# Efficient Keypoint Reduction for Document Image Matching

Thomas Konidaris[1], Volker Märgner[1,2], Hussein Adnan Mohammed[1] and H. Siegfried Stiehl[1,3]

[1]*Centre for the Study of Manuscript Cultures, Universität Hamburg, Hamburg, Germany*

[2]*Technische Universität Braunschweig, Braunschweig, Germany*

[3]*Department of Informatics, Universität Hamburg, Hamburg, Germany*

Keywords: Document Analysis, Word Spotting, SIFT Features, Image Matching.

Abstract: In this paper we propose a method for eliminating SIFT keypoints in document images. The proposed method is applied as a first step towards word spotting. One key issue when using SIFT keypoints in document images is that a large number of keypoints can be found in non-textual regions. It would be ideal if we could eliminate as much as irrelevant keypoints as possible in order to speed-up processing. This is accomplished by altering the original matching process of SIFT descriptors using an iterative process that enables the detection of keypoints that belong to multiple correct instances throughout the document image, which is an issue that the original SIFT algorithm cannot tackle in a satisfactory way. The proposed method manages a reduction over 99% of the extracted keypoints with satisfactory performance.

## 1 INTRODUCTION

Document image analysis has gained a lot of attention due to the increased need of exploitation of the several manuscript collections found worldwide. Furthermore, the technological advances in terms of image acquisition and digitization have given access to a vast amount of digital material. However, processing this information manually requires a significant amount of time as well as effort in order to complete the underlying tasks. Furthermore, it is not uncommon, due to this load, some of the tasks to be thought of as non-realistic to be processed in a manual way due to the time they require to complete. This resulted to the development of computer vision methods that can assist in a number of tasks and also, importantly, allow the batch processing of large amount of information in much less time than the manual non-computer based way. Tasks like, writer/scribe identification, manuscript indexing through word spotting and layout analysis are among the tasks that can be solved through the use of computer vision methods and have become active research areas on the field of computer vision and more specifically, on the area of document image analysis.

As mentioned above, one of the tasks that needs to be tackled is that of document indexing. Although, in contemporary documents Optical Character Recognition (OCR) methods have been dominant, this is not the case when we refer to historical document images. Various factors can severely affect the performance of OCR systems rendering the whole process as extremely challenging. Although there are attempts to use OCR in historical documents(Jenckel et al., 2016), there is still a lot of ground to cover until we claim that we have approached the problem in a satisfactory way. An alternative to OCR for historical document indexing is word spotting. Word spotting refers the the detection of the desired information directly on the document images without any OCR.

One of the most common ways to perform word spotting is using feature-based methods where from the query keywords and the target document images a, usually large, number of features is extracted. Some of these features belong to the information that we desire to detect while the rest, which are the majority, do not belong to the correct instances that we want to spot. A very popular keypoint detection and descriptor algorithm is Scale Invariant Feature Transform (SIFT)(Lowe, 2004). Usually, the number of SIFT keypoints that are extracted from document images is large and therefore, in this paper we present a method for eliminating SIFT keypoints from document images that do not belong to the areas that we want to detect. The application of the proposed method is presented as a first step of a word spotting

Konidaris, T., Märgner, V., Mohammed, H. and Stiehl, H.
Efficient Keypoint Reduction for Document Image Matching.
DOI: 10.5220/0007412006640670
In *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2019)*, pages 664-670
ISBN: 978-989-758-351-3

process.
The rest of the paper is organized as follows: In Section 2 we present related work and in Section 3 we give a detailed description of the proposed method. In Section 4 experimental results are shown and in Section 5 conclusions are drawn.

