Robust Edge-Enhanced Fragment Based Normalized Correlation Tracking in Cluttered and Occluded Imagery

Muhammad Imran Khan[1], Javed Ahmed[2], Ahmad Ali[3], and Asif Masood[4]

[1,2,4]*Department of Computer Science, NUST Military College of Signals, Rawalpindi, Pakistan* [3]*Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan*
{[1]mimran,[2]javed,[4]asifmasood}@mcs.edu.pk, [3]ahmadali1655@hotmail.com

Abstract

Correlation trackers are in use for the past four decades. Edge based correlation tracking algorithms have proved their strength for long term tracking, but these algorithms suffer from two major problems: clutter and slow occlusion. Thus, there is a requirement to improve the confidence measure regarding target and non-target object. In order to solve these problems, we present an "Edge Enhanced Fragment Based Normalized Correlation (EEFNC)" algorithm, in which we: (1) divide the target template into nine non-overlapping fragments after edge-enhancement, (2) correlate each fragment with the corresponding fragment of the template-size section in the search region, and (3) achieve the final similarity measure by averaging the correlation values obtained for every fragment. A fragment level template updating method is also proposed to make the template adaptive to the variation in the shape and appearance of the object in motion. We provide the experimental results which show that the proposed technique outperforms the recent Edge-Enhanced Normalized Correlation (EENC) tracking algorithm in occlusion and clutter.

Keywords: *fragment, correlation, template, occlusion, clutter, Kalman filter.*

1. Introduction

A visual tracking system automatically finds the location of target in the consecutive frames of a video. This task becomes difficult when the target is changing its orientation, shape and size. The presence of clutter (i.e. other objects near the target) and occlusion (i.e. other objects in front of the target) in tracking environment makes the problem even more difficult. Computational cost of the algorithm is also important for real time tracking applications.

While performing tracking in an environment where abrupt changes in the background are not expected, modeling of background is normally a preferred approach. For this purpose, Gaussians Mixture Model [4] technique is very successful but this technique has limited capabilities when used alone. Tracking of objects, when the background is not static and changing dynamically, becomes more challenging. In this situation, we cannot develop model of the background as new objects quickly become part of the background and then disappear. Furthermore, the moving target can change its orientation which is an additional problem. Lucas Kanade tracking algorithm [2] uses different feature points of object to be tracked in the next frame, but dynamic selection and then tracking of these feature points in real-world scenarios is very difficult, especially in case of illumination variation.

The histogram matching based trackers [10, 11, 12] can also work without requiring background model, but they suffer from the inherent problem with the histogram that two different images can have similar histograms because the histogram does not preserve the pixel location. To some extent, this problem has been addressed in [7] by dividing the object template into multiple non-overlapping fragments and using the histograms of those fragments in the matching process, but the same problem with the histograms can occur in the fragment level.

Since the correlation [1, 3] process does not lose the spatial information of the pixels, they are more robust to clutter than the histogram based trackers. Edge Enhanced Normalized Correlation (EENC) tracker [5] has significantly solved the real-world problems of orientation, illumination, obscuration, intermittent occlusion, complex object motion, object fading, and noise. In EENC, the template and the search image are edge-enhanced before performing normalized correlation between them. The best match of the template in the search image is found at the location corresponding to the location of the peak value in the correlation surface (matrix). The best match region and the current template are then linearly combined to prepare the new template to be searched for in the next frame. This technique of template updating plays a vital role in long term tracking as it caters for the variation in the object shape, appearance, and orientation. The EENC tracker also handles the intermittent fast occurring occlusion using Kalman filter [6], but it fails in case of slow occurring occlusion and strong clutter. In order to overcome these issues, we propose Edge-Enhanced Fragment Based Normalized Correlation (EEFNC), in which we divide the template into nine non-overlapping fragments. Then, we correlate every fragment of template with the corresponding fragment of the template-size section in the search region of the video frame. Furthermore, in order to address the varying appearance, shape, and orientation of the object and reject the effect of clutter and occlusion further, we update every fragment of the template independently.

The next section discusses in detail the proposed EEFNC tracking framework. Section III presents the experimental results. Finally, the conclusion is drawn in Section IV.

