

Improved Mean Shift Target Localization using True Background Weighted Histogram and Geometric Centroid Adjustment

Rashid MEHMOOD¹, Noor ul HUDA², Jinho SONG³, M. Moshin RIAZ², Naveed IQBAL¹,
Tae Sun CHOI³

¹National University of Science and Technology (NUST), Islamabad, Pakistan

²Center of Advanced Studies in Telecommunication, COMSATS, Islamabad, Pakistan

³Gwangju Institute of Science and Technology (GIST), Oryong-Dong, Buk-Gu, Gwangju 500-712, Republic of Korea

rashidmehmood@mcs.edu.pk, noorulhuda@comsats.edu.pk, sam9157@gist.ac.kr, mohsin.riaz@comsats.edu.pk,
naveedi@mcs.edu.pk, tschoi@gist.ac.kr

Manuscript received March 02, 2016

Abstract. *Mean Shift (MS) tracking using histogram features alone may cause inaccuracy in target localization. The problem becomes worst due to presence of mingled background features in target model representation. To improve MS target localization problem, this paper propose a spatio-spectral technique. The true background features are identified in target model representation using spectral and spatial weighting and then a transformation is applied to minimize their effect in target model representation for localization improvement. The target localization is further improved by adjusting the MS estimated target position through edge based centroid re positioning. The paper also propose method of target model update for background weighted histogram based algorithms followed by weighted transformation through online feature consistency data. The proposed method is designed for single object tracking in complex scenarios and tested for comparative results with existing state of the art techniques. Experimental results on numerous challenging video sequences verify the significance of proposed technique in terms of robustness to complex background, occlusions, appearance changes, and similar color object avoidance.*

Keywords

Target tracking, spatio-spectral technique, Mean Shift, guided filter, target localization

1. Introduction

Object tracking is an important and challenging aspect in computer vision applications. Tracking provides a continuous estimated trace of the moving object in camera's field of view. A good tracking algorithm should be computationally

efficient and perform well in controlled as well as dynamic environment. Summary of different tracking techniques is provided in [1–3]. Mean shift (MS) [4] is a commonly used target tracking technique due to its ease of implementation and real time response. MS is a deterministic iterative procedure for locating the maxima of a density function given discrete data samples [8]. In tracking applications it compares a target model density function with current frame to find out the most promising converging region [4]. However, in complex scenarios (i.e. same background and target features) it fails to represent the best nonparametric density estimate [3], [4], [6] hence causes error in target localization. In MS tracking, localization means to find a best target candidate by maximizing a similarity function. Due to irregular target shapes, the rectangular or elliptical search window also includes some background features that introduces the local maxima and localization error. Furthermore, the MS algorithm sometimes fail to properly estimate target center (due to presence of local maxima), resulting in incorrect initialization for the next frame hence, this insufficiency would often cause false convergence.

For improving MS localization, a Background-Weighted Histogram (BWH) transformation is proposed to minimize the effects of background features [7]. However, the BWH only add scaling effect to the MS iteration formula, which is invariant to the scale transformation of weights [7]. Corrected background weighted histogram (CBWH) transformation minimizes the effect of background features in target model [7]. However, in the natural video sequences where the background and target have overlapping features, the performance of CBWH decreases drastically. The issue is further discussed by [8] and proposed background feature minimization by enlarging the pixel weighting which maps to the pixels on the target. The method assumes that all the pixels on the target have the greater contribution which is not always true. [13] Introduce foreground feature saliency

concept into the background modeling to exploits salient features from both foreground and background. Recently [14], applies the structural local sparse representation method to analyze the background region around the target to reduce the probability of prominent features in the background. The above methods use only first frame for background weight transformation, whereas in natural videos, the background can change in immediate next frame or after few frames. To improve the MS localization, combination of target structural and spectral information has also been proposed [10]. Local binary pattern based structural feature integration in MS algorithm along spectral information improves target localization [10]. Histogram probabilistic multi hypothesis tracker (H-PMHT) using reciprocal pixel intensity measurement is developed to improve the results in intense clutter environment [17]. The concept of spatiogram [15], [16] for incorporating the values of the pixels as well as their spatial relationships fails in case of similar target and background color. Therefore, the MS tracker requires a true and consistent background feature identification for continuous weight updation. Due to dependence on histogram based features alone, MS tracking often produces false convergence [9] especially when similar color modes exist in the target neighborhood. Therefore, to make the MS tracker more robust for localization, necessary post processing using some structural features is also being applied. In [18], a post processing step is applied by using normalized cross correlation for handling meانشift target occlusion problem. In [11], target localization has been improved by applying post processing steps like edge based centroid estimation for convergence of tracker (on the true target center). Canny edge detection through non-maximum suppression and hysteresis thresholding produces well defined edges for centroid estimation. However, the method produces limited efficiency in high textured or edge concentrated targets.

In this paper, target localization is improved by two different approaches followed by a target model update criterion. Firstly, a true background weighted histogram (TBWH) transformation is proposed for MS algorithm to minimize significant background features effect in target model representation by evaluating their share in target features. In first frame, the true background features that are not prominent features of target itself are identified for weight minimization transformation. The transformation function is formed by considering the contribution of background features bins in target model by comparing the corresponding values and finding spatial displacement from the target center. The spatial displacement of bins is calculated by tracing the spatial position of bins elements (pixels) through Epanechnikov kernel weighting. As Epanechnikov kernel assigns the higher weights to pixels that are near to the target center and gradually decreases away from center. In this way each bin can be categories on suitable threshold for spatial belonging to background or target. The threshold ensures the selection of feature bins with majority of background pixels located away from target center. This transformation guarantees the se-

lection of true background features by intelligently recording their spectral and spatial share in target.

Secondly, to further improve the target localization, this paper also proposed a post processing step to relocate the MS centroid position using object geometric information. The canny edge detection is applied in search window around MS estimated centroid to extract the object edges. The smoothing stage in canny edge algorithm is modified and Guided filtering [12] is replaced in place of Gaussian filtering. The guided filter computes the filtering output by considering the content of a guidance image which is target template image in our case. We have demonstrated that application of canny edge detection on Guided filtering smoothed images, produces only prominent outer edges while smoothing the inner texture detail. The MS estimated target centroid is then updated by computing the geometric center within the boundary as explained in [11]. Target model update is important for MS tracking especially when it observe a serious change in target or background appearance. The model update is even more essential for the background weighted histograms based algorithms that are based on target as well as background appearances. Moreover, sometimes a transient change occurs in appearance and applying the weighted transformation in target model representation at this stage may introduce an erroneous update that continue to propagate even when the target recovered from transient change. A novel method is proposed to update the target model followed by weighted adjustment in transformation applied for background weighted histogram using online feature consistency data. The target model update is carried out by considering both the target and background appearance.

