Radial Line Fourier Descriptor for Handwritten Word Representation

Anders Hast and Ekta Vats
Department of Information Technology
Uppsala University, SE-751 05 Uppsala, Sweden
anders.hast@it.uu.se; ekta.vats@it.uu.se

ABSTRACT

Automatic recognition of historical handwritten manuscripts is a daunting task due to paper degradation over time. The performance of information retrieval algorithms depends heavily on feature detection and representation methods. Although there exist popular feature descriptors such as Scale Invariant Feature Transform and Speeded Up Robust Features, in order to represent handwritten words in a document, a robust descriptor is required that is not over-precise. This is because handwritten words across different documents are indeed similar, but not identical. Therefore, this paper introduces a Radial Line Fourier (RLF) descriptor for handwritten word feature representation, which is fast to construct and short-length with 32 elements only. The effectiveness of the proposed RLF descriptor is empirically evaluated using the VLFeat benchmarking framework (VLBenchmarks), and for handwritten word image representation using a historical marriage records dataset.

KEYWORDS

Radial Line Fourier descriptor, interest point detection, feature representation

ACM Reference format:
Anders Hast and Ekta Vats. 2017. Radial Line Fourier Descriptor for Handwritten Word Representation. In *Proceedings of , , ,* 6 pages.
https://doi.org/

1 INTRODUCTION

Automatic recognition of poorly degraded handwritten text is a daunting task due to complex layouts and paper degradations over time. Typically, an old manuscript suffers from degradations such as paper stains, faded ink and ink bleed-through. There is a variability in writing style, and the presence of text and symbols written in an unknown language. This hampers the document readability and make tasks like word spotting more challenging. However, the performance of information retrieval algorithms as well as other computer vision applications depends heavily on the appropriate selection of feature detection and representation methods [11].

Efforts have been made in the recent past towards research on feature detection and representation. Some popular methods include Scale Invariant Feature Transform (SIFT) [18], Speeded Up Robust Features (SURF) [4] and Histograms of oriented Gradients (HoG) [8]. SIFT and HoG contributed significantly towards the progress of several visual recognition systems in the last decade [12]. In a word spotting scenario, the performance of different features was evaluated using Dynamic Time Warping (DTW) [23] and Hidden Markov Models (HMMs) [24]. It was found that local gradient histogram features outperform other geometrical or profile-based features. These methods generally match features from evenly distributed locations over normalised words [22] where no nearest neighbour search is necessary as each point in a word has its corresponding point in the other word located in the very same position. Recently, a method based on feature matching of keypoints derived from the words was proposed [14], which requires a nearest neigbour search. In this case, a robust descriptor is required that is not too precise, since the handwritten words are not normalised. This is due to complex characteristic of handwritten words, unlike simple OCR text. Handwritten words across different documents are indeed similar, but not identical due to variability in writing styles.

This paper proposes a Radial Line Fourier (RLF) descriptor for handwritten word feature representation. In general, the RLF descriptor is based on the idea of radial lines integration. RLF descriptor is tailor-made for word spotting applications with fast feature representation and robustness to degradations. However, the RLF descriptor can be flexibly used in other applications for feature representation of challenging images with promising results. The VLFeat benchmarking framework called VLBenchmarks [17] is used to test the descriptor performance. The RLF descriptor is capable of handling viewpoint changes, scale-invariance to a limited extent, and conditions such as illumination, defocus and image compression. This paper evaluates the RLF descriptor on degraded word images and challenging scene images of varying image conditions. Also an elaborate comparison analysis is done using RLF and other popular methods such as SIFT and SURF using VLBenchmarks to demonstrate the effectiveness of the proposed RLF descriptor.

2 RELATED WORK

Appropriate selection of interest points and descriptors is indispensable for the performance of a word spotting system. This section discusses some popular interest point detection and feature representation methods with reference to word spotting systems.

