IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 18, 2019, accepted December 9, 2019, date of publication December 13, 2019,
date of current version December 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959326

Image Local Features Description Through Polynomial Approximation

FAWAD[1], (Student Member, IEEE), MUHIBUR RAHMAN[2], MUHAMMAD JAMIL KHAN[1], (Member, IEEE), MUHAMMAD ADEEL ASGHAR[1], YASAR AMIN[1], (Senior Member, IEEE), SALMAN BADNAVA[3], AND SEYED SAJAD MIRJAVADI[4]

[1]ACTSENA Research Group, Telecommunication Engineering Department, University of Engineering and Technology Taxila, Taxila 47050, Pakistan
[2]Department of Electrical Engineering, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada
[3]Department of Computer Science and Engineering, College of Engineering, Qatar University, Doha, Qatar
[4]Department of Mechanical and Industrial Engineering, College of Engineering, Qatar University, Doha, Qatar

Corresponding authors: Muhibur Rahman (muhibur.rahman@polymtl.ca) and Salman Badnava (sb1107439@qu.edu.qa)

This work was supported by the Qatar National Library.

ABSTRACT This work introduces a novel local patch descriptor that remains invariant under varying conditions of orientation, viewpoint, scale, and illumination. The proposed descriptor incorporate polynomials of various degrees to approximate the local patch within the image. Before feature detection and approximation, the image micro-texture is eliminated through a guided image filter with the potential to preserve the edges of the objects. The rotation invariance is achieved by aligning the local patch around the Harris corner through the dominant orientation shift algorithm. Weighted threshold histogram equalization (WTHE) is employed to make the descriptor in-sensitive to illumination changes. The correlation coefficient is used instead of Euclidean distance to improve the matching accuracy. The proposed descriptor has been extensively evaluated on the Oxford's affine covariant regions dataset, and absolute and transition tilt dataset. The experimental results show that our proposed descriptor can categorize the feature with more distinctiveness in comparison to state-of-the-art descriptors.

INDEX TERMS Covariant, descriptor, handcrafted feature, patch, textures.

I. INTRODUCTION

Feature descriptor is used to describe the image interest region in such a way that the description remains robust against geometric and photometric transformations of the image. In many computer vision, applications feature descriptors play a vital role to describe the local neighborhood of a feature point within the image. Image local patch descriptor has a wide variety of applications in object recognition, such as plant species [1], blood cell [2], and fingerprint [3]. Moreover, the image retrieval [4], 3D scene reconstruction [5], and panoramic stitching [6] problems also utilize the local patch descriptors. The local features are employed to track [7], localize [8] objects in the image. The feature representation is required to be invariant in the presence of variations, such as scale, orientation, viewpoint, and illumination. Without any information about image condition, it is a very challenging task to identify the correct visual correspondence.

The associate editor coordinating the review of this manuscript and approving it for publication was Gangyi Jiang.

Image local features have been described using histogram of oriented gradient, that are scale invariant feature transform (SIFT) [9], visual orientation inhomogeneity SIFT (V-SIFT) [10], and gradient location and orientation histogram (GLOH) [11]. Histogram of oriented gradient based descriptors are scale and rotation invariant [12]. However, they are computationally expensive [13], also in-case of illumination difference and image blur, their performance degrades. Dominant SIFT [14], speeded up robust features (SURF) [15], and DAISY [16] is developed to reduce the computational complexity and feature dimensionality [17]. However, their matching performance is lower than SIFT. The edge-oriented histogram scale-invariant feature transform (EOH-SIFT) [18] is developed for face recognition that provides a better result than SIFT, but requires pre-processing and filtering. Unlike SIFT and its variants, moment-based invariant feature transform (MIFT) [19] remains insensitive to complicated shape deformation.

Illumination robust description have also been achieved by using the intensity orders, in local binary pattern

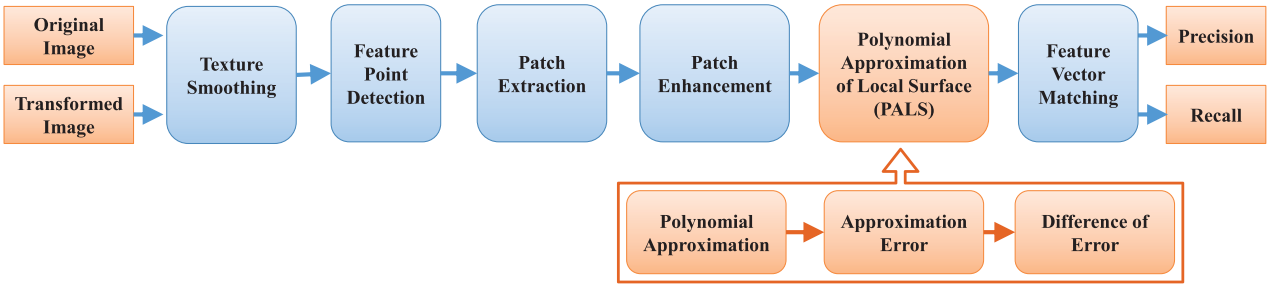


FIGURE 1. Flow chart of the proposed scheme.

(LBP) [20] orientation based descriptor, local intensity order pattern (LIOP) [21], histogram of relative intensities center-symmetric local ternary patterns (HRI-CSLTP) [22], local second-order statistics with soft-pooling (L2SSP) [23], and local tri-directional patterns (LTriDP) [24]. The intensity order based descriptors are very sensitive to noise in the input image. In [25] Overlapped Multi-oriented Tri-scale Local Binary Pattern (OMTLBP) is developed to bring invariance to illumination, scale and orientation for the noisy texture. However, the OMTLBP can not effectively represent the small local patches of the image.

The local patch representation get more robustness to variations when the deep learning method is adopted. In DeepDesc [26] the Siamese Network structure is trained by hard sample mining strategy to grab 28-dimensional feature vector from the local patch. Although DeepDesc provides excellent matching performance, it comes up with a substantial computational cost [5]. The visual geometry group (VGG) [27] integrate the convex optimization objective function with the sparse low-rank regularizer to describe the local image segment. The VGG descriptor is not suitable for realtime application due to its computational complexity. The L2 norm of the extracted features train the L2Net [28] by adopting progressive sampling strategy. The incremental sampling capacitate the L2Net to bring better performance in comparison to DeepDesc and VGG model. The overall limitation of deep learning mechanism is that it requires an enormous amount of ground truth training data in the form of matched and mismatched patches. Besides the training data requirement, the deep learning craves for expensive GPUs to meet their computational needs.

The descriptor is desired to have high discriminative power to retain its properties even in the presence of image transformations. Since natural images contain complex intrinsic structures and large extrinsic variations, therefore it a challenging problem to extract distinctive and robust features [29]–[31]. It is desirable that feature descriptors are invariant against the scale, rotation, translation, blur, illumination, and compression changes. This work focuses on the development of an improved image local descriptor, which is robust to compression noise, blurriness, illumination, and view-angle variations.

Object retrieval also requires stable key-points of the input image. Several feature detectors have been developed in the literature to detect corner, edges, and blobs. The detectors like Harris corner [32] are rotation invariant that are used to identify corners. The Harris-affine [33], and Hessian-affine [33] are also rotation invariant, but less efficient than Harris corner. The maximally stable extremal regions (MSER) [34] is a rotation, scale, and affine invariant, which is used to detect the feature points with high repeatability and efficiency, but it is sensitive to image blur. For rotation invariance, the Harris detector has the highest repeatability and localization accuracy as compared to other detectors.

The polynomial approximation can represent the image local interest region/patch verified by Savitzky-Golay filters [35] for both 1D and 2D signals. The difference of polynomial (DoP) [36] uses the approximation capability of the polynomial, has a better performance in different matching cases excluding JPEG compression. The work in [36] shows that the performance of DoP is better than SIFT and SURF for view angle variation, image blur and illumination difference in the image.

To bring improvement in the local patch representation under varying conditions of illumination and 3D view angle variation, image blur, and JPEG compression noise in the picture, we have introduced Polynomial Approximation of Local Surface (PALS) which provide a more distinctive representation of the local features. The feature vectors matching technique is modified to improve the classification accuracy.

II. PROPOSED DESCRIPTOR

In the proposed scheme, guided image filtering [37] is included to filter out the noise from the input image. Noise affects the performance of the orientation alignment algorithm. Harris corners are detected in the filtered image. A circular shape window instead of square shape window is used to extract the neighborhood of each key-point. The dominant orientation of the gradient aligns the patch orientation in the proposed method. The polynomial approximation approach is used to represent the extracted patch as shown in the flow diagram. The recall vs. (1-precision) plot is used to present and compare the performance of the proposed feature descriptor.