The improvements for MS target localization and online model update are proposed for single object tracking in complex scenarios and tested for comparative results with existing state of the art techniques. Experimental results on numerous challenging video sequences verify the significance of proposed technique in terms of robustness to complex background, occlusions, appearance changes, and similar color object avoidance. Following major contributions are proposed:

- 1) A robust and intelligent background weighted histogram method is formulated for MS algorithm by considering the spectral and spatial contribution of background features in target features for improving localization.
- 2) A geometric adjustment in MS estimated centroid is also proposed as a post processing step using guided filter smoothing and canny edge detection.
- 3) A novel method is proposed to update the target model through target and background appearances for background weighted histogram based algorithms followed by online feature consistency data weighting.
- 4) Performance comparison of proposed method with existing trackers on challenging video sequences.

This paper is organized as follows: the next section presents the related work, including brief introduction about meanshift, BWH and CBWH algorithms. Section 3 describes the proposed methodology that has been followed in order to improve Meanshift target localization. Finally, experimental results and conclusions are presented in Sections 4 and 5 respectively.

2. Related Work – Mean Shift, BWH and CBWH

The standard meanshift algorithm is an efficient non parametric tracking technique. The brief introduction of MS is as follows, whereas, the detailed mathematical derivations and discussions about standard meanshift algorithm can be found in [4]. MS is a deterministic iterative procedure for locating the maxima of a density function given discrete data samples. In tracking applications it compares a target model histogram with current frame target candidate histogram to find out the most promising converging region [4]. The target model histogram is represented by:

$$t_b = C \sum_{j=1}^J \xi(\|h_j\|^2) \delta[\zeta(h_j) - b] \tag{1}$$

where $C = 1/\sum_{j=1}^J \xi(\|h_j\|^2)$ is the normalization constant, $b = 1, 2, \dots, B$ be the color/intensity histogram bins, ξ is an Epanechnikov Kernel profile and ζ associates the pixel h_j to the b^{th} bin. Similarly, the target candidate histogram is represented as:

$$C_b(y) = C \sum_{j=1}^J \left\| \frac{y - h_j}{Bw} \right\|^2 \delta[\zeta(h_j) - b]. \tag{2}$$

Here Bw is the MS kernel bandwidth and y is the current location of MS centroid. The Bhattacharyya coefficient (BC) $\rho = \sum_{b=1}^B \sqrt{c_b t_b}$ is used to statistically measure the similarity between target model and target candidate histograms. The higher ρ means higher similarity and probability of target finding. The MS algorithm iteratively search for optimal value of BC and find the updated target position using MS vector define as:

$$y_{\text{new}} = \frac{\sum_{j=1}^J h_j \omega_j \xi\left(\left\| \frac{y-h_j}{Wb} \right\|^2\right)}{\sum_{j=1}^J \omega_j \xi\left(\left\| \frac{y-h_j}{Wb} \right\|^2\right)}, \tag{3}$$

$$\omega_i = \sum_{u=1}^m \delta[\zeta(h_j) - b] \sqrt{\frac{t_b}{c_b(y)}} \quad \text{as} \quad \xi(h) = -k'(x). \tag{4}$$

The standard meanshift algorithm compromises its performance in different real time applications. For example, in target tracking, the background information is often included in the detected target region. If the correlation between

target and background is high, the localization accuracy of the object will be decreased. To reduce the interference of salient background features in target localization, a representation model of background features was proposed by [4] as background-weighted histogram (BWH). The aim of BWH was to select the salient features from the background and reduce their effect in target feature presentation. The method adopt the standard MS algorithm and only add a multiplication step to the target model equation (1) and candidate histograms equation (2) through a background weighted transformation v_b .

$$v_b = \left\{ \min\left(\frac{\min_{b=1}^B(O_b)}{O_b}, 1\right) \right\}. \tag{5}$$

The O_b is target surrounding background histogram with area three times as of target area. Later [7] proved that proposed BWH transformation practically does not performed its goal rather it only add scaling effect to the MS iteration formula, which is invariant to the scale transformation of weights [7]. Proposed a corrected background weighted histogram (CBWH) transformation by applying transformation v_b to only target model equation (1) and not with candidate equation (2) as was done in BWH.

3. Proposed Methodology

To improve MS target localization, a spatio-spectral technique defines as true background weighted histogram (TBWH) is applied to locate and minimize the effect of background features in target model representation. The target localization is further improved by adjusting the MS estimated target position through edge based centroid adjustment. Meanshift method also requires model update to overcome the appearance change of target during tracking. As the proposed technique is based on target as well as background features, hence both are considered to be updated when require. Therefore, to get an effective model update strategy, this paper propose a novel method for background weighted histogram based algorithms to estimate the time and type of update. The method also proposes a damping transformation in target model through online feature consistency data to minimize transient disturbances. The detail methodology is explained in following sections.

3.1 True Background Weighted Histogram (TBWH)

In MS algorithm, usually the target is selected through a rectangular or elliptical search window. Due to irregular target shapes the window also includes some background features and unintentionally becomes the part of the target features. It is to be noticed that the main part of these features exists in target outer region. Therefore, if the target and background have same or mingled features, then the results of CBWH technique are not appropriate. This requires an intelligent method to first identify the true background features in

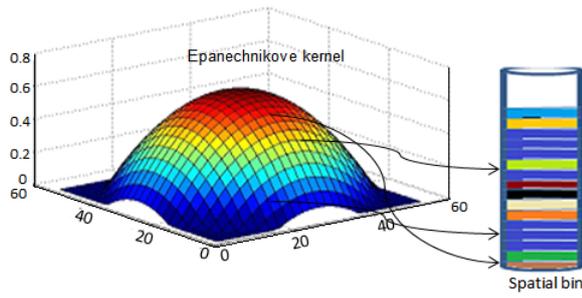


Fig. 1. Epanechnikov kernel and bin special pixel assignment.

target model and then apply the transformation. Moreover as the tracking involves different challenging situations, relying only on histograms feature is always helpful and there is need for MS method to compensate with some post processing step involving spatial features. Therefore, this paper proposed a modified transformation for background feature minimization in target representation as well as a post processing step using target geometric information for improvement in target localization.

