Projet ANR- Blanc International II -

SIMI 2 - Science informatique et applications

GUWENSHIBIE

Programme Blanc International II 2012

Work Package n°4

Deliverable n°13

Robust Strokes and Features Extraction

State of the art of existing free segmentation feature in computer vision and evaluation on Chinese handwritten documents.

Author: Thomas Konidaris

Delivery date: 30 October 2014

1 Introduction

There is a vast amount of document collections available nowadays and there has been a great effort in order to exploit their invaluable content. The analysis and processing of historical documents and in general historical document collections aims to provide access to the rich information they contain.

The first step that was taken towards the exploitation of information contained in historical documents was the digitization process and the creation of digital libraries. That way the content was made available into digital form. Furthermore, there has been research interest on the potential ways someone could use in order to retrieve the desired information. However, a number of factors makes the processing of such information a challenging as well as a tedious process. Image degradation, typesetting imperfections, digitization errors, unknown fonts and complex layouts are among the factors that can severely affect the performance of the methods that try to exploit their invaluable content.

Optical Character Recognition (OCR) is one of the traditional methods for processing digitized document collections. The aim is to produce computer recognizable encoding, thus providing the chance to analyse and retrieve the information in them. However, the majority of OCR methods perform well when applied in relatively modern and printed documents. Due to the aforementioned factors that are present in historical documents, OCR is not a very reliable way to go. Therefore, the documents need to undergo a pre-processing step which aims to improve their quality, minimize the factors that affect their efficient exploitation and in general bring them in a more suitable form for further processing. These pre-processing steps most often include binarization and quality improvement in order to be able to perform tasks such as layout analysis, document segmentation and others. However, there are cases where the amount of the degradations in the documents is so large that processes like binarization and document segmentation leading to significant information loss thus, failing to produce the desired results. In order to prevent information loss, there is increased research interest in segmentation-free methods. Methods that do not undergo any segmentation process. The document images are treated as a single entity. Furthermore, although there are many methods concerning binarization, the document images are at least grayscale, aiming to use as much information as possible.

2 Chinese Historical Manuscripts

China is among the very few countries that their great history and civilization goes thousands of years back. Their contribution to the world is undoubtedly huge and so is the amount of their historical document collections. The manuscripts are written using ideograms which are symbols that are entirely different than their Latin siblings. The structure of these ideograms vary significantly. There are ideograms with simple structure, other that are more complex, and others that can consist of two or more parts. Figure 1 illustrate a sample of Chinese ideograms with the afore mentioned properties. These properties make the processing of such documents a very challenging task.

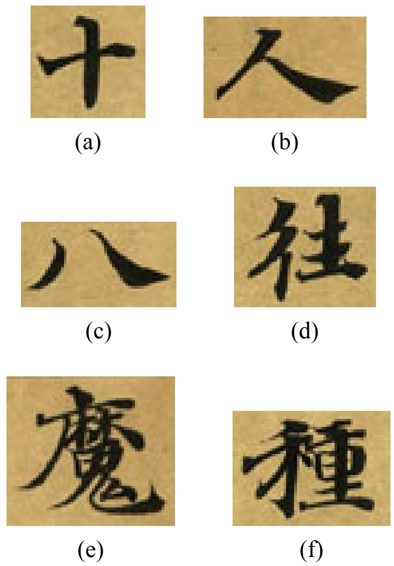


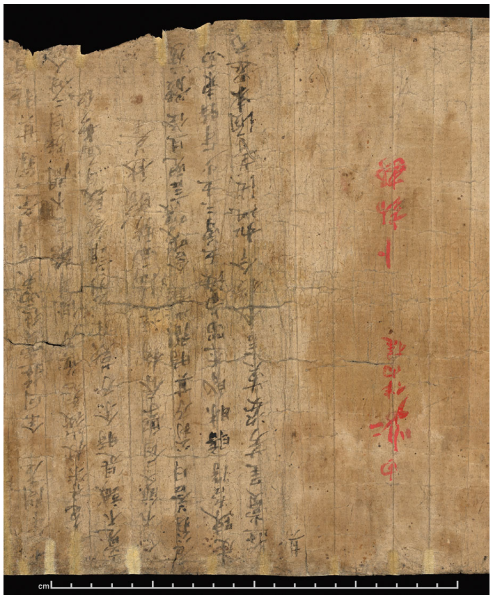
Figure 1: Samples of chinese ideograms.