2. EEFNC Tracking Framework

The proposed Edge-Enhanced Fragment Based Normalized Correlation (EEFNC) tracking framework consists of edge-enhancement, fragment based normalized correlation, fragment level template updating, and Kalman predictor.

2.1 Edge Enhancement

The most commonly used similarity measure is normalized correlation coefficient (NCC), when the images to be correlated are gray-level images. However, it has been reported in [5] that normalized correlation (NC) is more robust than NCC, when the images to be correlated are edge-enhanced images. Therefore, we use the latter technique. The edge-enhancement procedure comprises: (1) RGB to gray level conversion as in [9] for reducing computational cost without significantly affecting the tracking performance, (2) Gaussian smoothing with adaptive standard deviation parameter as in [5] for attenuating noise without introducing noticeable blur, (3) gradient magnitude computation using Sobel edge detector masks in x and y directions as in [5], and (4) normalization to stretch the pixel values in the gradient images in the whole range of 0 to 255 as in [5], so the object may stand out in low contrast imagery.

2.2 Fragment Based Normalized Correlation

As a result of conventional normalized correlation, a correlation surface is developed which provides matching values between template *t*, and search image *s*, when the template is placed at every pixel of search image as [5]:

$$C(m,n) = \frac{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1} s(m+i,n+j).t(i,j)}{\sqrt{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1} s^2(m+i,n+j)}\sqrt{\sum_{i=0}^{K-1}\sum_{j=0}^{L-1} t^2(i,j)}}, \quad (1)$$

$F_{(0,0)}$	$F_{(0,1)}$	$F_{(0,2)}$
$F_{(1,0)}$	$F_{(1,1)}$	$F_{(1,2)}$
$F_{(2,0)}$	$F_{(2,1)}$	$F_{(2,2)}$

Figure 1. Nine non overlapping fragments, $F_{(a,b)}$ with 2D

where $C(m,n)$ is an element of correlation surface (matrix) at row m and column n, where m = 0, 1, 2,…M-K, n = 0, 1, 2,…N-L, and K and L are the height and width of the template, respectively.

In order to make the edge-enhanced correlation tracker more robust to strong clutter and slow occlusion, we propose to divide the edge-enhanced template and template-size patch in the search region into nine non-overlapping fragments, $F(a,b)$, where a =0, 1, 2 and b = 0, 1, 2, as depicted in Figure 1.

Then, we propose to correlate every fragment of the edge-enhanced template with the corresponding fragment of the current template-size patch in the edge-enhanced search image, and compute the average value of all nine correlation results to get the final correlation value at the position (m, n) in the search image. Mathematically, the fragment based normalized correlation can be formulated as in (2):

$$C(m,n) = \frac{\sum_{a=0}^{2}\sum_{b=0}^{2} C_{a,b}(m,n)}{9}, \quad (2)$$

where a = 0, 1, 2, b = 0, 1, 2, and Ca,b is the correlation value corresponding to fragment $F(a, b)$ computed as in (3), where h_a and w_b are the height and width, respectively, of the fragment at (a, b), and the sign ‘^’ represents logical AND operation. After obtaining the correlation surface, $C(m, n)$, we get the best-match location in the search image by finding the (m*, n*) position of the peak value, c_{max}, in $C(m, n)$

The basic difference between EENC and EEFNC is that of normalization technique used in the correlation. In EEFNC, the normalization effect is local to each fragment, thus producing better results than EENC in cluttered imagery. Furthermore, by dividing the template into fragments, the effects of occlusion also become local to each fragment and fragments that are affected from occlusion could be separated from non affected fragments.

Therefore, instead of updating the whole template, fragment level updating (discussed in the next sub-section) supports in occlusion handling.

