Potential of SIFT, SURF, KAZE, AKAZE, ORB, BRISK, AGAST, and 7 More Algorithms for Matching Extremely Variant Image Pairs

Shaharyar Ahmed Khan Tareen
Pakistan Navy Engineering College (PNEC)
National University of Sciences and Technology (NUST)
Karachi, Pakistan
sakt.2015@pnec.nust.edu.pk

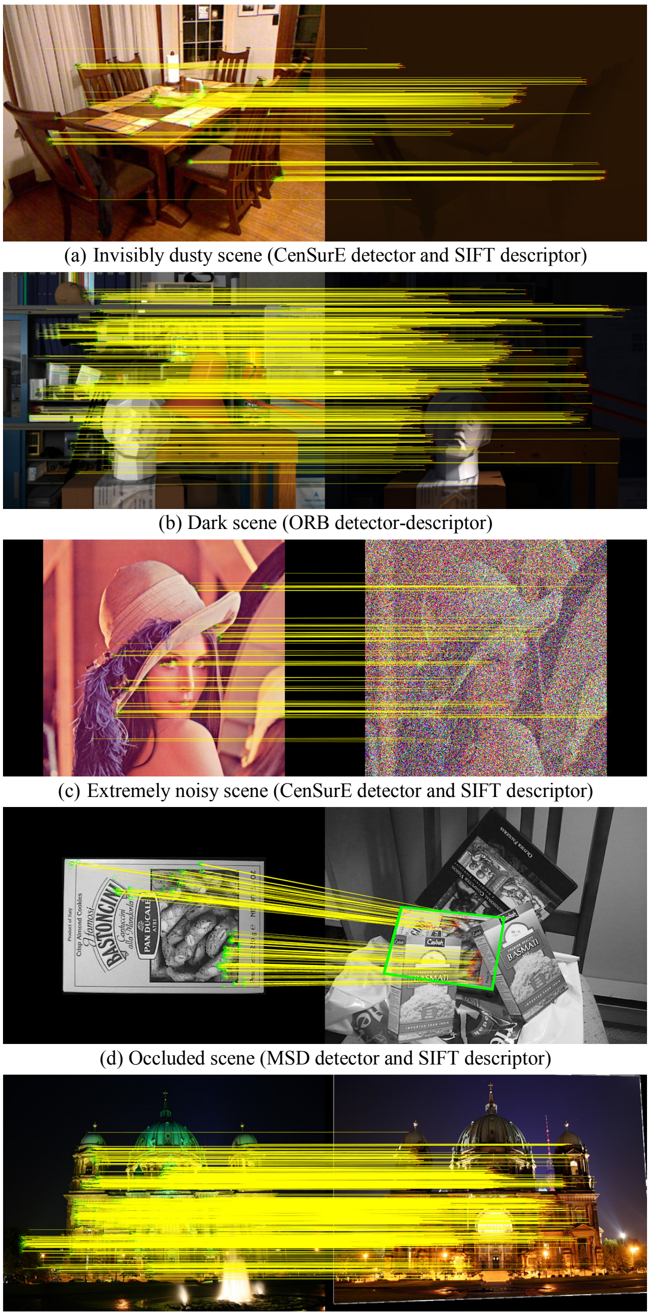
Rana Hammad Raza
Pakistan Navy Engineering College (PNEC)
National University of Sciences and Technology (NUST)
Karachi, Pakistan
hammad@pnec.nust.edu.pk

***Abstract*—Extremely variant image pairs include distorted, deteriorated, and corrupted scenes that have experienced severe geometric, photometric, or non-geometric-non-photometric transformations with respect to their originals. Real world visual data can become extremely dusty, smoky, dark, noisy, motion-blurred, affine, JPEG compressed, occluded, shadowed, virtually invisible, etc. Therefore, matching of extremely variant scenes is an important problem and computer vision solutions must have the capability to yield robust results no matter how complex the visual input is. Similarly, there is a need to evaluate feature detectors for such complex conditions. With standard settings, feature detection, description, and matching algorithms typically fail to produce significant number of correct matches in these types of images. Though, if full potential of the algorithms is applied by using extremely low thresholds, very encouraging results are obtained. In this paper, potential of 14 feature detectors: SIFT, SURF, KAZE, AKAZE, ORB, BRISK, AGAST, FAST, MSER, MSD, GFTT, Harris Corner Detector based GFTT, Harris Laplace Detector, and CenSurE has been evaluated for matching 10 extremely variant image pairs. MSD detected more than 1 million keypoints in one of the images and SIFT exhibited a repeatability score of 99.76% for the extremely noisy image pair but failed to yield high quantity of correct matches. Rich information is presented in terms of feature quantity, total feature matches, correct matches, and repeatability scores. Moreover, computational costs of 25 diverse feature detectors are reported towards the end, which can be used as a benchmark for comparison studies.**

***Keywords*—SIFT, SURF, KAZE, AKAZE, ORB, BRISK, AGAST, FAST, MSER, MSD, GFTT, Harris Corner Detector based GFTT, Harris Laplace Detector, Star Detector, CenSurE, SIFT Descriptor, Edge Detectors, Line Detectors, FLANN, K-D Trees, Multi-Probe LSH, Nearest Neighbors Distance Ratio.**

I. INTRODUCTION

Image matching is the process of matching images on the basis of common points between them. These points are also called corresponding points or correspondences [1]. Feature detectors are used to detect distinct keypoints in the images so that the correspondences between them can be established. Corresponding points are then used to derive Homography matrices (for applications with 2D outputs: image matching, object tracking, mosaicing, panorama stitching, etc.) and Fundamental/Essential matrices (for applications with 3D outputs: pose estimation, 3D reconstruction, visual odometry, visual SLAM, etc.). Feature detectors are designed to detect features on the basis of some mathematical functions so that they can detect same features repeatedly, even if the subject image undergoes geometric, photometric, or non-geometric-non-photometric transformation or deterioration. Therefore, different detectors have different strengths and weaknesses. Some of them are computationally efficient, some are robust



(a) Invisibly dusty scene (CenSurE detector and SIFT descriptor)

(b) Dark scene (ORB detector-descriptor)

(c) Extremely noisy scene (CenSurE detector and SIFT descriptor)

(d) Occluded scene (MSD detector and SIFT descriptor)

(e) Decorated scene (GFTT detector and SIFT descriptor)

Fig. 1. Matching of extremely variant image pairs by using different feature detectors and descriptors.