2.1 Interest Point Detection

Feature detection, or interest point detection refers to finding key points in an image that contain crucial information. The selection of interest point detectors has a great impact on the performance of an information retrieval algorithm. Several methods have been suggested in literature for interest point detection [27, 32, 34]. The

, ,

ACM ISBN .
https://doi.org/

Harris corner detector [13] is popularly used for corner points detection. It computes a combination of eigenvalues of the structure tensor such that the corners are located in an image. Shi-Tomasi corner detector [28] is a modified version of Harris detector. The minimum of two eigenvalues is computed and a point is considered as a corner point if this minimum value exceeds a certain threshold. The FAST detector [26], based on the SUSAN detector [29], uses a circular mask to test against the central pixel. MSER [19] detects keypoints such that all pixels inside the extremal region are either darker or brighter than all the outer boundary pixels.

In general, interest point based feature matching is done by using a single type of interest point detector. SIFT and SURF are the most popular ones that capture the blob type of features in the image. SIFT uses the Difference of Gaussians (DoG) that computes the difference between Gaussian blurred images using different values of σ, where σ defines the Gaussian blur from a continuous point of view. SURF computes the Determinant of the Hessian (DoH) matrix, that defines the product of the eigenvalues. Several different combinations of any number of interest point detectors can be chosen depending upon the application [32]. On the other hand, the RLF descriptor proposed in this work is independent of the choice of interest points selection. Any efficient interest point detection method can be flexibly employed with the RLF descriptor.

2.2 Feature Representation

After a set of interest points has been detected, a suitable representation of their values has to be defined to allow matching between a query word image and the document image. In general, a feature descriptor is constructed from the pixels in the local neighborhood of each interest point. Fixed length feature descriptors are most commonly used that generate a fixed length feature vector, which can be easily compared using standard distance metrics (e.g. the Euclidean distance). Sometimes, fixed length feature vectors are computed directly from the extracted features without the need of a learning step [11].

Gradient-based feature descriptors tend to be superior, and include SIFT [18], HoG [8] and SURF [4] descriptors. The 128 dimensional SIFT descriptor is formed from histograms of local gradients. SIFT is both scale and rotation invariant, and includes a complex underlying framework to ensure this. Similarly, HoG computes a histogram of gradient orientations in a certain local region. However, SIFT and HoG differ in the sense that HoG normalizes the histograms in overlapping blocks, and creates a redundant expression. The SURF descriptor is generally faster than SIFT, and is created by concatenating Haar wavelet responses in sub-regions of an oriented square window. SIFT and SURF are invariant to rotation changes, unlike HoG. There are several variants of these descriptors that have been effectively employed for word spotting [11, 23, 30].

The KAZE detector [2] uses a non linear scale space created using non linear diffusion filtering, instead of Gaussian Blurring. An accelerated version AKAZE [21] uses a faster method for creating the scale space and a binary descriptor. Many feature descriptors use the local image content in square areas around each interest point to form a feature vector [14]. Both scale and rotation invariance can be obtained in different ways [10]. The Fourier transform has been used to compute descriptors [6, 7, 33] that is illumination and rotation invariant, and scale-invariant to a certain extent. In order to overcome dimensionality issues, binary descriptors are introduced that are faster, but less precise, for example the BRISK descriptor [17] and FREAK [1, 17].

The choice of feature descriptor depends upon the target application. For handwritten words representation in a document, a fast and robust descriptor like RLF descriptor is required that is not over-precise. The RLF descriptor is discussed in detail as follows.

3 RADIAL LINE FOURIER DESCRIPTOR

Radial Line Fourier (RLF) Descriptor is inspired from a variant of Scale Invariant Descriptor (SID) [16], i.e. SID-Rot [31]. In general, the idea is to perform log-polar sampling in a circular neighborhood around each keypoint [15]. Then the Fourier transform is applied over scales, making it rotation sensitive (hence the name SID-Rot). However, a desirable property is that it will be less sensitive to scale changes. Nevertheless, in order to achieve this, the descriptor is dense with a very large radius, and a length of 3360.

The RLF descriptor addresses this issue and computes a fast and short-length feature vector of 32 dimensions, to be able to perform quick matching in the nearest neighbour search. The RLF descriptor is formed directly from the interest points extracted, without the need to involve a learning step. It characterizes an image region as a whole using a single feature vector of fixed size.