## 2 RELATED WORK

In this section we present the different works on the field of word spotting and in particular Query-by-Example(QBE) word spottting. The literature is divided into two main categories, namely, segmentation-based and segmentation-free. The former employs the segmentation of the document images into lines (Kolcz et al., 2000)(Marcolino et al., 2000)(Mondal et al., 2016), words(Rath and Manmatha, 2003) or characters (Kim et al., 2005), while the latter does not apply any kind of segmentation on the document images (Leydier et al., 2009)(Rusinol et al., 2011)(Konidaris et al., 2016)(Mhiri et al., 2018). Deep learning is nowadays widely used and word spotting has its share of attention. Various methods have been proposed that use CNNs and other deep representations(Barakat et al., 2018)(Sudholt and Fink, 2016)(Zhong et al., 2016)(Krishnan et al., 2018).
The proposed method is a feature-based method for word spotting and uses SIFT(Lowe, 2004) keypoints and descriptors. There are various methods that are also feature-based and try to tackle the task of word spotting. In (Zagoris et al., 2017) word spotting is performed using document oriented local features. HOG features are used in the methods proposed in (Bolelli et al., 2017)(Thontadari and Prabhakar, 2016). Bag-of-visual-words (BoVW) for word spotting is employed in (Aldavert and Rusiñol, 2018)(Shekhar and Jawahar, 2012)(Rothacker and Fink, 2015).
Here we will present the various works on word spotting that use SIFT features. Concerning SIFT keypoint reduction, the work presented in (Fujiwara et al., 2013) tries to tackle this specific task. However, the authors try to reduce the number of keypoints on an image assuming that if a number of keypoints have close similarity with a keypoint on the same image, than those are removed since they constitute a repeated pattern. Concering document images, in (Aldavert et al., 2015) word spotting in handwritten documents is presented. The authors use SIFT features with a Bag-of-Visual-Words (BoVW) representation. The method applies segmentation of the documents on word level. Another method that used SIFT and BoVW is proposed in (Rusinol et al., 2011). This work follows a segmentation-free approach although they apply a grid-based segmentation of the documents in order to extract the desired features. Yalniz and Manmatha(Yalniz and Manmatha, 2012) present a word spotting method based on SIFT descriptors applied to FAST(Rosten and Drummond, 2006) keypoints. The extracted features are further quantized using K-Means and the matching is performed using the Longest Common Subsequence (LCS) method. A Bag-of-Features (BoF) representation is presented in (Rothacker et al., 2013). The method uses SIFT descriptors to feed an HMM in order to apply word spotting in handwritten document following a segmentation-free approach. Ghosh and Valveny (Ghosh and Valveny, 2015) present a segmentation-free word spotting method based on word attributes. Query words are encoded using Fischer Vector representation and are used together with pyramidal histogram of characters labels (PHOC) to learn SVM-based attribute models. For the matching process a sliding window is applied. The word attributes used as the representation involve the calculation of SIFT descriptors. Another approach that uses PHOC, Fischer Vectors and densely extracted SIFT descriptors is proposed in (Almazán et al., 2014). The SIFT features are extracted by a variable size patch and their dimension is reduced using PCA. In the work of Sfikas et. al.(Sfikas et al., 2015) a Gaussian mixture model (GMM) is trained using SIFT descriptors. Fisher vectors are then calculated for each image as a function of their SIFT description and the gradients of the GMM with respect to its parameters. This results to a fixed-length, highly discriminative representation, that can be seen as an augmented bag of visual words description that encodes higher order statistics.