Preprint Version — Accepted in IEEE iCoMET-2023 Conference (DOI: https://doi.org/10.1109/iCoMET57998.2023.10099250)

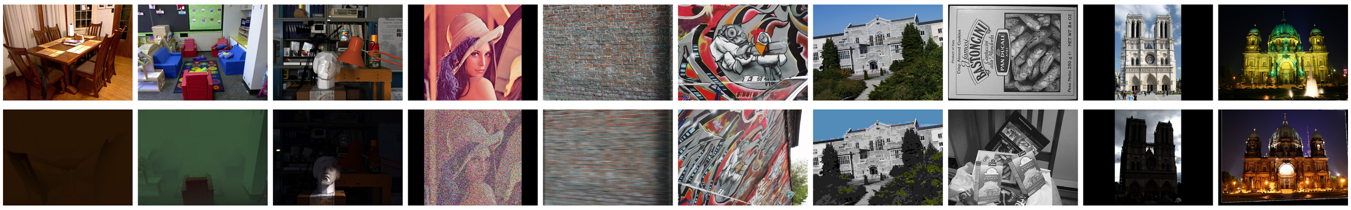


Fig. 2. Image pairs used for experiments: Top row shows the original scenes. Extremely variant versions are shown underneath. (a) Invisibly dusty. (b) Smoky. (c) Dark. (d) Noisy. (e) Motion-blurred. (f) Extremely affine. (g) JPEG compressed. (h) Occluded. (i) Sunny-shadowed. (j) Decorated with lighting effects.

against scale, rotation, affine, noise, blur, illumination, etc. while others are not [1]. Since visual data can experience various types of transformations, the knowledge about merits and demerits of the cutting-edge feature detectors against different image transformations carries high importance. It allows researchers to identify the most suitable algorithm(s) according to their application requirements, conveniently.

In this paper, 14 feature detectors: SIFT, SURF, KAZE, AKAZE, ORB, BRISK, AGAST, FAST, MSER, MSD, Good Features To Track (GFTT), Harris Corner Detector based GFTT (GFTT-H), Harris Laplace Detector (Harris-L), and CenSurE (also called Star Detector) have been evaluated for matching extremely variant and problematic image pairs. With strict or standard thresholds, algorithms generally fail to yield substantial number of corresponding points between extremely variant scenes. Therefore, to investigate their full potential, parameter thresholds of the algorithms have been kept extremely low (see Table I). The 10 image pairs used in this evaluation have been selected in the light of various studies including: robust pedestrian detection under tough visual conditions [2], challenging panorama stitching problems [3], turbid underwater image matching [4], visual odometry [5] and visual SLAM [6] in extreme environments with troublesome visibility scenarios. They consist of scenes which are: invisibly dusty, smoky, dark/dim lighted, noisy, motion-blurred, affine, JPEG compressed, occluded, sunny-shadowed, and decorated using lighting effects. Comparative results with rich information are presented in terms of feature quantity, total matches, correct matches, and repeatability scores (see Table II). In the end, detection timings of 25 diverse feature detectors (covering corners, blobs, straight lines, and edges) are also reported in order to highlight their relative computational efficiencies and costs (see Table III).

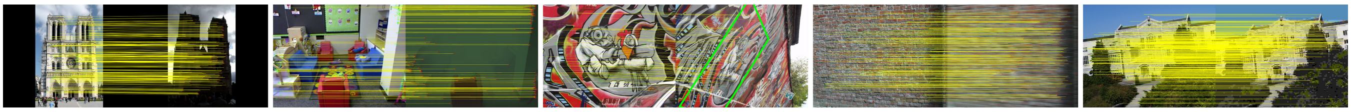
II. Literature Review

Feature based image matching finds its uses in numerous computer vision applications including image mosaicing, panorama generation, object tracking, augmented reality, structure from motion, thermal modeling, 3D reconstruction, visual odometry, and visual SLAM. Various studies have been conducted till date to evaluate detectors and descriptors for matching diverse scenes. In [7], a detailed comparison of 6 algorithms is performed for matching different scenes with normal parameter thresholds. L. Juan et al. evaluated the performance of SIFT, PCA-SIFT, and SURF for different image transformations in [8]. E. Karami et al. compared the normal performance of SIFT, SURF, and ORB for different image distortions including scale, rotation, smear, and fisheye effect [9]. Performance of SIFT and its variants (PCA-SIFT, GSIFT, CSIFT, SURF, ASIFT) was explored in [10] for scale, rotation, affine, and blur with normal thresholds. In [11], multiple feature detectors are evaluated mainly for matching translated and rotated images. SIFT, SURF, ORB, and AKAZE are studied in [12] for monocular visual odometry. Similarly, some popular feature detectors have been studied in [13] for visual SLAM. All such studies cover a limited number of algorithms (using normal thresholds) for matching scenes that are not extremely difficult. Therefore, the potential of algorithms for matching extremely problematic and challenging scenes is unknown and results of 14 feature detectors are presented in this paper to fill this knowledge gap.

SIFT [14] detector is based on Difference of Gaussians and its descriptor yields 128 dimensional vectors generated on the basis of orientations and magnitudes of the neighboring pixels around the detected keypoints. SURF [15] is an efficient version of SIFT, based on Box Filter and Determinant of Hessian Matrix. Its descriptor is based on Haar Wavelet responses. KAZE uses non-linear scale spaces to detect blobs on the basis of Determinant of Hessian [16]. Its descriptor is a modified form of M-SURF. AKAZE is the accelerated version of KAZE which uses Fast Explicit Diffusion to generate non-linear scale spaces [17]. AKAZE descriptor is a modified form of Local Difference Binary descriptor, hence called Modified-LDB descriptor. ORB is a combination of FAST detector and rotated version of BRIEF descriptor with improvements [18]. BRISK uses FAST detector with 9-16 configuration in a scale space pyramid [19]. BRISK descriptor generates N equally spaced points in the form of concentric circles around each detected keypoint. AGAST detects corners by improving the accelerated segment test of FAST [20]. FAST is a quick corner detector that identifies potential corners by comparing pixel-intensities in circular regions and then intensively tests them [21]. MSER detects stable regions (where intensity/color remains consistent over a large area) in the images [22]. MSD detects features which are highly dissimilar over a large area [23]. GFTT detects corners (Shi-Tomasi corners) that are very suitable for tracking [24]. GFTT-H (available in OpenCV Library) detects corners that are invariant to scale changes using Harris Corner Detection approach. Harris-L Detector is a combination of Harris Corner Detector and Laplacian based scale selection method [25]. CenSurE is also based on Harris Detector with multiple improvements which make it robust to scale and affine changes [26]. SIFT, SURF, KAZE, AKAZE, ORB, and BRISK detectors do have their own description algorithms whereas others do not. The descriptors of AKAZE, ORB, and BRISK are yielded as binary strings, that are different from the descriptors of SIFT, SURF, and KAZE.