Furthermore, as we have already stated, these manuscripts can be in bad condition due to ageing and their expose to the elements and due to their complex layout structure. Samples of documents that present some of the aforementioned difficulties are shown in Figure 2.

Having this in mind, we will review some of the most widely used algorithms for feature extraction when applied to historical Chinese manuscripts.

3 Features

One of the crucial parts of document processing is that of the feature extraction. Researchers try to come up with ideas for robust features that will be able to tackle the aforementioned document characteristics. Furthermore, affine invariance may also be the case when developing a feature extraction algorithm.



(a)

(b)

Figure 2: Samples of documents with a) significant degradations, b) complex layout.

3.1 Histogram of Oriented Gradients (HoG)

Features that are based on the Histogram of Oriented Gradients (HoG) were first introduced by Dalal and Triggs for human detection [2]. HoGs have been widely used in human and object detection and tracking [7][14]. Besides, the application of HoGs into the aforementioned areas, they are also used for text detection [9][13].

The image if first divided into a number of cells of predefined size. In each cell, the orientations of the gradient information are calculated and are clustered into bins. The cells are further clustered into blocks and their values are normalised using the entire block information. An example of HoG features that are extracted from a chinese character using a cell size of 5×5 pixels and a block size of 2×2 cells are shown in Figure 3.

3.2 Gabor Features

Gabor features are widely used in a broad spectrum of computer vision application such as texture analysis [11][15], biometrics [3], OCR [6], etc. They consist of a real and an imaginary part that are defined as follows:

