# Recognition of Degraded Handwritten Characters Using Local Features*

Markus Diem and Robert Sablatnig
Vienna University of Technology, Austria
diem@prip.tuwien.ac.at

## Abstract

*The main problems of Optical Character Recognition (OCR) systems are solved if printed latin text is considered. Since OCR systems are based upon binary images, their results are poor if the text is degraded. In this paper a codex consisting of ancient manuscripts is investigated. Due to environmental effects the characters of the analyzed codex are washed out which leads to poor results gained by state of the art binarization methods. Hence, a segmentation free approach based on local descriptors is being developed. Regarding local information allows for recognizing characters that are only partially visible. In order to recognize a character the local descriptors are initially classified with a Support Vector Machine (SVM) and then identified by a voting scheme of neighboring local descriptors. State of the art local descriptor systems are evaluated in this paper in order to compare their performance for the recognition of degraded characters.*

## 1 Introduction

Off-line character recognition systems generally consist of three steps [17]. First a preprocessing step is applied where the images are enhanced (e.g. background removal, line segmentation) and thresholded. Afterwards a segmentation of characters or words is performed. Finally, features of the fragments, characters or words are computed which are subsequently classified. This architecture has proven to give good results for general document recognition ($60.0\% - 99.3\%$) [17]. However, there exist alphabets like the glagolitic for which a generally designed character recognition system fails, since words are not separated and the manuscripts are degraded. In this paper Glagolitica, the oldest slavonic alphabet, is considered. More precisely, the Missale (Cod. Sin. slav. 5/N) is analyzed which was written in the 11$^{\text{th}}$ century and discovered in 1975 at the St. Catherine's Monastery [8].

In this script the space between characters is equidistant – independent to words – which excludes the use of word segmentation and dictionaries. In addition the manuscripts are degraded due to bad storage conditions and environmental effects. The partially faded-out ink cannot be handled with state of the art binarization methods [4, 14]. Figure 1 shows a sample of the glagolitic manuscript and the corresponding binary image when the Sauvola thresholding method is applied having tuned parameters $(k, R)$. In addition the correct glagolitic character is overlaid. Preliminary tests showed that this character can be recognized with the proposed system (see Figure 2).

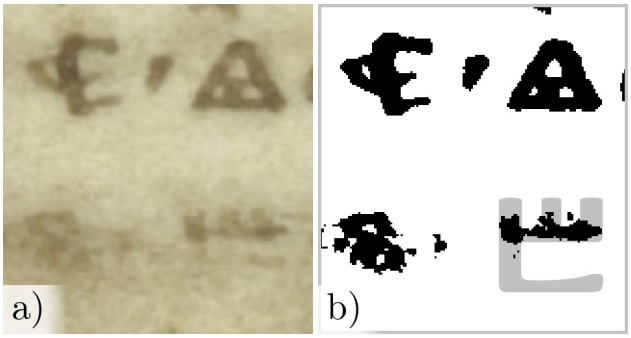


**Figure 1. A detail of a glagolitic manuscript page with degraded characters.**

The system presented in this paper is based on local features. Thus, the preprocessing step with the image binarization can be completely rejected which allows for a recognition of characters which are highly degraded. Having computed the local features for a whole manuscript page, they are classified using a Support Vector Machine (SVM). Subsequently, the characters are located by clustering the interest points. A voting scheme of the classified local descriptors finally recognizes the characters.

The detection of interest points is a crucial task, since the results of the subsequent feature matching or classification is directly related to its performance. Hence, if an interest point detector is chosen which has a low repeatability against certain geometrical distortions (e.g. scale change) that are present in the investigated images, then the feature matching performs poorly. As a consequence, interest

*This work was supported by the Austrian Science Fund under grant P19608-G12

points with no corresponding partner in the other image cannot be matched at all, since the same interest points need to be selected in both images in order to match them. Due to the previously mentioned importance of the interest point detection, it is a well investigated but still active research topic (see [12, 10, 6, 1]).

The principle of local descriptors is to find distinctive image regions such as corners and to analytically describe these regions independent of a predefined set of transformations (e.g. rotations). Local descriptors are able to recognize objects even if parts of the objects are occluded because solely local information is computed to establish the correspondence between the objects.

This paper is organized as follows. The subsequent section covers related work focusing on local feature systems. In Section 3 the proposed system is explained. Afterwards, in Section 4 selected local feature systems are evaluated in order to figure out the most convenient for the investigated dataset. Finally, a short discussion about the future work and a conclusion is given in Section 5.