A. TEXTURE SMOOTHING

In the proposed scheme shown in Fig. 1, the image is first subjected to guided image filtering [37] to remove the noise

and unwanted textures while preserving the edges. The textured portion of the image contains those structures which are similar in shape. When the texture is not smoothed the Harris corner detects many key-points from the repeated ''Textons'' of the textured region. The ultimate drawback of the patch associated with the textured region is that the descriptor represents each patch with the similar feature vector, and the overall matching performance gets poor. Moreover, the special filtering operation improves the localization accuracy and distinctiveness in-case of blur images. The work in [38] has used L0 gradient minimization [39] to make the descriptor blur invariant. The L0 gradient minimization based filtering was applied, but better results were obtained by using guided image filtering [40] in the proposed scheme. The Harris corner is sensitive to noise in the image; therefore filtered images were subjected to the detector to get precise key-points or interest points.

$$I_i = \sum_j h_{ij}(J)(J_j) \tag{1}$$

$$h_{ij}(J) = \frac{1}{N_i} \exp\left(\frac{-\|c_i - c_j\|^2}{\sigma_c^2}\right) \exp\left(\frac{-\|J_i - J_j\|^2}{\sigma_J^2}\right) \tag{2}$$

In Eq. (1) I_i represents the filtered pixel value with coordinate i, J_j is the pixel values of the input image with coordinates j, J is the input image that is used as a guided image. The $h_{ij}(J)$ in Eq. (2) represent the filter kernel centered at position i and overlapping the pixels j, c_i and c_j represent the location of the center pixel and the neighbor pixel and σ^2 represent the variance of the gaussian function. The N_i in Eq. (2) is the normalization parameter to ensure that $\sum_j h_{ij} = 1$.

B. HARRIS CORNER DETECTION

The local image features are desired to be invariant to scale, orientation, translation, and illumination variations. The corner points that are the junction point of two edges are the stable regions of the image. Harris identified the corner points by using the gradient information of the image. All feature points that have significant variations in the gradient values have been selected. Harris extracts eigenvalues from the Hessian matrix, and then identify the feature point by using threshold value. The feature points of the images in the first and second row of Fig. 3 have been marked by the help of blue and green colors respectively. The pictures of each row belong to a different dataset shown the Fig. 3.

C. PATCH EXTRACTION

The interest points or key-points were detected using Harris corner in the filtered image I. The patch I_k in Eq. (3) around each key-point was extracted using a square window of specified 62 pixel side length as shown in Fig. 2(a). The x_k and y_k in Eq. (3) are the pixel coordinates around the k_{th} key-point of the input image.

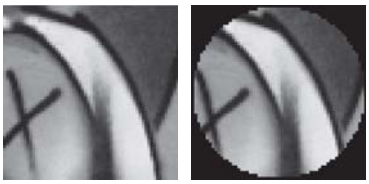


FIGURE 2. (a) Extracted patch around key-point with dimension 61 × 61 pixel (b) Extracted patch cropped using circular window of radius 30 Pixel.

$$I_k = I(x_k, y_k) \tag{3}$$

In the proposed scheme the patch around each key-point is cropped using circular window instead of using square window before the patch orientation detection. Due to the rotation of the image, the circular window can extract the same pixel at any orientation of the image as shown in Fig. 2(b). Circular window improves the orientation detection capability of the patch alignment scheme. The circular window discards the corner pixels of the square patch. This modification in the patch extraction scheme improves the performance of the feature descriptor. The structural representation of the feature exclude all corner pixels of the patch. The pixel at a uniform distance around the key-point is used to represent the local image patch.

D. PATCH ORIENTATION ALIGNMENT

In the proposed scheme rotation invarience is achieved by aligning patch using dominant orientation shift [48], [49]. The extracted patches are aligned using the dominant orientation of the gradient. The gradient of the cropped patch is calculated to see at which angle in the cropped patch there is a maximum intensity variation. The gradient magnitude and orientation of the k_{th} patch of the image is calculated using Eqs. (4) and (5), as shown at the bottom of this page. The m_k in Eq. (4) represents the magnitude and θ_k in Eq. (5) represent the orientation angle of the gradient. Histogram of gradient magnitude against angle θ_k is used to calculate the orientation of the patch. All gradient magnitude values that have the same orientation are added together to create the histogram bins. In Fig. 4, the normalized weighted gradient has a magnitude above the threshold level of 80% at angle 160 degrees, so the extracted patch will be described after rotating it by -160 degree. The patch is described for all

$$m_k = \sqrt{\Big(I_k(x+1,y) - I_k(x-1,y)\Big)^2 + \Big(I_k(1,y+1) - I_k(x,y-1)\Big)^2} \tag{4}$$

$$\theta_k = \tan^{-1}\left(\frac{I_k(1,y+1) - I_k(x,y-1)}{(I_k(x+1,y) - I_k(x-1,y)}\right) \tag{5}$$

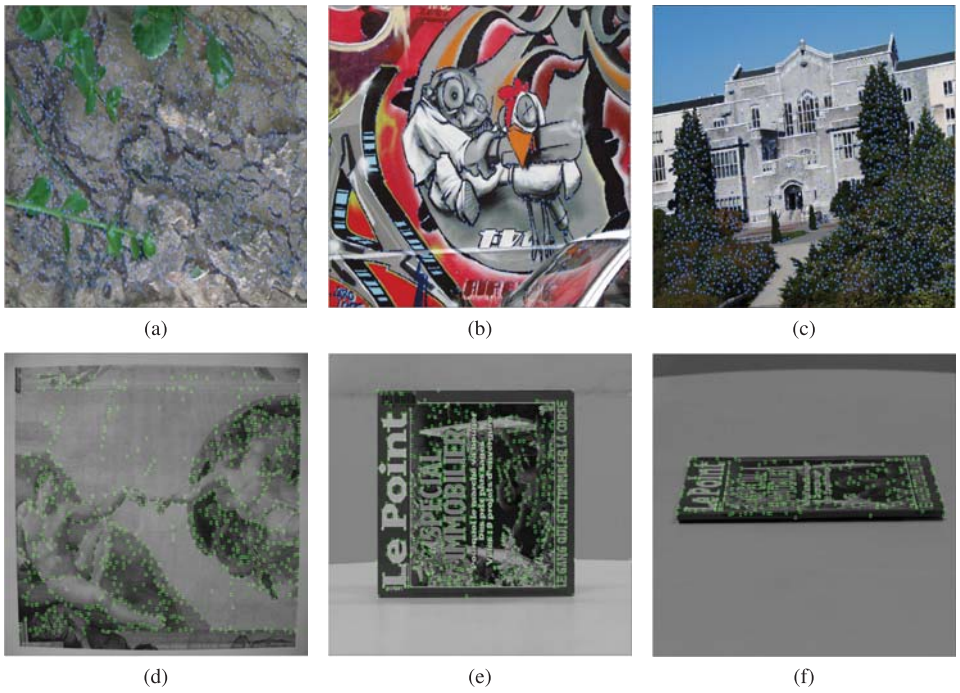


FIGURE 3. 1st row is the affine covariant features test dataset [11], while second row is absolute and transition tilts test dataset [47] (a) Ubc image (b) Graffiti (c) Buildings (e) Painting zoom x1 (f) Magazine zoom x4 (g) Magazine tilt t: 4.

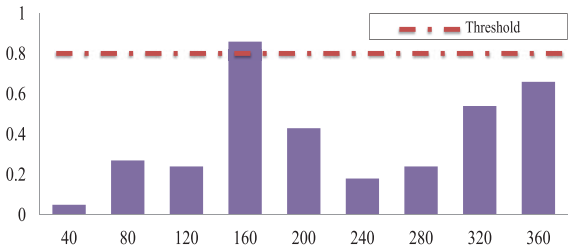


FIGURE 4. Histogram of dominant gradient orientation.

dominant orientations, in the case where more than one orientation has the gradient magnitude value above the defined threshold. An optimum value of the threshold is selected for orientation alignment. If we reduce the amount of the threshold, then we get many orientations for the patch alignment, which is not a realistic approach. The orientation of the patch is not changed, in the case where no gradient magnitude value is above the defined threshold value.

E. PATCH ENHANCEMENT

The extracted patches of the image are equalized by weighted threshold histogram equalization (WTHE) [41]. The Equalization through WTHE bring more invariance within the proposed PALS against illumination change. The local patches around the detected Harris key-point a equalized individually before the feature extraction process. The light intensity within the extracted patch get normalized and the similarity between identical structures of image increases by enhancing through WTHE scheme. Among various equalization techniques tested for the patch enhancement, the WTHE brings more robustness and distinctiveness in the proposed ASLA representation. The local patch was also subjected to layered difference representation (LDR) [42], contextual and variational contrast enhancement (CVC) [46], and histogram equalization (HE), but better results were obtained using WTHE. WTHE has improved the matching results even in the presence of illumination difference in the images. The local image patch is enhanced by the weighted threshold histogram equalization approach before the feature extraction process. Each input pixel intensity I_k, is enhanced individually through Eq. (6). The illumination difference within the patches is equalized through this approach to get similar patch representation even in different lighting conditions.

$$\acute{I}_k = (M-1).\mathcal{C}^{wt}(I_k) \tag{6}$$

$$\mathcal{C}^{wt}(I_k) = \sum_{j=0}^{I_k} P^{wt}(j) \tag{7}$$

$$P^{wt}(I_k) = \begin{cases} T_u, & \text{if} \quad P(I_k) > T_u \\ \left(\frac{P(I_k)-T_l}{T_u-T_l}\right)^r \times T_u, & \text{if} \quad T_l \le P(I_k) \le T_u \\ 0, & \text{if} \quad P(I_k) < T_l \end{cases} \tag{8}$$

The P^{wt} in Eq. (8) is the weighted and threshold probability distribution function, $\mathcal{C}^{wt}$ in Eq. (7) is the cumulative distribution function and $\acute{I}_k$ in Eq. (6) is the resulting enhanced intensity of the patch coordinate x_k. The T_u and T_l in Eq. (8) are the upper and lower threshold level of the probability.