The true background weighted histogram transformation for MS algorithm is proposed to minimize significant background features effect in target model representation by evaluating their share in target features. The MS and CBWH approaches are adopted in this paper as baseline algorithms. The main factor for consideration is to identify the true prominent background features in target model representation that are not part of target. The transformation function is formed by considering the contribution of background features bins in target model by comparing their corresponding values and finding spatial displacement from the target center. For this purpose the transformation defined in (5) is modified as follows:

$$v_b = \begin{cases} \min\left(\frac{\min_{b=1}^B(O_b)}{S_b}, 1\right) & \text{if } S_b > 0, \\ 1 & \text{otherwise.} \end{cases} \quad (6)$$

The O_b is target surrounding background histogram with area three times as of target area. In modified v_b we determined significant background feature bins S_b based on spectral and spatial share of background features in target area. The function S_b contains the background feature bins with higher spectral (histogram) values to that of corresponding target feature bins and having majority of pixels located near target boundary. The spatial displacement of bins is calculated by tracing the spatial position of bins elements (pixels) through Epanechnikov kernel weighting. Epanechnikov kernel assigns the higher weights to pixels that are near to the target center and gradually decreases as move away from center. In this way each bin can be categorized on suitable threshold for spatial belonging to background or

target. The true background identity factor S_b for modified transformation v_b is calculated as:

$$S_b = \begin{cases} O_b & \text{if } O_b - t_b > 0 \text{ \& } \tilde{\xi}_b > T, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

where $\tilde{\xi}_b$ is number of pixels in bin b having Epanechnikov Kernel profile value less than 0.3 and T is the threshold value for bin spatial verification as true background bins. Here T is kept as $0.5 \times n_b$, n_b is total number of pixels in any specific bin.

The Epanechnikov Kernel profile and pixel spatial weighting bin assignment technique can be seen in Fig. 1. The pixel with weight less than 0.3 have come from outer target region and if a bin includes more than its half pixels through this region are spatially identified as outer region bin and if it also have prominence in background area it is selected as true background bin. The identity factor will ensure that only those background feature bins are selected that have higher values than target and are composed of majority pixels lying away from target center i.e belong to background. The transformation v_b on the basis of S_b ensures that the weights of background features that are not salient features of target are being minimized. The target model feature histogram with modified background weighted transformation is computed as $t_b(m)$, and used in standard MS algorithm for tracking applications.

$$t_b(m) = v_b * C \sum_{j=1}^J \xi(\|h_j\|^2) \delta[\zeta(h_j) - b]. \quad (8)$$

An example histograms for target model and its background are shown in Fig. 2(a) and 2(b) respectively. A comparative display of modified target histograms by CBWH and TBWH against original target and background histograms is shown in Fig. 2(c) and 2(d) respectively. The CBWH method blindly transforms all the histogram features without considering their participation in target model hence decreases some prominent target features whereas our proposed TBWH method only transform the true background features while keeping the prominent target features intact.

3.2 Geometric Centroid Adjustment

Histogram as a feature, only describes the global color distribution and ignores the structure of the object, it often causes false convergence especially when similar color modes exist in the target neighborhood. Therefore, in addition to the improvement in target localization through background weighted histogram method, a post processing step for adjustment of MS estimated centroid on the basis of object features. Canny edge based centroid estimation for MS algorithm proposed in [11] is modified by applying guided filter preprocessing for edge preserved smoothing. We set the

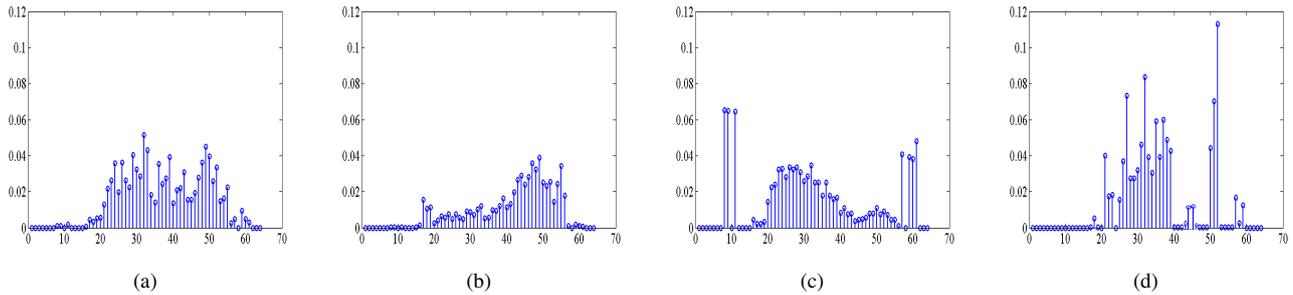


Fig. 2. (a) Target model, (b) Background model, (c) CBWH [7], (d) Proposed TBWH.

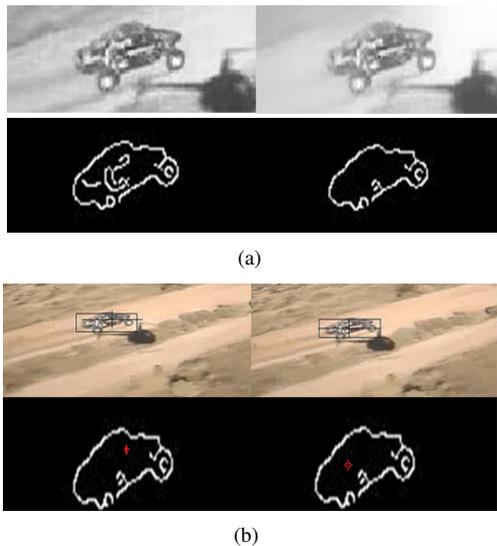


Fig. 3. MS Geometric centroid adjustment (a) Canny edge detection with and without Guided filter (b) MS Centroid adjustment

target model as template for guided filtering that smoothed out small unwanted structural information and preserve the prominent structures in target area.

After application of edge detection the centroid is calculated using finite set of points along horizontal and vertical axis within object boundaries. The tracking window is then placed on the new center point. The detail implementation is similar to that proposed in our earlier work [11]. Edge based centroid adjustment provides quick and accurate reference for MS algorithm to search true local maxima in next frame. The number of MS iterations has been drastically reduce due to localization improvement and hence compensate the extra computational cost of edge based centroid adjustment.

Figure 3 shows an example of localization improvement through geometric centroid adjustment using guided filter smoothing technique. Figure 3(a) shows the comparison between canny edge detection with and without guided filter as smoothing operator. The guided filter smoothing provided all the major structural information whereas smoothed out all unnecessary edges as seen in edge image hence improves the object geometric centroid marking and update. Figure 3(b) shows the centroid adjustment procedure. The biased mean-

shift centroid on left side is updated through true geometric centroid on right side of figure.

