performance gain of using color in terms of AUC. The best color model for each base tracker is given in the parentheses.

2) Gain from Encoding Color Information: The combination of different color representations and base trackers generates $160 = 16 \times 10$ trackers (including the base ones). We run all of them on TColor-128 to study the effects of encoding color for tracking. The performances are plotted in details in Figures 4 and 5, containing respectively success plots (based on AUC) and precision plots. The plots show clearly that, when appropriate color models (e.g., Opponent, LAB and HSV) are used, encoding color information always boosts the performance of base trackers. Since different base trackers favor different color models, we create a set of "best" color-enhanced trackers as

$$\begin{aligned}\mathcal{T}_{\text{best}} = \{&\text{MEEM(LAB)}, \text{Struck(HSV)}, \text{KCF(LAB)},\\ &\text{Frag(HSV)}, \text{ASLA(LAB)}, \text{MIL(LAB)}, \text{OAB(LAB)},\\ &\text{L1APG(OPP)}, \text{SemiT(OPP)}, \text{LOT(OPP)},\\ &\text{FCT(OPP)}, \text{CSK(OPP)}, \text{CT(HSV)},\\ &\text{CPF(HSV)}, \text{DFT(TRGB)}, \text{IVT(OPP)}\}.\end{aligned}$$

To further understand the gain of using color information, we calculate the performance gain (in terms of AUC) achieved by the best color-enhanced trackers in $\mathcal{T}_{\text{best}}$. For example, for MEEM, the gain is $0.069 = 0.5 - 0.431$ achieved by LAB. Such performance gains are summarized in Fig. 6. It clearly shows that for most base trackers, color information can significantly improve tracking performance by more than 10%. Among all the trackers, LOT benefits most by using the concatenated pixel values from each channel of the Opponent space. Fig. 8 shows some example frames where color helps. Considering that only straightforward ways are used to encode color information into base trackers, the improvement is very promising and we expect that further performance gains can be achieved by more carefully encoding color information in the trackers.

3) Comparison with Other Color Trackers: As discussed in the related works Section, several notable tracking algorithms have recently been proposed that explicitly take color

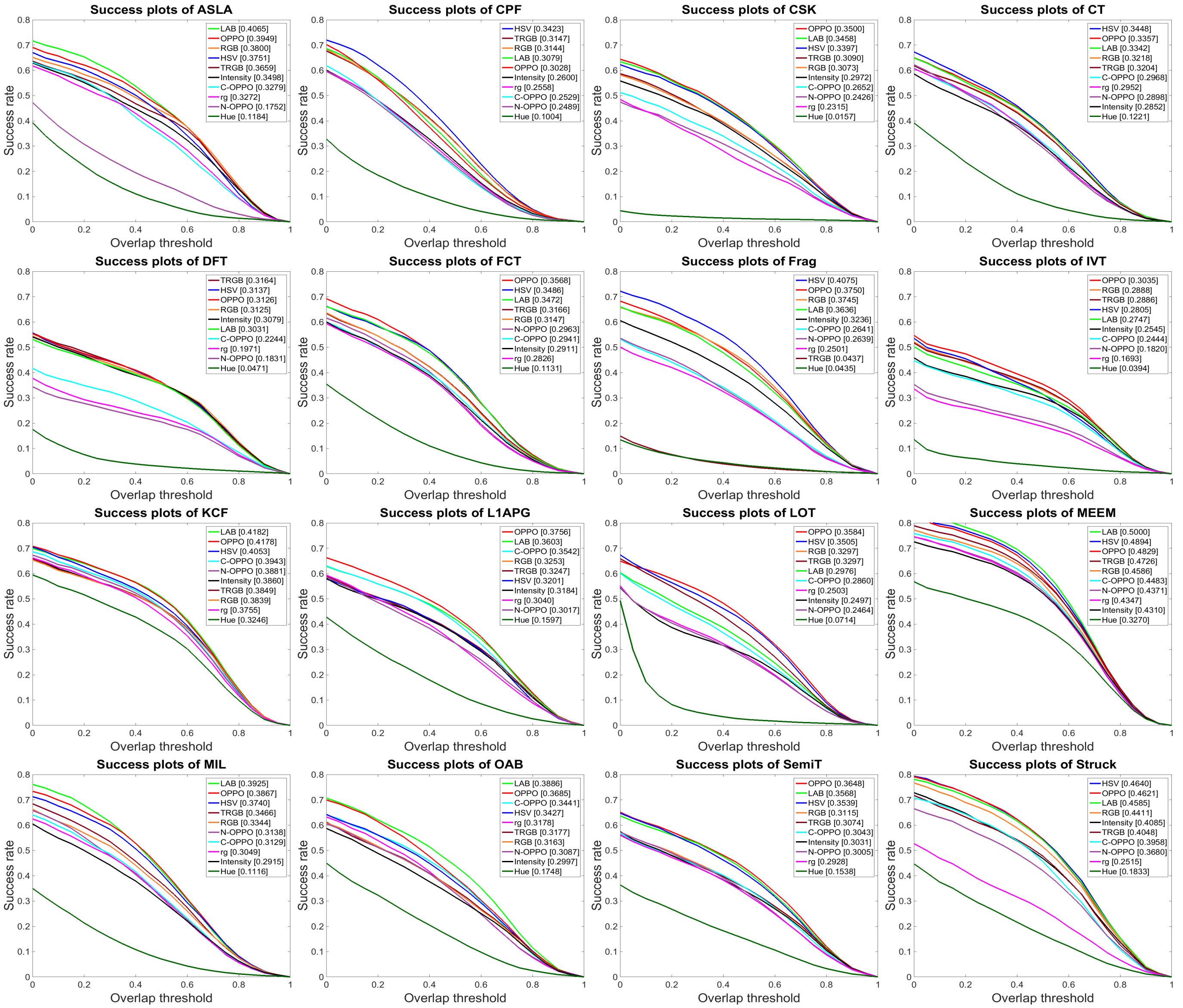


Fig. 4. Comparison of color models in color tracking for the base trackers in $\mathcal{T}_{\text{base}}$ using *success plots*. The AUC score of each color model for a given tracker is shown in the legend.

information into consideration. It is imperative to evaluate them together with the color enhanced trackers proposed in this paper. We include two such trackers in our experiments, namely CN$_2$ [2] and VTD [31]. The two trackers are evaluated on the TColor-128 benchmark together with the 16 color enhanced trackers (same as listed in Fig. 6).