The M in Eq. (6) represents the maximum gray-scale value in the input patch.

F. POLYNOMIAL APPROXIMATION

The $\acute{I}_k$ represents the enhanced k_{th} patch detected through Harris corner. The enhanced patch is approximated by using polynomial $f_r(x, y)$ of different degrees as shown in Eq. (9). The subscript r in Eq. (10) represents the degree value of the approximated polynomial.

$$f_r(x, y) = \left\{ \sum_{m+n=0}^{r} G_{m,n} x^m y^n \right\}_{r=0}^{N} \tag{9}$$

$$E_r = \left\{ \frac{1}{RC} \sum_{i=0}^{R-1} \sum_{j=0}^{C-1} \left(\acute{I}_k(x_i, y_j) - f_r(x_i, y_j) \right)^2 \right\}_{r=0}^{N} \tag{10}$$

$$d_{r-1,r} = \left\{ E_{r-1} - E_r \right\}_{r=1}^{N} \tag{11}$$

$$F = \left\{ d_{r-1,r} \right\}_{r=1}^{N} \tag{12}$$

The $G_{m,n}$ in Eq. (9) are the coefficients of the approximated polynomial. The mean square error E_r of the original interest region and the approximated patch is calculated for the polynomials of different degree as shown in Eq. (10). R and C in Eq. (10) represent the number of rows and columns of the patch respectively. The consecutive difference of the mean square error $d_{r-1,r}$ in Eq. (11) is used to represent the PALS feature vector F.

G. MATCHING

The normalized cross-correlation coefficient (NCC) is used instead of using the square sum of difference (SSD) to improve the matching accuracy. The reason for choosing the NCC is that it has a defined range of values that are [−1 1]. Whenever the features perfectly match each other, the NCC results into 1, similarly when features perfectly mismatch each other, the result is −1. The Nearest neighbor threshold selection for better matching results is quite simple in such a short span of values for the similarity index. However, in the case of Euclidean distance, the perfect match is represented by 0, while complete mismatch has no specific limited value. The Nearest neighbor threshold selection in such a case is quite tricky. Therefore we have used NCC in the nearest neighbor threshold matching process of the proposed descriptor, with a selected threshold of 0.75. Feature vectors of both images are matched using transformed normalized cross-correlation coefficient ρ^t.

$$\rho^t = \frac{1 - \rho\left(F_i^1, F_j^2\right)}{2} \tag{13}$$

The ρ in Eq. (13) is the normalized cross-correlation coefficient F_i^1 is the i_{th} feature vector of first image and F_j^2 is the j_{th} feature vector of the second image. 0 value of ρ^t represent perfect match and 1 value of ρ^t represent perfect mismatch. The matched pair is decided with the help of nearest neighbour matching ratio threshold. The ρ^t has lowest value for F_i^1 and F_j^2, while ρ^t has the second lowest value for F_i^1 and F_k^2. The F_k^2 in Eq. (14) represent all the k_{th} features extracted from the second image excluding the j_{th} feature vector only.

$$\mathcal{R}_{nn}\left(F_i^1, F_j^2\right) = \frac{\rho^t\left(F_i^1, F_j^2\right)}{\min\left\{\rho^t\left(F_i^1, F_k^2\right)\right\}} \quad \text{for} \quad j \neq k \tag{14}$$

The key-points of the respective feature vectors are matched pairs if the nearest neighbor ratio $\mathcal{R}_{nn}$ of the transformed correlation coefficient ρ^t in Eq. (14) is below the threshold τ. The τ is kept as 0.75 to decide the matching pair of the key-points within the images. The Fig. 7 the τ has been varied from the range 0 to 1, and the recall vs. (1-precision) plots are drawn, it is observed that 0.75 is an optimum value of the threshold which provides a better matching. The τ if further reduced, will increase the False Negatives in the matching set, similarly if the tau is increased, it increases False Positives in the matching set.

$$M(i, j) = \begin{cases} \left(KP_i^1, KP_j^2\right), & \text{if } \mathcal{R}_{nn} < \tau \\ 0, & \text{if } \mathcal{R}_{nn} \geq \tau \end{cases} \tag{15}$$

In Eq. (15), the KP represents the detected key-point or commonly known as the feature points. The subscript i and j of the term KP in Eq. (15) denote the index of the feature point, while the superscript 1 and 2 pointing the first and second image respectively.

H. EVALUATION

A matched pair $M(i, j)$ in Eq. (15) is the correct match if the transformed patch associated with the key-points KP_i^1 and the patch I_j^2 associated with KP_j^2 has overlap above 50%. The coordinates of the patch I_i^1 are transformed by the 3×3 homography matrix of the database to get the transformed patch. In the proposed scheme the evaluation is performed using recall vs. (1-precision) as shown in Fig. 7. The evaluation parameters, that are precision and recall have been described with Eqs. (16) and (17) respectively.

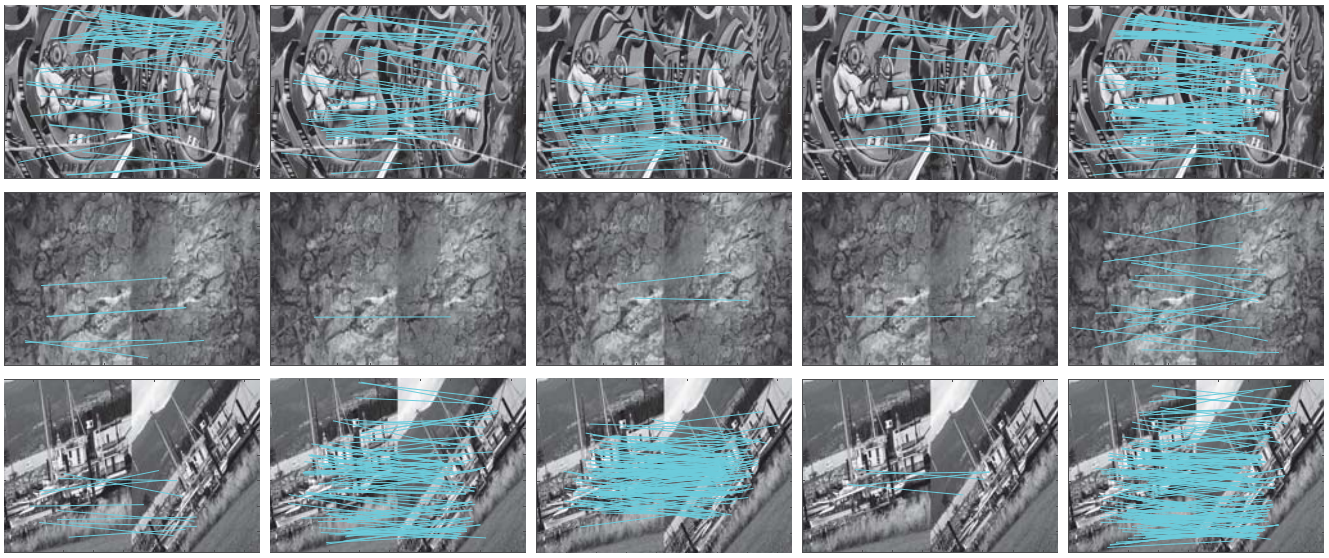
$$precision = \frac{\#correctmatchings}{\#matchings} \tag{16}$$

$$recall = \frac{\#correctmatchings}{\#correspondences} \tag{17}$$

The correct matches are all those estimated matched pairs, that can satisfy the homographic transformation of the dataset. Feature homography transformation matrix can be obtained by random sample consensus (RANSAC) algorithm, if not available with the test dataset. Number of correspondence in Eq. (17) is the total of matched pairs that exist in the test image pair.

III. SIMULATIONS AND RESULTS

The recall vs. (1-precision) plots is used as the descriptor evaluation criterion [11]. (1-precision) is the ratio of incorrect matches to the total number of matches, and recall is the ratio of the correct match to the correspondences. Correspondence is calculated using the homographic transformation matrix



(a) DoG (b) Fast (c) BRISK (d) Shi-Tomasi (e) Harris Corner

FIGURE 5. Matching performance of the proposed ASLA descriptor with various combinations of feature detectors tested on Oxford affine covariant dataset [11], the first row is the Graf sequence with 20 degrees view angle difference, the second row is the matching results on bark sequence with difference in view angle and scale, the third row is the boat sequence with 20 degrees orientation and 2x zoom level.

associated with each image pair. The key-points coordinate of the image is transformed using the homography matrix. For feature correspondence, the patch overlap should be higher than 50%. The recall vs. (1-precision) plot is calculated by varying the threshold value τ between 0 and 1. Each value of threshold results in a new value of precision and recall that are used to plot the performance graph of the feature descriptor. The descriptor is evaluated using two benchmark datasets, that are the affine covariant regions and absolute and transition tilt dataset. The execution time of the proposed descriptor is 407 ms when tested on 1000 key-points. The proposed approach has 41ms, and 1397 ms lower computation time than SIFT and LIOP, respectively.