3.3 Model Update and Online Feature Consistency Weighting

A major limitation of standard meanshift algorithm is lack of model update strategy to cope with change in target or background appearance. The model update technique is required to indicate the due time and type of update. As our proposed technique is based on both target as well as background features, hence requires update in both. The model update event can occur in any of the situation:

- (1) A serious change in target appearance is observed while the background is not much change. In this case we have to update the target model only.
- (2) A serious change in background appearance is observed while the target is not much effected. In this case we have to update the background only.
- (3) A serious change is observed in appearance of both target and background. In this case we have to change the target as well as background models.

The Bhattacharyya coefficient (BC) ρ is used as similarity measure and threshold to apply model update for these events. We define a similarity threshold η_1 associated with similarity measure (ρ_t) between current target window and target template and a dissimilarity threshold η_2 associated another similarity measure (ρ_b) exist between current background and last updated background. We have also defined a τ function to specify time or number of frames to observe for temporary or permanent appearance change. The τ is based on target motion vector and complexity of background. In our case we set the τ as five frames. The following updates conditions are proposed to indicated model update in TBWH method:

Target model update condition: $\rho_t < \eta_1$ and $\rho_b < \eta_2$ for consecutive τ -frames.

Background model update condition: $\rho_t > \eta_1$ and $\rho_b > \eta_2$ for consecutive τ -frames.

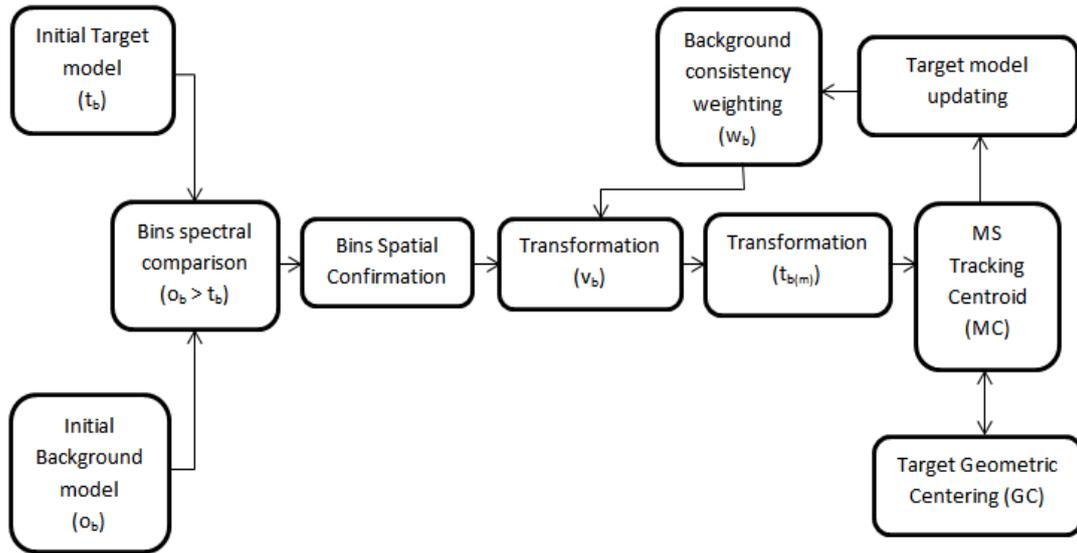


Fig. 4. Block diagram of TBWH target model update meanshift tracking.

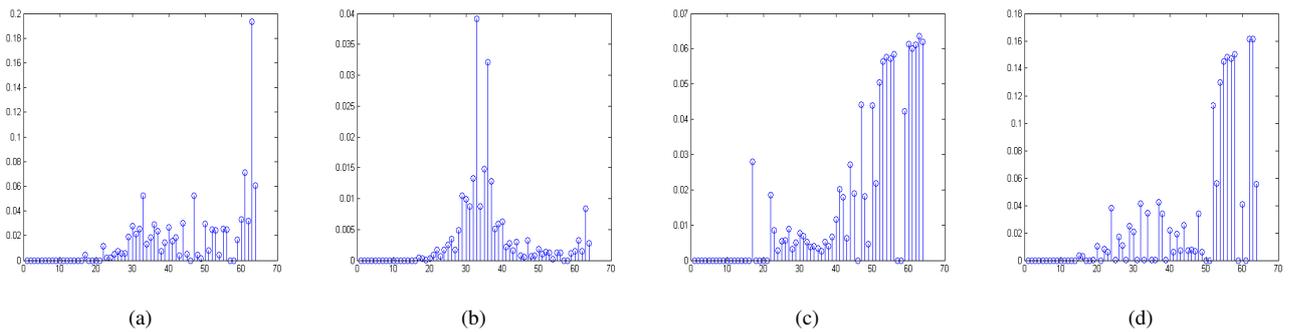


Fig. 5. Simple background: Ping pong ball histogram comparison (a) Target model, (b) Background model, (c) CBWH [7], (d) Proposed TBWH.

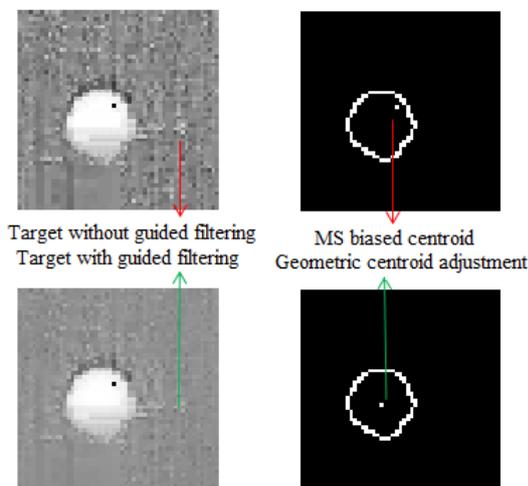


Fig. 6. Geometric centroid adjustment: Ping pong ball.

Target and Background model update condition: $\rho_t < \eta_1$ and $\rho_b > \eta_2$ for consecutive τ -frames.

Moreover, applying weight updating directly to transformation v_b may results in erroneous outcome because of transient changes. Thus it requires an online mechanism to store the previous background feature bins affiliation and consistency so that a weighted update is applied to the transformation v_b after target model update. When the target model is updated, the transformation v_b is also updated with an online feature consistency criterion to add a damping weights on the basis of online observation. Moreover the damping is also applied after recovery from transient event. The criterion ensures that in target model update, the transformation v_b is updated not only on current appearance changes but also involves the effect of online historic data.