$$C_{a,b}(m,n)=\begin{cases}\dfrac{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}s(m+i,n+j)\,t(i,j)}{\sqrt{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}s^2(m+i,n+j)}\sqrt{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}t^2(i,j)}}, & if\ (a=0)\wedge(b=0)\\ \dfrac{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}s(m+i,n+j+bw_{b-1})\,t(i,j+bw_{b-1})}{\sqrt{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}s^2(m+i,n+j+bw_{b-1})}\sqrt{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}t^2(i,j+bw_{b-1})}}, & if\ (a=0)\wedge(b>0)\\ \dfrac{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}s(m+i+ah_{a-1},n+j)\,t(i+ah_{a-1},j)}{\sqrt{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}s^2(m+i+ah_{a-1},n+j)}\sqrt{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}t^2(i+ah_{a-1},j)}}, & if\ (a>0)\wedge(b=0)\\ \dfrac{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}s(m+i+ah_{a-1},n+j+bw_{b-1})\,t(i+ah_{a-1},j+bw_{b-1})}{\sqrt{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}s^2(m+i+ah_{a-1},n+j+bw_{b-1})}\sqrt{\sum_{i=0}^{h_a-1}\sum_{j=0}^{w_b-1}t^2(i+ah_{a-1},j+bw_{b-1})}}, & otherwise\end{cases} \quad (3)$$

2.3 Fragment Level Template Updating

In order to make the template adaptive to the variation in the object shape and appearance in the real world scenarios, we must update the template. In [5], the template is updated as:

$$t[n+1]=\begin{cases}\lambda c_{\max}b[n]+(1-\lambda c_{\max})t[n] & if\ c_{\max}>\tau_t\\ t[n] & otherwise\end{cases}, \quad (4)$$

where $t[n]$ is the current template, $t[n+1]$ is the updated template for next iteration, $b[n]$ is the current best match section, and $C_{\max}$ is the peak value in the correlation surface. The value of λ is 0.16, and τ_t is the threshold of which value is 0.84.

We propose to update the fragments independently instead of updating the whole template, as:

$$F_{(a,b)}[n+1]=\begin{cases}\lambda f_{(a,b)\max}b_{(a,b)}[n]+(1-\lambda f_{(a,b)\max})F_{(a,b)}[n], & if\ (f_{(a,b)\max}>\tau_t)\wedge(c_{\max}>\tau_t)\\ F_{(a,b)}[n], & otherwise\end{cases}, \quad (5)$$

where $F_{(a,b)}[n]$ is fragment at (a,b) position of the current template, $b_{(a,b)}[n]$ is the fragment at (a,b) position of the current best-match, $f_{(a,b)\max}$ is the correlation value between $F_{(a,b)}[n]$ and $b_{(a,b)}[n]$, and $F_{(a,b)}[n+1]$ is the fragment at (a,b) position of the updated template to be used in the next iteration.

There are two differences between (4) and (5). Firstly, in (4) the updating is performed at template level, while in (5) the updating is performed at fragment level. Secondly, in (4) the template is updated if the best-match correlation value $C_{\max}$ is greater than threshold, τ_t, but in (5) the fragment level correlation value is also considered. This way the fragment containing major portion of the target is updated with higher weight while the fragment containing short-term background clutter or occluding object is updated with lower weight (or even not

updated). For better understanding, consider a situation when a slow moving object is occluding the target. In this situation, $C_{\max}$ will be higher than the threshold τ_t and the whole template will be allowed to update in case of EENC (in which case the occluded object will become part of the template, resulting in target loss later). However, in fragment level updating, the fragments that have been occluded will not be updated, because the fragment level correlation value is dropped below τ_t for the occluded fragments.

2.4 Kalman Predictor

EENC [5] has been strengthened by the use of Kalman predictor [6]. When the process of normalized correlation produces the peak value below τ_t, the target coordinates estimated by it are disregarded the target coordinates predicted by the Kalman filter in the previous iteration are utilized in the current iteration, and the process of template updating is bypassed. This technique provides support in case of occlusion. This advantage of Kalman predictor has also been exploited in the proposed EEFNC. We have used constant acceleration with random walk model with six states: position, velocity and acceleration in x and y directions. Furthermore, the position and the dimensions of the search window for the next iteration are also dynamically updated using the predicted position and its error, as in [5], to reduce the computational complexity and cater for object maneuvering with variable velocity.

3. Experimental Results

In this section, we compare EENC with EEFNC using different publicly available standard image sequences. We will further analyze the behavior of both the algorithms in the presence of clutter and occlusion using post regression analysis technique [14], which compares the calculated target coordinates with the ground truth target coordinates and provides three parameters m (regression slope), b (regression Y-intercept), and R (regression correlation coefficient). The ideal values of these parameters (when the calculated and the ground truth coordinates match perfectly) are $m = 1.0$, $b = 0.0$, and $R = 1.0$.

