Expanding histogram of colors with gridding to improve tracking accuracy

Article · July 2015

DOI: 10.1109/MVA.2015.7153234

Kourosh Meshghi
RIKEN Center for Advanced Intelligence Project (AIP)

MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

Expanding Histogram of Colors with Gridding to Improve Tracking Accuracy

Kourosh MESHGI, Shin ISHII
Graduate School of Informatics
Kyoto University, Kyoto, Japan
meshgi-k@sys.i.kyoto-u.ac.jp, ishii@i.kyoto-u.ac.jp

Abstract

Using color information in object tracking is a prudent choice, but the vast variety of choices and difficulties of obtaining a desirable stable result, unnerves many scholars. Color histograms, as a compact and robust representation is the center of attention while it suffers from lack of spatial information about colors. Besides, comparison and updating such histograms in a meaningful and efficient manner is challenging. In this paper, we proposed the idea of gridding for color histogram, which grants specific statistical property to the histogram through a decomposition phase followed by a recombination stage. Additionally, a thorough comparison of the modern similarity functions and model update techniques in RGB colorspace is presented. This comparison reveals that our proposed method in combination with established similarity measures, enhances the tracking performance.

1 Introduction

Histogram-based local descriptors are pervasive tools in many computer vision tasks, such as image retrieval, object tracking, and object recognition. Such descriptors compress the original distribution, which save storage and computation time and grant a certain perceptual robustness to the matching [1]. Being both compact and invariant to specific changes in the image, color histograms gained popularity in object tracking, e.g. the famous mean-shift tracker [2] and color-based particle filter is built upon this feature.

Histogram of color (hereafter **HOC**), however, suffers from lack of spatial information and cannot differentiate patterns of colors and just considers the portion of them in an image. For that, several other cues are used along with them in systems to improve tracking accuracy such as texture, motion, shape etc. Furthermore, the quantization effect and perceptual difference of colors inaugurate more challenges in using this feature. Nevertheless, HOC is usually in charge of making most out of the color information. To alleviate the shortcomings of this feature, researchers use various color spaces with special characteristics, employ different similarity measures to compare color histograms, and try to embed necessary spatial information into the feature whenever possible.

This article focuses on RGB color histograms and scrutinizes the use of gridding as a simple mathematical trick to boost the performance of the feature while preserving all of the advantages of a HOC. Presenting a gist of numerous ways in which HOC is used in RGB space, the manuscript then compare these ideas and enhance them by proposed scheme, to show the effectiveness of the gridding idea.

2 Literature Review

2.1 Histogramming

Histograms, by default are a partition of the space into equal bins. Applied on RGB data, each channel is partitioned into same-size bins. The result is a regular histogram of color or R-HOC in which n governs the coarseness of the bins. Being simple and interpretable, this type of HOC is the very popular. However, regular binning for high dimensional data often yields poor performance: fine binning leads to fluctuations due to statistically insignificant sample sizes for most bins, while coarse binning diminishes resolving power [1].