The proposed descriptor, in combination with various types of feature point detectors, is tested on the Affine covariant dataset, as shown in Fig. 5. The neighborhood pixels of the detector like difference of Gaussian (DoG) [9], features from accelerated segment test (FAST) [43], binary robust invariant scalable keypoints (BRISK) [44], Shi-Tomasi [45], and Harris corner [32] are described to evaluate the better performance in each case. Fig. 5 shows the Shi-Tomasi and DoG both results in a poor performance in the case of orientation and scale changes. Moreover, in case of view angle changes, the performance of FAST and BRISK also fails to provide satisfactory performance. The Harris corner outperforms the other detector, as shown in Fig. 5.

The influence of parameters such as patch radius, matching threshold, and the degree of the polynomial, on the feature matching performance, is examined. The matching of the proposed descriptor is evaluated for various radii of the patch. The mean value of the correct correspondence achieved in the presence of a view angle difference from 10 to 50 degrees is presented in Fig. 8. The highest number of mean correct matched pairs have been found at the 25-pixel size of the patch radius. The total number of correctly matched pairs for various degrees of the proposed polynomial is shown in Fig. 9. It can be seen that the polynomial with degrees 0-4 provides the highest correct matching in comparison to other variants of the proposed descriptor. Due to the overfitting problem, the polynomials with a degree higher than four produce less number of correct matches. The higher degree polynomials overfit the surface of the local image patch, which reduces the distinctiveness in the feature representation.

A. ROBUSTNESS TO DISTORTIONS

The descriptor is evaluated using the images of affine covariant regions database [11]. The robustness of the descriptor to the 1) viewpoint change 2) image blur 3) illumination changes and 4) JPEG compression is studied. The key-points are detected using Harris corner, the patches with radius size 31 pixel is used for the feature extraction. The matching results of the proposed scheme are compared with DAISY, DoP, GLOH, HRI-CSLTP, L2SSP, LIOP, SIFT, and VSIFT as shown in Fig. 7. The proposed descriptor of 0-4 degree polynomial with dimension 200 variables provides better results in comparison to other feature descriptors. The descriptors evaluated in recall vs. (1-precision) plot have different similarity measures. As mentioned in the literature HRI-CSLTP, LIOP, GLOH, SIFT, V-SIFT, and SURF use Euclidean distance-based matching, while L2SSP, DAISY, and DoP use the correlation-based matching technique.

To evaluate the robustness of the descriptor on blur images, we have used the blur image sequence ''bikes'' and ''trees'' of

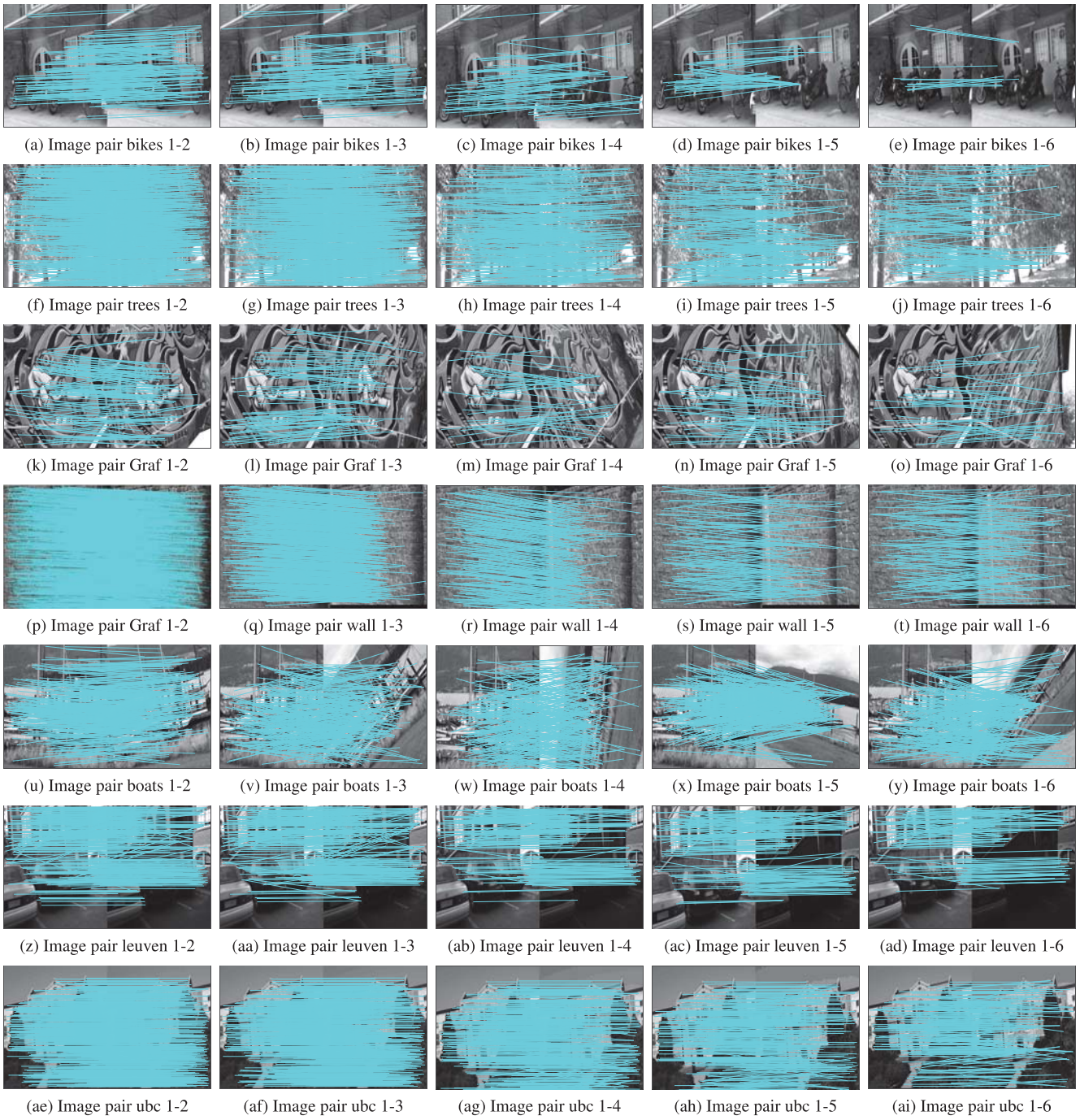
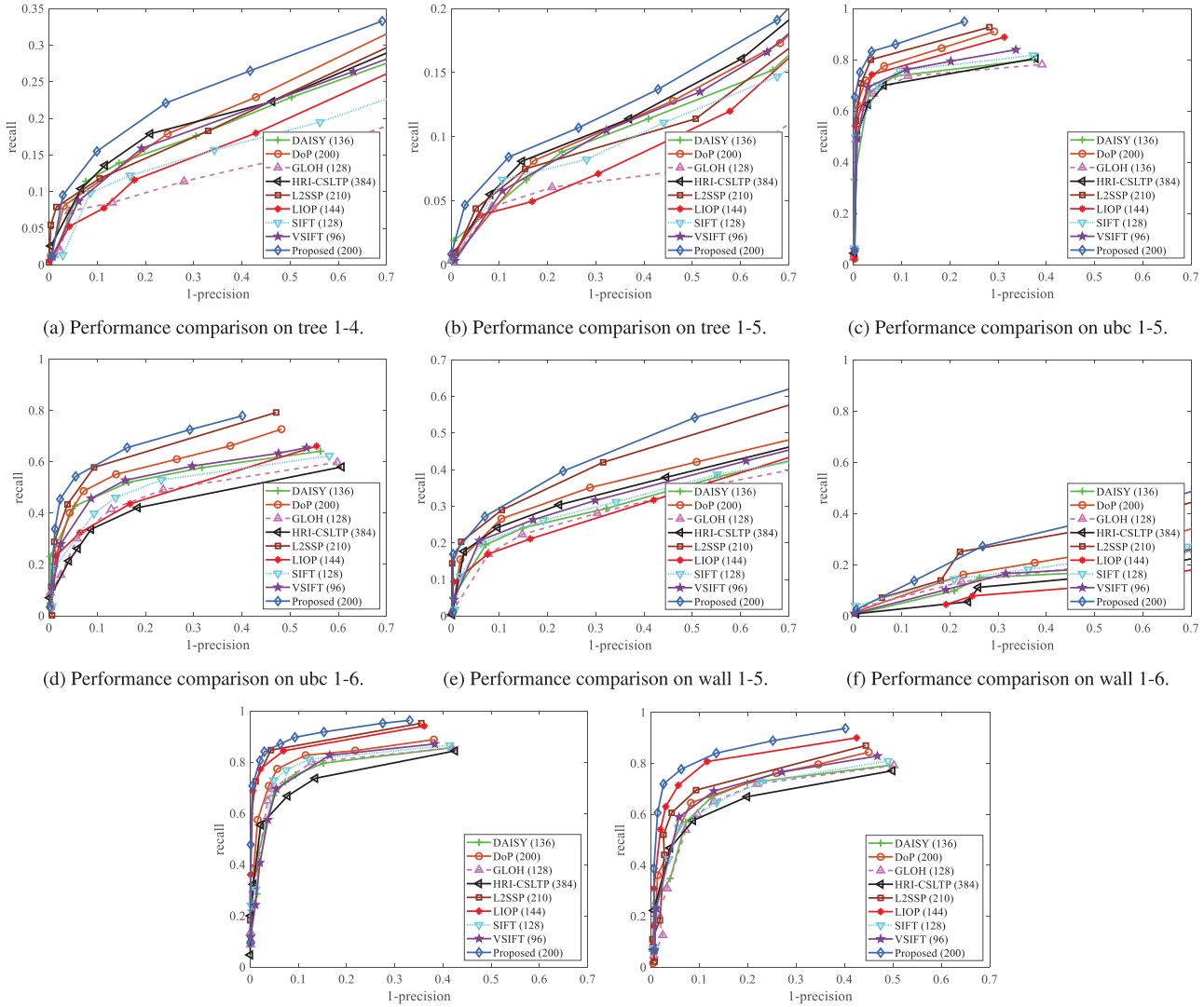