III. Methodology

The evaluation process starts from feature detection in the images with extremely low thresholds (see Table I). Out of the detected features, duplicate features are discarded. Afterwards, feature description and matching are performed. SIFT, SURF, KAZE, AKAZE, ORB, and BRISK detectors are used with their own descriptors whereas others (which do



(a) Sunny-Shadowed (b) Smoky (c) Extremely affine (d) Motion-blurred (e) JPEG compressed

Fig. 3. Feature based matching of additional diverse problematic image pairs. Zoom-in for improved view and clear observation.

TABLE I. THRESHOLD SETTINGS USED FOR FULL POTENTIAL OF THE FEATURE DETECTION AND MATCHING ALGORITHMS

METHOD	OpenCV's Constructor Settings
SIFT	cv.SIFT('NOctaveLayers',4,'ConstrastThreshold',1e-9,'EdgeThreshold', 100, 'Sigma',0.5)
SURF	cv.SURF('NOctaveLayers',4,'HessianThreshold',1e-9)
KAZE	cv.KAZE('NOctaveLayers',4,'Threshold',1e-9)
AKAZE	cv.AKAZE('NOctaveLayers',4,'DescriptorType','MLDB','Threshold', 1e-9)
ORB	cv.ORB('MaxFeatures',1e+6,'FastThreshold',1)
BRISK	cv.BRISK('Octaves',4,'Threshold',1)
AGAST	cv.AGAST(Image,'Type','AGAST_7_12d','NonmaxSuppression',false, 'Threshold',1)
FAST	cv.FAST(Image,'NonmaxSuppression',false,'Threshold',1)
MSER	cv.MSER('MaxEvolution',1e+4,'EdgeBlurSize',3)
MSD	cv.MSDDetector('ThSaliency',1,'NMSRadius',0,'ComputeOrientation', true,'NScales',3)
GFTT	cv.GFTTDetector('MinDistance',0.1,'MaxFeatures',1e+6,'QualityLevel', 1e-6)
GFTT-H	cv.GFTTDetector('HarrisDetector',true,'K',1e-3,'MaxFeatures',1e+6, 'QualityLevel',1e-6)
Harris-L	cv.HarrisLaplaceFeatureDetector('CornThresh',1e-6,'DOGThresh',1e-6, 'MaxCorners',1e+6)
CenSurE	cv.StarDetector('ResponseThreshold',1e-3,'LineThresholdProjected',50, 'LineThresholdBinarized',50,'SuppressNonmaxSize',0)
K-D Trees	cv.DescriptorMatcher('FlannBasedMatcher','Index',{'KDTree','Trees',4})
MP-LSH	cv.DescriptorMatcher('FlannBasedMatcher','Index',{'LSH', 'TableNumber',6,'KeySize',15, 'MultiProbeLevel',1})

not have their own description method) are used with the SIFT descriptor. For string like descriptors (SIFT, SURF, KAZE) K-Dimensional Trees based matching is performed whereas for matching the binary descriptors (AKAZE, ORB, BRISK) Multi-Probe LSH (MP-LSH) method is used. Both these methods are part of Fast Library for Approximate Nearest Neighbors (FLANN). Nearest Neighbors Distance Ratio (NNDR) [1], [7] is opted as the descriptor matching strategy (for both methods) with a threshold ratio of **"0.9"**. Out of the total matches obtained, the correct matches are then filtered on the basis of ground-truth Homographies by using Euclidean distance with a tolerance threshold of **"2.5 pixels"**. In real time applications when the ground-truth is not known, any of the RANSAC, PROSAC, and MSAC techniques can be applied for outlier rejection. Fig. 4 provides a generic overview of the performance evaluation process.

In this paper, **"repeatability"** is defined as the ratio of "quantity of features repeated in the deteriorated image" to the "quantity of features detected in the original image". A feature in the deteriorated image has been marked as **"repeated"** if its centre point lies within the tolerance of **"2.5 pixels"** of Euclidean Distance from the ground-truth based projected centre point of the feature detected in the original image. Please note that this definition of repeatability is different from [27], which was based on overlapping errors between the features. This novelty is opted because in vision based applications (panorama stitching, 3D reconstruction, etc.) results are affected more by the error between the centre points of the matched features as compared to their overlapping error. A detector is more invariant to a particular transformation if it has higher repeatability for it. In case of geometrically transformed scenes (extremely affine, occluded, and sunny-shadowed), features detected only within the overlapping regions of the image pairs have been considered in measuring the repeatability scores. Table II presents the performance of 14 algorithms for matching extremely variant image pairs. In the end, Table III shows the detection timings of 25 diverse feature detection algorithms (including corner detectors, blob detectors, straight line detectors, and edge detectors) that can be used to get an idea about their **"relative"** computational efficiencies and costs.

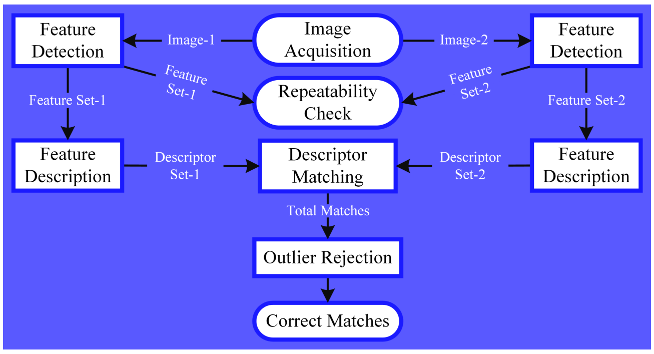


Fig. 4. An overview of the performance evaluation and matching process.

The invisibly dusty scene is Image-551 (Clear Image, Type-9 Image) and the smoky scene is Image-325 (Clear Image, Type-5 Image) from [28]. Left camera images (Fluorescent-1, Lamps-1) of [29] represent the dark scene. The noisy scene is synthetically rendered in MATLAB by adding Gaussian Noise to the image "Lena.jpg" with mean = 0 and variance = 1. Similarly, motion-blurred scene is rendered by convolving "Wall-1.ppm" image from [30] with a horizontal motion filter having kernel = 45 and angle = 0. The extremely affine (Graffiti-1, Graffiti-6) and JPEG Compressed (UBC-1, UBC-6) scenes are also collected from [30]. The occluded scene is taken from OpenCV Library and additional occlusion has been added. The sunny-shadowed and decoration (using lighting effects) scenes are picked from [31]. Additional darkness has been added to the shadowed scene to make it complex. MATLAB Executable (MEX) files of OpenCV Library have been used in this evaluation from [32]. All the experiments are conducted in MATLAB R2020a on a general purpose Desktop (Core i7-4790 CPU @ 3.60 GHz) with 8 GB RAM and 64-bit Operating System (Windows-7 Professional). MATLAB codes and the 10 extremely variant image pairs used in this research work are provided in [33].