Let O_b^+ represent total number of times the bin O_b is selected as significant background feature bin and O_b^-

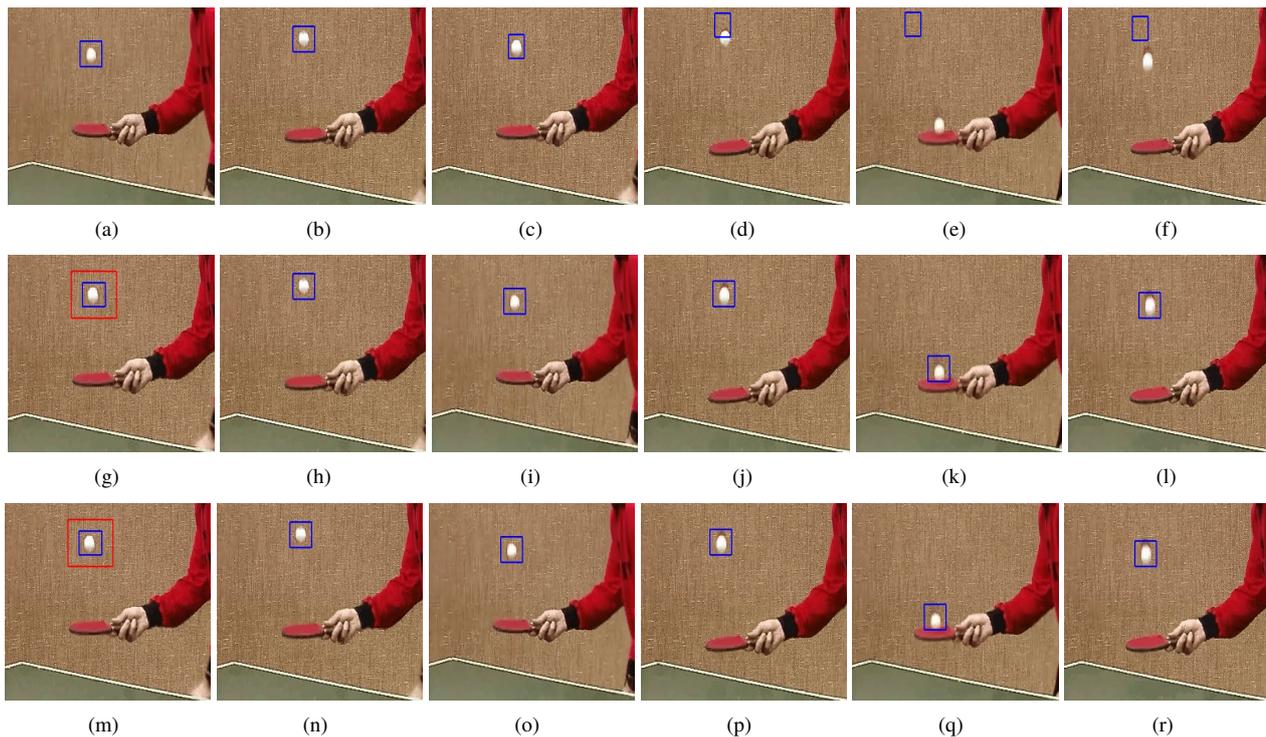


Fig. 7. Comparative tracking results on simple background object: MS (Fig. a-f), CBWH (Fig. g-l) and TBWH (Fig. m-r).

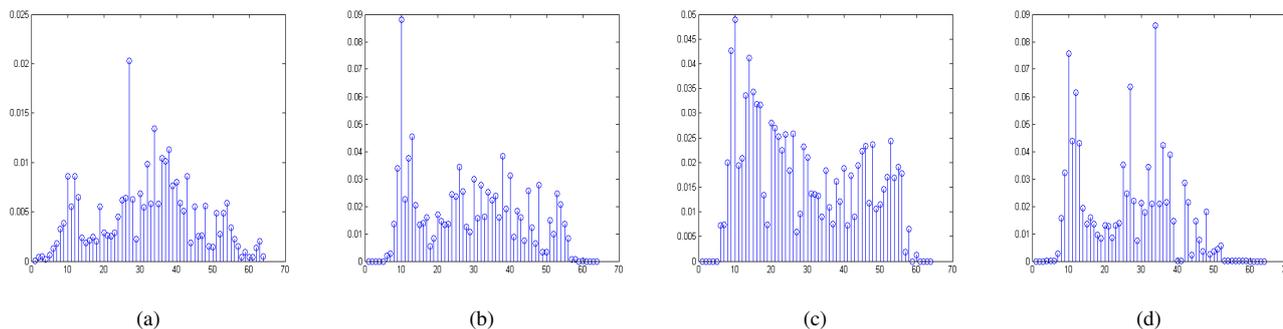


Fig. 8. Complex background: Gymnastic girl sequence histogram comparison (a) Target model, (b) Background model, (c) CBWH [7], (d) Proposed TBWH.

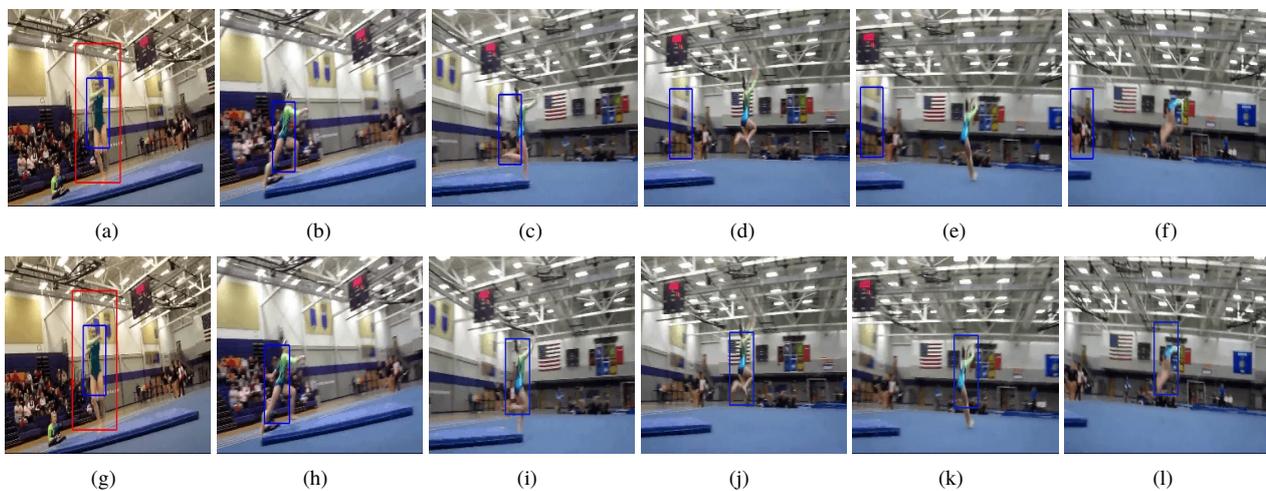


Fig. 9. Comparative tracking results on complex background object: CBWH (Fig. a-f) and TBWH (Fig. g-l).

represent total number of times the bin O_b value is not selected as significant background bin (from the n initial frames). A weighting factor w_b is computed for last n consecutive frames,

$$w_b = \frac{O_b^+}{O_b^+ + O_b^-}. \quad (9)$$

The updated \hat{v}_b on the basis of online background feature consistency is computed as $\hat{v}_b = v_b \times 1/w_b$. The O_b with maximum weight will assign lower values in transformation v_b and hence lower weightage in target model. The integrated block diagram of improved Meanshift target tracking using true background weighted histogram along with geometric centroid adjustment and model updating through background consistency weighting is shown in Fig. 4.