## 3 PROPOSED METHOD

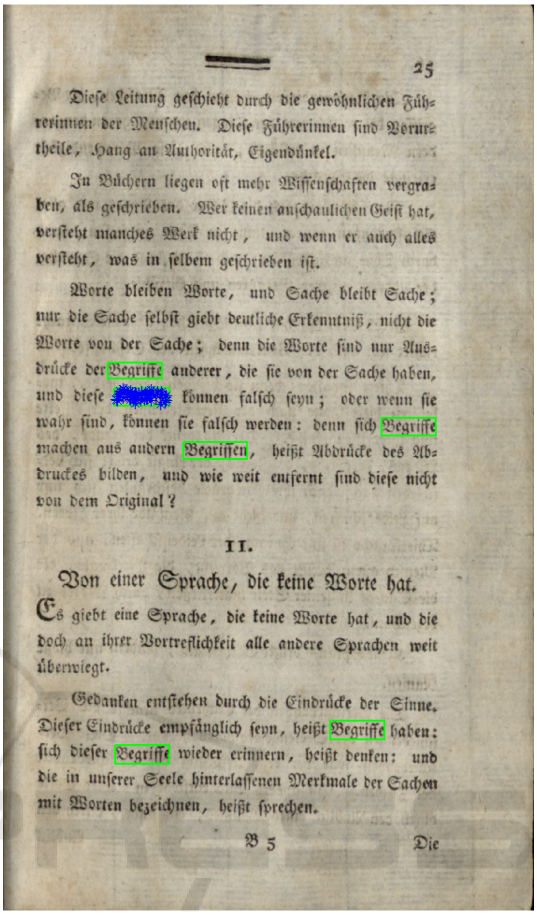
In this paper we propose an alternative matching scheme for matching SIFT descriptors as applied as a first step towards word spotting. The task is to find relevant keypoints on a target document image when keypoints of a query keyword image are matched against it. Ideally, the detected keypoints on the target document image will be part of the correct instance of the query keyword. The propose method aims to reduce the amount of keypoints detected in the images but on the same time keeping the keypoints that are most relevant to the keypoints of the query keyword image. The method uses an iterative process to reduce the keypoints following a different approach than the one applied by the SIFT algorithm as proposed by

Lowe(Lowe, 2004). Instead of getting only a single keypoint matched in the target image for every keypoint in the query keyword image as it occurs when using the original SIFT, we try to get all the strong keypoints in the target image for every keypoint in the query keyword image using the same matching criterion as proposed in the original SIFT algorithm.
In the original paper, every descriptor in the query image is compared with all the descriptors in the target image. If the ratio of the two closest descriptors for every query descriptor is less than a certain threshold $t$ than the keypoint that correspond to the closest descriptor is kept as a valid keypoint. The distance ratio threshold is shown in Equation 1.

$$\frac{d_1}{d_2} < t \tag{1}$$

where $d_1$ and $d_2$ are the distances of the two closest keypoint descriptors of the document image to a query keyword keypoint descriptor and threshold $t = 0.8$. The original paper claims that this matching scheme with $t = 0.8$ eliminates 90% of the false keypoint matches while discarding 5% of the correct keypoint matches. Although this may be the case where there is a single object located in the images, when it comes to document images there is a very important issue that needs to be taken under consideration. In document images the desired information that we need to detect may have multiple correct instances in the same page. This is a very common scenario in word spotting methods. Consider the following example as illustrated in Figure 1. We want to detect the keyword "Begriffe" on a document image. The case follows a segmentation-free word spotting approach. In this particular example, the query keyword image is one of the instances found on the page. The original SIFT matching algorithm manages to successfully detect the word on the document image, since it is the query keyword used for matching, but fails to do so with the rest of the correct instances. The bounding boxes on the image indicate the correct instances of the query keyword as found in the ground truth.

In the proposed method, concerning historical document images, we follow a different approach. The idea lies on the fact that it is very common to have multiple instances of the information we want to locate. For that reason we want to be able to use the matching of the descriptors as proposed in the original SIFT algorithm but also enable the process to find all the relevant keypoints that belong to other correct instances of the query keyword image. Based on that thought, we introduce an iterative process that enables us to succeed in the aforementioned task. For every query keypoint we perform the original SIFT as defined in Equation 1. If there are keypoints that meet the threshold criterion, these are kept as valid for further processing and the process is repeated. At the next iteration, we use the same query keypoint but this time, without taking into consideration the already detected valid keypoints. This will enable the algorithm to detect other strong keypoints for the same query keypoint. The assumption is that there might be other keypoints on the document image that satisfy the threshold criterion for the same query keypoint that belong to other correct instances of the the query keyword image. This iterative process overcomes the limitation of the original SIFT algorithm to detect multiple instances of the correct information on a document image, especially when a document image contains the query keyword image and other correct instances as shown in Figure 2. Algorithm 1 shows the various steps of the proposed method.