IV. RESULTS & DISCUSSION

The experiments of matching extremely variant image pairs by using 14 feature detectors have shown a lot of encouraging results and discoveries. Performance of detectors and descriptors is superior or inferior for different types of image transformations because they are either robust against or sensitive to a particular type of deterioration. For instance SIFT detector gives highest repeatability (99.76%) in the noisy image pair but its descriptor loses its discriminative power, hence only 25 correct matches are retrieved out of the 54203 ground-truth correspondences. Prominent findings are explicitly described in the underneath sections. However, researchers can deduce additional information from Table II and Table III as per their needs and application requirements.

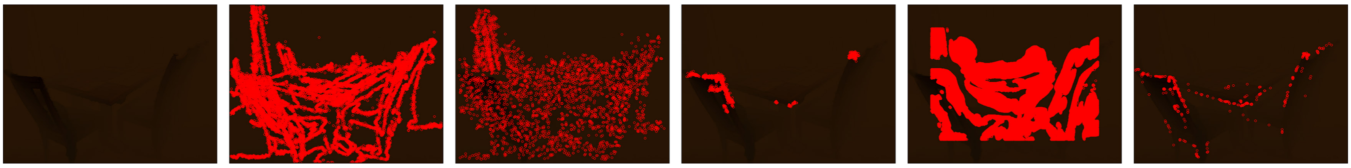
A. *Quantitative Comparison*

MSD has the highest potential to detect huge quantity of features and it detected more than **1 million** features in image-1 of the motion-blurred scene. MSER and CenSurE also detected more than **300,000** keypoints in some images.

TABLE II. PERFORMANCE OF FEATURE DETECTORS AND DESCRIPTORS FOR MATCHING EXTREMELY VARIANT AND PROBLEMATIC SCENES. HIGHEST VALUES ARE BOUNDED BY GREEN WHILE LOWEST VALUES ARE BOUNDED BY RED IN EACH FIELD.

ALGORITHM	Features Detected		Repeatability (%)	Total Matches	Correct Matches	Features Detected		Repeatability (%)	Total Matches	Correct Matches
	Image-1	Image-2				Image-1	Image-2			
	INVISIBLY DUSTY SCENE					**SMOKY SCENE**				
SIFT	28507	13612	26.81	6018	73	32205	29266	49.82	7911	518
SURF	4481	2808	18.68	1047	12	3816	3110	31.08	1192	121
KAZE	8435	3051	16.40	1873	71	7693	5393	41.54	2424	488
AKAZE	5727	2255	19.09	1288	37	5176	3404	38.70	1453	164
ORB	27622	272	0.86	4018	4	21228	1522	6.22	6350	90
BRISK	21179	244	0.81	3609	3	15565	1494	7.02	3003	85
AGAST	90214	459	0.50	7349	86	68582	3371	4.88	8380	1918
FAST	98634	670	0.67	6719	111	77537	4767	6.12	9116	2062
MSER	190205	14679	6.11	41417	131	169770	31815	9.28	46088	1339
MSD	417704	1967	0.46	5549	162	322839	29400	9.04	24313	3923
GFTT	13984	3559	16.08	2231	180	12491	7919	41.16	3327	1083
GFTT-H	6117	0	0.00	0	0	4398	137	2.34	1032	63
Harris-L	3577	19	0.17	264	5	4556	19	0.29	391	8
CenSurE	91209	66253	50.86	4266	329	87032	49080	42.56	4913	592
	DARK SCENE					**NOISY SCENE**				
SIFT	37533	61899	92.03	9957	1226	54334	100453	99.76	8424	25
SURF	4073	5048	45.15	1413	391	4182	12267	60.50	608	11
KAZE	7884	8186	57.70	3472	1508	7179	8851	46.25	1538	50
AKAZE	5534	5786	59.87	2128	927	4929	6711	53.78	1095	65
ORB	20520	6177	25.05	5418	854	33217	54290	98.38	5408	258
BRISK	13659	4709	23.38	3127	596	25313	32398	88.56	3834	90
AGAST	60019	12633	18.75	15798	5907	88256	119946	98.30	823	149
FAST	69259	16857	22.17	16809	6713	87133	99843	92.51	958	223
MSER	240872	206718	74.33	76002	10215	142461	152236	70.50	43170	29
MSD	352957	95637	26.35	63681	17115	403281	442169	87.37	19025	2691
GFTT	13879	16238	71.99	5028	2349	14312	17126	75.26	1605	29
GFTT-H	4414	535	8.13	1190	187	6209	14288	68.77	657	29
Harris-L	3470	303	4.24	512	54	3768	10900	45.99	926	87
CenSurE	105219	93583	79.01	11801	4637	77365	56958	55.82	1092	83
	MOTION BLURRED SCENE					**EXTREMELY AFFINE SCENE**				
SIFT	136929	78050	42.77	17280	42	101759	91592	32.57	21310	44
SURF	14148	8161	18.84	3784	25	8273	9742	21.79	2231	8
KAZE	21413	16727	30.99	3819	70	13116	14030	26.07	2980	6
AKAZE	16365	12444	33.20	2570	88	9792	10453	25.63	2238	4
ORB	123976	45172	17.00	15191	267	70815	78506	39.14	12567	22
BRISK	66891	41342	24.67	5624	129	49052	49869	38.16	7425	34
AGAST	342273	121354	2.20	3990	152	173970	193259	41.24	7347	3
FAST	335035	141523	3.15	3958	205	179875	191138	39.51	7302	12
MSER	454384	152448	27.53	100575	19	348622	301351	29.81	105135	69
MSD	1088016	817681	70.85	38779	4164	796158	759985	31.16	37433	79
GFTT	35967	29879	31.85	2629	9	26216	24760	29.08	6331	7
GFTT-H	27733	6695	16.04	1709	3	13360	16086	32.56	3063	6
Harris-L	32220	2791	4.52	2621	102	6624	7787	10.33	796	0
CenSurE	363605	305633	80.66	2798	108	224028	237983	29.29	9693	19
	JPEG COMPRESSED SCENE					**OCCLUDED SCENE**				
SIFT	103118	18917	17.19	17983	1079	16092	51837	25.15	3723	162
SURF	11647	5422	22.45	3947	768	1454	4389	18.43	401	20
KAZE	16057	8501	29.74	6440	2095	2103	5573	24.74	599	12
AKAZE	12218	6124	28.97	3954	1281	1058	3743	27.03	223	8
ORB	74643	20294	25.07	17489	4966	7156	22818	33.02	1195	37
BRISK	43381	15458	27.29	7278	1483	5214	16556	30.44	776	29
AGAST	189153	50303	25.94	23198	8564	25100	57515	23.41	1034	26
FAST	182564	62394	32.83	20298	7423	24592	56166	20.43	1078	57
MSER	214018	143797	46.50	57710	818	35697	85589	29.65	9801	102
MSD	694851	359699	51.36	118400	56928	101218	297395	29.03	5431	361
GFTT	25579	6461	16.60	4893	1398	3322	9923	24.38	628	5
GFTT-H	15685	5350	21.58	3485	1237	2308	3770	24.87	337	7
Harris-L	11520	6607	22.67	3648	827	475	909	8.21	75	6
CenSurE	192576	204867	93.74	17647	8843	11380	54839	33.21	678	46
	SUNNY-SHADOWED SCENE					**DECORATED SCENE**				
SIFT	51648	44882	71.89	11643	346	113576	149572	85.67	22324	277
SURF	4131	4291	33.57	1145	74	12561	16097	37.78	3208	30
KAZE	7235	7086	43.21	2038	222	19355	18279	40.67	4448	137
AKAZE	5019	4971	46.18	1315	131	15050	15503	47.45	2889	120
ORB	36883	22034	51.83	7957	463	57878	72466	78.94	9941	342
BRISK	20704	16860	56.90	3631	250	31697	43180	71.53	4091	196
AGAST	97108	57201	51.92	8350	1648	146659	155930	80.93	11247	1520
FAST	101835	62134	53.26	8435	1676	151807	156937	81.01	11024	1325
MSER	156203	161131	75.29	48747	2529	306600	226035	47.22	96978	2338
MSD	376804	331564	79.35	28447	3199	538929	837685	92.65	38857	3181
GFTT	13104	12013	56.57	3260	410	29397	31689	60.30	5706	514
GFTT-H	8193	3040	21.49	2058	230	10928	11197	47.97	2718	436
Harris-L	5344	904	4.61	897	24	6868	8168	40.16	960	44
CenSurE	92854	75631	60.89	9498	1289	260175	250051	73.70	22191	1131