Fig. 7 gives the success and precision plots of the evaluation. It shows that MEEM+LAB outperforms other color trackers by a noticeable margin. Moreover, it shows that some grayscale trackers (e.g., Frag and ASLA), after boosting with color information, outperforms recently proposed color trackers. Such observations again confirm the effectiveness of using color for visual tracking and implies there is still room for improvement.

4) Analysis and Discussion: **Comparison of color representations:** While different base trackers favor different color representations, some representations bring more advantages than others. For a quantitative comparison, we calculate the average rank of each color representation by averaging its ranks of all color-enhanced trackers, and the result is shown in Table III. In addition, we calculate the mean performance gain by averaging the gain measured using AUC of all the evaluated trackers. The statistics are visualized in Fig. 9, together with standard deviations.

The results show that Opponent, HSV, LAB and RGB are in general helpful for visual tracking. It is worth noting that these four representations do not have any photometric invariance properties, but they have strong discriminative capability. The performance gains from other color models are either insignificant or negative.

Performance gain of color representations with respect to challenge factors: Taking benefit of the challenge factor annotation in TColor-128, we study the specific performance of each color representation with respect to each challenge

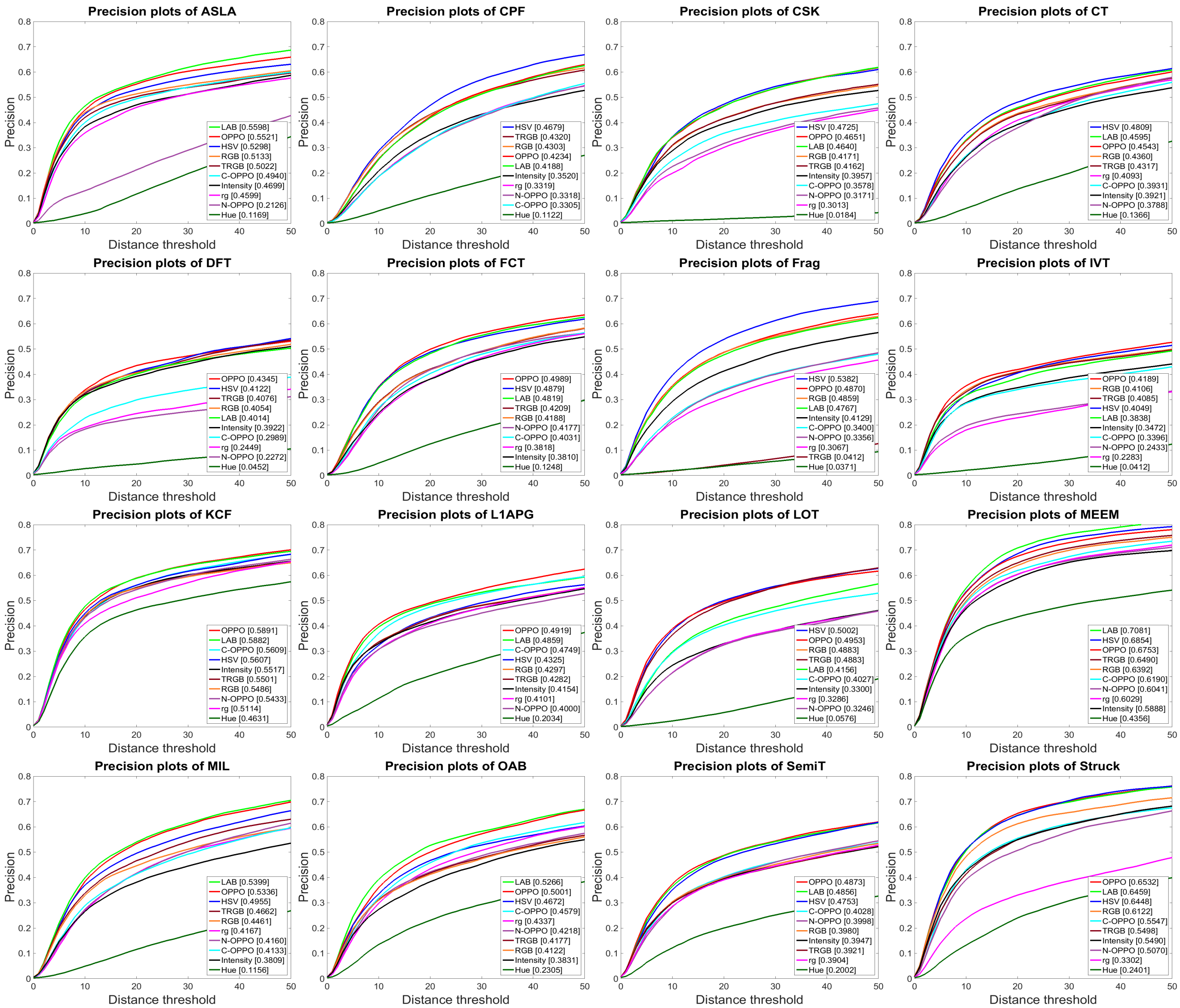


Fig. 5. Comparison of color models in color tracking for the base trackers in $\mathcal{T}_{\text{base}}$ using *precision plots*. The precision obtained at threshold 20 is shown in the legend.