## 2 Related Work

It is not intended to give an exhaustive overview on current manuscript recognition systems in this section, since the presented one has a different architecture. In fact, an overview on local descriptors is given where the systems evaluated in Section 4 are introduced.

**Interest Point Detectors:** Local interest points for stereo image matching tasks were first introduced by Moravec [12] in 1981. Harris and Stephens improved the repeatability of the Moravec detector using the second moment matrix. In order to compensate the lack of scale-invariance, Mikolajczyk et al. [10] combined the Harris corner detector with a Laplacian. Smith and Brady [16] proposed the Smallest Univalue Segment Assimilating Nucleus (SUSAN) which is a fast corner and edge detector based on non-linear filtering. Another method that focuses on real-time corner detection rather than finding corners accurately invariant to a given set of distortions was recently proposed by Rosten and Drummond [13] and is called Features from Accelerated Segment Test (FAST). In contrast to the previously mentioned methods Lowe [6] does not localize features with the Harris function but by computing the Difference-of-Gaussians (DOG) which detects blob-like image regions.

**Local Descriptors:** While image matching with local features was at the beginning solely used in stereo vision tasks, Schmid and Mohr [15] proposed to use feature matching for image retrieval tasks. Currently local descriptors are applied to solve general image processing tasks such as wide baseline stereo vision [7], object recognition [3] and object localization [2].

Lowe [6] proposed the Scale Invariant Feature Transform SIFT for object recognition tasks. Mikolajczyk and Schmid [9] extended the SIFT approach in order to gain more robustness and distinctiveness. To achieve this, they use a log-polar location grid instead of a Cartesian grid to compute the so-called Gradient Location Orientation Histogram GLOH. Ke and Sukthankar [5] improved the SIFT descriptor. They compute the PCA in order to reduce the dimensionality of the feature vector. Bay et al. [1] designed a new descriptor called Speeded Up Robust Features (SURF) for on-line application focusing on computational speed.

## 3 Methodology

The recognition system presented consists of two basic steps. First the local descriptors are computed at locations found by the interest point detector. Afterwards, they are classified with a SVM. In the second step a clustering is applied and the characters are recognized with a voting scheme based on the preliminary descriptor classification.

**Interest Point Detector:** The DOG detector is used for the localization of image regions where local descriptors are computed. It was chosen by reason of the consecutively given advantages which were gathered by studies of Mikolajczyk [11], Lowe [6] and comparisons of interest point detectors on the investigated dataset (see Section 4.1).

The DOG detects blobs in a scale-space which allows for a scale invariant feature extraction that is needed for a character recognition of mutual manuscripts. Compared to the LOG, the DOG is faster to compute but produces similar results, as it is a close approximation to the LOG. For viewpoint changes below $50°$ the DOG has a higher repeatability than Harris based interest point detectors. Additionally, the DOG produces more interest points than comparable methods such as MSER which reduces the human effort of training the SVM.

**Local Descriptor:** For each interest point, detected by the DOG, a descriptor is computed which considers the structure of the neighborhood of a given interest point. The size of the considered neighborhood depends on the scale factor $\sigma$ which was determined by the scale selection. The aim of a local descriptor is to maximize its distinctiveness, while at the same time maximizing its robustness against a certain set of image distortions. Obviously, the distinctiveness of a descriptor decreases, when increasing its robustness against image transformations.

A comprehensive test of state of the art local descriptors was performed on the investigated dataset in order to choose the one with the best performance. According to the literature [11, 6] and to the evaluation of local descriptors which is further explained in Section 4.2 SIFT was chosen.

**Classification:** As mentioned in the introduction, each local descriptor is classified in our approach. Therefore a

SVM is constructed which has one RBF kernel per character class. In order to train the SVM, 20 different sample images per character are extracted of a given codex. Afterwards, each local descriptor of a manuscript image is classified by means of one-against-all tests with the respective RBF kernels. This results in a look-up table containing probabilities for the local descriptor of belonging to a class.

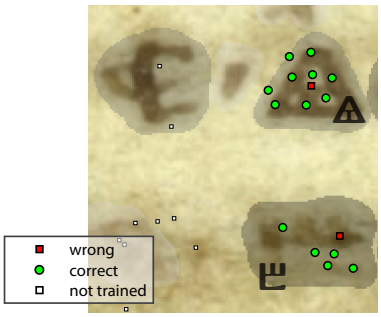


**Figure 2. Classification result of local descriptors**

The classification results are presented in Figure 2. Gray blobs illustrate manually tagged ground truth data. Correctly classified interest points are illustrated by circular markers, where rectangles indicate falsely classified features. Small white rectangles illustrate features outside the evaluated domain.