25

Dieſe Leitung geſchieht durch die gewöhnlichen Führerinnen der Menſchen. Dieſe Führerinnen ſind Vorurtheile, Hang an Authorität, Eigendünkel.

In Büchern liegen oft mehr Wiſſenſchaften vergraben, als geſchrieben. Wer keinen anſchaulichen Geiſt hat, verſteht manches Werk nicht, und wenn er auch alles verſteht, was in ſelbem geſchrieben iſt.

Worte bleiben Worte, und Sache bleibt Sache; nur die Sache ſelbſt giebt deutliche Erkenntniß, nicht die Worte von der Sache; denn die Worte ſind nur Ausdrücke der Begriffe anderer, die ſie von der Sache haben, und dieſe können falſch ſeyn; oder wenn ſie wahr ſind, können ſie falſch werden: denn ſich Begriffe machen aus andern Begriffen, heißt Abdrücke des Abdruckes bilden, und wie weit entfernt ſind dieſe nicht von dem Original?

II.

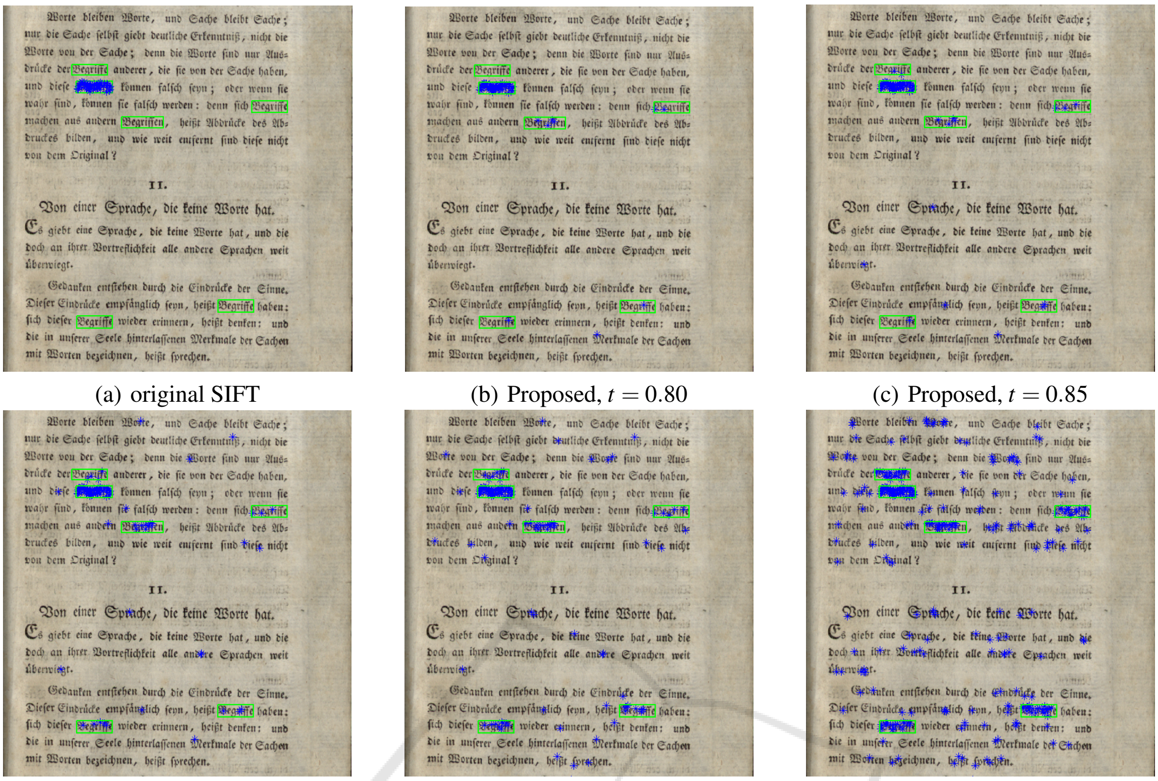
Von einer Sprache, die keine Worte hat.