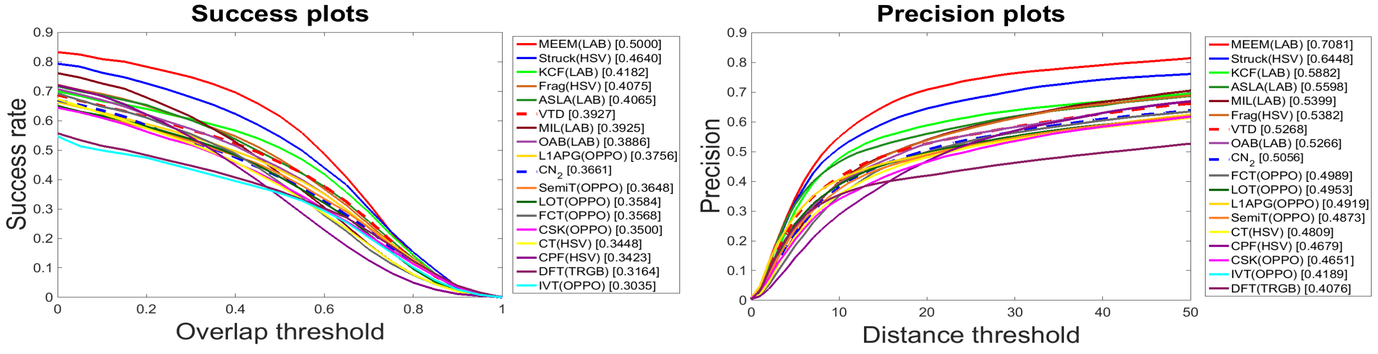
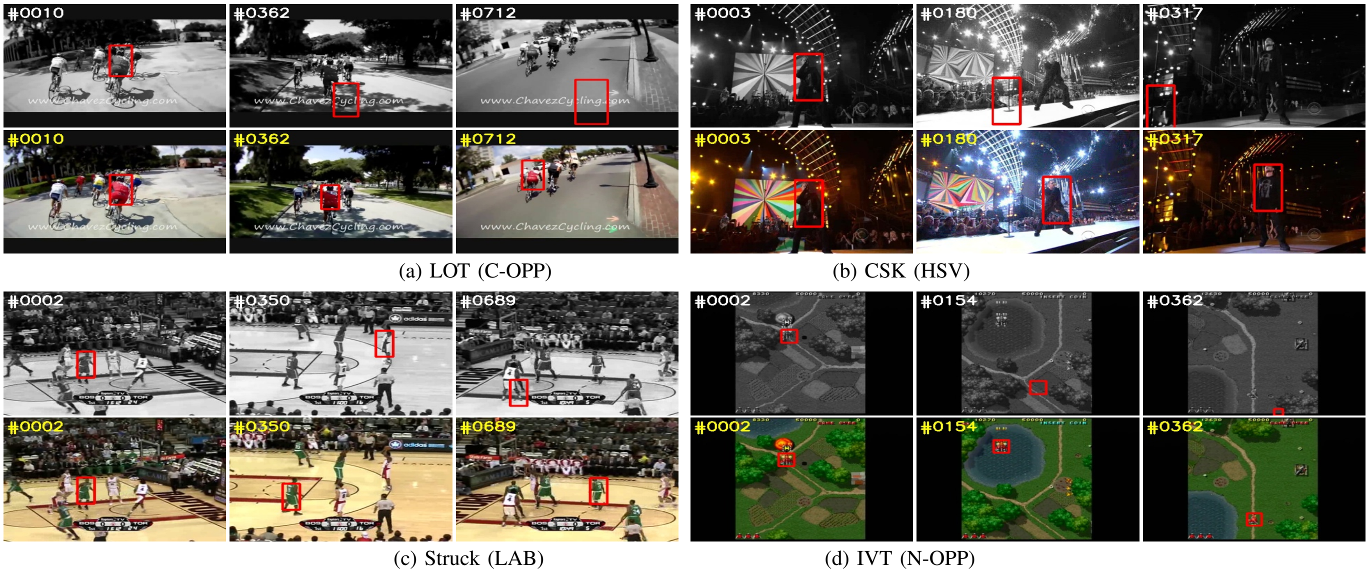


Fig. 7. Success and precision plots for all color-enhanced trackers and come recently proposed color trackers on TColor-128.

factor, and such a study can provide further understanding of why and how color helps in visual tracking.

For each of the 11 challenge factors (see Sec. IV), we construct a subset of TColor-128 containing all sequences that involve the factor. Then, the performances of all color encoded trackers on such subsets are summarized in terms of AUC. The results and rankings are given in Table IV. From the results, we can see the performances of different



(a) LOT (C-OPP)

(b) CSK (HSV)

(c) Struck (LAB)

(d) IVT (N-OPP)

Fig. 8. Examples where color helps tracking. For each sequence, the results are from the tracker (shown under the frames) that gains the most by using color.

TABLE III
THE PERFORMANCES (AUC) AND RANKINGS (IN PARENTHESES) OF DIFFERENT COLOR REPRESENTATIONS FOR **DIFFERENT TRACKERS**. THE FIRST AND SECOND BESTS ARE INDICATED BY RED AND BLUE RESPECTIVELY.

	Avg. rank	ASLA	CPF	CSK	CT	DFT	FCT	Frag	IVT	KCF	L1APG	LOT	MEEM	MIL	OAB	SemiT	Struck
OPP	1.94	0.395(2)	0.303(5)	0.350(1)	0.336(2)	0.313(3)	0.357(1)	0.375(2)	0.304(1)	0.418(2)	0.376(1)	0.358(1)	0.483(3)	0.387(2)	0.368(2)	0.365(1)	0.462(2)
HSV	2.63	0.375(4)	0.342(1)	0.340(3)	0.345(1)	0.314(2)	0.349(2)	0.408(1)	0.281(4)	0.405(3)	0.320(6)	0.350(2)	0.489(2)	0.374(3)	0.343(4)	0.354(3)	0.464(1)
LAB	2.75	0.406(1)	0.308(4)	0.346(2)	0.334(3)	0.303(6)	0.347(3)	0.364(4)	0.275(5)	0.418(1)	0.360(2)	0.298(5)	0.500(1)	0.393(1)	0.389(1)	0.357(2)	0.459(3)
RGB	4.31	0.380(3)	0.314(3)	0.307(5)	0.322(4)	0.312(4)	0.315(5)	0.374(3)	0.289(2)	0.384(8)	0.325(4)	0.330(3)	0.459(5)	0.334(5)	0.316(7)	0.312(4)	0.441(4)
TRGB	4.63	0.366(5)	0.315(2)	0.309(4)	0.320(5)	0.316(1)	0.317(4)	0.044(9)	0.289(3)	0.385(7)	0.325(5)	0.330(4)	0.473(4)	0.347(4)	0.318(6)	0.307(5)	0.405(6)
C-OPP	6.06	0.328(7)	0.253(8)	0.265(7)	0.297(6)	0.224(7)	0.294(7)	0.264(6)	0.244(7)	0.394(4)	0.354(3)	0.286(6)	0.448(6)	0.313(7)	0.344(3)	0.304(6)	0.396(7)
Intensity	6.94	0.350(6)	0.260(6)	0.297(6)	0.285(9)	0.308(5)	0.291(8)	0.324(5)	0.254(6)	0.386(6)	0.318(7)	0.250(8)	0.431(9)	0.291(9)	0.300(9)	0.303(7)	0.409(5)
N-OPP	7.75	0.175(9)	0.249(9)	0.243(8)	0.290(8)	0.183(9)	0.296(6)	0.264(7)	0.182(8)	0.388(5)	0.302(9)	0.246(9)	0.437(7)	0.314(6)	0.309(8)	0.300(8)	0.368(8)
rg	8.00	0.327(8)	0.256(7)	0.231(9)	0.295(7)	0.197(8)	0.283(9)	0.250(8)	0.169(9)	0.375(9)	0.304(8)	0.250(7)	0.435(8)	0.305(8)	0.318(5)	0.293(9)	0.251(9)
Hue	10.00	0.118(10)	0.100(10)	0.016(10)	0.122(10)	0.047(10)	0.113(10)	0.043(10)	0.039(10)	0.325(10)	0.160(10)	0.071(10)	0.327(10)	0.112(10)	0.175(10)	0.154(10)	0.183(10)