SIFT severely outperforms SURF in detecting huge feature quantity with the low parameter thresholds. AKAZE detects less features than KAZE for all the 10 scenes. An estimated order of the potential to detect huge feature quantity is: **MSD, MSER, CenSurE, FAST, AGAST, SIFT, ORB, BRISK, GFTT, GFTT-H, KAZE, AKAZE, SURF, and Harris-L**.



Invisibly Dusty Scene SIFT (13612) SURF (2808) MSD (1967) CenSurE (66253) AGAST (459)

Fig. 5. Diverse types of features detected in extremely dusty image (almost invisible to human eye). Increase screen brightness and zoom-in for improved view.

TABLE III. RELATIVE COMPUTATIONAL COSTS OF 25 ALGORITHMS. SUPERIOR VALUES BOUNDED BY GREEN, INFERIOR VALUES BY RED.

ALGORITHM	Detected Points/Pixels	Detection Time (ms)	Description Time (ms)
DARK SCENE (Image-1) (Image Size = 480 × 640)			
SIFT	37533	165.26	261.09
SURF	4073	53.83	131.67
KAZE	7884	327.44	298.94
AKAZE	5534	50.13	70.32
ORB	20520	64.96	106.21
BRISK	13659	236.42	165.93
AGAST	60019	92.85	1263.60
FAST	69259	99.62	1415.26
MSER	240872	758.96	929.84
MSD	352957	2325.11	5228.06
GFTT	13879	34.58	126.87
GFTT-H	4414	19.44	48.68
Harris-L	3470	342.21	64.69
CenSurE	105219	268.58	21655.43
Direct Pixel Selection (DPS)	307200	1.68	–
Pixel Intensity Filtering (PIF)	296356	2.59	–
Simple Blob Detector	4239	1864.89	–
Line Segment Detector (LSD)	41923	45.19	–
Fast Line Detector (FLD)	66375	207.56	–
Canny Edge Detector	57689	13.62	–
Approx. Canny Edge Detector	87983	6.39	–
Sobel Edge Detector	90903	3.04	–
Prewitt Edge Detector	91445	3.13	–
Roberts Edge Detector	96126	4.57	–
LOG Edge Detector	55509	14.79	–

B. Matching of Extremely Variant Image Pairs

Feature detectors can detect keypoints even in the images that are complex, corrupted, distorted, deteriorated, and invisible to human sight. Therefore, feature based matching of scenes with extreme variations due to dust, fire, smoke, extreme fog, darkness, lighting effects, rain, snowfall, motion-blur, noise, underwater turbidity, occlusion, camera lens glare, shadows, affine changes, data corruption, etc. is generally possible.

• *Invisibly Dusty Scene:* CenSurE detects 66253 features (highest) in image-2 followed by MSER and SIFT. GFTT-H detected no feature while Harris-L detected only 19 features. CenSurE has 50.86% repeatability (highest) and it also gives 329 correct matches (highest). SIFT, KAZE, AGAST, FAST, MSER, MSD, GFTT produce adequate correct matches. Moreover, GFTT-H, BRISK, ORB, Harris-L, SURF, and AKAZE yield fewer correct matches (0, 3, 4, 5, 12, and 37).

• *Smoky Scene:* CenSurE detects 49080 features (highest) in image-2 followed by MSER, MSD, and SIFT (29266). SIFT, CenSurE, and KAZE show higher repeatabilities (49.82%, 42.56%, and 41.54%). MSD gives 3923 correct matches (highest) followed by FAST, AGAST, MSER, and GFTT. Moreover, GFTT-H, BRISK, and ORB yield lower correct matches: 63, 85, and 90. Performance of Harris-L is overall unsatisfactory and it produces only 8 correct matches.

• *Dark Scene:* MSER detects 206718 features (highest) in image-2 followed by MSD, CenSurE, and SIFT. SIFT shows highest repeatability (92.03%) followed by CenSurE, MSER, and GFTT. MSD gives 17115 correct matches (highest) followed by MSER (10215), FAST, AGAST, and CenSurE (4637). GFTT-H and SURF produce only 187 and 391 correct matches. Harris-L detects 303 features in image-2, shows low repeatability (4.24%) and yields lowest correct matches (54).