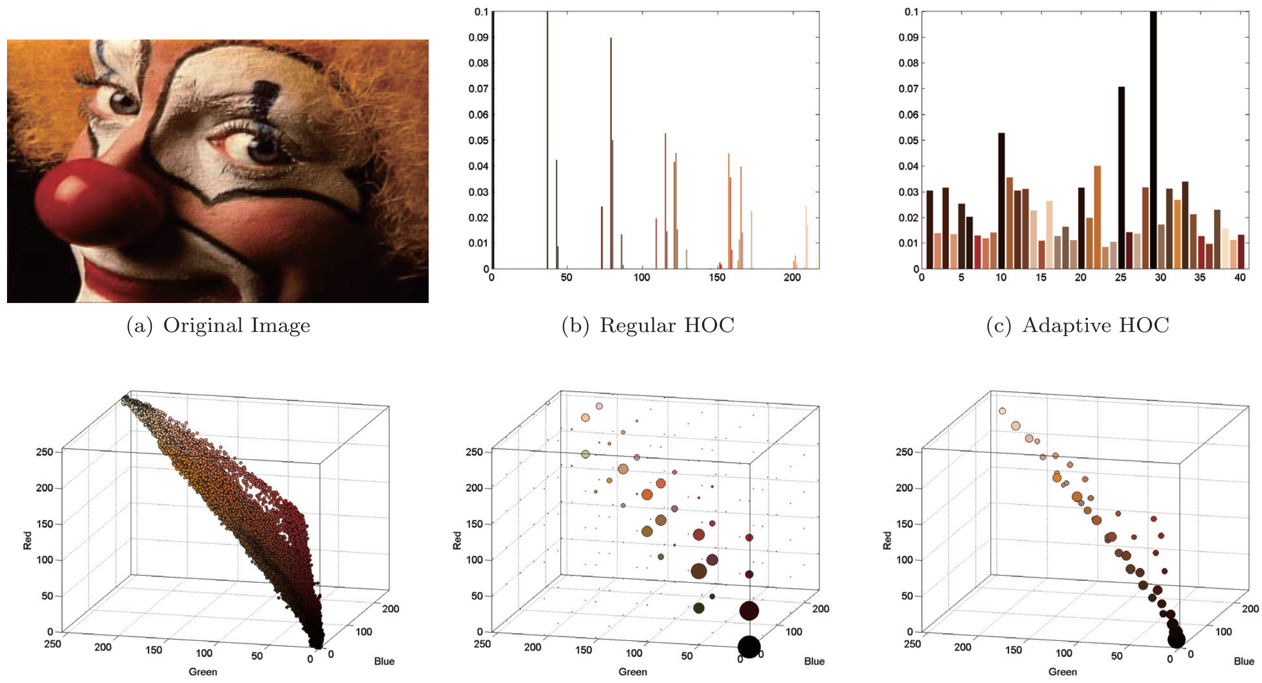
A solution to this issue is to employ adaptive binning whereby the bins are adapted to the color distribution. Clustering methods such as k-means provides such binning, which are then used to make an adaptive histogram of colors (or A-HOC). The number of bins in this method is parameterized by color clusters. Figure 1 illustrates the difference between regular and adaptive binning to construct color histograms.

2.2 Similarity Functions

In order to match histograms, a similarity function must be coupled with the color feature. Being an empirical estimate of a probability distribution of feature, histograms can be treated similar to distributions, thus a similarity function, enjoys a wealth of techniques to compare two distributions. Following [1], we distinguish four categories for dissimilarity measures.

First category involves heuristic histogram distances that are mostly proposed in the image retrieval context, but is also popular in other computer vision tasks. The Minkowski distance family (L_p) is arguably the most used heuristic for dissimilarity. L_1 computes the sum of absolute dissimilarity, L_2 penalizes the larger errors more severely, and L_∞ as the limiting case measures the maximal difference. Having the normalized histograms in hand, Cosine distance (CO) reduces to inner product of two histograms and indicates the distance between them as the cosine of angle in between. The Pearson correlation coefficient (CR) indicates the linear dependence between random variables.

The second category is derived from non-parametric test statistics, which examine the hypothesis that two empirical distributions are generated from the same true distribution. The Kolmogorov-Smirnov distance (KS) is defined as maximal discrepancy between two cumulative distributions, Match distance (MA) equals the sum of absolute distance between their cumulative distances. Cramer-von Mises statistics (CM) penalize the discrepancy of two cumulative histograms quadratically as it sums them. The measures based on



(a) Original Image (b) Regular HOC (c) Adaptive HOC

(d) Color Point Cloud (e) Regular Binning ($6 \times 6 \times 6$ bins) (f) Adaptive Binning (40 bins)

Figure 1. Regular and adaptive binning to construct HOCs. While the former, assign color points to predefined bins, the latter cluster colors to find the best bin centers for histogramming. The Regular HOC, is sparse and is sensitive to quantization level of color channels. The Adaptive HOC employs a more compact representation of the color distribution in the image, and is more intuitive. In this figure the color histograms are normalized, and the scale of color distributions are altered for better visualization.

cumulative histograms are all sensitive to bin ordering. The χ^2 statistics (CS) is adopted from Pearson's chi-sqaure test statistics and was introduced with an asymmetric formulation in literature. Later a slight modification addressed the issue and improved the performance of the metric. Zweng et al. [3] reported that a bounded version of this function has best performance in a tracking task in several color spaces. However, this function is sensitive to quantization effect and shape deformation [4]. Another distance in this category is Bhattacharyya distance (BH), which uses Bhattacharyya coefficient to measure the dissimilarity of two distributions.