TABLE IV
THE PERFORMANCES (AUC) AND RANKINGS (IN PARENTHESES) OF DIFFERENT COLOR REPRESENTATIONS UNDER DIFFERENT **CHALLENGE FACTORS**. THE FIRST AND SECOND BESTS ARE INDICATED BY RED AND BLUE RESPECTIVELY.

	IV	OPR	SV	OCC	DEF	MB	FM	IPR	OV	BC	LR
OPP	0.362(2)	0.360(1)	0.350(1)	0.335(2)	0.360(2)	0.322(2)	0.353(2)	0.352(1)	0.299(2)	0.362(2)	0.318(1)
HSV	0.365(1)	0.355(3)	0.340(3)	0.327(3)	0.363(1)	0.306(3)	0.339(3)	0.349(2)	0.287(3)	0.366(1)	0.313(2)
LAB	0.359(3)	0.358(2)	0.347(2)	0.336(1)	0.359(3)	0.323(1)	0.353(1)	0.346(3)	0.304(1)	0.356(3)	0.300(4)
RGB	0.317(5)	0.325(4)	0.317(4)	0.308(4)	0.332(4)	0.286(4)	0.320(4)	0.316(4)	0.265(4)	0.328(4)	0.300(3)
TRGB	0.301(6)	0.307(6)	0.297(6)	0.291(5)	0.310(5)	0.263(6)	0.297(6)	0.295(6)	0.242(6)	0.307(5)	0.272(6)
C-OPP	0.320(4)	0.320(5)	0.298(5)	0.285(6)	0.290(7)	0.283(5)	0.301(5)	0.314(5)	0.252(5)	0.299(7)	0.245(7)
Intensity	0.298(8)	0.296(7)	0.294(7)	0.284(7)	0.295(6)	0.262(7)	0.296(7)	0.288(7)	0.239(8)	0.306(6)	0.278(5)
rg	0.298(7)	0.295(8)	0.276(8)	0.265(8)	0.274(9)	0.244(8)	0.271(8)	0.288(8)	0.233(9)	0.273(9)	0.217(9)
N-OPP	0.291(9)	0.289(9)	0.275(9)	0.256(9)	0.278(8)	0.244(9)	0.271(9)	0.283(9)	0.239(7)	0.276(8)	0.219(8)
Hue	0.143(10)	0.138(10)	0.129(10)	0.127(10)	0.121(10)	0.119(10)	0.131(10)	0.137(10)	0.099(10)	0.115(10)	0.082(10)

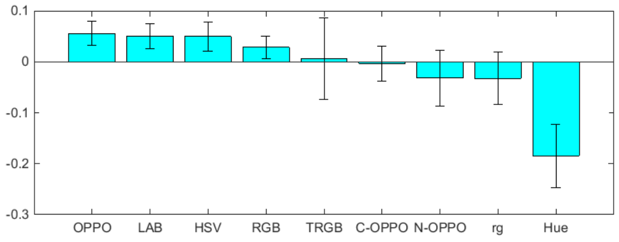


Fig. 9. Average performance gains for different color representations in AUC.

color representations under different challenge factors have very similar trends as those under general cases (i.e., for all sequences). OPP, HSV, LAB and RGB again rank top four for all challenge factors (except for illumination variation (IV) on which RGB ranks the fifth), though none of them is fully invariant to highlights or shadow-shading. On the other hand, even for sequences with significant illumination variation (IV), the color representations with strong invariance properties, such as Hue and N-OPP, do not perform well. These observations imply that the discriminative power is more important for the success of a color representation than are invariance properties.

For further understanding the gain using color information, for each sequence, we also record the best performance, named AUC_0, achieved over $\mathcal{T}_{\text{best}}$. Then the gain for this sequence is calculated as the difference between AUC_0 and the AUC achieved by the corresponding base tracker. The

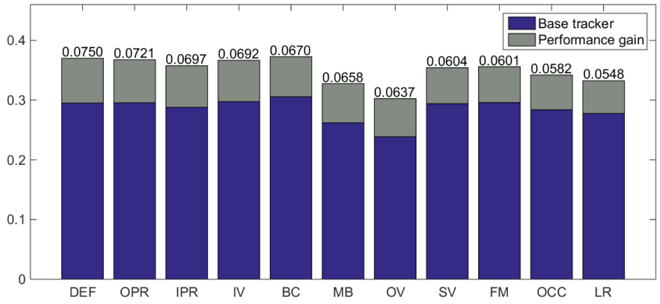


Fig. 10. Performance gain of using color information with respect to different challenge factors, in terms of AUC. The challenge factors (x-axis) are ordered according to performance gain.

average gain for each challenge factor is then estimated for all sequences involve the challenge factor. Fig. 10 summarizes these performance gains. The figure shows that, while using color helps improving tracking for all challenge factors, deformation (DEF), in plane rotation (IPR), out of plane rotation (OPR) and illumination variation (IV) benefit relatively more than other attributes. By contrast, the performance gains for occlusion (OCC), fast motion (FM) and low resolution (LR) are relatively small. This can be explained by the instability or unavailability of appearance information under these challenges.