4. Results and Discussion

The existing MS [4], CBWH [7] and proposed tracking algorithms are evaluated on different challenging video sequences including videos from Bonn Benchmark on tracking (BoBoT) and vivid tracking dataset. For comparison of proposed TBWH technique with standard MS and CBWH techniques following online benchmark tracking videos have been selected based on content complexities and tracking challenges. One is the Ping-Pong ball test sequence already used in evaluating different tracking techniques including MS and CBWH. The second is the Gymnastic girl sequence and third is the bike sequence. The first sequence does not include challenge of background interference and will serve as benchmark to test the algorithms basic tracking performance. The latter two video sequences have complex background and incorporates different challenges to evaluate the proposed TBWH tracker as well as comparison with standard MS and CBWH techniques. The average iteration numbers of each tracker is measured for convergence speed and qualitatively measure the tracking performance by considering whole video sequence results. The results and discussion section include three main subsections to introduce the problem statement and comparison between existing and proposed technique.

4.1 Tracking Results in Simple Background Environment

In the first experiment, the Ping-Pong ball sequence has 52 frames of 352×240 pixels. In test Ping-Pong ball video, the tracking window contains white colored ball and some part of brown colored background. It can be seen in Fig. 5 that the target model pdf t_b and target background pdf O_b have clear discrimination in color feature bins. The pdf distribution of target window t_b is shown in Fig. 5(a) which is clearly reflecting the presence of said histogram. In similar way the background window has mid concentrated histogram reflecting the presence of few similar colors present in target window. The background weighted histogram transforma-

tion applied by the CBWH and TBWH scheme is shown in Fig. 5(c) and Fig. 5(d) respectively. The fig shows that the outcome of both methods are looking almost same with some minor differences in histogram shape.

The results shows that all the three algorithms could track the target well because the target appearance is simple and almost have no appearance change. At the point when a little complexity in the form of ball touching the racket, the standard MS method could not maintained to track the target while the other two algorithms CBWH and TBWH continued to track the target to the rest of the sequences. Since CBWH and TBWH transform the target model by minimizing the background features so it well discriminated the target from background while the simple MS algorithm could not performed well. The average numbers of iterations are 2.41 for MS, 3.35 for CBWH and 3.12 for TBWH respectively.

Figure 6 shows the geometric centroid adjustment in ping pong ball video sequence. The first column (of Fig. 6) shows the smoothing results obtained with guided filtering, whereas the second column shows the adjustment of biased MS centroid position to the true object center. It can be seen that new centroid is more appropriate than the MS estimated centroid. The tracking results are shown in Fig. 7 for MS, CBWH and TBWH respectively.

4.2 Tracking Results in Complex / Multicolored Background Environment

In the second experiment we have selected two video sequences, the challenging Gymnastic girl sequence has 105 frames of 320×240 pixels and a more complex bike sequence with 795 frames of 320×240 pixels. These two videos are having different challenges including the multi colored background, similar color objects in surrounding, scale and orientation change, and high speed maneuvering etc. By considering the histogram as shown in Fig. 8, for target model and target background, it can be seen that both of these have complex and multicolored histogram. These two videos will serve as the test case to analyze the CBWH and TBWH results. The results of standard MS technique are not shown for brevity and because its performance is not appropriate on these challenging videos hence can not be compared. The histogram transformation due to CBWH and proposed TBWH can be seen in Fig. 8.

As the background histogram also have color distribution of target window colors, it is bit difficult for CBWH to discriminate between true target colors and hence it applies blind weight minimization transformation. The resulted histogram of CBWH transformation have also minimized the prominent target features while minimizing the background features weights. Whereas the modified TBWH scheme have intelligently applied the weighted transformation to the target histogram and only the prominent background features are minimized Fig. 8(d). Similarly the histogram comparison of both technique on bike video have almost same results.

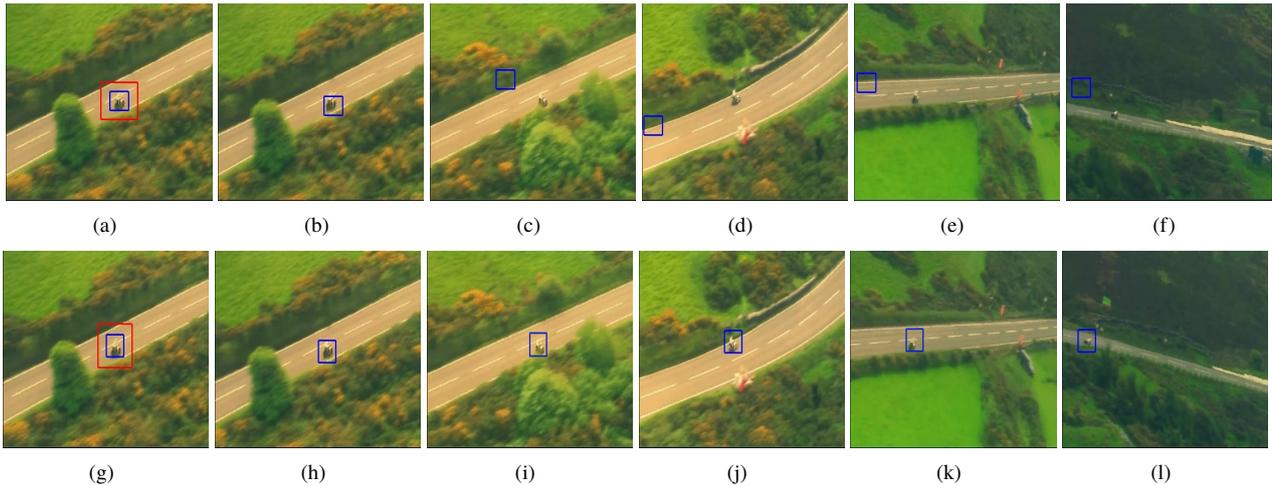


Fig. 10. Comparative tracking results on complex background object: CBWH (Fig. a-f) and TBWH (Fig. g-l).