Table 1. Post regression results

Tracker	m	b	R
Walking Woman Sequence			
EENC	0.5825	75.03	0.6882
EEFNC	0.7530	41.78	0.8790
Three Men Crossing Sequence			
EENC	0.6428	43.74	0.6232
EEFNC	0.9710	7.98	0.9759
Shop Assistant Sequence			
EENC	-0.1026	148.98	-0.3732
EEFNC	0.9344	12.72	0.9309
F16 Take-off Sequence			
EENC	0.0514	95.6092	0.0830
EEFNC	0.9955	-10.349	0.9404

3.1. "Walking Woman" Sequence

The first experiment is performed on a publicly available *Walking Woman* image sequence [7]. The tracking results from both the algorithms are visually almost same, as shown in Figure 2. However, the accuracy of the target coordinates provided by the EEFNC is better than that of the target coordinates provided by the EENC, as illustrated in Table 1.

Figure 2. The first row presents the results of EENC and the second row presents the results of EEFNC. Both algorithms have visually performed well.

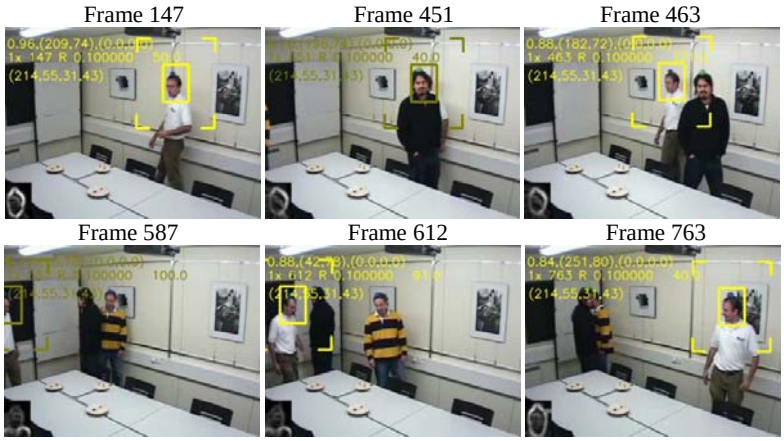

Figure 3. Tracking results of EEFNC for Three Men Crossing sequence

Figure 4. Tracking results of EENC for Three Men Crossing sequence

Figure 5. Tracking result of EENC is shown in the first row and tracking result of EEFNC is shown in the second row

3.2. "Three Men Crossing" Sequence

In the second experiment we test the trackers on *Three Men Crossing* sequence from AV16.3 v6 dataset[13], EEFNC has survived the occlusion and clutter, and has provided much longer tracking than EENC, as shown in Figs. 3 and 4. In Figure 3, the target (face of the person with white shirt) is being tracked by EEFNC successfully even during occlusion in Frame 451 and Frame 587. When an occlusion event is sensed automatically because the peak correlation value is dropped below the threshold, the color of the overlaying text and reticule is changed to golden from yellow for demonstration purpose. The occlusion is then handled using Kalman filter and the proposed fragment level template updating method. Once the object comes into view gain, the tracking is resumed in normal mode. Moreover, in Frame 763, the target is passing through background clutter; even then the tracking is not disturbed. If the tracking is performed using EENC along with its Kalman filter and template updating method, the tracking is lost after Frame 587, when the target is partially out of view, as shown in Figure 4. The regression analysis in Table 1 for *Three Men Crossing* Sequence also illustrates that the EEFNC tracker outperforms the EENC tracker also for this sequence.

3.3. "Shop Assistant" Sequence

Third experiment is performed on Shop Assistant sequence from the CAVIAR database [8]. Figure 5 (upper row) illustrates, that while tracking the person with dark shirt, the EENC did not survive the occlusion produced by the person with white shirt. EEFNC is, however, able to track the target object successfully even during and afterthe occlusion, as shown in Figure 5 (lower row). Table 1 also illustrates that the EEFNC performs much better than EENC in terms of tracking accuracy.