Third category of distances is inspired from information-theoretic divergences. The Kullback-Leibler divergence (KL) measures "how inefficient on average it would be to code one histogram using the other as the true distribution for coding". This measure is not bounded and is sensitive to empty bins in the histogram. A symmetric and numerically stable version of KL-divergence is presented in Jeffrey Divergence (JD).

The final category is composed of distance measures empowered with cross-bin information. In applying the bin-to-bin functions presented in the last three categories, it is often assumed that the domain of the histograms is aligned. However, in practice, such an assumption can be violated due to various factors, such as shape deformation, non-linear lighting change, and heavy noise. The ground distance, is defined as the amount of difference between elements of the feature vector. With the aggregation of all ground distances in a positive-definite symmetric similarity matrix $\mathbf{M}$, the weighted L_2 distance will be transformed into Quadratic distance for color histograms (QD). An extension to Quadratic distance, introduced in [4], reduces the effect of large bins. In this hybrid, Quadratic-Chi distance (QC), the normalization power of chi-square is utilized along with cross-bin relationship presented by quadratic distance. Another member of this category is Diffusion distance (DF), which models the problem as a temperature field and considers the diffusion process on it [5]. Earth Movers Distance [6] is another family member that models the histogram distance with different bin sizes and bin centroids as a transportation problem (EM). Many approaches try to improve the speed of this method. For example $EMD - L_1$ exploits the structure of L_1 distance and $\widehat{EMD}$ [7] used a graph theoretic approach with thresholded ground distance to linearize the computational complexity.

2.3 Model Update

The tracking task, involves lots of dynamics in the subject ranging from rigid and non-rigid transformations to abrupt changes in pose, motion, illumination, camera parameters, temporal clutter, noise and occlusion. Robust trackers compensate for these changes by updating their model across the time, to keep the model close to the designated subject. Several model update (**MU**) approaches are popular in the literature.

The simplest approach is to average all of the observations from the beginning of the sequence. Although updates the model for new changes, this method lacks enough flexibility to accommodate drastic changes in the model. In the second method a leaky memory scheme is employed, i.e., all of the observations contribute to the template model based on their recentness. The result of this approach is over-smoothed and may lose the clarity after a while. The third approach uses an extra buffer to update the model. Essentially an extension of the leaky memory model, this method updates the model with small forgetting rate for every observation. Then after a certain number of intervals, the model is updated with a larger forgetting factor that helps the removal of the remainders of initial observations if not useful anymore. Finally, the forth approach is to use only the recent observations. For this a queue of a fixed size is utilized and each new observation is enqueued in it. The template in each frame is the average of all models in the queue.

3 Gridding for HOC: Divide and Conquer

This section elaborates the proposed feature.

3.1 Challenges for Color Histogram

It has been stated in the literature that tracking using region based descriptors take two extremes: blob-based trackers and pixel-wise template trackers. Generally a blob-based tracker accumulates less information about the target, since it discards all its spatial information. A middle level to this spectrum has been established by the use of kernel histogram methods. While the initial method based on mean shift formulation [2] share more similarities to blob trackers, some modifications (such as those in [8]) drift them more toward the template tracker end of spectrum.

However, even with adaptive binning based on the overall distribution of all images in a tracking sequence, often for specific patches of image only a small fraction of the bins in a histogram contain significant information. Fine quantization for the histogram as a solution is highly inefficient. On the other hand for an image characterized by many color details, a coarse binning for histogram is not adequate. Traditional histograms as fixed-size structures can hardly balance efficiency and expressiveness. Several methods have been proposed to enhance the color representation, including variable-size histograms called signatures [1], embedding color spatial relations into color corellograms [9] and expansion of histogram with higher order spatial moment called spatiograms [10]. These approaches are changing the nature of color histograms thus reducing its robustness to rotation, deformations and illumination changes while improving the performance in the normal condition.