Failure cases: Though color information helps improving the tracking performance, tracking in general remains a challenging problem. In the following we discuss some typical failure cases observed in our experiments (Fig. 11).

As shown in Fig. 10, occlusion (OCC) and out of view (OV) are two key sources for tracking failure. When a target is seriously occluded or leaves the view, it is hard for trackers to catch it when it reappears, as illustrated in Fig. 11(b) and Fig. 11(e). One reason for the failure is the contamination of target model from the background when part or whole of the target is missing. Meanwhile, when the target reappears, it may fall out of the search range of the current tracker.

Fast motion (FM) and motion blur (MB) are another pair of challenge factors that bother tracking algorithms, as shown in Fig. 11(c). When there is a large target movement between two frames, the target may fall out of the search range of a tracker and then trigger a failure. Using color information helps little in such a scenario, though color can bring discriminative information for motion blurred target appearance.

Another factor that often causes problems is low target resolution (LR), as shown in Fig. 11(a). When the target is too small, it is hard to capture enough appearance information to distinguish the target from background, even with color information encoded.

From the above discussion, we see that OV, OCC, MB, FM and LR contribute to a large portion of tracking failures, and they can not be addressed by simply using color information. Improving the search strategy may be an option for them, though often at the sacrifice of run time efficiency. In addition to these factors, we also show some failure cases due to scale variation (SV) and background clutters (BC) in Fig. 11(d) and Fig. 11(f), respectively.

Degree of difficulty of sequences in TColor-128. For the newly proposed TColor-128 benchmark, it is to quantitatively analyze the degree of difficulty of the sequences in terms of visual tracking. We formally derive the *degree of difficulty* (DoD) for each sequence based on the above evaluations. Specifically, for each sequence, the AUCs of the 16 best color-enhanced trackers in $\mathcal{T}_{\text{best}}$ are recorded. Then, the tracking degree of difficulty is formally defined as

$$\text{DoD} = 1 - (\text{average AUC of the top 5 results}).$$

The degrees of difficulty for some sequences are listed in Fig. 2. The results show that many newly included sequences are very challenging for visual tracking, especially compared with previously used ones. We will disclose the degrees of difficulty for all sequences when sharing the TColor-128 dataset.

B. Validation on Princeton Tracking Benchmark

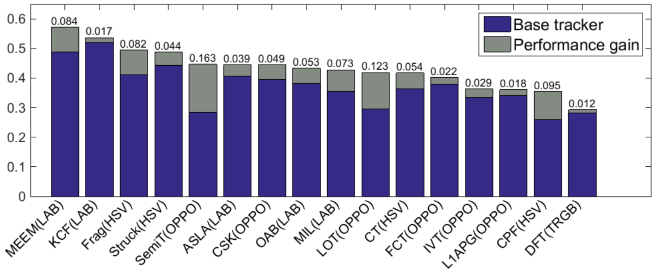
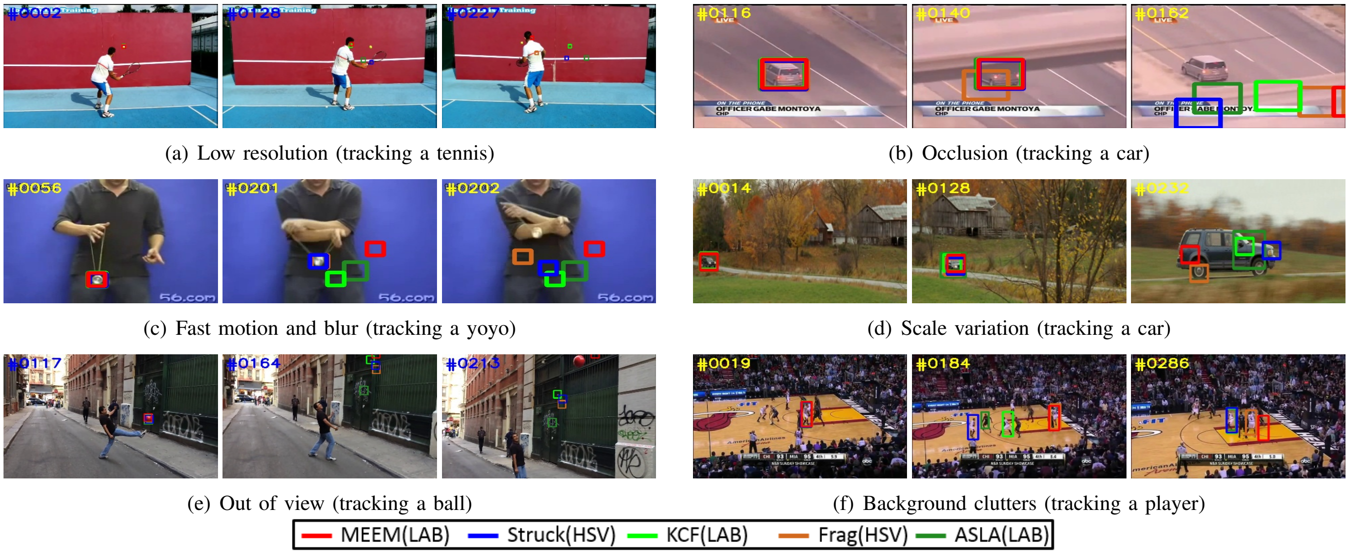


Fig. 12. Performance gain of using color in terms of overall success rate on the Princeton Tracking Benchmark. The best color model for each base tracker is given in the parentheses.

Princeton Tracking Benchmark (PTB) [8] is a recently proposed dataset containing 100 RGBD sequences. Among the 100 sequences, the ground truth of 5 sequences are released for parameter tuning, and the rest 95 sequences are withheld for evaluation. The color components of PTB make it suitable for testing color trackers. That said, we use PTB for validation rather than for the main evaluation since PTB is designed for evaluating RGBD trackers, and the collection of the sequences is specific for environments where the depth information plays an important role. In addition, due to the limitation in current depth acquisition techniques, the sequences are limited to indoor environments and targets are close to cameras.