FIGURE 6. **Matching performance of the proposed ASLA descriptor on Oxford affine covariant dataset [11].**

the Oxford dataset [11]. The blur sequences in the dataset are acquired by changing the camera focus. The image sequence contains blur images with dimension 1000 × 700 pixel. The proposed descriptor is evaluated by using complete set of image pairs available within the dataset as shown in Fig. 6(a) - (j). Each image pair has a different intensity of blurriness specified by the image label. The matched pair of the feature point has been highlighted through the cyan color line with both ends pointing towards the corresponding feature coordinates of the image pair. The performance of the proposed method is superior to all other state of the art methods as shown in the Fig. 7(a) and (b) for the tree sequence 1-4 and 1-5.

The invariance of the proposed descriptor to view angle variation is evaluated by using the complete ''wall'' and ''graf'' sequence from the Oxford dataset. The dataset contains six different images with camera viewpoint varies from a fronto-parallel view to one with significant foreshortening at



(a) Performance comparison on tree 1-4.

(b) Performance comparison on tree 1-5.

(c) Performance comparison on ubc 1-5.

(d) Performance comparison on ubc 1-6.

(e) Performance comparison on wall 1-5.

(f) Performance comparison on wall 1-6.

(g) Performance comparison on Leuvin 1-2.

(h) Performance comparison on Leuvin 1-3.

FIGURE 7. Descriptor performance evaluation on Oxford dataset [11].

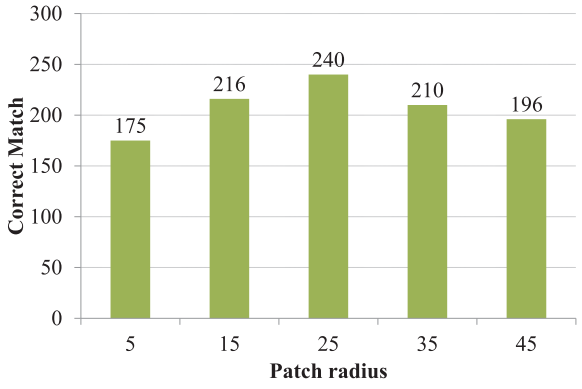


FIGURE 8. Proposed descriptor's average matching results on different radii of the input patch using the from graf sequence of Oxford database.

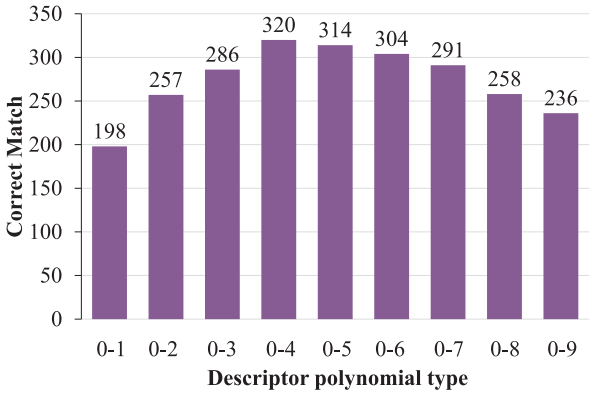
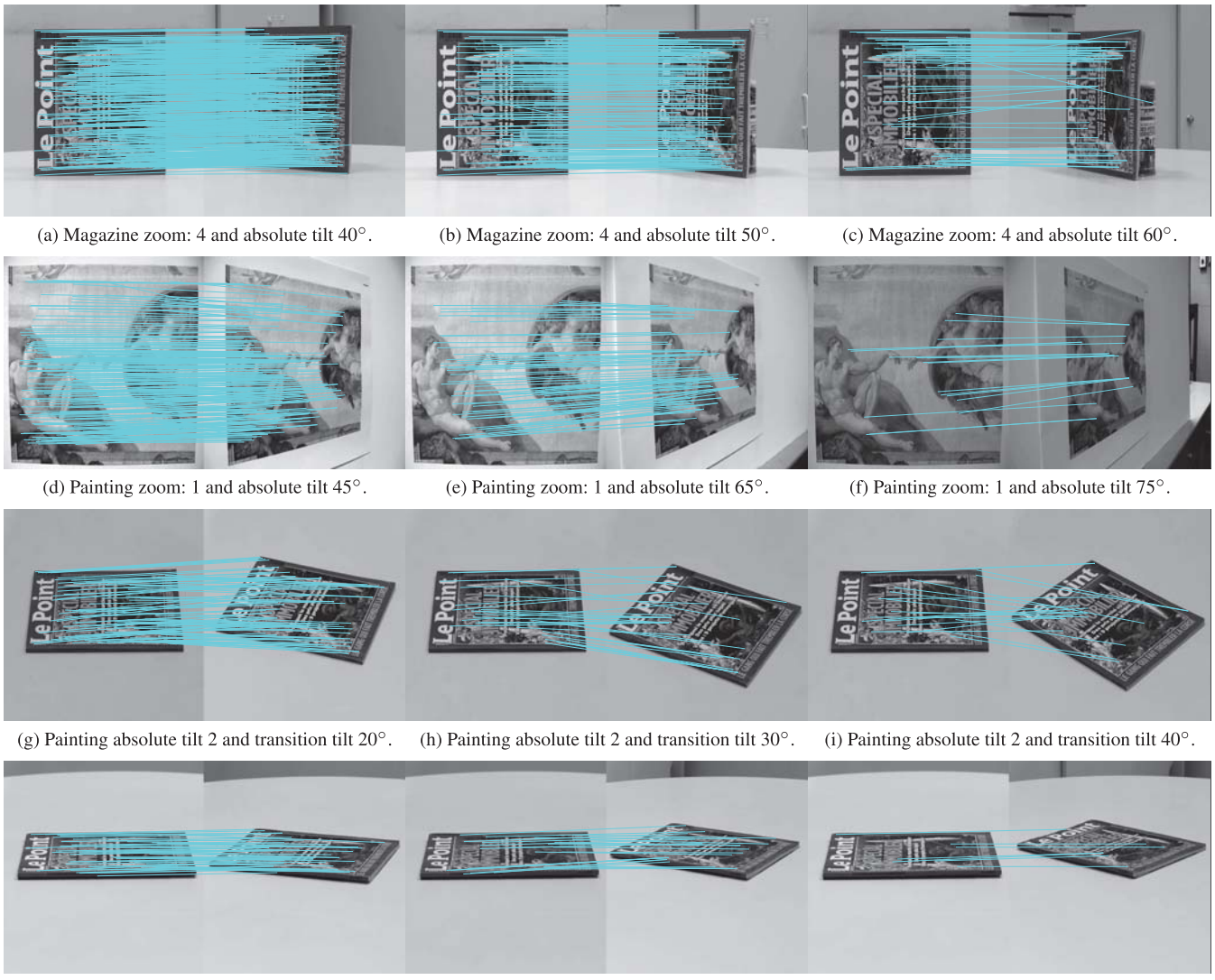


FIGURE 9. The matching performance on different degrees of the polynomial is presented in this figure.

approximately 60 degrees to the camera. The complete image set from "wall" and "graf" sequence of Oxford dataset shown in the Fig. 6(k) - (t) is used to evaluate the descriptor. The image sequence Wall 1-5 and 1-6 have a view angle



(a) Magazine zoom: 4 and absolute tilt 40°. (b) Magazine zoom: 4 and absolute tilt 50°. (c) Magazine zoom: 4 and absolute tilt 60°.

(d) Painting zoom: 1 and absolute tilt 45°. (e) Painting zoom: 1 and absolute tilt 65°. (f) Painting zoom: 1 and absolute tilt 75°.

(g) Painting absolute tilt 2 and transition tilt 20°. (h) Painting absolute tilt 2 and transition tilt 30°. (i) Painting absolute tilt 2 and transition tilt 40°.

(j) Painting with absolute tilt 4 and transition tilt 20°. (k) Painting with absolute tilt 4 and transition tilt 30°. (l) Painting with absolute tilt 4 and transition tilt 40°.