• *Noisy Scene:* SIFT shows highest repeatability (99.76%) followed by ORB (98.38%), AGAST, FAST, BRISK, and MSD (87.37%). However, SIFT yields only 25 correct matches which shows that SIFT descriptor is sensitive to extreme noise. ORB showed lower feature quantity and total matches than SIFT but yields 258 correct matches (higher than SIFT). Therefore, ORB detector and descriptor are more robust than SIFT against image noise. MSD detected 442169 features (highest) and also produced 2691 correct matches (highest) using SIFT descriptor. SURF, GFTT, and GFTT-H produced only 11, 29, and 29 correct matches, respectively.

• *Motion Blurred Scene:* MSD detects highest and huge feature quantities in both image-1 (1 million+) and image-2 (817681). It also yields 4164 correct matches (highest). CenSurE shows 80.66% repeatability (highest) followed by MSD and SIFT. GFTT-H, GFTT, MSER, and SURF produced only 3, 9, 19, and 25 correct matches. AGAST, FAST, ORB, BRISK, Harris-L, and CenSurE provided moderate numbers of correct matches. SIFT, KAZE, and AKAZE could not yield high numbers of correct matches.

• *Extremely Affine Scene:* MSD and MSER show highest feature quantities (759985 and 301351) and highest correct matches (79 and 69). AGAST shows highest repeatability (41.24%) but only 3 correct matches. Harris-L shows lowest repeatability (10.33%) and yields not a single correct match. Relatively higher numbers of correct matches are produced by SIFT (44), BRISK (34), ORB (22), and CenSurE (19).

• *JPEG Compressed Scene:* MSD detects 359699 features (highest), followed by CenSurE (204867) and MSER (143797). CenSurE gives 93.74% repeatability (highest), followed by MSD (51.36%) and MSER (46.50%). MSD yields 56928 correct matches (highest) while SURF, MSER, and Harris-L yield lowest correct matches (768, 818, and 827). CenSurE, AGAST, FAST, ORB, and KAZE show relatively high correct matches whereas SIFT, AKAZE, BRISK, GFTT, and GFTT-H show moderate correct matches. SURF shows higher repeatability than SIFT only in this scene.

• *Occluded Scene:* CenSurE shows highest repeatability (33.21%) followed by ORB, BRISK, MSER, MSD, and AKAZE (27.03%). MSD yields 361 correct matches (highest) followed by SIFT (162) and MSER (102). GFTT, Harris-L, GFTT-H, AKAZE, KAZE, and SURF produce very low correct matches. FAST, CenSurE, ORB, BRISK, and AGAST show relatively medium amounts of correct matches.

• *Sunny-Shadowed Scene:* MSD shows highest values in all fields with a repeatability of 79.35%. MSD yields 3199 correct matches (highest) followed by MSER (2529), FAST (1676), AGAST (1648), and CenSurE (1289). Harris-L and SURF show relatively lower values. Harris-L gives lowest repeatability (4.61%). SIFT outperforms SURF in all fields.

- *Decorated Scene:* MSD tops all the fields by showing highest values with 92.65% repeatability. CenSurE detects 250051 features (second highest) followed by MSER, FAST, AGAST, and SIFT. MSD yields 3181 correct matches (highest) and MSER yields 2338 correct matches (second highest). SURF and Harris-L give lowest correct matches i.e. 30 and 44, respectively. SIFT outperforms SURF in all fields.

C. Computational Costs

Direct Pixel Selection, Pixel Intensity Filtering, and Edge detectors have considerably low computational costs as compared to other algorithms and they all can detect relatively high quantity of keypoints (pixels). MSD detects very high feature quantity but its computational cost is highest. MSER also detects very high quantity of points (considering pixels of the detected regions) but with a reasonable computational cost. Cost of computing SIFT descriptors for MSER is less as compared to MSD. CenSurE detects high feature quantity but the cost of computing SIFT descriptors is very high. Edge detectors are computationally efficient compared to blob, corner, and straight line detectors.

V. CONCLUSION

This paper evaluates the potential of 14 feature detectors for matching 10 extremely problematic scenes: dusty, smoky, dark, noisy, motion-blurred, affine, compressed, occluded, shadowed, and decorated. Extremely variant images are difficult to match with their originals due to deterioration of the view. With normal thresholds, feature detection and matching algorithms typically fail to match them. Therefore, in order to obtain maximum performance at full throttle, the parameter thresholds of detectors are kept extremely low. Many important findings have been discovered and reported. MSD, MSER, and CenSurE detect very high quantity of features. Computational cost of MSD is highest among all. Performance of Harris-L, GFTT-H, and SURF is overall towards the lower side. All popular edge detectors generally yield high quantity of points (pixels) and are computationally efficient than blob, corner, and line detectors. SIFT emerged as the most versatile algorithm showing good repeatabilities for most of the scenes. MSD, MSER, SIFT, KAZE, GFTT, and CenSurE give robust performance in matching dark or dim lighted scenes. Moreover, SIFT severely outperforms SURF, KAZE, AKAZE, GFTT, etc. in producing high feature quantity for all the scenes with these low thresholds.

In future there is a need to focus on the development of robust computer vision techniques for image matching, panorama stitching, mosaicing, object tracking, visual odometry, visual SLAM, etc. that are not just computationally efficient but also capable enough to handle troublesome or highly variant visual data. Deep learning architectures, multi-modal matching, advanced feature detection algorithms, and hybrid techniques (that use a combination of approaches) can also be developed to detect and match extremely challenging scenes.

REFERENCES

[1] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. on Pattern Anal. & Mach. Intell., vol. 27, no. 10, pp. 1615-1630, 2005.

[2] S. Cygert and A. Czyzewski, "Toward robust pedestrian detection with data augmentation," IEEE Access, vol. 8, pp. 136674-136683, 2020.

[3] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," Int. J. of Comput. Vision, vol. 74, no. 1, pp. 59-73, 2007.

[4] R. Garcia and N. Gracias, "Detection of interest points in turbid underwater images," in OCEANS 2011 IEEE, Spain, 2011, pp. 1-9.

[5] L. Li et al., "Visual odometry for planetary exploration rovers in sandy terrains," Int. J. of Advanced Robotic Syst., vol. 10, no. 5, 2013.

[6] K. Ebadi et al., "Present and future of slam in extreme underground environments," arXiv preprint, arXiv: 2208.01787, 2022.

[7] S. A. K. Tareen and Z. Saleem, "A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK," in IEEE Int. Conf. on Computing, Math. & Eng. Technologies, Sukkur, iCoMET, 2018, pp. 1-10.

[8] L. Juan and O. Gwun, "A comparison of SIFT, PCA-SIFT and SURF," Int. J. of Image Process., vol. 3, no. 4, pp. 143-152, 2009.