**Clustering:** In general characters are localized by means of a binarization method. However, when degraded manuscripts are regarded, the localization based on gray-values has two major drawbacks. If the binarization is robust against noise, washed out characters are not detected too. On the other hand, background clutter is mistaken with characters if a sensible binarization method is applied. That is why, the character localization is based on the distribution of detected interest points in the proposed system. Therefore, simply a $k$Means clustering is applied on the interest points' coordinates. The number of clusters $k$ is estimated by means of the interest points' distribution in scale. This approach is motivated by two observations when manuscript images are regarded:

- A character is covered by one feature having a specific scale.
- The interest points' scale distribution approximates a Poisson distribution with a small $\lambda$.

These observations allow for the extraction of approximately one interest point per character which are used to initialize the $k$Means Clustering. Preliminary tests of this approach reveal promising results. Nevertheless, an exhaustive evaluation still needs to be performed.

Subsequently a voting scheme is applied so that the character class of a given cluster is determined.

## 4 Results

The robustness of different local feature methods is evaluated against two types of image transformations (scale and rotation) which need to be considered when recognizing characters (see Figure 3). The robustness is evaluated with four test panels which are synthetically distorted according to the defined image transformations. The performance of each interest point detector is computed with its classification rate which is evaluated using manually tagged ground truth data. Hence, 84 characters are used as validation set in order to compute the classification rate.

### 4.1 Interest Point Detectors

Four state of the art interest point detectors (namely: DOG, FAST, MSER, SUSAN) are trained on 10 character classes of the given dataset in order to evaluate their performance.

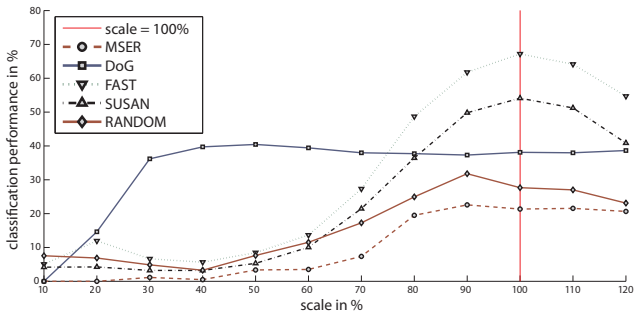


**Figure 4. Comparison of interest point detectors with varying image size ($10\% - 120\%$).**

**Scale:** Each test panel was resampled 12 times with a step size being $10\%$ of the original image size. It can be seen that FAST performs best arround $100\%$ of the original image size. Since all interest point detectors are not computed in a scale invariant manner except for the DOG, it outperforms all other methods. The DOG has a constant performance if the image size is more than $30\%$ of the original image size. Eventhough the radius of the FAST detector could be changed, it could never be used to extract features in a scale invariant manner.

**Rotation:** The invariance against rotation of the interest point detectors was tested by rotating each test panel from $0° - 180°$. The step size was chosen to be $20°$ so that image degradations caused by interpolations are minimized. Table 1 shows the classification performance of each interest point detector tested with increasing rotation angles.