FIGURE 10. **Proposed descriptor's performance on absolute tilt and transition dataset [47].**

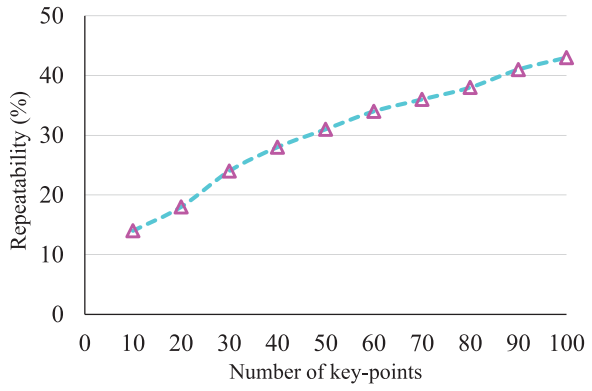


FIGURE 11. **The influence of key-points on the performance of the method.**

difference of 50 and 60 degrees. The cyan color line shows the feature correspondence. The performance graph of the proposed descriptor shown in the Fig. 7(e) and (f), out-forms the state of the art descriptors. The matching results of the "boat" sequence are evaluated in the presence of scale and rotation variations in the images. The scale changes to a factor of four with orientation difference up to 60 degrees in the image pair shown in Fig. 6 (u) - (y).

The Leuvin sequence of the Oxford dataset is used to measure the illumination invariance of the proposed descriptor. The dataset [11] consists of six images with different intensity of light. The light variation in the dataset is introduced by varying the camera aperture. The image sequence of Leuvin with all images pairs collected from the Oxford dataset has been used to evaluate the proposed descriptor. The matched key-point coordinates represented using cyan color lines shown in the Fig. 6(z) - (ad). The performance graph shown in Fig. 7(g) and (h) represents the proposed method gives better performance as compared to other techniques.

TABLE 1. Total number of correct matching in Oxford dataset [11].

Categories		Descriptors									
		SURF (64)	SIFT (128)	V-SIFT (96)	DoP (200)	GLOH (128)	DAISY (136)	HRI-CSLTP (384)	LIOP (144)	L2SSP (201)	Proposed (200)
Image blur	Image 1-2	303	367	395	453	344	381	401	394	447	**498**
	Image 1-3	283	338	389	399	327	348	356	351	397	**438**
Bikes Sequence	Image 1-4	213	266	275	298	218	285	299	292	326	**378**
	Image 1-5	147	190	198	202	156	224	238	229	278	**317**
	Image 1-6	107	143	151	173	124	168	177	165	197	**243**
Image blur	Image 1-2	412	457	472	513	422	425	412	438	441	**536**
	Image 1-3	340	357	363	402	342	332	336	341	387	**449**
Trees Sequence	Image 1-4	237	273	278	311	208	285	289	283	333	**389**
	Image 1-5	153	197	208	227	163	232	242	232	289	**323**
	Image 1-6	129	157	166	182	133	175	182	171	188	**255**
Viewpoint	Image 1-2	355	352	366	373	359	379	390	370	386	**399**
	Image 1-3	188	207	203	218	215	217	219	221	224	**231**
Graf Sequence	Image 1-4	32	51	64	77	64	73	72	76	78	**84**
	Image 1-5	4	8	9	13	7	9	8	11	13	**17**
	Image 1-6	2	1	2	3	2	2	3	3	4	**5**
Viewpoint	Image 1-2	347	372	370	387	362	381	384	389	391	**403**
	Image 1-3	176	200	205	213	203	214	216	218	220	**226**
Wall Sequence	Image 1-4	28	49	66	73	63	70	69	72	71	**78**
	Image 1-5	3	6	8	10	6	8	7	9	10	**13**
	Image 1-6	0	0	1	2	1	1	2	2	3	**4**
Zoom + Rotate	Image 1-2	280	266	294	342	263	270	313	284	342	**351**
	Image 1-3	232	212	239	315	234	246	282	249	327	**329**
Boat Sequence	Image 1-4	183	168	205	158	128	135	143	129	206	**207**
	Image 1-5	117	150	148	162	116	164	108	101	178	**182**
	Image 1-6	87	53	91	89	94	99	107	101	117	**119**
Illumination	Image 1-2	294	329	348	345	301	363	354	432	433	**531**
	Image 1-3	235	295	325	356	288	307	298	364	372	**404**
Leuven Sequence	Image 1-4	193	227	244	235	206	224	214	238	254	**274**
	Image 1-5	180	208	231	256	196	228	203	229	232	**261**
	Image 1-6	148	178	191	189	165	184	172	198	203	**218**
JPEG compression	Image 1-2	630	858	980	868	823	881	840	920	987	**1040**
	Image 1-3	598	692	792	727	640	690	650	734	796	**820**
Ubc Sequence	Image 1-4	401	469	543	487	411	488	443	504	549	**558**
	Image 1-5	182	277	285	289	230	298	281	287	298	**372**
	Image 1-6	98	137	132	156	104	154	128	162	187	**193**

The robustness of the proposed descriptor to compression noise is evaluated on JPEG compressed images. The JPEG image pairs of the ''ubc'' 1-5 sequence collected from Oxford dataset shown in the Fig. 6(ae) - (ai) are used to evaluate the proposed descriptor. The JPEG sequence is generated by changing the image quality parameter from 40% to 2% through a standard xv image browser. The image sequence contains compressed images with dimension 800×640 pixel. The matched features have been highlighted through the cyan color line with both ends pointing towards the corresponding feature points. The proposed descriptor provides high recall vs. (1 - precision) in comparison to DAISY, DoP, GLOH, HRI-CSLTP, L2SSP, LIOP, SIFT, and VSIFT as shown in the Fig. 7(c) and (d).

The homographic transformation matrix is associated with each image pair in the Oxford dataset, that is used to identify the correct match in the estimated correspondence. The Table 1 shows the comparison of various feature descriptors. The proposed approach provides more recognition as compared to other techniques. In Table 1 the proposed technique is compared with SURF, SIFT, V-SIFT, DoP, GLOH, DAISY, HRI-CSLTP, LIOP, and L2SSP, with feature dimensionality 64, 128, 96, 200, 128, 136, 384, 144, and 201 variables respectively. The number of Harris corner points depends on the threshold value of the corner response. The larger the threshold value, fewer feature points will be detected, and vice versa. The key-points identified through Harris corner in Fig. 3, show the smooth region of the image lacks the presence of key-points. A uniform set of key-points has been detected by changing the threshold value on the Harris corner response. The feature repeatability of the ''Wall'' sequence presented in Fig. 11 shows an increase with the increase in feature points.

B. INVARIANCE TO ABSOLUTE AND TRANSITION TILTS

The descriptor performance is also evaluated using tilt test dataset [47]. The circular patch with radius size 61 pixel is

TABLE 2. Total number of correct matched feature points in absolute tilt test suit [47].

θ (°)	Zoom × 1					Zoom × 10				
	SURF	SIFT	V-SIFT	DoP	Proposed	SURF	SIFT	V-SIFT	DoP	Proposed
+45	96	153	173	177	**184**	69	95	115	116	**129**
-45	74	108	120	125	**129**	74	118	128	133	**137**
+65	44	56	58	61	**65**	11	14	12	12	**20**
-65	48	56	74	76	**79**	3	4	8	9	**12**
+75	4	8	17	22	**23**	2	3	3	4	**5**
-75	5	10	23	24	**25**	1	2	3	4	**4**
+80	2	2	3	5	**6**	0	3	1	2	**3**
-80	3	5	3	4	**4**	0	2	1	2	**2**

TABLE 3. Total correct matched feature points in transition tilt test suit [47].

ϕ (°)	t = 2			t = 4		
	SIFT	V-SIFT	Proposed	SIFT	V-SIFT	Proposed
10	166	175	**190**	15	23	**31**
20	25	25	**29**	11	15	**18**
30	4	4	**6**	3	4	**7**
40	2	4	**5**	1	1	**4**
50	1	0	**3**	1	1	**3**
60	2	2	**3**	0	0	**1**
70	1	1	**2**	0	0	**0**
80	0	0	**1**	0	0	**0**
90	2	2	**2**	0	0	**0**

represented using the polynomial of 0-4 degree. The feature correspondence of the images shown in Fig. 10 belongs to absolute tilt and transition test suits of the image dataset (systematic evaluation of robustness to absolute and transition tilts). The correspondence map of the proposed descriptor shown in Fig. 10 have variations in absolute tilt and transition angle. In Table 2 the total number of correct matching of the proposed approach is shown in comparison to SIFT and V-SIFT by using the absolute tilt test suit of the dataset. In Table 2 the results have been analyzed for the tilt angle $\theta = \pm 45°, \pm 65°, \pm 75°$, and $\pm 80°$ along with optical zoom of 1 and 10. In all cases shown in Table 2 the proposed feature representation techniques give the highest number of correct identification of the corresponding feature points. In transition tilt test suit, two fixed latitude angle θ, with t = 2 and 4 have been set to capture test images with longitude angle ϕ varies from 0° to 90°. The absolute tilt t = 2 and 4 in the transition tilt test suit corresponds to latitude angle $\theta = 60°$ and 75° respectively. Table 3 show that even the performance of all the descriptors degrades with an increase in longitude angle ϕ, but the proposed descriptor provides better recognition in comparison to SURF, SIFT, V-SIFT, DoP, GLOH, DAISY, HRI-CSLTP, LIOP, and L2SSP.