To begin with, the Fourier Feature Transform (FFT) is simplified as it is rather slow and requires $O(N\ log(N))$ computations for a discrete series $f(n)$ with N elements. The modification is such that each element needed will be computed using the Discrete Fourier Transform (DFT) $f(n)$. Therefore, rewriting using Euler's formula, the computation required is

$$\mathcal{F}[f(n)](k) = \sum_{n=0}^{N-1} f(n)\cos(2\pi nk/N) - i(f(n)\sin(2\pi nk/N)). \quad (1)$$

The value of k determines the frequency used to compute the Fourier element, where $k \in 0, 2, 4...$ Since noise has higher frequencies as compared to the main structures in the image with lower frequencies, the second ($k = 2$) and third ($k = 4$) elements of the Fourier transform are selected to form a descriptor. Hence, now the algorithm requires only $O(N)$ computations. Note that the Discrete Cosine (DC) component is obtained for $k = 0$ and is less informative. The trigonometric functions in the DFT do not have to be computed for each step as they can be efficiently computed using a few additions and multiplications by the Chebyshev recurrence relation [3, 5], just as is done in the case of FFT.

The RLF descriptor is constructed by computing the amplitude:

$$|\mathcal{F}[f(n)](k)| = \sqrt{\Re(\mathcal{F}[f(n)](k))^2 + \Im(\mathcal{F}[f(n)](k))^2}. \quad (2)$$

Forming the descriptor using only $k = 2$ suffices very well and the descriptor is very short. However, adding a second element for $k = 4$ improves the quality of the subsequent matching noticeably, even if the feature vector will be twice as long. The advantage is, however, that it makes it possible to sample in a smaller neighborhood, while still getting the same number of corresponding matches, as it is more accurate. Nevertheless, adding a third element for $k = 6$ did not improve the accuracy significantly, and is found to be not worth the extra computational effort.

When sampling is done in a log-polar fashion, some kind of interpolation is required as coordinates seldom are in pixel centres. One could for instance use bilinear interpolation to achieve higher accuracy. However, interpolation in a 3x3 neighborhood using a Gaussian is chosen instead.

The RLF descriptor is illumination and rotation invariant, and also scale-invariant to a limited extent. These are important characteristics a feature vector must possess to handle different kind of words with varying size, shape, slant characters etc. RLF descriptor is resistant to high frequency changes, such as due to residuals from neighboring words, as it is based on the low frequency content in the local neighborhood. Nevertheless, it is insensitive to small differences in form and shape as long as they are more or less the same, i.e. the low frequencies are sufficiently similar. The RLF descriptor present a clear advantage over other feature representations as has been experimentally validated in the next section.

4 EXPERIMENTAL RESULTS

This section describes the datasets used in the experiments and empirically evaluates the proposed RLF descriptor.

4.1 Esposalles Dataset

In order to evaluate the performance of RLF descriptor in representing a word in a degraded historical document, a subset of the Barcelona Historical Handwritten Marriages (BH2M) database [9] i.e. Esposalles dataset [25] is used for the experiments. The Esposalles dataset consists of handwritten marriages records stored in the archives of Barcelona cathedral, written between 1617 and 1619 by a single writer in old Catalan. In total, there are 174 handwritten pages corresponding to the volume 69. For the experiments, 50 pages written by a single author are selected from the 17th century.

4.2 VGG Affine Dataset

VGG Affine dataset consists of a set of test images under varying imaging conditions [20]. It consists of eight scene types i.e. graffiti, wall, boat, bark, bikes, trees, ubc and leuven, where each of the categories contain images with different conditions. It is employed in the VLBenchmarks [17] framework for testing image feature detectors and descriptors. This dataset effectively helps in testing the performance of the RLF descriptor with reference to a variety of test images, and for comparing with other feature descriptors such as SIFT and SURF.

The descriptor is evaluated under five different imaging conditions: viewpoint angle change, scale change, image blur, JPEG compression and illumination change. The same change in imaging conditions is applied in case of viewpoint change, scale change and blur for two different scene categories. This means that the effect of varying the image conditions can be separated from the effect of varying the scene type. Structured scene category consists of homogeneous regions with distinctive edge boundaries (e.g. graffiti and buildings), and the textured scene category consists of repeated textures of different forms [20].