Since our focus is on 2D color tracking, we exclude the trackers using 3D depth information such as those proposed in [8]. We run the 16 best color-encoded trackers in $\mathcal{T}_{\text{best}}$ and their corresponding base trackers on all sequences in PTB. Then, following the protocol in [8], we submit the tracking results to PTB website for evaluation.

The evaluation results, together with those reported in [8], are summarized in Table V. In [8], the sequences are divided into several categories according to target type, target size, movement, occlusion and motion type. In Table V, the trackers are ranked according to the average success rate, which is calculated by thresholding the overlap between the estimated bounding box of the target and the ground truth. In addition, similar to analysis on TColor-128, we summarize in Fig. 12 the performance gains achieved by integrating color information for the 16 base trackers.



(a) Low resolution (tracking a tennis)

(b) Occlusion (tracking a car)

(c) Fast motion and blur (tracking a yoyo)

(d) Scale variation (tracking a car)

(e) Out of view (tracking a ball)

(f) Background clutters (tracking a player)

Fig. 11. Typical failures observed in our experiments involving different challenge factors as listed in the subtitles.

TABLE V

EVALUATION RESULTS ON THE PRINCETON TRACKING BENCHMARK. SUCCESS RATE (SR) AND CORRESPONDING RANKINGS (IN PARENTHESES) ARE GIVEN UNDER DIFFERENT CATEGORIZATIONS.

Algorithm	Avg. rank	all SR	target type			target size		movement		occlusion		motion type	
			human	animal	rigid	large	small	slow	fast	yes	no	passive	active
MEEM(LAB)	1.73	0.572	0.510(1)	0.510(7)	0.680(2)	0.580(1)	0.560(1)	0.680(1)	0.530(1)	0.460(1)	0.720(2)	0.690(1)	0.530(1)
KCF(LAB)	2.45	0.536	0.420(5)	0.540(2)	0.680(1)	0.510(3)	0.550(2)	0.670(2)	0.480(2)	0.400(5)	0.720(1)	0.660(2)	0.490(2)
KCF	4.27	0.519	0.420(4)	0.500(10)	0.650(3)	0.480(7)	0.540(3)	0.650(5)	0.470(3)	0.410(2)	0.670(4)	0.650(3)	0.470(3)
Struck(HSV)	5.18	0.488	0.400(7)	0.520(3)	0.580(6)	0.490(5)	0.490(4)	0.620(7)	0.430(4)	0.350(7)	0.680(3)	0.590(6)	0.450(5)
Frag(HSV)	6.27	0.494	0.410(6)	0.460(18)	0.610(4)	0.520(2)	0.470(7)	0.650(4)	0.430(5)	0.400(4)	0.620(8)	0.630(4)	0.440(7)
MEEM	6.45	0.488	0.430(2)	0.490(13)	0.560(7)	0.500(4)	0.480(5)	0.640(6)	0.430(7)	0.410(3)	0.600(13)	0.570(7)	0.460(4)
CN2	9.00	0.471	0.420(3)	0.500(9)	0.510(13)	0.480(6)	0.460(11)	0.570(17)	0.430(6)	0.350(8)	0.640(6)	0.520(14)	0.450(6)
ASLA(LAB)	9.64	0.445	0.340(14)	0.540(1)	0.520(11)	0.430(12)	0.460(10)	0.650(3)	0.360(14)	0.330(10)	0.610(10)	0.540(11)	0.410(10)
CSK(OPP)	10.64	0.445	0.360(9)	0.510(6)	0.500(14)	0.430(13)	0.450(13)	0.620(8)	0.380(12)	0.330(11)	0.610(11)	0.550(9)	0.400(11)
SemiT(OPP)	11.27	0.446	0.340(13)	0.420(22)	0.590(5)	0.430(11)	0.460(8)	0.570(15)	0.400(8)	0.350(6)	0.570(18)	0.590(5)	0.390(13)
Struck	11.36	0.444	0.350(12)	0.470(16)	0.530(10)	0.450(10)	0.440(14)	0.580(12)	0.390(10)	0.300(17)	0.640(5)	0.540(10)	0.410(9)
VTD	12.91	0.430	0.310(19)	0.490(11)	0.540(8)	0.390(17)	0.460(9)	0.570(16)	0.370(13)	0.280(18)	0.630(7)	0.550(8)	0.380(16)
CT(HSV)	14.09	0.417	0.320(17)	0.510(5)	0.480(17)	0.400(15)	0.430(15)	0.590(10)	0.350(15)	0.270(21)	0.610(9)	0.490(17)	0.390(14)
MIL(LAB)	14.18	0.428	0.350(11)	0.470(15)	0.490(15)	0.380(19)	0.470(6)	0.520(27)	0.390(9)	0.310(13)	0.590(14)	0.480(19)	0.410(8)
OAB(LAB)	14.36	0.434	0.340(15)	0.440(21)	0.540(9)	0.460(9)	0.410(18)	0.540(23)	0.390(11)	0.310(16)	0.610(12)	0.530(12)	0.400(12)
LOT(OPP)	15.91	0.418	0.360(10)	0.440(20)	0.480(18)	0.430(14)	0.410(17)	0.590(11)	0.350(17)	0.310(15)	0.560(20)	0.510(15)	0.380(18)
ASLA	16.73	0.406	0.280(25)	0.500(8)	0.510(12)	0.350(25)	0.450(12)	0.600(9)	0.330(19)	0.270(22)	0.590(15)	0.480(20)	0.380(17)
Frag	17.64	0.412	0.390(8)	0.410(26)	0.440(21)	0.460(8)	0.370(24)	0.580(14)	0.350(16)	0.330(12)	0.520(23)	0.460(27)	0.390(15)
CSK	18.82	0.396	0.310(20)	0.460(17)	0.460(19)	0.390(18)	0.400(19)	0.550(18)	0.330(20)	0.280(19)	0.560(19)	0.490(18)	0.360(20)
FCT(OPP)	19.27	0.401	0.310(22)	0.520(4)	0.440(23)	0.380(20)	0.420(16)	0.550(22)	0.340(18)	0.270(25)	0.580(17)	0.470(26)	0.370(19)
OAB	22.45	0.381	0.310(21)	0.450(19)	0.430(25)	0.360(24)	0.400(20)	0.530(25)	0.320(21)	0.270(24)	0.540(22)	0.470(25)	0.350(21)
FCT	22.73	0.379	0.280(27)	0.490(12)	0.430(26)	0.370(23)	0.390(22)	0.550(21)	0.310(24)	0.230(33)	0.580(16)	0.480(22)	0.340(24)
CT	23.73	0.364	0.310(18)	0.470(14)	0.370(30)	0.390(16)	0.340(30)	0.490(29)	0.310(22)	0.230(30)	0.540(21)	0.420(29)	0.340(22)
TLD	23.73	0.359	0.290(23)	0.350(33)	0.440(20)	0.320(27)	0.380(23)	0.520(28)	0.300(25)	0.340(9)	0.390(31)	0.500(16)	0.310(26)
CPF(HSV)	23.91	0.355	0.280(26)	0.410(24)	0.410(28)	0.370(22)	0.350(29)	0.580(13)	0.270(28)	0.310(14)	0.420(29)	0.470(23)	0.310(27)
L1APG(OPP)	24.00	0.360	0.240(29)	0.370(31)	0.490(16)	0.320(30)	0.390(21)	0.550(20)	0.290(27)	0.280(20)	0.470(28)	0.520(13)	0.300(29)
IVT(OPP)	24.18	0.363	0.290(24)	0.400(28)	0.430(24)	0.350(26)	0.370(25)	0.550(19)	0.290(26)	0.270(23)	0.490(25)	0.480(21)	0.320(25)
MIL	25.55	0.355	0.320(16)	0.370(30)	0.380(29)	0.370(21)	0.350(28)	0.460(30)	0.310(23)	0.260(26)	0.490(24)	0.400(31)	0.340(23)
L_1APG	27.00	0.342	0.230(30)	0.420(23)	0.440(22)	0.320(29)	0.360(27)	0.530(26)	0.270(29)	0.240(29)	0.490(26)	0.430(28)	0.310(28)
IVT	28.27	0.334	0.220(33)	0.410(27)	0.420(27)	0.300(32)	0.360(26)	0.530(24)	0.260(30)	0.230(31)	0.480(27)	0.470(24)	0.290(30)
DFT(TRGB)	31.82	0.293	0.210(34)	0.390(29)	0.330(32)	0.310(31)	0.280(33)	0.460(31)	0.230(32)	0.220(34)	0.400(30)	0.320(33)	0.280(31)
LOT	31.91	0.296	0.250(28)	0.360(32)	0.310(34)	0.320(28)	0.280(34)	0.450(32)	0.230(33)	0.230(32)	0.380(33)	0.370(32)	0.270(33)
SemiT	32.18	0.283	0.220(31)	0.330(34)	0.330(31)	0.240(35)	0.320(31)	0.380(35)	0.240(31)	0.250(27)	0.330(34)	0.420(30)	0.230(35)
DFT	32.73	0.281	0.190(35)	0.410(25)	0.320(33)	0.260(34)	0.290(32)	0.440(33)	0.220(34)	0.200(35)	0.390(32)	0.300(35)	0.270(32)
CPF	33.64	0.260	0.220(32)	0.280(35)	0.300(35)	0.270(33)	0.250(35)	0.430(34)	0.190(35)	0.240(28)	0.290(35)	0.310(34)	0.240(34)