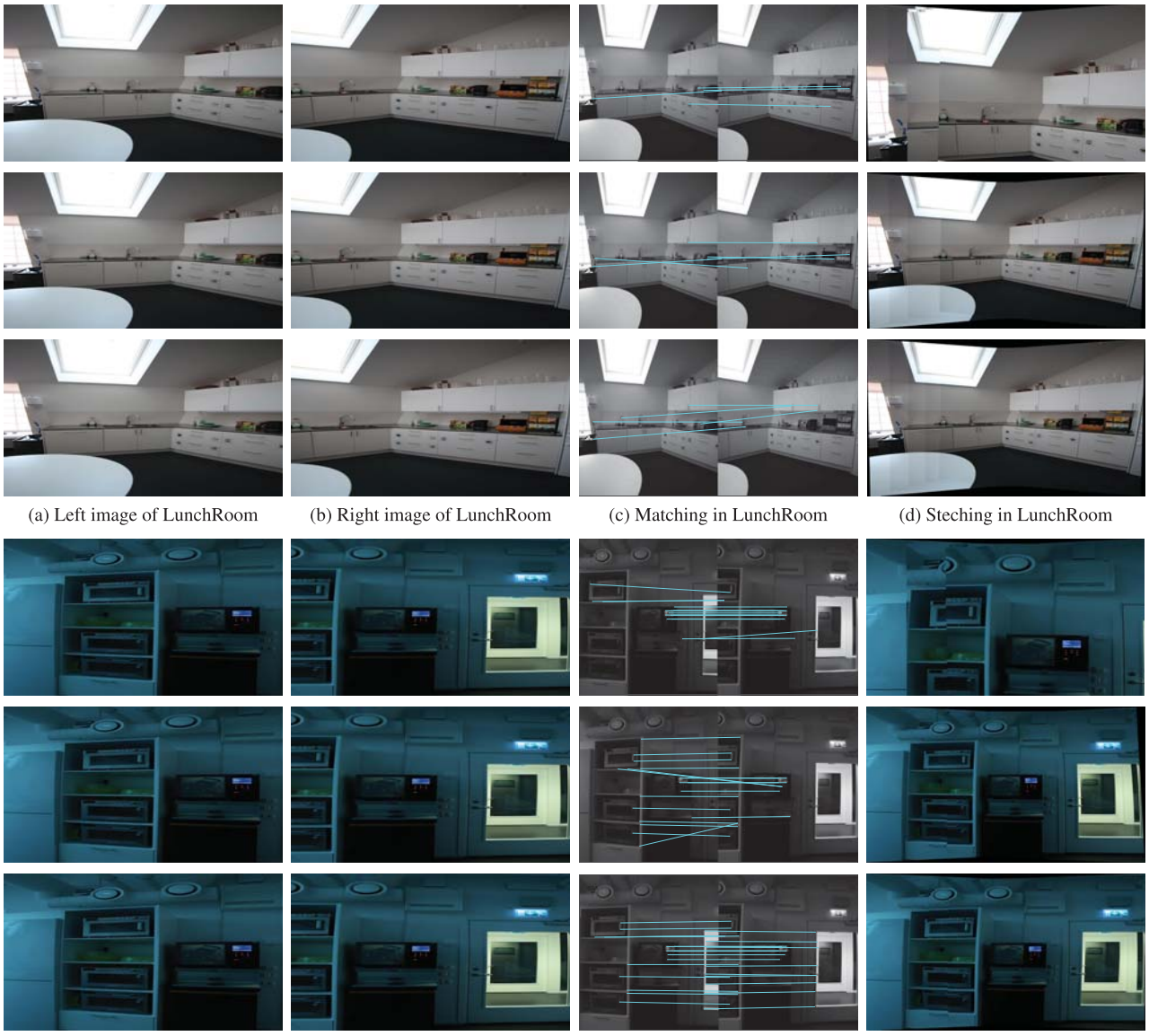
Table 4 compare the proposed descriptor feature dimensionality with other recently reported works. The proposed descriptor has a dimensionality of 200 variables within its feature vector. The proposed feature vector has better performance while lower dimension as compared to HRI-CSLTP, and L2SSP. The proposed descriptor provides more recognition in comparison to SIFT, V-SIFT, SURF, DoP, GLOH, DAISY, and LIOP; however, the feature dimension of the proposed descriptor is higher than the mentioned techniques.

TABLE 4. A comparison of the feature dimensionality with state-of-the-art methods.

Descriptor	Dimensionality	Description time (ms) (1000 key-points)
HRI-CSLTP [22]	384	725
L2SSP [23]	201	848
LIOP [21]	144	1804
DAISY [16]	136	409
GLOH [11]	128	922
SIFT [9]	128	848
V-SIFT [10]	96	731
SURF [15]	64	419
Proposed	**200**	**407**

IV. APPLICATIONS

The local features are essential in various applications of computer vision. The visual trackers, panorama stitching, and 3D reconstruction of the visual scene all require a robust descriptor. The proposed ASLA descriptor is tested on PAnorama Sparsely STructured Areas (PASSTA) [50] dataset for the panorama stitching of translated images. Fig. 12, shows the achieved results after stitching. Fig. 12(a)-(d) show the left, right, matched, and stitched images. The images harris key-points are matched through the proposed descriptor, and the matching results are used to calculate the homography by Random sample consensus (RANSAC). The set of matched key-points and the homographic transformation matrix are used to stitched the translated images together. The panorama stitching results of the GLOH, and SIFT descriptor in comparison to the proposed descriptor shown in Fig. 12. The images in the first three rows are from Lunchroom, while the remaining three rows are from the LunchroomBlue sequence of the PASSTA dataset. The stitching performance of GLOH, SIFT, and the proposed descriptor is shown in the first, second, and third row of Fig. 12. The GLOH descriptor stitching has discontinuities demonstrated in the last column of both sequences. However, the proposed descriptor stitching is smooth in comparison to SIFT and GLOH.



(a) Left image of LunchRoom (b) Right image of LunchRoom (c) Matching in LunchRoom (d) Steching in LunchRoom

(e) left image of LunchRoomBlue (f) Right image of LunchRoomBlue (g) Matching in LunchRoomBlue (h) Steching in LunchRoomBlue

FIGURE 12. A comparison of various descriptor in panorama stitching validated on PASSTA dataset. [50].

V. CONCLUSION

A novel local feature descriptor, Polynomial Approximation of Local Surface (PALS), is developed in this work. The proposed PALS is invariant to the changes in orientation, scale, viewpoint, and illumination. The pre-processing through WTHE, and guided filtering provide robustness against blurriness, illumination difference, and JPEG compression noise. The PALS crops the ROI using a circular mask followed by a dominant orientation shift that imparts rotation invariance to the local features representation. The feature descriptor matching is archived with transformed normalize correlation coefficient instead of using the square sum of difference, that has improved the matching accuracy. When evaluated on Oxford Affine Covariant and Absolute Tilt and Transition dataset the proposed PALS descriptor offers a high recognition rate under conditions of geometric and photometric variations. The proposed PALS holds potential for improvement across a broad array of applications related to local feature matching including panorama stitching, 3D reconstruction, facial landmark representation, and so on.

ACKNOWLEDGMENT

S. S. Mirjavadi also appreciates the help from the Fidar Project Qaem Company (FPQ).

REFERENCES

[1] J. Wäldchen and P. Mäder, "Plant species identification using computer vision techniques: A systematic literature review," *Arch. Comput. Methods Eng.*, vol. 25, no. 2, pp. 507–543, 2018.

[2] D. Lopez-Puigdollers, V. J. Traver, and F. Pla, "Recognizing white blood cells with local image descriptors," *Expert Syst. Appl.*, vol. 115, pp. 695–708, Jan. 2019.