In the test for viewpoint angle change, the camera varies from frontal parallel view to significant foreshortening at approximately 60 degrees to the camera. The illumination variations are introduced by changing the camera aperture. The scale change is acquired by varying the camera zoom and it changes by about a factor of four. The blur sequences are acquired by varying the camera focus. The JPEG compression sequence is generated using a standard xv image browser with the image quality parameter varying from 40% to 2%. Each of the test sequences contains six images with a gradual geometric or photometric transformation. All images are of medium resolution (approximately 800 x 640 pixels).

Table 1: Comparison between RLF and HoG descriptors for finding three occurrences of the query words. The average is reported and the number of asterisks (*) denotes the count of words not present due to a low number of matching points.

Query words	No. of Matched points		No. of inliers		Inlier ratio	
	RLF	HoG	RLF	HoG	RLF	HoG
reberé	120	61	114	57	0.95	0.93
pages	109	84	98	81	0.89	0.96
habitant	122	79**	102	69**	0.83	0.87
fill	81	***	73	***	0.90	-
Barna	94	59*	83	53*	0.88	0.89

4.3 Results

To evaluate the performance of the RLF descriptor for word image representation, the number of matched feature points and the number of inliers are calculated. Let $matchedPointsNum$ indicate the number of matched points, $inliersNum$ indicate the number of inliers i.e. true points, then the inlier ratio can be defined as:

$$InlierRatio = \frac{matchedPointsNum}{inliersNum} \tag{3}$$

In the ideal case, the inlier ratio should be 1. In the first set of experiments, the RLF descriptor is compared with the HoG descriptor as it is most widely used in word spotting applications [23, 30]. For the same query word, the number of matching interest points found using RLF and HoG descriptors are calculated and quantitatively evaluated as in Table 1. In order to perform word matching and find all inliers, a preconditioner based simple clustering method is employed. Figure 1 present the sample results obtained using RLF and HoG feature descriptors for the two variants of example query word *reberé*. The matching keypoints (i.e. inliers) are in green and the matches that are discarded (i.e. outliers) by the preconditioner are in red. It is clearly observed that the number of inliers found using the RLF descriptor is more than the number of inliers found using the HoG descriptor, and has been quantitatively evaluated in Table 1. The matching algorithm divides the word into three parts in order to avoid mismatching the same letter occurring in several places. In the table one can note that HoG produces noticeably less matching points. This causes the algorithm to miss some words. In the experiments, three occurrences were found in the search using RLF, while some were missed by HoG, and in one case none was found. This could partly be solved by relaxing the threshold. However, then some incorrect words are found instead. This clearly shows the advantage of RLF over HoG, since it is less precise in the sense that the neighborhood forming the descriptor can be non equal, yet similar while being robust enough and not causing too many mismatches, i.e. yielding a high inlier ratio.

(a) Matched points represented with RLF descriptor.

(b) Matched points represented with HoG descriptor.

Figure 1: Example results obtained using RLF and HoG descriptors for an example query word *reberé*. The matching keypoints (inliers) are in green and the matches discarded (outliers) by the preconditioner are in red. Figure best viewed in color.

Table 2: Number of matches evaluated using VLBenchmarks on VGG Affine dataset.

Images	Imaging conditions	SIFT	SIFT+RLF64	SIFT+RLF32	SURF	SURF+RLF64	SURF+RLF32
Image2	viewpoint	1450	1408	1401	1082	1304	1294
	JPEG compression	1276	1222	1219	1271	1337	1331
	decreasing light	992	933	937	544	575	565
	increasing blur	1532	1514	1509	517	569	570
Image3	viewpoint	1126	1132	1110	744	1006	987
	JPEG compression	1067	1025	1023	1134	1208	1209
	decreasing light	931	883	894	402	405	407
	increasing blur	1441	1412	1398	422	472	467
Image4	viewpoint	733	702	644	436	584	559
	JPEG compression	837	806	803	959	1055	1060
	decreasing light	869	808	807	292	316	316
	increasing blur	1003	971	972	254	314	313
Image5	viewpoint	372	302	282	178	243	226
	JPEG compression	551	523	523	685	817	818
	decreasing light	814	749	751	205	218	213
	increasing blur	794	731	734	190	220	218
Image6	viewpoint	27	24	19	18	22	23
	JPEG compression	346	281	282	403	490	498
	decreasing light	688	639	633	134	156	157
	increasing blur	573	455	459	126	129	129

Table 3: Descriptor matching scores evaluated using VLBenchmarks on VGG Affine dataset.