3.2 Proposed Method

Here we introduce the idea of gridding, which while preserving the very nature of color histograms, addresses the above-mentioned issues. The intuition behind this method is that with breaking a big block of data into smaller chunks, processing them and aggregating them again, the data become more tractable and is augmented with the underlying process. The agglomeration scheme in turn grants desired properties to the result. Specifically in this case, we break the input image, into spatially-regular grid segments, calculate the histogram of color to each of sub-images and combine obtained histograms to make one final

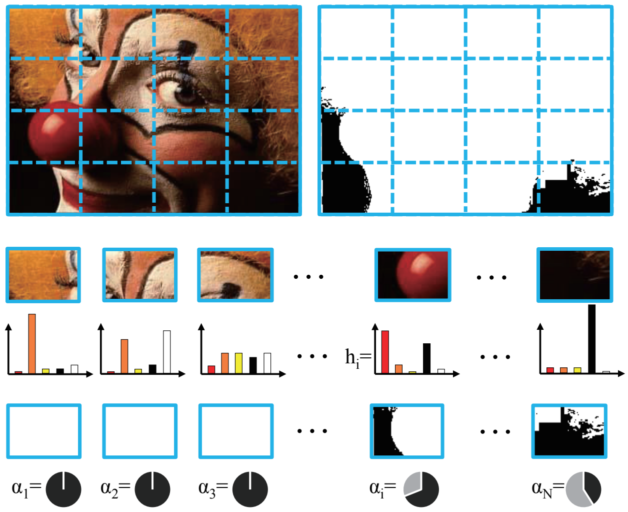


Figure 2. Gridding to enhance HOC. The Image and its background mask (if available), is divided into spatially-regular grid. Then, a HOC is constructed for each cell, and the weight of cell (i.e., ratio of foreground to all pixels) is calculated. Finally, these histograms are agglomerated to produce a HOC with the similar length of each of cell HOCs.

histogram. We propose two agglomeration techniques: *(i)* averaging in which the final HOC is smoothed to avoid a few bins overload the total histogram if placed only in a part of image; *(ii)* linear combination in which each grid cell is analyzed and given an importance coefficient and the resulting histogram is the linear combination of grid histograms weighted by these coefficients. We call the former method grid-driven average HOC (G_a) which requires no additional information. In case of the latter, grid-driven weighted HOC (G_w), we define the importance of the grid cell proportional as the ratio of foreground pixels to all pixels in the grid cell. This requires background subtraction, which is not feasible in many tasks, and the importance criteria can be arbitrary function which assigns a weight to the grid cells (Figure 2). The procedure of constructing G_a and G_w is elaborated in Algorithm 1. The binning strategy is arbitrary in line 4 of this algorithm.

It is important to contrast our method with spatial pyramid matching [11]. This method performs pyramid matching [12] in image space and use clustering in feature space. It concatenates the results of different levels of pyramid, which yield a large sparse feature vector. This holistic method is suitable for scene classification. It is not suitable for tracking task in which background clutter challenges the tracker and object transformations should be handled. Also exponentially shrinking cells of this method, increase the chance of outlier problem in the histogram. In contrast, our proposed feature manages object transformations by averaging and accounts for background.

4 Evaluation

In this paper we investigate the performance of different similarity measures under various histogramming paradigms, and show the efficiency of the proposed gridding method to enhance almost all of them. Another experiment is dedicated to examine the com-

Algorithm 1: Constructing Grid-driven HOC

input : Gridding granularity n, Image Patch I
output: Grid-driven HOCs G_a and G_w

$B \leftarrow$ *Calculate Image Background Mask*
$I_{i,j} \leftarrow$ *Divide* I *into* $n \times n$ *cells*
$B_{i,j} \leftarrow$ *Divide* B *into* $n \times n$ *cells*
$C \leftarrow$ *Calculate Histogram Centers*
$\alpha_{i,j} \leftarrow$ *Calculate ratio of foreground pixels to all in* $B_{i,j}$
$h_{i,j} \leftarrow$ *Calculate HOC for* $I_{i,j}$ *using* C
$G_a \leftarrow \frac{1}{n^2} \sum_{(i,j)=(1,1)}^{(n,n)} h_{i,j}$
$G_w \leftarrow \frac{1}{n^2} \sum_{(i,j)=(1,1)}^{(n,n)} \alpha_{i,j} h_{i,j}$