Es giebt eine Sprache, die keine Worte hat, und die doch an ihrer Vortreflichkeit alle andere Sprachen weit überwiegt.

Gedanken entſtehen durch die Eindrücke der Sinne. Dieſer Eindrücke empfänglich ſeyn, heißt Begriffe haben: ſich dieſer Begriffe wieder erinnern, heißt denken: und die in unſerer Seele hinterlaſſenen Merkmale der Sachen mit Worten bezeichnen, heißt ſprechen.

B 5 Die

Figure 1: Matching the query keyword "Begriffe" with the document image. The keyword is taken from the document image. SIFT algorithm manages to detect only this word omitting the other correct instances found in the document image.



(a) original SIFT (b) Proposed, $t = 0.80$ (c) Proposed, $t = 0.85$

(d) Proposed, $t = 0.90$ (e) Proposed, $t = 0.95$ (f) Proposed, $t = 0.99$

Figure 2: The proposed method as compared with the original matching of the SIFT algorithm. The query word is taken directly from the illustrated page. (a) is the result of the original SIFT matching where the algorithm correctly spots the query word image but fails to spot the additional correct instances. (b)-(e) the results of the proposed method with $t = \{0.80, 0.85, 0.90, 0.95, 0.99\}$ respectively. The green bounding boxes indicate correct instances of the word as found in the ground truth. For clarity only a portion of the document image is shown.

Algorithm 1: Iterative SIFT Matching.

```
Calculate descriptors for query and document
  image
init valid_pts
∀ i ∈ query_descr
while matched_pts do
    dist = ||query_descr_i, image_descr||_2
    sort(dist)
    if $\frac{d_1}{d_2} \leq t$
        $idx = argmin(d_1)$
        matched_pts.append($image\_pts(idx)$)
        remove $image\_descr(idx)$
    else
        matched_pts = false
    end
```

## 4 EXPERIMENTAL RESULTS

The experiments involve checking whether the reduction of the keypoints is efficient in the sense that the remaining keypoints are located inside the ground truth bounding boxes. For the purpose of the experiments, the setup proposed in (Konidaris et al., 2016) is used. There are 100 document images(von Eckartshausen, 1778) and 100 keyword images. We have used five different values for threshold $t$, 0.80, 0.85, 0.90, 0.95 and 0.99, respectively. The idea behind the various threshold values is that if for a query keypoint the two closest keypoints are valid, this means that their threshold ratio will be high. According to our experiments this seems to hold. The experiments concern the mean average recall ($mAR$), the average number of remaining keypoints on the document images and the ratio of the remaining keypoints found inside the ground truth bounding boxes over the number of the remaining keypoints. The experiments do not evaluate the proposed method as a complete word spotting method. Rather, the idea is to use the proposed method as a first step towards word spotting where the keypoints that remain after the elimination can be further processed yielding the final results.

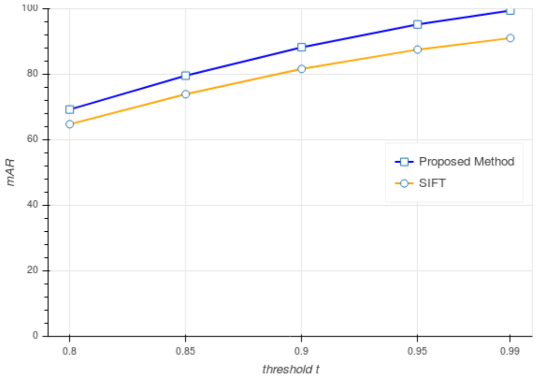


Figure 3: Mean average recall (mAR) for the different values of *t* for all query keywords.

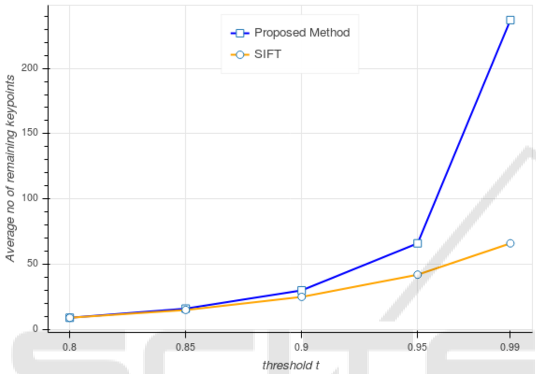


Figure 4: Average number of remaining keypoints per page for the different values of *t*.