3.4. "F16 Take Off" Sequence

The fourth experiment has been performed on F-16 Take-off sequence, which has been used in [5], which proved robustness of EENC in heavily cluttered imagery. The same sequence has been used here to test EEFNC based tracking. It is observed that the robustness of EENC depends on how accurately the template is initialized. Typically, when we selected the template from initial frame and started the tracking session, the EENC tracker was disturbed by the clutter (white roof of small shed), and the track is lost, as shown in the first row in Figure. 3.6. However, when we selected the template of the same size from the same place in the initial frame, and started the EEFNC tracker, the F16 airplane was tracked

successfully throughout the whole image sequence, as shown in the last two rows in Figure. 3.6. Post regression results of EEFNC are also better than those of EENC as illustrated in Table 3.1.

4. Conclusion

We presented Edge Enhanced Fragment Based Normalized Correlation (EEFNC) algorithm and fragment level template updating method accompanied with Kalman filter to address the problems of strong clutter and

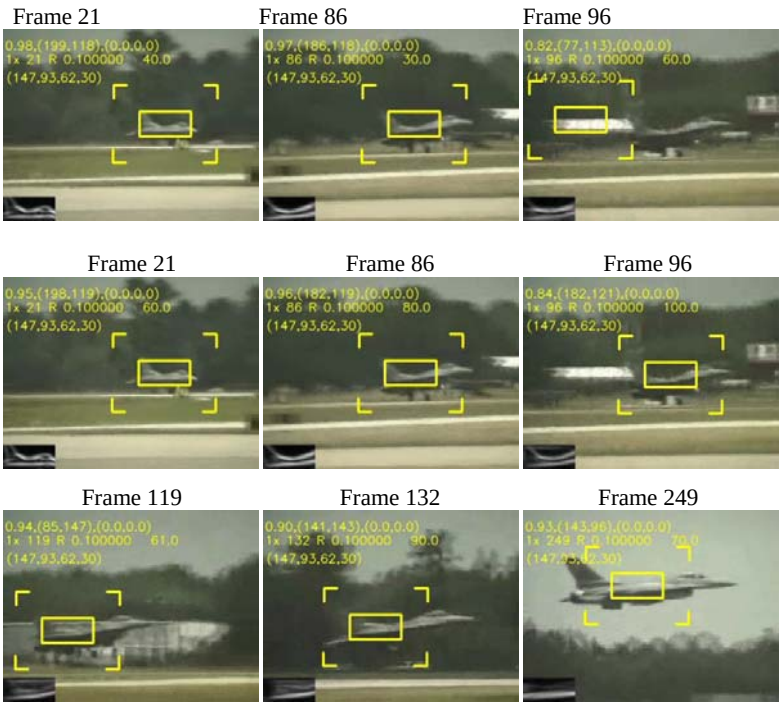

Figure 6. Tracking result of EENC is shown in first row and tracking result of EEFNC is shown in second row and third row

slow occlusion that the recent Edge Enhanced Normalized Correlation (EENC) method could not reliably handle. As far as the computational speed is concerned, the EENC works at the speed of about 75 fps [5] when the template size is typically 25×25 pixels. However, the proposed EEFNC for the same size template is about 25 fps, which was achieved when the search was performed using pyramid search technique up to two course levels. Although 25 fps is enough for the standard PAL cameras, the speed can be further increased using optimization techniques, if the higher frame rate cameras are used.

The concept of BMRA (Best Match Rectangle Adjustment) is presented in [15]. This technique adjusts the template size while minimizing background from the template and improves the tracking performance significantly. Use of BMRA with the proposed EEFNC algorithm can further enhance its performance. Furthermore, the exploitation of color components instead of single gray scale component can further make the EEFNC algorithm more robust to complex situations in which the clutter object look exactly same as the target object even when their color is different.