patibility of histogramming method with similarity measure and model update. All of the sequences are in RGB color space and for all experiment related to a histogramming schemes, the bins are preserved the same. Following the style of [3], we used 50 consecutive frames from `S2.L1.12-34.View001` sequence of PETS 2009 dataset to simulate a tracking task. The three pedestrians present in the sequence are represented with a bounding box, manually pre-segmented and a background mask is prepared to block the effect of other factors on the evaluations [3].

Table 1. Combinations of $\langle HOC, DIST \rangle$ for tracking robustness, row best is in **bold**, column best is underlined, and best value is shaded (A: adaptive, R: regular))

HOC	Similarity Measures (%)									
	L2	**CR**	**CS**	**BH**	**KL**	**DF**	**KS**	**CM**	**QC**	**EM**
R	69.7	62.6	70.4	72.1	**86.2**	71.0	69.6	48.8	64.2	67.8
A	72.6	72.0	74.5	72.9	**88.2**	72.4	66.8	54.5	64.0	66.4
R $G_a(2\times2)$	70.3	63.5	71.0	72.4	**86.4**	71.4	70.1	49.8	65.9	68.0
A $G_a(2\times2)$	73.0	72.9	74.5	73.1	**88.9**	73.3	71.1	54.5	65.6	66.6
R $G_a(3\times3)$	70.0	63.3	71.0	71.9	**86.2**	70.9	69.9	49.6	65.7	67.5
A $G_a(3\times3)$	72.7	72.0	75.4	72.5	**90.4**	73.4	69.9	51.9	65.3	66.2
R $G_a(5\times5)$	69.5	62.8	70.9	71.7	**86.0**	70.7	69.3	48.4	64.9	66.9
A $G_a(5\times5)$	72.9	72.1	73.4	71.8	**88.7**	72.4	68.4	54.2	64.6	66.4
R $G_w(2\times2)$	69.9	62.7	70.5	70.9	**86.5**	69.8	69.3	51.1	63.8	65.0
A $G_w(2\times2)$	72.6	73.2	75.4	72.1	**88.8**	71.9	64.9	40.3	64.0	63.7
R $G_w(3\times3)$	70.0	63.3	71.0	71.9	**86.2**	70.9	69.9	49.6	65.7	67.5
A $G_w(3\times3)$	73.6	74.6	76.2	73.0	**90.8**	73.6	72.7	60.9	65.4	66.2
R $G_w(5\times5)$	69.5	62.8	70.9	71.7	**86.0**	70.7	69.3	48.4	64.9	66.9
A $G_w(5\times5)$	72.8	72.1	74.5	72.5	**88.9**	73.4	69.4	61.1	64.6	66.1

4.1 Experiment I

In the first experiment, different combinations of histogramming method and similarity measures are examined on the sequence. The similarity of each person in each frame to that of last frame, *intra-similarity*, is calculated throughout all of the sequence. Furthermore, the mutual similarity of each person to other subjects in each frame, the *inter-similarity*, is calculated for each frame as well. The inter- and intra-similarity values are obtained from $a_{ij} = 1 - (d_{ij} \div d_{max})$ in which d_{ij} is the distance between bounding boxes i and j, and d_{max} is the maximum of all distances measured for this dissimilarity function (e.g. CS). Intuitively, a similarity function should maximize the intra-similarity, while minimizing inter-similarity. To facilitate comparison between different combinations of $\langle HOC, DIST \rangle$, we introduce a score $S = \sqrt{\alpha(1-\beta)}$ in which α is the average of all intra similarity values and β is the average of all inter-similarity values. Higher S values (values closer to one), indicates a better combination in distinguishing target and non-target bounding boxes.