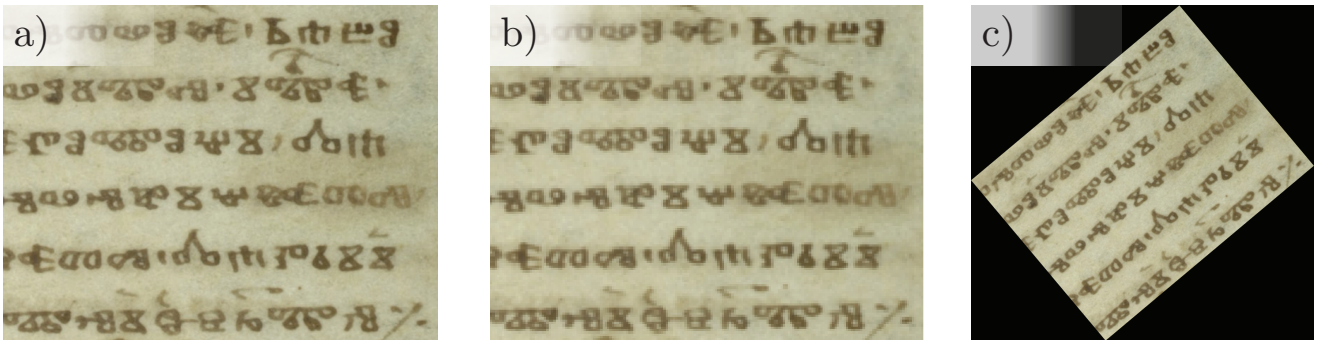


**Figure 3. Synthetic transformations. Test panel (a), $30\%$ scale (b) and $40^\circ$ rotation (c).**

It is clear that the FAST detector closely followed by the SUSAN detector outperform the other interest point detectors. Nevertheless, the performance of FAST decreases with increasing angles ($67.2\%$ at $0^\circ$ and $51.5\%$ at $160^\circ$). The mean performance which is $\emptyset = 36.3\%$ of the DOG is weaker than those of FAST and SUSAN. This can be traced back to the fact that the DOG is computed with a scale-space where features of a coarse scale level are mistaken with features of a fine scale level. A more important fact than the classification rate is the stability of the DOG compared to the other interest point detectors. MSER has a weak performance since it detects fewer keypoints on the characters which results in a worse training of the classifier.

| DETECTOR | # KEYPOINTS | MEAN | STD ($\sigma$) |
|---|---|---|---|
| MSER | 124 | $19.8\%$ | $4.25\%$ |
| DOG | 289 | $36.3\%$ | $\mathbf{1.09}\%$ |
| FAST | 249 | $59.6\%$ | $5.19\%$ |
| SUSAN | 200 | $56.2\%$ | $2.93\%$ |
| RANDOM | 216 | $29.1\%$ | $2.47\%$ |

**Table 1. Number of keypoints per test panel, mean and standard deviation of the performance with respect to rotation.**

## 4.2 Local Descriptors

Similar to the interest point detectors, the performance of five state of the art local descriptors (namely: SIFT, SURF, GLOH, PCA-SIFT and gradient moments) is evaluated on the investigated dataset. This evaluation has the same test setup as the comparison of interest point detectors described in Section 4.1. In order to demonstrate the effect of the feature vector on the classification performance, the interest points were pre-computed using the DOG detector. Since training and testing for all local descriptors is done with the same keypoints, the varying results can be traced back to the weaknesses and strengths of the different local descriptors. Indeed, the classifier could possibly influence the results. In order to minimize this effect the classifier's parameters ($\gamma$, $C$) are estimated individually by means of a 3 fold cross validation.

**Scale:** In general all descriptors have a similar robustness against scale changes. This can be attributed to the fact that the robustness against scale mainly depends on the scale selection scheme which is implemented in the interest point detector algorithm.

SIFT has the highest classification rate which is $\overline{x} = 35.6\%$. The 128 dimensional PCA-SIFT descriptor performs similarly. In contrast, the 36 dimensional PCA-SIFT has a worse performance ($\overline{x} = 20.4\%$) which is not significantly different to the second low-dimensional local descriptor (gradient moments). GLOH adapts slower to the scale changes and reaches its mean performance at $50\%$ of the original scale regarding the other descriptors which adapt at $30\%$. The worst results are obtained by SURF on the investigated dataset. This can be traced back to the fact that the descriptor has a high dependence to the proposed Fast-Hessian detector [1] (see Section 4.3).

**Rotation:** All evaluated descriptors are invariant against rotation (maximum standard deviation: $\sigma = 1.01\%$). The ranking of the mean classification performance is headed by SIFT ($\overline{x} = 38.8\%$) and PCA-SIFT 128 ($\overline{x} = 33.7\%$). In the center span the low-dimensional descriptors gradient moments ($\overline{x} = 23.2\%$) and PCA-SIFT 36 ($\overline{x} = 22.6\%$) are located. Contrary to the expectations GLOH performs poorly within the proposed system and in combination with the DOG having a mean classification performance of ($\overline{x} = 15.0\%$). But the worst results are achieved with SURF ($\overline{x} = 3.3\%$) in combination with the DOG.

## 4.3 Local Descriptor Systems

The previous evaluation was chosen to show exactly the characteristics of different local descriptors if embedded in the proposed system. Due to the strong dependence of some

descriptors (e.g. SURF) to the proposed interest point detectors, an additional evaluation was done, where the whole systems are tested on glagolitic manuscript images.