From the results, we can see that all the evaluated trackers benefit from integrating color information on PTB, which is consistent with the observation on TColor-128. MEEM, when enhanced by LAB, achieves the best performance in terms of the average rank, which is expected given MEEM's outstanding performances on the CVPR2013 benchmark [6] and TColor-128. A somewhat surprising observation is the excellent performance by Frag (enhanced by HSV). An explanation is that more than half of the sequences in PTB involve occlusion, and Frag is specifically developed to handle occlusion by using multiple image patches to represent the tracking target. Compared with results on TColor-128, the performance gain on PTB has larger variances. In particular, SemiT benefits the most from using color, with 0.163 increase in success rate (over 60%); by contrast, the performance gain for DFT is only 0.012.

VI. Conclusion and Future work

In this paper, we present a study on using color information for visual tracking. On one hand, a new large tracking benchmark is constructed containing 128 color sequences associated with annotations. On the other hand, 16 state-of-the-art visual trackers have been enhanced with 10 different color models for evaluation. As the first comprehensive color tracking benchmark, our study systematically demonstrates the effectiveness of encoding color information for visual tracking. In particular, our results show that, when appropriate color models are used, the performances of existing grayscale trackers are consistently improved. Some trackers even outperform recently proposed color trackers.

More detailed analysis has been conducted for deeper understanding. First, it has been shown that different trackers are in favor of different color models. That said, Opponent, HSV, LAB, and RGB are in general very helpful for boosting tracking performance, though none of them possesses strong invariance properties. Second, it has been observed that color information is particularly helpful for addressing challenges due to deformation, in plane rotation, out of plane rotation and illumination variation. By contrast, out of view, occlusion, fast motion, motion blur and low resolution remain to be main challenges for visual tracking.

A limitation of encoding color information in our straightforward framework is the sacrifice of efficiency. It would be interesting to investigate how to make use of color information in a more efficient way, for example, utilizing dimensionality reduction technique as [2]. Also, we focus on general effects of integrating color in tracking, no tracker-specific strategy has been developed for enhancing grayscale trackers. It would therefore be attractive to explore how to encode color information in a more tracker-aware way, e.g., allowing decisions from different color channels to collaborate with each other as [31]. Given the promising results delivered in this paper and by sharing the resources in our study, we expect our study to provide the guidance, motivation and benchmark for future work on encoding color in visual tracking.

Acknowledgment

The authors would like to thank Semir Elezovikj for proof reading the manuscript, and anonymous reviewers for valuable comments and suggestions.

References

[1] S. Hare, A. Saffari, and P. H. Torr, "Struck: Structured output tracking with kernels," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 263–270, 2011.