Table 1 gathers a comprehensive list of all combinations of histogramming methods and similarity measure combinations. [1]In this table, two types of HOC, Regular with 5×5×5 bins and Adaptive with 40 bins, two types of gridding with different grid sizes ($G_a n\times n$ and $G_w n\times n$), and 16 similarity measures are compared. For the cross-bin similarities, a matrix $\mathbf{M}$ is constructed for both regular and adaptive histogramming methods based on CIEDE2000 color distance and is shared between all measures. The project code can be found at https://github.com/meshgi/Histogram_of_Color_Advancements and EM distance is based on $\widehat{EMD}$ [7]. From Table 1 it is apparent that KL-divergence provides the best score for tracking which means the object has higher intra-similarity throughout the sequence, while the rest of the objects have large distance to that. It is also evident that gridding-enhanced histograms have better performance than the normal condition. In this particular example the 3×3 grid-driven weighted color histogram outperforms other schemes for most of the possible similarity measures, especially with KL-divergence that achieves the best performance of $S = 90.83\%$.

Many other patterns are evident in this table, interesting for further investigation. For instance, it is apparent that the performance of the grid-based average HOC along with cross-bin similarity measures is better than the other distances. On the other hand the statistic test distances get along with weighted version of gridding. This is intuitive as the weighted gridding emphasizes the statistics in each sub-frame proportional to their significance in the total result. On the other hand, cross-bin measures emphasize on the distance of the colors and the amount of correspondent bin-to-bin difference which is smoothed by the process of averaging. Interestingly, QC as a bridge between these two families bends in favor of averaging. Moreover, there is an optimized grid size for the sequence in which most of the similarity measures showed the best performance. This size is object-specific and in our case, for the pedestrians a 3×3 grid works well (left/middle/right + head/torso/legs) despite the out-of-plane rotation of the subjects in the sequence.

In addition to our proposed features, we examined the performance of pyramid matching [12] and spatial pyramid matching [11] in this experiment. The former achieved $S = 79.7\%$ while the latter, which is designed as a holistic feature for scene understanding, obtained only $S = 51.3\%$.

4.2 Experiment II

In a strive to find the best $\langle HOC, DIST, MU \rangle$ combination, we attempt another experiment in which all new bounding boxes are matched to a template. This

[1]Due to space limitation only 10 of the measures are presented, the rest are available in the author webpage: http://ishiilab.jp/member/meshgi-k/r-tracking.html

experiment is particularly useful to evaluate the performance of this combination in top-down trackers. In this experiment the target in the first frame acts as an initialization for the template, which is updated in the successive frames using surveyed methods. All of the similarity values for each tuple of $\langle HOC, DIST, MU \rangle$ are averaged to make a final score. In this implementation, the forgetting factor for leaky memory is set to 0.1, the queue size for storing the latest templates are set to 5, and the forgetting factor for the second layer of buffer is set to 0.4. Table 2 contains the best result of model update schemes.

Interestingly, all of similarity measures, except CO, took their maximum value from last 5 frame MU scheme. This indicates the role of incorporating recent changes into the model. It also suggests that higher forgetting rates for the leaky memory schemes may be beneficial in this scenario. It is also noteworthy that all of the test cases prefer a model update to no update case, which clarifies the role of model update in dynamics of tracking. In Table 2 it is clear that CM has the best template matching performance, but as it is inferred from Table 1 it does not have adequate inter-object similarity rejection and thus is not a good candidate for the histogram-based template matching. KL-divergence, on the other hand, has on average 4% less template matching performance under model update case. Yet, it enjoys an exponential punishment method for other objects, which makes it as an excellent choice among all the others.

Table 2. Combinations of $\langle HOC, DIST, MU \rangle$ for best template matching performance, row best is in **bold**, column best is underlined, and best value is shaded (A: adaptive, R: regular))