[9] E. Karami et al., "Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images," arXiv preprint, arXiv: 1710.02726, 2017.

[10] J. Wu et al., "A Comparative Study of SIFT and its Variants," Measurement Science Review, vol. 13, no. 3, 2013.

[11] J. Bian et al., "Image matching: An application-oriented benchmark," arXiv preprint, arXiv: 1709.03917, 2017.

[12] H. J. Chien et al., "When to use what feature? SIFT, SURF, ORB, or A-KAZE features for monocular visual odometry," in IEEE Int. Conf. on Image and Vision Computing New Zealand, IVCNZ, pp. 1-6, 2016.

[13] J. Hartmann et al., "A comparison of feature descriptors for visual SLAM," in IEEE European Conf. on Mobile Robots, ECMR, 2013, pp. 56-61.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. of Comput. Vision, vol. 60, no. 2, pp. 91-110, 2004.

[15] H. Bay et al., "Speeded-up robust features (SURF)," Comput. Vision and Image Understanding, vol. 110, no. 3, pp. 346-359, 2008.

[16] P. F. Alcantarilla et al., "KAZE features," in European Conf. on Computer Vision, ECCV, 2012, pp. 214-227.

[17] P. F. Alcantarilla and T. Solutions, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," IEEE Trans. on Pattern Anal. & Mach. Intell., vol. 34, no. 7, pp. 1281-1298, 2011.

[18] E. Rublee et al., "ORB: An efficient alternative to SIFT or SURF," in IEEE Int. Conf. on Comput. Vision, ICCV, 2011, pp. 2564-2571.

[19] S. Leutenegger et al., "BRISK: Binary robust invariant scalable keypoints," in IEEE Int. Conf. on Comput. Vision, ICCV, 2011, pp. 2548-2555.

[20] E. Mair et al., "Adaptive and generic corner detection based on the accelerated segment test," in European Conf. on Comput. Vision, ECCV, 2010, pp. 183-196.

[21] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in European Conf. on Comput. Vision, ECCV, 2006, pp. 430-443.

[22] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (MSER) tracking," in IEEE Comput. Society Conf. on Comput. Vision and Pattern Recognition, CVPR, 2006, pp. 553-560.

[23] F. Tombari and L. Di Stefano, "Interest points via maximal self-dissimilarities," in Asian Conf. on Comput. Vision, 2014, pp. 586-600.

[24] J. Shi, "Good features to track," in IEEE Conf. on Comput. Vision and Pattern Recognition, 1994, pp. 593-600.

[25] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," Int. J. of Comput. Vision, vol. 60, pp. 63-86, 2004.

[26] M. Agrawal et al., "Censure: Center surround extremas for realtime feature detection and matching," in European Conf. on Comput. Vision, ECCV, 2008, pp. 102-115.

[27] K. Mikolajczyk et al., "A comparison of affine region detectors," Int. J. of Comput. Vision, vol. 65, no. 1-2, pp. 43-72, 2005.

[28] N. Silberman et al., "Indoor segmentation and support inference from rgbd images," in European Conf. on Comput. Vision, ECCV, 2012, pp. 746-760.

[29] M. Peris et al., "Towards a simulation driven stereo vision system," in Proc. IEEE 21st Int. Conf. on Pattern Recognition, ICPR, 2012, pp. 1038-1042.

[30] Visual Geometry Group, "Affine covariant regions datasets," robots.ox.ac.uk. [Online]. Available: https://www.robots.ox.ac.uk/~vgg/data/affine. [Accessed Nov. 21, 2022].

[31] D. C. Hauagge and N. Snavely, "Image matching using local symmetry features," cs.cornell.edu. [Online]. Available: https://www.cs.cornell.edu/projects/symfeat. [Accessed Nov. 11, 2022].

[32] K. Yamaguchi, "Mexopencv," github.com, Oct. 8, 2018. [Online]. Available: https://github.com/kyamagu/mexopencv. [Accessed Nov. 15, 2022].

[33] S. A. K. Tareen, "Matching of extremely variant image pairs," github.com. [Online]. Available: https://github.com/saktx.

APPENDIX

TABLE IV. PERFORMANCE OF FEATURE DETECTORS AND DESCRIPTORS FOR MATCHING EXTREMELY VARIANT IMAGE PAIRS WITH NORMAL THRESHOLDS. HIGHEST VALUES ARE SHOWN AS GREEN WHILE LOWEST VALUES ARE SHOWN AS RED IN EACH FIELD.