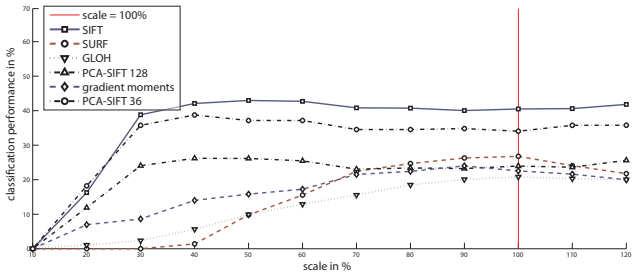


**Figure 5. Comparison of different local descriptor systems with varying image size.**

**Scale:** Even though, the local descriptors are computed with their particular interest point detector, SIFT still performs best on this dataset. GLOH and gradient moments have – in combination with their detector – even a lower scale adaption (at $\approx 70\%$) compared to the previous tests which were carried out using the DoG. In return, GLOH performs better at scales nearby the trained scale (max: $20.82\%$ compared to max: $15.05\%$ in the previous test). As mentioned before, SURF performs significantly better (up to $22.07\%$) in combination with the proposed Fast-Hessian detector. Due to the approximations made (e.g. integral image) the Fast-Hessian detector is not scale invariant, but robust against scale changes.

**Rotation:** The Fast-Hessian, which was used for the computation of SURF, appears to be solely invariant for orthogonal angles. Therefore, the classification performance decreases significantly when applying image rotations with non-orthogonal angles.

## 5 Conclusion & Future Work

The performance tests introduced on glagolitic manuscript images showed, that SIFT performs best for the given task and is robust against common image transformations. This can be attributed to the fact that SIFT is a high-dimensional and therefore highly distinctive descriptor.

The proposed system can recognize degraded handwritten characters as it is not dependent to a successful binarization of a document image. Owing to local descriptors, the system can handle partially visible and connected characters. It can be easily adapted to other alphabets and languages as no dictionary is incorporated for a performance improvement.

However, preliminary tests showed that the character localization – currently performed by clustering – still needs to be improved. A verification process could be introduced after the voting. Additionally generative models will be tested in order to improve the character localization. Finally, exhaustive tests and comparisons with current state of the art handwriting recognition systems will be performed.

## References


[1] H. Bay, T. Tuytelaars, and L. J. V. Gool. SURF: Speeded Up Robust Features. In *Proceedings of the European Conference on Computer Vision*, pages 404–417, 2006.

[2] P. Carbonetto, G. Dorkó, C. Schmid, H. Kück, and N. de Freitas. A Semi-supervised Learning Approach to Object Recognition with Spatial Integration of Local Features and Segmentation Cues. In *Toward Category-Level Object Recognition*, pages 277–300, 2006.

[3] R. Fergus, P. Perona, and A. Zisserman. Object Class Recognition by Unsupervised Scale-Invariant Learning. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 264–271, 2003.

[4] B. Gatos, I. Pratikakis, and S. J. Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39(3):317–327, 2006.

[5] Y. Ke and R. Sukthankar. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 506–513, 2004.

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

[7] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image Vision Comput.*, 22(10):761–767, 2004.

[8] H. Miklas and et al. St. Catherine' s Monastery on Mount Sinai and the Balkan-Slavic Manuscript-Tradition. In *Slovo: Towards a Digital Library of South Slavic Manuscripts.*, pages 13–36, Sofia, Bulgaria, 2008.

[9] K. Mikolajczyk. *Detection of local features invariant to affine transformations*. PhD thesis, Institut National Polytechnique de Grenoble, France, 2002.

[10] K. Mikolajczyk and C. Schmid. Indexing Based on Scale Invariant Interest Points. In *Proceedings of the International Conference on Computer Vision*, pages 525–531, 2001.

[11] K. Mikolajczyk and C. Schmid. A Performance Evaluation of Local Descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630, 2005.

[12] H. P. Moravec. Rover Visual Obstacle Avoidance. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, pages 785–790, 1981.

[13] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *ECCV (1)*, pages 430–443, 2006.

[14] J. J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.

[15] C. Schmid and R. Mohr. Local Grayvalue Invariants for Image Retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(5):530–535, 1997.

[16] S. M. Smith and J. M. Brady. Susan - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997.

[17] A. Vinciarelli. A survey on off-line cursive word recognition. *Pattern Recognition*, 35(7):1433–1446, 2002.