HOC	Similarity Measures (%)									
	L2	CR	CS	BH	KL	DF	KS	CM	QC	EM
R	86.0	95.5	93.4	82.8	93.2	84.5	86.4	**96.8**	90.8	80.4
A	86.5	94.5	92.9	83.1	92.8	83.7	88.8	**96.4**	90.4	80.5
R $G_a(2\times2)$	85.9	95.4	93.2	82.7	93.1	84.3	86.3	**96.7**	90.6	80.4
A $G_a(2\times2)$	86.0	94.4	93.0	83.2	92.8	83.8	84.1	**95.2**	90.2	80.7
R $G_a(3\times3)$	86.0	95.4	93.2	82.8	93.1	84.5	86.4	**96.8**	90.7	80.6
A $G_a(3\times3)$	86.7	94.4	92.6	82.9	92.4	83.5	88.6	**96.8**	90.3	80.2
R $G_a(5\times5)$	86.1	95.5	93.2	82.8	93.0	84.5	86.7	**96.8**	90.9	80.7
A $G_a(5\times5)$	87.1	94.6	93.0	83.1	92.8	84.1	88.0	**96.9**	90.3	80.4
R $G_w(2\times2)$	84.1	95.0	92.5	81.8	92.3	83.4	84.4	**96.2**	89.8	80.6
A $G_w(2\times2)$	84.0	93.4	91.9	82.2	91.7	82.6	85.8	**96.2**	89.4	81.2
R $G_w(3\times3)$	86.0	95.4	93.2	82.8	93.1	84.5	86.4	**96.8**	90.7	80.6
A $G_w(3\times3)$	85.6	94.3	92.9	83.6	92.8	83.4	88.4	**96.8**	90.4	80.8
R $G_w(5\times5)$	86.1	95.5	93.2	82.8	93.0	84.5	86.7	**96.8**	90.9	80.7
A $G_w(5\times5)$	86.5	94.4	92.9	83.3	92.8	83.6	88.5	**96.1**	90.5	80.6

5 Conclusion

In this paper, A new tool to improve visual tracking accuracy using color cues is introduced. The idea involves decomposing a bounding box into a regular grid, in order to incorporate more spatial information and embed local salient color details into the histogram. Two gridding schemes are also proposed to accommodate different requirements, when the background data is/is not available. Then a comprehensive study over the existing (dis)similarity measures is conducted and many ideas from other computer vision realms such as image retrieval and texture analysis are brought into object tracking.

Using two simulation experiments, we first study the performance of (dis)similarity functions in finding relevant target using color histogram, while discarding other objects to successfully track the designated object. The results signify the usefulness of gridding to improve the tracking performance. It showed that almost any combination of histograms and similarity measures are boosted by gridding. Based on the results it is advisable to use grid-based weighted HOC when the background data is available, but the other version is also superior to normal HOC in most of the cases. To employ these features in other trackers, the degree of gridding, n can be calculated using cross-validation.

In the second experiment we emulate a tracking scenario and illustrate the compatibility of different template update schemes for various dissimilarity measures. It has been found that while all model updates surveyed in this paper improves the performance of template matching, none of them is completely suitable for proposed gridding schemes. Also the results revealed that, calculating template as the average of the last 5 observations, outperformed other methods.

In future, we plan to extend the scope of study to other color spaces, irregular grids, optimal griddings, and uncontrolled tracking scenarios.

Acknowledgment

The authors wish to thank Andreas Zweng for providing the annotated dataset.

References

[1] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," *CVIU*, 2001.

[2] D. Comaniciu and V. Ramesh, "Mean shift and optimal prediction for efficient object tracking," in *ICIP'00*, vol. 3. IEEE, 2000, pp. 70–73.

[3] A. Zweng, T. Rittler, and M. Kampel, "Evaluation of histogram-based similarity functions for different color spaces," in *CAIP'11*. Springer, 2011, pp. 455–462.

[4] O. Pele and M. Werman, "The quadratic-chi histogram distance family," in *ECCV'10*. Springer, 2010.

[5] H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *CVPR'06*. IEEE, 2006.

[6] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *IJCV*, vol. 40, no. 2, pp. 99–121, 2000.

[7] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *ICCV'09*. IEEE, 2009, pp. 460–467.

[8] G. D. Hager, M. Dewan, and C. V. Stewart, "Multiple kernel tracking with ssd," in *CVPR'04*. IEEE, 2004.

[9] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *CVPR'09*. IEEE, 1997, pp. 762–768.

[10] S. T. Birchfield and S. Rangarajan, "Spatiograms versus histograms for region-based tracking," in *CVPR'05*, vol. 2. IEEE, 2005, pp. 1158–1163.

[11] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR'06*, vol. 2, 2006.

[12] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *ICCV'05*, vol. 2, 2005, pp. 1458–1465.