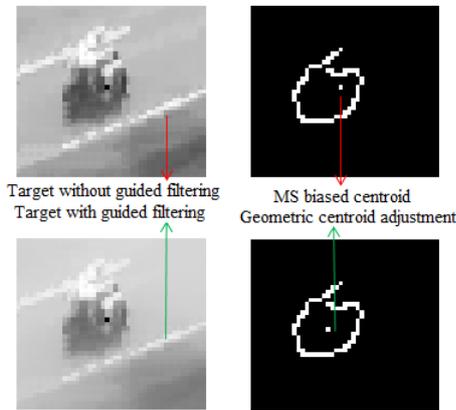


Fig. 11. Geometric centroid adjustment: Bike sequence.

The tracking results are shown in Fig. 9 for CBWH and TBWH respectively. In the start of Gymnastic girl sequence, all the technique perform well but very soon in frame 22 the MS method lost the target due to target abrupt motion as the simple MS tracking can not perform well when the object motion becomes greater than its meanshift vector. However both the CBWH and TBWH also include the outer search region they continued tracking. When the target features became mixed with the same colors object in background and to the people sitting in background, at that time CBWH fails to track the target whereas our proposed TBWH continue to track the object to the end of the sequence even in such a challenging sequence.

The similar results are obtained on bike sequences which is even more challenging than previous one. The tracking results are shown in Fig. 10 for CBWH and TBWH respectively. Since the background transformation is not applied in standard MS tracker, it begins to drift off the target in very early i.e in frame 15 due to above mentioned challenges and is unable to recover the failure. CBWH tracker also drifts off the target when it comes near the similar color short time occlusion in the frame 78 and is also unable to re

track the object. In comparison, the proposed method performed extremely well on this very challenging video due to incorporation of intelligently defined background weighted transformation and centroid based post processing step to enhance the target localization. The main reason in standard MS and CBWH failure was the start of localization drift which was allowed to be continued and at the end both were resulted in failure where as in proposed approach the localization drift was corrected by centroid adjustment technique.

Figure 11 shows the geometric centroid adjustment in bike video sequence. The biased MS estimated centroid shown in top left where it is very close to the boundary of object. Similarly its edge image is shown in top right that is adjusted by applying geometric centroid updation. Guided filtering is applied before canny edge detection for edge preserved smoothing and the difference in smoothing can be seen in bottom left. The centroid adjustment in edge image can be seen in Fig. 11 bottom right. It can be seen that new centroid is more appropriate than the MS estimated centroid.

We also included a non chromatic video sequence in IR mode to demonstrate the performance of proposed TBWH method with emphasis on centroid adjustment technique. The video mainly contains a similar target-background challenge and some nearby objects with similar appearance to the target.

The tracking results are shown in Fig. 12 and it can be seen that the target in TBWH results is more localized as compared to CBWH where beside tracking the sequence the localization is not appropriate. This error in localization resulted in failure of tracking for CBWH when target passed nearby a similar color and shape object. The localization drift became two high for the CBWH to recover back to original target and become stick to the wrong one where as due to centroid adjustment, the TBWH readjust the MS centroid hence the localization is well maintained throughout the tracking sequence.

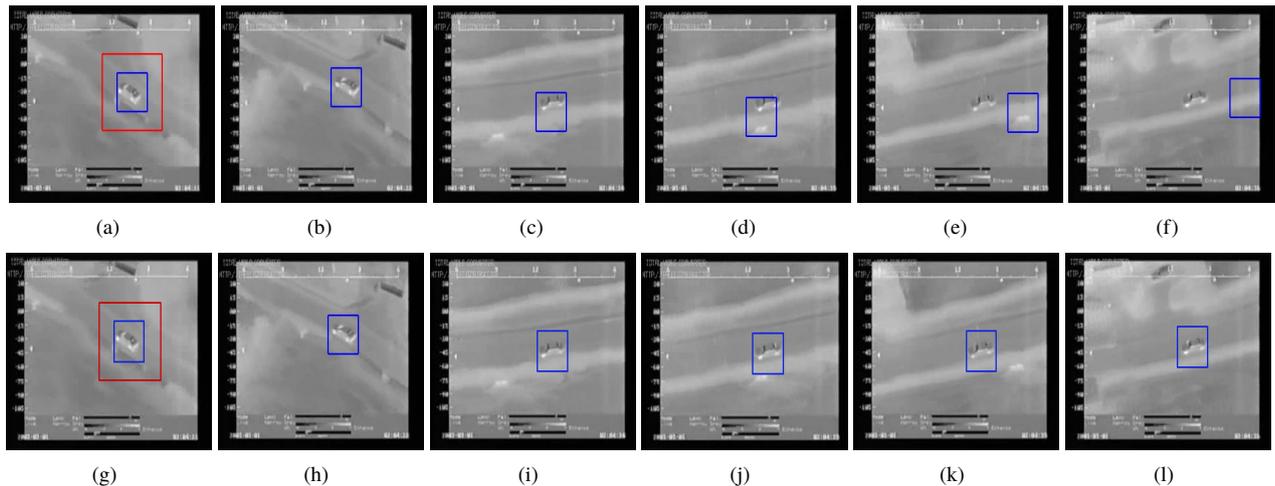


Fig. 12. Comparative tracking results on IR targets with similar background object: CBWH (Fig. a-f) and TBWH (Fig. g-l).

In comparison to existing MS [4] and CBWH [7] schemes, the proposed algorithm was able to track target in whole sequences (by efficiently handling all the above mentioned challenges). Moreover, the MS localization is greatly improved which provides the accurate starting point to tracker. This results in fast convergence with drastic decrease in MS iterations therefore, the computational complexity of proposed method is comparable with original MS and CBWH schemes. the proposed technique is equally applicable for color and monochromatic videos.

5. Conclusion

The paper presents a robust tracker based on modified MS algorithm. The proposed method effectively reduces background interference for target localization in presence of complex/multi colored environment without compromising the tracking efficiency. The true background features are identified in target model representation using spectral and spatial cues and then a transformation is applied to minimize their effect in target model representation. The target localization is further improved by adjusting the MS estimated target position through geometric centroid estimation. A novel method is also proposed to update the background weighted histogram based algorithms for target model and weight transformation update using an online feature consistency data. Experimental results on numerous challenging video sequences verify the significance of proposed technique. Currently the proposed TBWH scheme is designed for single object tracking and can be extended for multiple targets as future work.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Gov-

ernment (MSIP) (NRF-2015R1A2A1A15053854) and National University of Science and Technology (NUST) Pakistan, Research fund.