[3] R. Pollak and R. Richter, "Fingerprint indexing via BRIEF minutia descriptors," *Proc. IJBM*, vol. 11, no. 2, pp. 101–112, 2019.
[4] Q. B. Dang, M. Coustaty, M. M. Luqman, and J. M. Ogier, "A comparison of local features for camera-based document image retrieval and spotting," *Int. J. Document Anal. Recognit.*, vol. 22, no. 3, pp. 247–263, 2019.
[5] B. Fan, Q. Kong, X. Wang, Z. Wanga, S. Xiang, C. Pan, and P. Fua, "A performance evaluation of local features for image-based 3D reconstruction," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 4774–4789, Oct. 2019.
[6] J. Qi, G. Li, Z. Ju, D. Chen, D. Jiang, B. Tao, G. Jiang, and Y. Sun, "Image stitching based on improved SURF algorithm," in *Proc. Int. Conf. Intell. Robot. Appl.* Shenyang, China: Springer, 2019, pp. 515–527.
[7] F. Pernici and A. Del Bimbo, "Object tracking by oversampling local features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2538–2551, Oct. 2014.
[8] H. Taira, M. Okutomi, T. Sattler, M. Cimpoi, M. Pollefeys, J. Sivic, T. Pajdla, and A. Torii, "InLoc: Indoor visual localization with dense matching and view synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7199–7209.
[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[10] S. H. Zhong, Y. Liu, and Q. C. Chen, "Visual orientation inhomogeneity based scale-invariant feature transform," *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5658–5667, 2015.
[11] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[12] Y. Liu, D. Yu, X. Chen, Z. Li, and J. Fan, "TOP-SIFT: The selected SIFT descriptor based on dictionary learning," *Vis. Comput.*, vol. 35, no. 5, pp. 667–677, 2019.
[13] E. Karami, S. Prasad, and M. Shehata, "Image matching using SIFT, SURF, BRIEF and ORB: Performance comparison for distorted images," 2017, *arXiv:1710.02726*. [Online]. Available: https://arxiv.org/abs/1710.02726
[14] A. T. Tra, W. Lin, and A. Kot, "Dominant SIFT: A novel compact descriptor," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 1344–1348.
[15] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 404–417.
[16] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.
[17] P. Li, Q. Wang, H. Zeng, and L. Zhang, "Local log-Euclidean multivariate Gaussian descriptor and its application to image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 803–817, Apr. 2017.
[18] A. Vinay, G. Kathiresan, D. A. Mundroy, H. N. Nandan, C. Sureka, K. B. Murthy, and S. Natarajan, "Face recognition using filtered eoh-sift," *Procedia Comput. Sci.*, vol. 79, pp. 543–552, 2016.
[19] H. Z. Zhang, D. W. Kim, T. K. Kang, and M. T. Lim, "MIFT: A moment-based local feature extraction algorithm," *Appl. Sci.*, vol. 9, no. 7, p. 1503, 2019.
[20] Y. K. Shen and C. T. Chiu, "Local binary pattern orientation based face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 1091–1095.
[21] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 603–610.
[22] R. Gupta, H. Patil, and A. Mittal, "Robust order-based methods for feature description," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 334–341.
[23] T. Song, F. Meng, Q. Wu, B. Luo, T. Zhang, and Y. Xu, "L2SSP: Robust keypoint description using local second-order statistics with soft-pooling," *Neurocomputing*, vol. 230, pp. 230–242, Mar. 2017.
[24] M. Verma and B. Raman, "Local tri-directional patterns: A new texture feature descriptor for image retrieval," *Digit. Signal Process.*, vol. 51, pp. 62–72, Apr. 2016.
[25] Fawad, M. J. Khan, M. A. Riaz, H. Shahid, M. S. Khan, Y. amin, J. Loo, and H. Tenhunen "Texture representation through overlapped multi-oriented tri-scale local binary pattern," *IEEE Access*, vol. 7, no. 1, pp. 66668–66679, Dec. 2019.
[26] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, and F. Moreno-Noguer, "Discriminative learning of deep convolutional feature point descriptors," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 118–126.
[27] K. Simonyan, A. Vedaldi, and A. Zisserman, "Learning local feature descriptors using convex optimisation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, 1573-1585, Aug. 2014.
[28] Y. Tian, B. Fan, and F. Wu, "L2Net: Deep learning of discriminative patch descriptor in Euclidean space," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 661–669.
[29] G.-S. Xia, G. Liu, X. Bai, and L. Zhang, "Texture characterization using shape co-occurrence patterns," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 5005–5018, Oct. 2017.
[30] T. Song, H. Li, F. Meng, Q. Wu, and J. Cai, "LETRIST: Locally encoded transform feature histogram for rotation-invariant texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1565–1579, Jul. 2018.
[31] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognit.*, vol. 60, pp. 135–160, Feb. 2017.
[32] C. G. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, pp. 10–5244.
[33] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.
[34] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004.
[35] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.
[36] B. Zheng, Y. Sun, J. Takamatsu, and K. Ikeuchi, "A feature descriptor by difference of polynomials," *IPSJ Trans. Comput. Vis. Appl.*, vol. 5, pp. 80–84, 2013.
[37] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2004, pp. 228–241.
[38] M. H. Lee and I. K. Park, "Blur-invariant feature descriptor using multidirectional integral projection," *ETRI J.*, vol. 38, no. 3, pp. 502–509, 2016.
[39] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, p. 174, 2011.
[40] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 6, pp. 1397–1409, Jun. 2013.
[41] Q. Wang and R. K. Tan, "Fast image/video contrast enhancement based on weighted thresholded histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 757–764, May 2007.
[42] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.
[43] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, Jan. 2010.
[44] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2548–2555.
[45] C. S. Kenney, M. Zuliani, and B. S. Manjunath, "An axiomatic approach to corner detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 191–197.
[46] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3431–3441, Dec. 2011.
[47] G. Yu and J. M. Morel, "A fully affine invariant image comparison method," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2009, pp. 1597–1600.
[48] R. Caldelli, I. Amerini, and A. Costanzo, "SIFT match removal and keypoint preservation through dominant orientation shift," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, 2015, pp. 2062–2066.
[49] Y. Tao, Y. Xia, T. Xu, and X. Chi, "Research progress of the scale invariant feature transform (sift) descriptors," *J. Converg. Inf. Technol*, vol. 5, no. 1, pp. 116–121, 2010.
[50] G. Meneghetti, M. Danelljan, M. Felsberg, and K. Nordberg, "Image alignment for panorama stitching in sparsely structured environments," in *Proc. Scand. Conf. Image Anal.* Scandinavian: Springer, 2015, pp. 428–439.

FAWAD received the B.S. degree in telecommunication from the National University of Modern Languages, Pakistan, in 2011, the M.S. degree from the National University of Science and Technology, Pakistan, in 2016, which was followed by his Graduate Assistant job in Ghulam Ishaq Khan Institute of Engineering Sciences and Technology. He is currently pursuing the Ph.D. degree in telecommunication engineering with the University of Engineering and Technology Taxila, Pakistan. In 2012, he joined Zong China Mobile Company as a Transmission Engineer. His research interests include computer vision, pattern recognition, and image processing.

MUHIBUR RAHMAN received the bachelor's degree in electrical (communication) engineering from the University of Engineering and Technology, Peshawar, Pakistan, in September 2014, and the M.S. degree in electrical engineering from NUST Islamabad, Pakistan, in March 2016. He is currently pursuing the Ph.D. degree with Polytechnique Montreal, Canada. He worked as a Research Assistant with the Lahore University of Management Sciences, and Dongguk University, Seoul, South Korea. He has published number of index journals and conference proceedings and taken various patents. He is also the active peer reviewer and his name is placed in Top quality peer reviewer in Engineering by Publon. His current research interests include microwave electronics, linear and nonlinear transmission lines, material characterization, and algebraic topology, deep learning, and mm-wave antennas.

MUHAMMAD JAMIL KHAN received the B.Sc. degree in computer engineering, the M.Sc. degree in telecommunication engineering, and the Ph.D. degree in computer engineering from the University of Engineering and Technology Taxila, Taxila, Pakistan, in 2005, 2009, and 2016, respectively. He is currently an Assistant Professor and the Director of the Embedded Systems and Digital Signal Processing Laboratory, University of Engineering and Technology Taxila. He is also the founder of the Virtual Reality Simulation Laboratory at the University. He has authored or coauthored numerous technical articles in well-known international journals and conferences. His current research interests include multimedia content analysis, RF identification, and machine learning.

MUHAMMAD ADEEL ASGHAR was born in Taxila, Pakistan. He received the B.Sc. degree in electrical engineering from HITEC University, Taxila, Pakistan, in 2013, and the M.S. degree in electrical engineering from HITEC University. He is currently pursuing Ph.D. degree in telecommunication engineering with the University of Engineering and Technology Taxila, Pakistan. He served as an Operation Manager at RWR PK limited. He designed and developed a 32 channel wireless EEG system for research area and public interest. His current research interests include computer vision, machine learning, digital signal processing, and brain–computer interface.

YASAR AMIN received the B.Sc. degree in electrical engineering with specialization in telecommunication and the M.B.A. degree in innovation and growth from the Turku School of Economics, University of Turku, Finland, and the M.Sc. degree in electrical engineering with specialization in system on chip design and the Ph.D. degree in electronic and computer systems from the Royal Institute of Technology (KTH), Sweden, with the research focus on printable green RFID antennas for embedded sensors. He is currently an Associate Professor and the Chairman of the Telecommunication Engineering Department, University of Engineering and Technology Taxila, Pakistan. He also serves as the Director of Embedded Systems Research and Development Centre. He is the founder of Agile Creative Technologies for Smart Electromagnetic Novel Applications (ACTSENA) Research Group. He has authored or coauthored more than 100 international technical articles in conferences and journals. His research interests include the design and application of multiple antenna systems for next generation mobile communication systems, millimeter-wave and terahertz antenna array, implantable & wearable electronics, and inkjet printing technology in microwave applications. He is a member of more than a dozen international professional societies and the Fellow of PAE.

SALMAN BADNAVA is currently a Researcher with Qatar University and the Co-Founder of ARVEX VR. His field of expertise is virtual/augmented/mixed reality, image processing, neural networks, and machine vision.

SEYED SAJAD MIRJAVADI received the B.Sc. and M.Sc. degrees from the University of Tehran. He has been a Researcher with Qatar University and the University of Tehran for more than seven years. He is currently working on three dimensional reconstruction and image processing, UWB MIMO antennas, machine vision, and deep learning. He is the author or coauthor of more than 66 peer-reviewed journals and four books and inventor of six patents.