Images	Imaging conditions	SIFT	SIFT+RLF64	SIFT+RLF32	SURF	SURF+RLF64	SURF+RLF32
Image2	viewpoint	60.29	58.57	58.28	52.25	62.93	62.45
	JPEG compression	70.42	67.44	67.27	80.65	84.84	84.45
	decreasing light	63.06	59.31	59.57	67.16	70.99	69.75
	increasing blur	56.53	55.83	55.64	73.86	81.29	81.43
Image3	viewpoint	49.32	49.58	48.62	38.04	51.43	50.46
	JPEG compression	58.89	56.57	56.46	73.83	78.65	78.71
	decreasing light	58.26	55.26	55.94	60.27	60.72	61.02
	increasing blur	52.38	51.35	50.84	69.18	77.38	76.56
Image4	viewpoint	38.32	36.75	33.72	25.69	34.43	32.96
	JPEG compression	46.19	44.48	44.32	60.85	66.94	67.26
	decreasing light	54.65	50.85	50.79	52.42	56.73	56.73
	increasing blur	37.37	36.14	36.17	59.76	73.88	73.65
Image5	viewpoint	20.76	16.85	15.74	11.09	15.14	14.08
	JPEG compression	30.41	28.86	28.86	43.46	51.84	51.90
	decreasing light	51.03	47.11	47.23	46.17	49.10	47.97
	increasing blur	29.69	27.33	27.44	59.56	68.97	68.34
Image6	viewpoint	1.84	1.64	1.30	1.50	1.83	1.91
	JPEG compression	19.09	15.51	15.56	27.38	33.29	33.83
	decreasing light	43.79	40.70	40.32	38.40	44.70	44.99
	increasing blur	22.88	18.16	18.32	58.60	60.00	60.00

In the next set of experiments, the performance of the RLF descriptor is evaluated using the VLBenchmarks framework. Table 2 and Table 3 present the number of matches and match scores, respectively, obtained using six different combinations of feature detectors and descriptors, i.e. SIFT descriptor, 64-dimensional RLF descriptor with SIFT keypoints (SIFT+RLF64 with 32 radial lines), 32-dimensional RLF descriptor with SIFT keypoints (SIFT+RLF32 with 16 radial lines), SURF descriptor, 64-dimensional RLF descriptor with SURF keypoints (SURF+RLF64) and 32-dimensional RLF descriptor with SURF keypoints (SURF+RLF32). The varying imaging conditions taken into account include viewpoint angle change, JPEG compression, decreasing light (illumination changes) and increasing blur (defocus). The RLF descriptor is scale invariant to a certain extent, and will be further investigated in future work.

Table 4: Matching scores obtained using MSER and DoG keypoints with different descriptors.

Images	Imaging conditions	MSER keypoints				DoG keypoints			
		SIFT	RLF64	RLF32	SURF	SIFT	RLF64	RLF32	SURF
Image 2	viewpoint	64.07	49.81	49.35	40.90	53.59	46.29	45.63	35.74
	JPEG compression	53.93	40.43	39.82	41.10	54.18	44.19	43.96	36.84
	decreasing light	71.08	53.77	53.35	48.68	54.32	41.62	41.75	38.59
	increasing blur	66.59	51.48	51.70	51.95	48.21	39.93	39.73	37.11
Image 3	viewpoint	58.48	42.06	41.28	33.48	46.19	36.67	34.92	27.19
	JPEG compression	47.85	34.31	34.18	34.02	45.67	35.17	34.92	28.64
	decreasing light	68.39	49.48	49.78	42.92	53.02	38.65	38.55	35.80
	increasing blur	63.53	43.87	45.30	44.67	50.74	42.16	42.10	39.46
Image 4	viewpoint	49.18	27.85	25.71	21.36	36.55	22.46	19.88	15.84
	JPEG compression	38.33	25.81	25.70	22.52	34.08	23.28	23.10	17.87
	decreasing light	67.74	48.19	46.85	39.07	50.20	35.93	36.11	30.58
	increasing blur	53.99	30.81	31.04	38.39	53.90	45.87	45.18	43.12
Image 5	viewpoint	29.91	12.10	10.32	8.63	21.91	8.21	7.01	5.56
	JPEG compression	27.69	15.45	14.95	14.56	30.51	18.19	18.01	15.13
	decreasing light	53.99	30.81	31.04	38.39	53.90	45.87	45.18	43.12
	increasing blur	50.28	23.58	23.30	33.52	55.85	43.84	43.99	45.09
Image 6	viewpoint	4.73	1.51	1.24	1.09	2.38	0.76	0.54	0.55
	JPEG compression	18.57	8.27	8.20	9.41	26.48	13.08	13.02	12.68
	decreasing light	70.04	46.27	46.91	41.04	47.48	32.61	32.69	26.30
	increasing blur	47.62	19.20	22.00	34.68	55.96	40.22	41.35	45.62