[2] M. Danelljan, F. Shahbaz Khan, M. Felsberg, and J. Van de Weijer, "Adaptive color attributes for real-time visual tracking," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014.

[3] K. E. Van De Sande, T. Gevers, and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, 2010.

[4] G. J. Burghouts and J.-M. Geusebroek, "Performance evaluation of local colour invariants," *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 48–62, 2009.

[5] I. Everts, J. van Gemert, and T. Gevers, "Evaluation of color stips for human action recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2850–2857, 2013.

[6] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2411–2418, 2013.

[7] Y. Pang and H. Ling, "Finding the best from the second bests-inhibiting subjective bias in evaluation of visual tracking algorithms," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 2784–2791, 2013.

[8] S. Song and J. Xiao, "Tracking revisited using rgbd camera: Unified benchmark and baselines," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 233–240, 2013.

[9] X. Jia, H. Lu, and M.-H. Yang, "Visual tracking via adaptive structural local sparse appearance model," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1822–1829, 2012.

[10] S. Stalder, H. Grabner, and L. Van Gool, "Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition," *Proc. Workshop Online Learning in Computer Vision*, 2009.

[11] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," *Proc. European Conf. Computer Vision*, pp. 661–675, 2002.

[12] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," *Proc. European Conf. Computer Vision*, pp. 702–715, 2012.

[13] K. Zhang, L. Zhang, and M.-H. Yang, "Real-time compressive tracking," *Proc. European Conf. Computer Vision*, pp. 864–877, 2012.

[14] T. B. Dinh, N. Vo, and G. Medioni, "Context tracker: Exploring supporters and distracters in unconstrained environments," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1177–1184, 2011.

[15] L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1910–1917, 2012.

[16] K. Zhang, L. Zhang, and M. Yang, "Fast compressive tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2002–2015, 2014.

[17] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 798–805, 2006.

[18] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int'l J. Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008.

[19] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, 2015.

[20] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust l1 tracker using accelerated proximal gradient approach," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1830–1837, 2012.

[21] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2259–2272, 2011.

[22] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1940–1947, 2012.

[23] B. Liu, J. Huang, L. Yang, and C. Kulikowsk, "Robust tracking using local sparse appearance model and k-selection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1313–1320, 2011.

[24] J. Zhang, S. Ma, and S. Sclaroff, "MEEM: robust tracking via multiple experts using entropy minimization," *Proc. European Conf. Computer Vision (2014)*, 2014.

[25] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2042–2049, 2012.

[26] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting." *Proc. Conf. British Machine Vision*, pp. 47–56, 2006.

[27] R. T. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1631–1643, 2005.

[28] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1838–1845, 2012.

[29] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," *Proc. European Conf. Computer Vision*, pp. 234–247, 2008.

[30] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, 2012.

[31] J. Kwon and K. M. Lee, "Visual tracking decomposition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1269–1276, 2010.

[32] J. Kwon and K. M. Lee, "Tracking by sampling trackers," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 1195–1202, 2011.

[33] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, 2006.

[34] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–577, 2003.

[35] H.-T. Chen and T.-L. Liu, "Trust-region methods for real-time tracking," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 717–722, 2001.

[36] J. Van De Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Trans. Image Processing*, vol. 18, no. 7, pp. 1512–1523, 2009.

[37] M. Yang, Y. Wu, and G. Hua, "Context-aware visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, 2009.

[38] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, F. Porikli, A. Khajenezhad, A. Salahledin, A. Soltani-Farani, A. Zarezade, A. Petrosino *et al.*, "The visual object tracking vot2013 challenge results," *IEEE Workshop on visual object tracking challenge*, 2013.

[39] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: an experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, 2014.

[40] R. Collins, X. Zhou, and S. K. Teh, "An open source tracking testbed and evaluation web site," *IEEE Int'l Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 17–24, 2005.

[41] F. S. Khan, J. van de Weijer, and M. Vanrell, "Modulating shape features by color attention for object recognition," *Int'l J. Computer Vision*, vol. 98, no. 1, pp. 49–64, 2012.

[42] J. Van De Weijer and C. Schmid, "Coloring local feature extraction," *Proc. European Conf. Computer Vision*, pp. 334–348, 2006.

[43] F. S. Khan, R. M. Anwer, J. van de Weijer, A. D. Bagdanov, A. M. Lopez, and M. Felsberg, "Coloring action recognition in still images," *Int'l J. Computer Vision*, vol. 105, no. 3, pp. 205–221, 2013.

[44] F. Shahbaz Khan, R. M. Anwer, J. van de Weijer, A. D. Bagdanov, M. Vanrell, and A. M. Lopez, "Color attributes for object detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3306–3313, 2012.

[45] J. von Kries, "Influence of adaptation on the effects produced by luminous stimuli," *Sources of color vision*, pp. 109–119, 1970.

[46] S. A. Shafer, "Using color to separate reflection components," *Color Research & Application*, vol. 10, no. 4, pp. 210–218, 1985.

[47] Z. Hong, C. Wang, X. Mei, D. Prokhorov, and D. Tao, "Tracking using multilevel quantizations," *Proc. European Conf. Computer Vision*, 2014.

[48] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," *Proc. European Conf. Computer Vision*, 2014.

[49] C. Bailer, A. Pagani, and D. Stricker, "A superior tracking approach: Building a strong tracker through fusion," *Proc. European Conf. Computer Vision*, 2014.

[50] H. Nam, S. Hong, and B. Han, "Online graph-based tracking," *Proc. European Conf. Computer Vision*, 2014.

[51] Y. Lu, T. Wu, and S.-C. Zhu, "Online object tracking, learning, and parsing with and-or graphs," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014.

[52] N. Wang, S. Li, A. Gupta, and D.-Y. Yeung, "Transferring rich feature hierarchies for robust visual tracking," *arXiv preprint arXiv:1501.04587*, 2015.

[53] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking," *Int'l J. Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[54] F. Porikli, "Integral histogram: A fast way to extract histograms in cartesian spaces," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 829–836, 2005.

[55] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1619–1632, 2011.

[56] P. Viola and M. J. Jones, "Robust real-time face detection," *Int'l J. Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.