ALGORITHM	Features Detected		Repeatability (%)	Total Matches	Correct Matches	Features Detected		Repeatability (%)	Total Matches	Correct Matches
	Image-1	Image-2				Image-1	Image-2			
	INVISIBLY DUSTY SCENE					SMOKY SCENE				
SIFT	1269	0	0.00	0	0	1252	1	0.07	0	0
SURF	1968	1	0.05	0	0	1626	42	1.41	463	6
KAZE	1093	0	0.00	0	0	1241	2	0.08	584	1
AKAZE	723	0	0.00	0	0	833	0	0.00	0	0
ORB	3084	0	0.00	0	0	3693	0	0.00	0	0
BRISK	1075	0	0.00	0	0	1358	0	0.00	0	0
AGAST	2662	0	0.00	0	0	2328	14	0.47	313	10
FAST	2657	0	0.00	0	0	2408	19	0.71	564	14
MSER	149302	22772	9.24	34379	54	130301	21549	14.51	31983	1602
MSD	338	0	0.00	0	0	445	0	0.00	0	0
GFTT	1717	1839	17.18	480	92	1683	638	21.27	558	240
GFTT-H	351	0	0.00	0	0	490	34	4.69	126	20
Harris-L	524	0	0.00	0	0	1305	0	0.00	0	0
CenSurE	258	0	0.00	0	0	265	0	0.00	0	0
	DARK SCENE					NOISY SCENE				
SIFT	1177	73	0.76	335	1	988	859	7.39	176	5
SURF	1811	243	4.58	533	42	1521	8369	49.24	226	8
KAZE	1129	126	0.97	171	0	768	1592	17.58	160	19
AKAZE	883	91	0.91	200	1	559	2151	24.69	184	19
ORB	4442	362	1.91	1310	2	4096	46346	97.17	746	52
BRISK	1555	113	0.96	300	0	1429	27746	91.18	220	16
AGAST	2231	220	4.93	423	46	4460	27792	85.56	66	11
FAST	2280	249	5.79	497	66	3753	26844	96.16	61	7
MSER	233906	173869	64.94	72514	8511	105594	18672	7.41	28320	28
MSD	440	45	0.23	21	0	275	2222	14.91	80	8
GFTT	1777	133	2.76	363	8	2950	17278	80.37	342	12
GFTT-H	356	33	0.56	84	0	652	14859	75.92	64	3
Harris-L	811	43	0.25	28	1	1235	8350	34.41	302	23
CenSurE	317	12	0.00	19	0	189	79	2.12	12	0
	MOTION BLURRED SCENE					EXTREMELY AFFINE SCENE				
SIFT	9118	65	0.18	1279	6	2264	4087	20.45	574	4
SURF	10326	2231	6.14	2673	12	4778	5940	17.30	1273	5
KAZE	4494	15	0.09	2524	0	3381	3118	15.91	578	2
AKAZE	4168	2	0.05	12	0	2409	2306	15.65	609	2
ORB	63298	1	0.00	0	0	8993	17206	36.35	1892	6
BRISK	22834	1	0.00	0	0	3507	6754	30.04	728	6
AGAST	45909	24	0.04	1269	1	7501	16023	41.66	629	0
FAST	41277	31	0.07	1589	1	7241	14345	39.04	599	1
MSER	396039	90238	16.44	76410	18	258765	241350	23.81	77048	51
MSD	3978	0	0.00	0	0	1124	1579	20.90	123	1
GFTT	30537	24710	48.02	2074	5	3076	6890	36.60	659	3
GFTT-H	11850	605	1.84	661	4	868	1988	31.65	145	1
Harris-L	15164	138	0.45	784	9	2627	3742	6.00	228	0
CenSurE	1659	0	0.00	0	0	885	852	11.07	91	1
	JPEG COMPRESSED SCENE					OCCLUDED SCENE				
SIFT	4165	2525	23.36	1067	287	487	857	18.48	112	13
SURF	5959	3811	28.11	2222	623	997	1894	16.25	257	16
KAZE	2787	2932	63.33	1926	1300	553	1026	20.80	115	6
AKAZE	2140	2190	50.46	1266	822	383	766	18.54	112	3
ORB	25295	15267	38.84	7389	3263	3575	6510	27.72	774	20
BRISK	9528	11480	41.39	2301	984	1661	2811	23.78	382	13
AGAST	25071	5615	15.84	7575	2816	2889	5508	26.20	153	5
FAST	21359	3530	12.47	5878	2273	2622	4324	24.14	154	9
MSER	161965	145533	48.43	43382	625	35697	85589	29.65	9801	102
MSD	2054	1510	25.41	909	427	349	523	7.74	52	2
GFTT	12880	6198	24.35	2848	1145	1482	2265	21.73	255	8
GFTT-H	2121	2014	43.52	908	619	620	878	15.48	111	3
Harris-L	1947	2347	52.85	1085	507	287	475	12.89	46	2
CenSurE	897	863	48.27	612	388	112	307	18.75	13	3
	SUNNY-SHADOWED SCENE					DECORATED SCENE				
SIFT	2505	248	2.08	479	0	3209	2327	17.79	726	34
SURF	3059	980	8.85	844	18	4024	4453	25.35	1082	22
KAZE	2572	462	1.96	439	1	2957	2671	12.58	595	9
AKAZE	2037	245	2.36	537	0	2388	2001	13.90	640	11
ORB	15683	1337	4.50	4157	5	20586	17265	36.84	3946	78
BRISK	6654	574	3.62	1303	0	6930	5774	25.18	1267	15
AGAST	9235	1586	9.70	1964	183	13808	12735	45.94	2077	358
FAST	9286	1583	10.88	1950	201	12993	11994	48.94	1911	360
MSER	109311	112134	70.93	33284	2341	137310	168189	60.44	43161	1872
MSD	1030	180	1.59	109	0	1360	1093	8.90	252	16
GFTT	5628	526	4.02	670	6	7223	5766	33.24	1780	222
GFTT-H	2632	265	2.28	270	0	1989	1593	14.63	432	21
Harris-L	2468	206	0.60	312	2	1323	1374	14.36	209	1
CenSurE	869	75	0.80	76	0	1037	785	9.06	182	18

APPENDIX

TABLE V. NORMALTHRESHOLD SETTINGS USED TO EVALUATE THE TYPICAL PERFORMANCE OF THE FEATURE DETECTION ALGORITHMS

METHOD	OpenCV’s Constructor Settings
SIFT	cv.SIFT('NOctaveLayers',3,'ConstrastThreshold',0.04,'EdgeThreshold',10,'Sigma',1.6)
SURF	cv.SURF('NOctaveLayers',3,'HessianThreshold',100)
KAZE	cv.KAZE('NOctaveLayers',4,'Threshold',0.001)
AKAZE	cv.AKAZE('NOctaveLayers',4,'DescriptorType','MLDB','Threshold',0.001)
ORB	cv.ORB('MaxFeatures',1e+6,'FastThreshold',20)
BRISK	cv.BRISK('Octaves',4,'Threshold',30)
AGAST	cv.AGAST(Image,'Type','AGAST_7_12d','NonmaxSuppression', true,'Threshold',10)
FAST	cv.FAST(Image,'NonmaxSuppression',true,'Threshold',10)
MSER	cv.MSER('MaxEvolution',200,'EdgeBlurSize',5)
MSD	cv.MSDDetector('ThSaliency',250,'NMSRadius',5,'ComputeOrientation',true,'NScales',3)
GFTT	cv.GFTTDetector('MinDistance',1,'MaxFeatures',1e+6,'QualityLevel',0.01)
GFTT-H	cv.GFTTDetector('HarrisDetector',true,'K',0.04,'MaxFeatures',1e+6,'QualityLevel',0.01)
Harris-L	cv.HarrisLaplaceFeatureDetector('CornThresh',0.01,'DOGThresh',0.01, 'MaxCorners',1e+6)
CenSurE	cv.StarDetector('ResponseThreshold',30,'LineThresholdProjected',10, 'LineThresholdBinarized',8,'SuppressNonmaxSize',5)

COMPARISON OF IMAGE MATCHING WITH NORMAL THRESHOLDS & EXTREMELY LOW THRESHOLDS

Table IV shows the results of image matching by using normal parameter thresholds of the feature detection algorithms whereas Table V describes the threshold values we have used. In Table II, we have presented the matching results which were based on extremely low parameter thresholds of the detectors and Table III describes the applied extremely low threshold values. If feature detectors with normal thresholds are applied, the quantity of detected features in the deteriorated images (image-2 of each scene) and the number of correct matches are substantially decreased (see Table IV). Moreover, repeatability scores of majority of the feature detectors decline when normal thresholds are used. With extremely relaxed thresholds (full throttle), a clear increase in quantity of detected features, number of correct matches, and repeatability scores is observed (see Table II).