The average number of keypoints per page in the document images used for the experiments is 16154. Figure 3 shows the results concerning the *mAR* which corresponds to the number of keypoints being found in the ground truth bounding boxes. For this calculation a true positive requires at least one keypoint to be found inside the ground truth. Furthermore, the ground truth follows an extended format in the sense that not only exact match words are included but also words that include the query keyword in whole as their part. This is primarily because the documents have not undergone any segmentation and the correct instance of a word can be found anywhere on the document images including when it is part of larger words too.

It is clear that for the various values of threshold *t*, the proposed method outperforms the original SIFT matching process. This justifies the assumption that a query keypoint may have other strong keypoint matches on the document images that are detected using the iterative process proposed in this paper.

Figure 4 shows the mean average number of remaining keypoints for all queries in the document images for the different values of *t*. SIFT manages to have less remaining keypoints but this is reasonable since it does not perform any kind of iterations, contrary to the proposed method where for each query keypoint a number of iterations is performed in order to detect all the keypoints that satisfy the matching criterion.

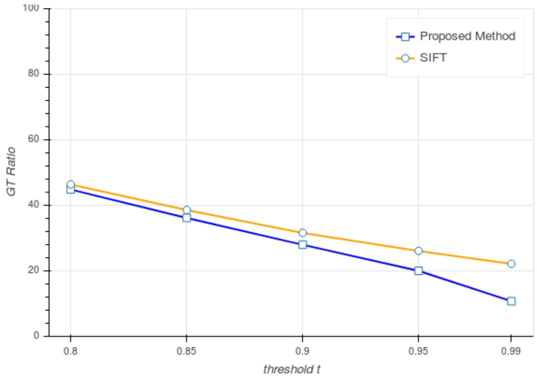


Figure 5: The ratio of the remaining keypoints inside the ground truth bounding boxes over the total number of remaining keypoints.

Table 1: Performance of the proposed method compared to the original SIFT for the different values of the threshold *t*.

| t | $\mathbf{mP_r}$ | GTRatio | mAR |
|---|---|---|---|
| | (Prop./SIFT) | (Prop./SIFT) | (Prop./SIFT) |
| 0.80 | 9 / 9 | 44.86% / 46.41% | 69.26% / 64.77% |
| 0.85 | 16 / 15 | 36.20% / 38.61% | 79.58% / 73.93% |
| 0.90 | 30 / 25 | 28.01% / 31.63% | 88.24% / 81.63% |
| 0.95 | 66 / 42 | 20.04% / 26.13% | 95.20% / 87.54% |
| 0.99 | 237 / 66 | 10.79% / 22.20% | 99.45% / 94.01% |

Figure 5 illustrates the ratio of remaining keypoints found in the ground truth bounding boxes to all the remaining keypoints. The matching scheme of the SIFT algorithm seems to have a better ratio but this can be justified by the less number of remaining keypoints than the proposed method as shown in the previous diagram. As we have already mentioned, in the original SIFT algorithm for every query keyword keypoint we get only a single matching keypoint on the target document image. Table 1 summarizes the experiments performed between the matching scheme proposed in this paper and the original matching scheme of the original algorithm.

where, $mP_r$ is the mean average remaining points, *GTRatio* is the average ground truth to remaining points ratio, and *mAR* is the mean average recall for all the query keywords.

Through the above experiments we provided an insight on how SIFT matching can be altered in order all the relevant keypoints to be extracted for every query keyword keypoint. The selection of the threshold lies solely upon the needs of the underline task. Lower values of $t$ will result in faster processing but lower accuracy, while larger values of $t$ will yield better results but more keypoints which some may not belong to areas of interest. The proposed method shows better performance than the original SIFT matching scheme.