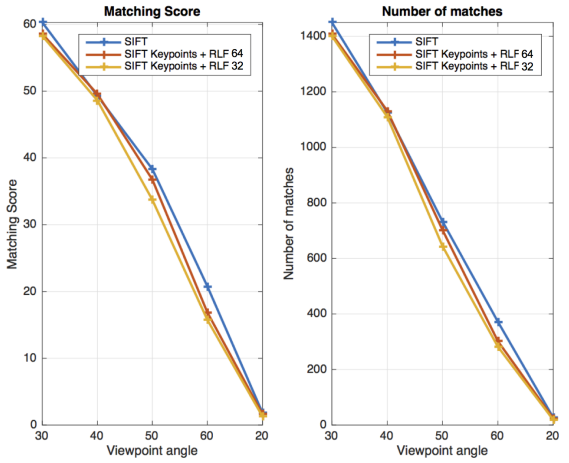


Figure 2: Matching performance comparison between SIFT and RLF descriptor on VGG Affine dataset with varying viewpoint angle. Figure best viewed in color.

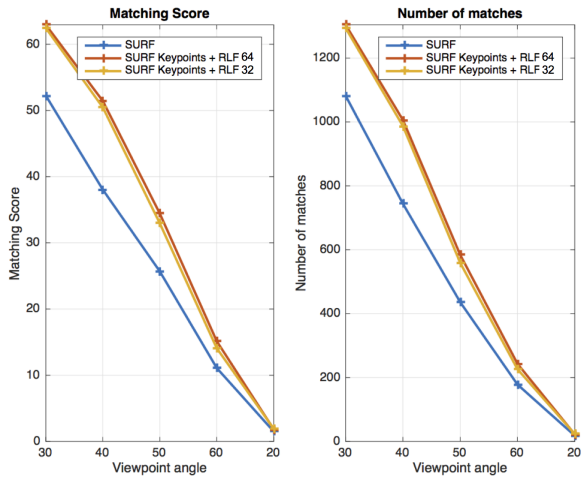
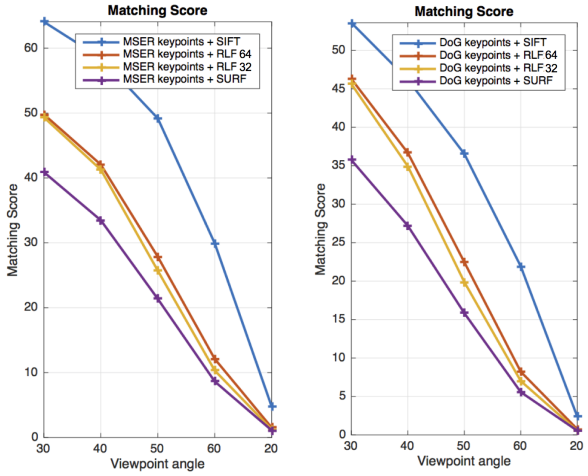


Figure 3: Matching performance comparison between SURF and RLF descriptor on VGG Affine dataset with varying viewpoint angle. Figure best viewed in color.