References

- [1] SMEULDERS, A. W., CHU, D. M., CUCCHIARA, et al. Visual tracking: an experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, vol. 36, no. 7, p. 1442–1468. DOI: 10.1109/TPAMI.2013.230
- [2] LIU, Q., ZHAO, X., HOU, Z. Survey of single-target visual tracking methods based on online learning. *IET Computer Vision*, 2014, vol. 8, no. 5, p. 419–428. DOI: 10.1049/iet-cvi.2013.0134
- [3] YILMAZ, A., JAVED, O., SHAH, M. Object tracking: A Survey. *ACM computing surveys*, 2006, vol. 38, no. 4, article 13. DOI: 10.1145/1177352.1177355
- [4] COMANICIU, D., RAMESH, V., MEER, P. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, vol. 25, no. 5, p. 564–575. DOI: 10.1109/TPAMI.2003.1195991
- [5] CHENG, Y. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, vol. 17, no. 8, p. 790–799. DOI: 10.1109/34.400568
- [6] KHALID, M. S., MALIK, M. B. Biased nature of Bhattacharyya coefficient in correlation of gray-scale objects. In *IEEE International Symposium on Image and Signal Processing and Analysis*, 2005, p. 209–214. DOI: 10.1109/ISPA.2005.195411
- [7] NING, J., ZHANG, L., ZHANG, D., et al. Robust mean-shift tracking with corrected background-weighted histogram. *IET Computer Vision*, 2012, vol. 6, no. 1, p. 62–69. DOI: 10.1049/iet-cvi.2009.0075
- [8] KEN, C., KANGKANG, S., CHOI, K., et al. Optimized meanshift target reference model based on improved pixel weighting in visual tracking. *Journal Of Electronics (CHINA)*, 2013, vol. 30, no. 3, p. 283–289. DOI: 10.1007/s11767-013-2168-5
- [9] XU, D., WANG, Y., AN, J. Applying a New Spatial Color Histogram in Mean-Shift Based Tracking Algorithm. In *Conference on Image and Vision Computing*, Univ. of Otago (New Zealand), 2005.

- [10] XIAORONG, P., ZHIHU, Z. A more robust mean shift tracker on joint color-CLTP histogram. *International Journal on Image, Graphics and Signal Processing*, 2012, vol. 12, p. 34–42. DOI: 10.5815/ijgisp.2012.12.05
- [11] MEHMOOD, R., ALI, M. U., TAJ, I. A. Applying centroid based adjustment to kernel based object tracking for improving localization. In *International Conference on Information and Communication Technologies*, 2009, Karachi (Pakistan), p. 209–214. DOI: 10.1109/ICICT.2009.5267188
- [12] HE, K., SUN, J., TANG, X. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, vol. 35, no. 6. DOI: 10.1109/TPAMI.2012.213
- [13] WANG, D., SUN, W., YU, S., et al. A novel background-weighted histogram scheme based on foreground saliency for mean-shift tracking. *Multimedia Tools and Applications*, 2015. DOI: 10.1007/s11042-015-3078-3
- [14] ZHAO, H., XIANG, K., CAO, S., et al. Robust visual tracking via CAMShift and structural local sparse appearance model. *Journal of Visual Communication and Image Representation*, 2016, vol. 34, p. 176–186. DOI: 10.1016/j.jvcir.2015.11.008
- [15] STANLEY, T., BIRCHFIELD, RANGARAJAN, S. Spatial histograms for region-based tracking. *ETRI Journal*, 2007, vol. 29, no. 5, p. 697–699. DOI: 10.4218/etrij.07.0207.0017
- [16] CHU, H., SONG, Q., YUAN, H. Research of Mean Shift target tracking with spatiogram corrected background-weighted histogram. In *IEEE International Conference on Information and Automation*, 2015, Lijiang (China). DOI: 10.1109/ICInfA.2015.7279606
- [17] PAKFILIZ, A.G. Video tracking for visual degraded aerial vehicle with H-PMHT. *Radioengineering*, 2015, vol. 24, no. 4, p. 1091–1098. DOI: 10.13164/re.2015.1091
- [18] MEHMOOD, R., NAWAZ, R., IQBAL, N. Occlusion handling in meanshift tracking using adaptive window normalized cross correlation. In *International Bhurban Conference on Applied Sciences and Technology (IBCAST)*. Islamabad (Pakistan), 2014. DOI: 10.1109/IBCAST.2014.6778134

About the Authors ...

Rashid MEHMOOD was born in 1978. He obtained MS Electronic Engineering in 2009 from Muhammad Ali Jinnah University (MAJU), Islamabad, Pakistan. He is working towards the Ph.D. degree in Military College of Signals, National University of Sciences and Technology, Islamabad, Pakistan. Currently he is research internee at Signal and Image Processing Lab (SIPL), Gwangju Institute of Science and Technology (GIST), Korea. His research interests include

image and video processing, computer vision, 3D shape recovery.

Noor ul HUDA was born in 1990. She obtained BS in Electrical Engineering from University of Engineering and Technology (UET), Taxila, Pakistan and MS Signal Processing from Center for Advance Studies in Engineering (CASE), Islamabad, Pakistan. Currently, she is working as lecturer at Center of Advanced Studies in Telecommunication (CAST), CIIT, Islamabad. Her research interests include signal and image processing, pattern recognition.

Jinho SONG was born in 1991. He obtained his BS in Mathematics from University of California Irvine, USA in 2014. Currently he is working towards the MS degree in Mechatronics at Gwangju Institute of Science and Technology (GIST), Republic of Korea. His research interests include signal and image processing, 3D shape recovery.

M. Mohsin RIAZ was born in 1985. He obtained MS Electronic Engineering in 2009 from International Islamic University, Islamabad, Pakistan. He has completed his PhD in Electrical Engineering in 2013 from National University of Sciences and Technology (NUST), Islamabad. Currently, he is working as Assistant Professor at Center of Advanced Studies in Telecommunication (CAST), CIIT, Islamabad. His research interests include signal and image processing, pattern recognition, evolutionary computing and fuzzy logic.

Naveed IQBAL RAO received his MS degree in Electrical Engineering from College of EME, National University of Sciences and Technology, Pakistan in 2003 and his PhD degree in Computer Software Engineering from Tsinghua University, Beijing, China 2007. He is currently Assistant Professor in Military College of Signals, National University of Science and Technology (NUST), Pakistan. His research interests include image processing, computer vision, machine learning.

Tae Sun CHOI (Corresponding Author) received his MS degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), in 1979 and his PhD degree in Electrical Engineering from the State University of New York in 1993. He is currently a Professor in the Mechatronics Department at Gwangju Institute of Science and Technology (GIST), Republic of Korea. His research interests include image processing, machine/robot vision, and visual communications.