## 5 CONCLUSIONS

In this paper we propose an alternative method for matching SIFT keypoints using their descriptors. The method manages to reduce the amount of keypoints used for further processing on the document images. The proposed method applies an iterative process that manages to eliminate more than 99% of the keypoints while, on the same time, the remaining keypoints are located in the areas of interest. The method in this paper is suggested as a first step for word spotting applications that follow a segmentation-free approach. It allows the reduction of the keypoints significantly, which can lead to less document areas to be searched, thus speeding up the entire process. The proposed method as mentioned throughout this paper is not a complete word spotting method. This is not the idea behind it. Therefore, it could be possible to apply it on other research areas where SIFT is used and there is the need to discard non-relevant keypoints so as to speed-up the entire process. The value of the threshold $t$, can be chosen based on the needs of the underlying task. The various values of $t$ allows finding keypoints with stronger relations between them, thus leading to keypoints that belong to correct word instances, as far as word spotting is concerned, or any other type of information we need to locate on an image.

## ACKNOWLEDGEMENTS

This work has been funded by the German Research Foundation (DFG) within the scope of the Collaborative Research Centre (SFB 950) at the Centre for the Study of Manuscript Cultures (CSMC) at Hamburg University.

## REFERENCES

Aldavert, D. and Rusiñol, M. (2018). Synthetically generated semantic codebook for bag-of-visual-words based word spotting. In *13th IAPR International Workshop on Document Analysis Systems, DAS 2018, Vienna, Austria, April 24-27, 2018*, pages 223–228.

Aldavert, D., Rusiñol, M., Toledo, R., and Lladós, J. (2015). A study of bag-of-visual-words representations for handwritten keyword spotting. *IJDAR*, 18(3):223–234.

Almazán, J., Gordo, A., Fornés, A., and Valveny, E. (2014). Word spotting and recognition with embedded attributes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(12):2552–2566.

Barakat, B. K., Alaasam, R., and El-Sana, J. (2018). Word spotting using convolutional siamese network. In *13th IAPR International Workshop on Document Analysis Systems, DAS 2018, Vienna, Austria, April 24-27, 2018*, pages 229–234.

Bolelli, F., Borghi, G., and Grana, C. (2017). Historical handwritten text images word spotting through sliding window HOG features. In *Image Analysis and Processing - ICIAP 2017 - 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part I*, pages 729–738.

Fujiwara, Y., Okamoto, T., and Kondo, K. (2013). SIFT feature reduction based on feature similarity of repeated patterns. In *International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS 2013, Naha-shi, Japan, November 12-15, 2013*, pages 311–314.

Ghosh, S. K. and Valveny, E. (2015). A sliding window framework for word spotting based on word attributes. In *Pattern Recognition and Image Analysis - 7th Iberian Conference, IbPRIA 2015, Santiago de Compostela, Spain, June 17-19, 2015, Proceedings*, pages 652–661.

Jenckel, M., Bukhari, S. S., and Dengel, A. (2016). anyocr: A sequence learning based OCR system for unlabeled historical documents. In *23rd International Conference on Pattern Recognition, ICPR 2016, Cancún, Mexico, December 4-8, 2016*, pages 4035–4040.

Kim, S., Park, S., Jeong, C., Kim, J., Park, H., and Lee, G. (2005). Keyword spotting on korean document images by matching the keyword image. In *Digital Libraries: Implementing Strategies and Sharing Experiences*, volume 3815, pages 158-166.

Kolcz, A., Alspector, J., Augusteijn, M., Carlson, R., and Popescu, G. V. (2000). A line-oriented approach to word spotting in handwritten documents. *Journal of Pattern Analysis and Applications*, 3(2):153-168.

Konidaris, T., Kesidis, A. L., and Gatos, B. (2016). A segmentation-free word spotting method for historical printed documents. *Pattern Analysis and Applications*, 19(4):963–976.

Krishnan, P., Dutta, K., and Jawahar, C. V. (2018). Word spotting and recognition using deep embedding. In *13th IAPR International Workshop on Document Analysis Systems, DAS 2018, Vienna, Austria, April 24-27, 2018*, pages 1–6.