Figures 2 and 3 graphically illustrate the matching performance in terms of matching scores and the number of matches obtained using different combinations of feature detectors and descriptors with varying viewpoint angles. In Fig. 2, it can be seen that the RLF performs fairly well with varying viewpoint angle changes, in comparison with the SIFT descriptor. Figure 3 highlights the performance of the RLF descriptor with viewpoint change in comparison with the SURF descriptor. It is clearly observed that the RLF descriptor performs better than the SURF descriptor (length 64) in terms of matching scores and the number of matches obtained.

Furthermore, tests are conducted using MSER and DoG keypoint detectors with different feature descriptor combinations, i.e. SIFT, RLF64, RLF32 and SURF. Figure 4 graphically presents the matching performance comparison between MSER keypoints represented using SIFT, RLF and SURF descriptors, and DoG keypoints represented using the same set of descriptors in varying viewpoint conditions. Table 4 quantitatively evaluates the performance of various descriptors and presents the set of results obtained using MSER and DoG keypoints with these descriptors in challenging imaging conditions. It can be seen that RLF outperforms SURF in nearly all the categories and varying conditions. However, it performs fairly in comparison with SIFT, not always better, depending upon the input images, and can be further improved as future work.



(a) Matching scores obtained using MSER keypoint detector and different descriptors.

(b) Matching scores obtained using DoG keypoint detector and different descriptors.

Figure 4: Performance evaluation results obtained by using MSER and DoG keypoints with different descriptors.

5 CONCLUSIONS

This paper presented a fast and robust Radial Line Fourier descriptor. The state-of-the-art feature descriptors such as SIFT, HoG and SURF include a rather complicated framework which is not needed in all applications such as word spotting. Therefore, a much simpler, yet effective descriptor is proposed that is not too precise for handwritten words representation. The novelty of the proposed descriptor lies in lesser computation time, shorter length of feature vector, invariance to rotation, viewpoint angles, and scale to a certain extent, and other issues such as illumination, defocus and image compression. The experimental results on a historical marriage records dataset and VGG Affine dataset demonstrate the effectiveness of the proposed descriptor in handwritten word image representation and test scene images from the VLBenchmarks framework. As future work, the ideas presented herein will be scaled to aid word feature representation for heavily degraded archival databases.

REFERENCES

[1] A. Alahi, R. Ortiz, and P. Vandergheynst. 2012. FREAK: Fast Retina Keypoint. In *Computer Vision and Pattern Recognition, 2012 IEEE Conference on*. 510–517.

[2] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J. Davison. 2012. KAZE Features. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part VI (ECCV'12)*. Springer-Verlag, Berlin, Heidelberg, 214–227.

[3] T. Barrera, A. Hast, and E. Bengtsson. 2004. Incremental Spherical Linear Interpolation. In *Sigrad 2004*. 7–10.

[4] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. 2008. Speeded-Up Robust Features (SURF). *Computer vision and image understanding* 110, 3 (2008), 346–359.

[5] R. L. Burden and J. D Faires. 2001. *Numerical Analysis Brooks*. Cole, Thomson Learning. 507–516 pages.

[6] Gustavo Carneiro and Allan D. Jepson. [n. d.]. In *In European Conference on Computer Vision (ECCV), Date-Added = 2017-07-18 08:53:17 +0000, Date-Modified = 2017-07-18 08:53:17 +0000, Pages = 282–296, Title = Phase-based local features, Year = 2002.*

[7] Gustavo Carneiro and Allan D. Jepson. 2003. Multi-scale phase-based local features. In *Proceedings of the 2003 IEEE computer society conference on Computer vision and pattern recognition (CVPR'03)*. 736–743.

[8] Navneet Dalal and Bill Triggs. 2005. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1. IEEE, 886–893.

[9] David Fernández-Mota, Jon Almazán, Núria Cirera, Alicia Fornés, and Josep Lladós. 2014. Bh2m: The barcelona historical, handwritten marriages database. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 256–261.

[10] Steffen Gauglitz, Tobias Höllerer, and Matthew Turk. 2011. Evaluation of interest point detectors and feature descriptors for visual tracking. *International journal of computer vision* 94, 3 (2011), 335–360.

[11] Angelos P Giotis, Giorgos Sfikas, Basilis Gatos, and Christophoros Nikou. 2017. A survey of document image word spotting techniques. *Pattern Recognition* 68 (2017), 310–332.

[12] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587.