References

[1] A.J. Lipton, H. Fujiyoshi, R.S. Patil "Moving Target Classification and Tracking from Real-time Video," IEEE Workshop on Applications of Computer Vision, 1998
[2] B. Lucas, and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," 7th International Joint Conference on Artificial Intelligence (IJCAI), pp. 674-679, 1981.
[3] S. Wong, "Advanced Correlation Tracking of Objects in Cluttered Imagery," Proceedings of SPIE, Vol. 5810, 2005.
[4] C. Stauffer and W. Grimson, "Learning Patterns of Activity Using Real Time Tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 747-767, Aug. 2002
[5] J. Ahmed, M. N. Jafri, M. Shah, and M. Akbar, "Real-Time Edge-Enhanced Dynamic Correlation and Predic-tive Open-Loop Car Following Control for Robust Tracking," Machine Vision and Applications Journal, Vol. 19, No. 1, pp. 1–25, January 2008.
[6] R. E. Kalman, and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory,"Transactions of the ASME - Journal of Basic Engineering, Vol. 83, 1961.
[7] Adam, E. Rivlin, I. Shimshoni,. "Robust Fragments-based Tracking using the Integral Histogram", IEEE Conference on Computer Vision and Pattern Recognition, 17-22 June, 2006.
[8] Caviar datasets available at http://groups.inf.ed.ac.uk/vision/caviar/caviardata1/
[9] J. Ahmed, M. N. Jafri, J. Ahmad, and M. I. Khan, "Design and Implementation of a Neural Network for Real-Time Object Tracking," International Conference on Machine Vision and Pattern Recognition in Conjunction with 4th World Enformatika Conference, Istanbul, 2005.
[10] F. Porikli, "Integral histogram a fast way to extract histograms in cartesian spaces," IEEE Conference on Computer Vision and Pattern Recognition, 2005.
[11] D. Comaniciu, R. Visvanathan, P. Meer, "Kernel based object tracking". IEEE Trans. Pattern Anal. Mach. Intell. 25(5), 564–575, 2003.
[12] D. Comaniciu, V. Ramesh, P. Meer, "Real-time tracking of non rigid objects using mean shift". Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, vol. 1, pp. 142–149, 2000.
[13] G. Lathoud, J. Odobez, and D. Gatica-Perez, "AV16.3: An Audio-Visual Corpus for Speaker Localization and Tracking," IDIAP, Martigny, Switzerland, IDIAP-RR 28, 2004; MLMI 2004;LNCS 3361, pp. 182–195, Springer-Verlag Berlin Heidelberg, 2005.
[14] MATLAB 7.0 On-line Help Documentation
[15] J. Ahmed, M. N. Jafri, "Best-Match Rectangle Adjustment Algorithm for Persistent and Precise Correlation Tracking," Proc. IEEE International Conference on Machine Vision, Islamabad, Pakistan, on 28-29 December, 2007.

Authors

Mr Muhammad Imran Khan did his B.Sc.(Hons) in Computer Science from University of Engineering & Technology (UET), Lahore, Pakistan, in 2003 Currently he is involved in his MS Software Engineering from National University of Science & Technology (NUST), Islamabad, Pakistan. His current areas of interest are Image Processing, Computer Vision, Visual Tracking and algorithms.

Javed Ahmed received his BE (Electronics) from NED-UET, Karachi (Pakistan) in 1994. He did MSc (Systems Engg.) with 2nd position from PIEAS, Islamabad (Pakistan) in 1997. Then, he obtained his PhD in Electrical Engg. (Machine Vision) from NUST, Islamabad (Pakistan) in 2008. He has conducted 8-month joint research at Computer Vision Lab, UCF, USA, and served ICMV-07 conference as a member of its organizing committee and a co-chair of its technical program committee. He has got 11 papers published in premier conferences and journals (incl. CVPR-08, AAAI-07, and MVA). He has reviewed numerous papers submitted to various conferences and journals. His current research areas are image processing, machine vision, signal processing, and soft computing. He is a member of IEEE since 2005.

Mr. Ahmad Ali did his bachelor degree in Computer Sciences (Hons.) from University of Engineering & Technology (UET), Lahore, Pakistan. He completed his M.Sc. in System Engineering from Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan. His areas of interest are image processing, computer vision, object tracking, artificial intelligence, and speech processing.

Dr Asif Masood did his Software Engineering from National University of Sciences and Technology (NUST), Pakistan. He completed his MSc and PhD from University of Engineering and Technology (UET) Lahore, Pakistan in 2007. Currently he is working as Assistant Professor in Military College of Signals in NUST. There he is working on various projects related to image processing and computer vision. His areas of interest are image processing, computer graphics, and computational geometry.