Leydier, Y., Ouji, A., LeBourgeois, F., and Emptoz, H. (2009). Towards an omnilingual word retrieval system for ancient manuscripts. *Pattern Recognition*, 42(9):2089-2105.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91-110.

Marcolino, A., Ramos, V., Ramalho, M., and Pinto, J. C. (2000). Line and word matching in old documents. In *Fifth IberoAmerican Symposium on Pattern Recognition (SIAPR)*, pages 123-135.

Mhiri, M., Abuelwafa, S., Desrosiers, C., and Cheriet, M. (2018). Hierarchical representation learning using spherical k-means for segmentation-free word spotting. *Pattern Recognition Letters*, 101:52–59.

Mondal, T., Ragot, N., Ramel, J., and Pal, U. (2016). Flexible sequence matching technique: An effective learning-free approach for word spotting. *Pattern Recognition*, 60:596–612.

Rath, T. M. and Manmatha, R. (2003). Features for word spotting in historical manuscripts. In *International Conference of Document Analysis and Recognition*, pages 218-222.

Rosten, E. and Drummond, T. (2006). Machine learning for high-speed corner detection. In *European Conference on Computer Vision*, volume 1, pages 430-443.

Rothacker, L. and Fink, G. A. (2015). Segmentation-free query-by-string word spotting with bag-of-features hmms. In *13th International Conference on Document Analysis and Recognition, ICDAR 2015, Nancy, France, August 23-26, 2015*, pages 661–665.

Rothacker, L., Rusiñol, M., and Fink, G. A. (2013). Bag-of-features hmms for segmentation-free word spotting in handwritten documents. In *2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, August 25-28, 2013*, pages 1305–1309.

Rusinol, M., Aldavert, D., Toledo, R., and Lladós, J. (2011). Browsing heterogeneous document collections by a segmentation-free word spotting method. In *11th International Conference on Document Analysis and Recognition (ICDAR'11)*, pages 63-67, China.

Sfikas, G., Giotis, A. P., Louloudis, G., and Gatos, B. (2015). Using attributes for word spotting and recognition in polytonic greek documents. In *13th International Conference on Document Analysis and Recognition, ICDAR 2015, Nancy, France, August 23-26, 2015*, pages 686–690.

Shekhar, R. and Jawahar, C. V. (2012). Word image retrieval using bag of visual words. In *10th IAPR International Workshop on Document Analysis Systems, DAS 2012, Gold Coast, Queenslands, Australia, March 27-29, 2012*, pages 297–301.

Sudholt, S. and Fink, G. A. (2016). Phocnet: A deep convolutional neural network for word spotting in handwritten documents. In *15th International Conference on Frontiers in Handwriting Recognition, ICFHR 2016, Shenzhen, China, October 23-26, 2016*, pages 277–282.

Thontadari, C. and Prabhakar, C. J. (2016). Scale space co-occurrence HOG features for word spotting in handwritten document images. *IJCVIP*, 6(2):71–86.

von Eckartshausen, C. (1778). *Aufschlüsse zur Magie aus geprüften Erfahrungen über verborgene philosophische Wissenschaften und verdeckte Geheimnisse der Natur*. Bavarian State Library.

Yalniz, I. Z. and Manmatha, R. (2012). An efficient framework for searching text in noisy document images. In *Proceedings of Document Analysis Systems (DAS)*, pages 48-52.

Zagoris, K., Pratikakis, I., and Gatos, B. (2017). Unsupervised word spotting in historical handwritten document images using document-oriented local features. *IEEE Trans. Image Processing*, 26(8):4032–4041.

Zhong, Z., Pan, W., Jin, L., Mouchère, H., and Viard-Gaudin, C. (2016). Spottingnet: Learning the similarity of word images with convolutional neural network for word spotting in handwritten historical documents. In *15th International Conference on Frontiers in Handwriting Recognition, ICFHR 2016, Shenzhen, China, October 23-26, 2016*, pages 295–300.