[13] C. Harris and M. Stephens. 1988. A Combined Corner and Edge Detector. In *Proceedings of The Fourth Alvey Vision Conference*. 147–151.

[14] Anders Hast and Alicia Fornés. 2016. A Segmentation-free Handwritten Word Spotting Approach by Relaxed Feature Matching. In *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on*. IEEE, 150–155.

[15] Iasonas Kokkinos, Michael Bronstein, and Alan Yuille. 2012. *Dense Scale Invariant Descriptors for Images and Surfaces*. Research Report RR-7914. INRIA.

[16] Iasonas Kokkinos and Alan L. Yuille. 2008. Scale invariance without scale selection. In *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[17] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. 2011. BRISK: Binary Robust Invariant Scalable Keypoints. In *Proceedings of the 2011 International Conference on Computer Vision (ICCV '11)*. IEEE Computer Society, 2548–2555.

[18] D. G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 2 (2004), 91–110.

[19] J. Matas, O. Chum, M. Urban, and T. Pajdla. 2002. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In *Proceedings of the British Machine Vision Conference*. BMVA Press, 36.1–36.10.

[20] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, Jiri Matas, Frederik Schaffalitzky, Timor Kadir, and Luc Van Gool. 2005. A comparison of affine region detectors. *International journal of computer vision* 65, 1-2 (2005), 43–72.

[21] Adrien Bartoli Pablo Alcantarilla, Jesus Nuevo. 2013. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In *Proceedings of the British Machine Vision Conference*.

[22] A. Papandreou and B. Gatos. 2014. Slant Estimation and Core-region Detection for Handwritten Latin Words. *Pattern Recognition Letters* 35 (2014), 16–22.

[23] Jose A Rodriguez and Florent Perronnin. 2008. Local gradient histogram features for word spotting in unconstrained handwritten documents. *Frontiers in Handwriting Recognition (ICFHR), 1st International Conference on* (2008), 7–12.

[24] José A Rodríguez-Serrano and Florent Perronnin. 2009. Handwritten word-spotting using hidden Markov models and universal vocabularies. *Pattern Recognition* 42, 9 (2009), 2106–2116.

[25] VeróNica Romero, Alicia FornéS, NicoláS Serrano, Joan Andreu SáNchez, Alejandro H Toselli, Volkmar Frinken, Enrique Vidal, and Josep LladóS. 2013. The ESPOSALLES database: An ancient marriage license corpus for off-line handwriting recognition. *Pattern Recognition* 46, 6 (2013), 1658–1669.

[26] Edward Rosten and Tom Drummond. 2006. Machine Learning for High-Speed Corner Detection. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part I (ECCV'06)*. Springer-Verlag, Berlin, Heidelberg, 430–443.

[27] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. 2000. Evaluation of Interest Point Detectors. *International Journal of Computer Vision* 37, 2 (June 2000), 151–172.

[28] Jianbo Shi and C. Tomasi. 1994. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR, 1994 IEEE Computer Society Conference on*. IEEE, 593–600.

[29] Stephen M. Smith and J. Michael Brady. 1997. SUSAN - A New Approach to Low Level Image Processing. *Int. J. Comput. Vision* 23, 1 (May 1997), 45–78.

[30] Kengo Terasawa and Yuzuru Tanaka. 2009. Slit style HOG feature for document image word spotting. In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*. IEEE, 116–120.

[31] Eduard Trulls, Iasonas Kokkinos, Alberto Sanfeliu, and Francesc Moreno-Noguer. 2013. Dense Segmentation-Aware Descriptors. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*. 2890–2897.

[32] Tinne Tuytelaars and Krystian Mikolajczyk. 2008. Local Invariant Feature Detectors: A Survey. *Foundations and Trends in Computer Graphics and Vision* 3, 3 (July 2008), 177–280.

[33] I. Ulusoy and E. R. Hancock. 2007. A statistical approach to sparse multi-scale phase-based stereo. *Pattern Recogn.* 40, 9 (Sept. 2007), 2504–2520.

[34] M. Zuliani, C. Kenney, and B. S. Manjunath. 2004. A Mathematical Comparison of Point Detectors. In *2nd IEEE Image and Video Registration Workshop*. 172–178.