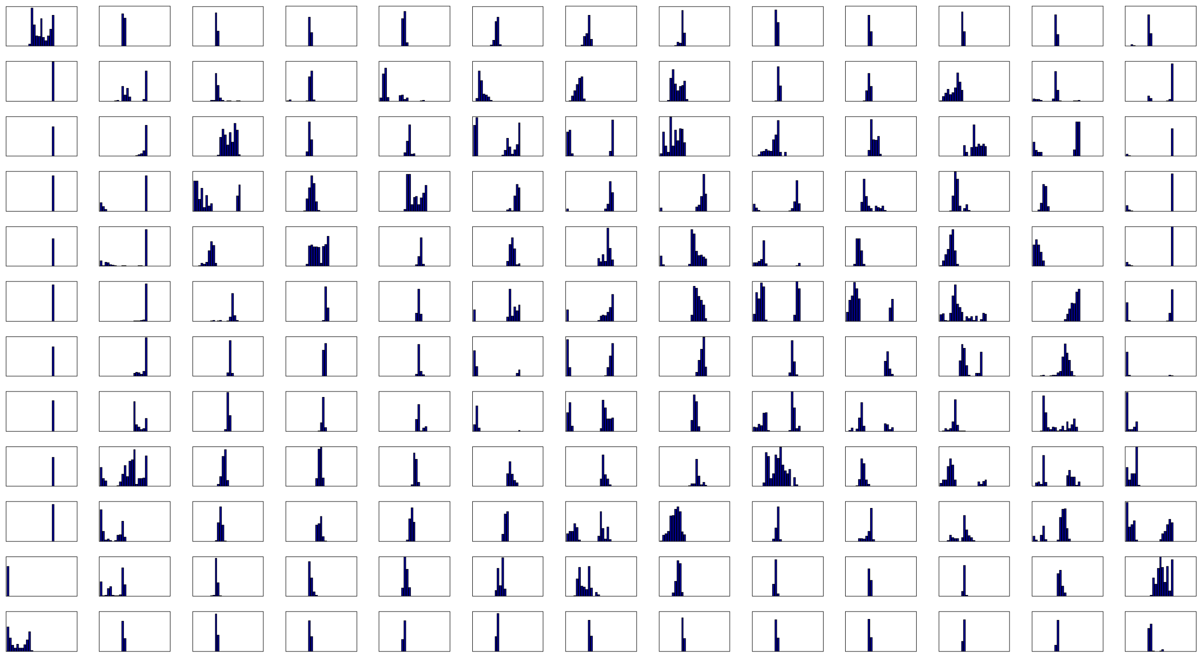
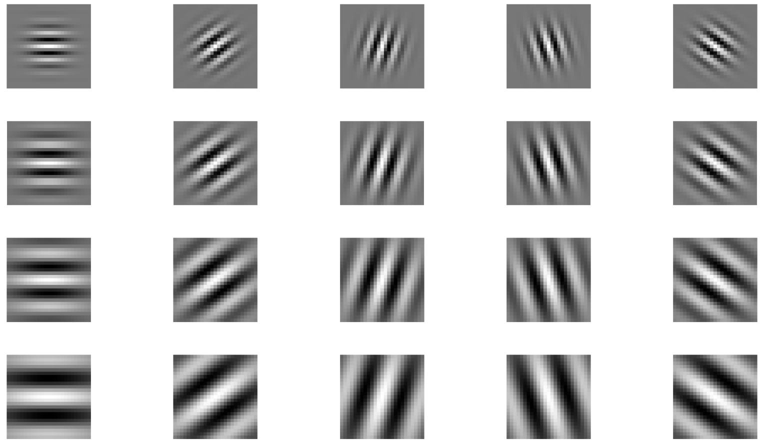
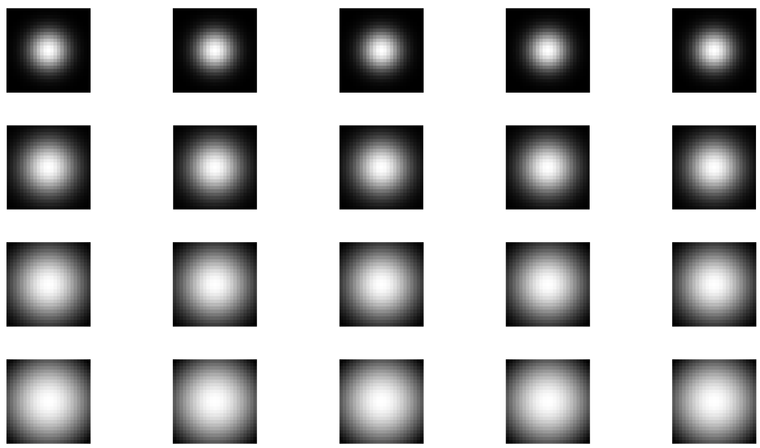


Figure 3: HoG features of a chinese character.

The concept behind Gabor features is to convolve the image with a series of Gabor filters called a filter bank. The filter bank covers multiple scales and rotations creating the Gabor space. The convolution of the filter bank with an image extracts the desired features. Figure 6 illustrates a Gabor filter bank of four scales and five rotations and their corresponding magnitudes.



(a)

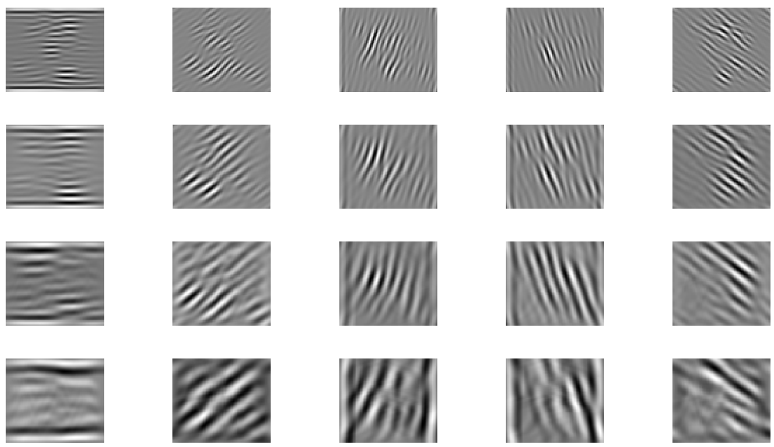


(b)

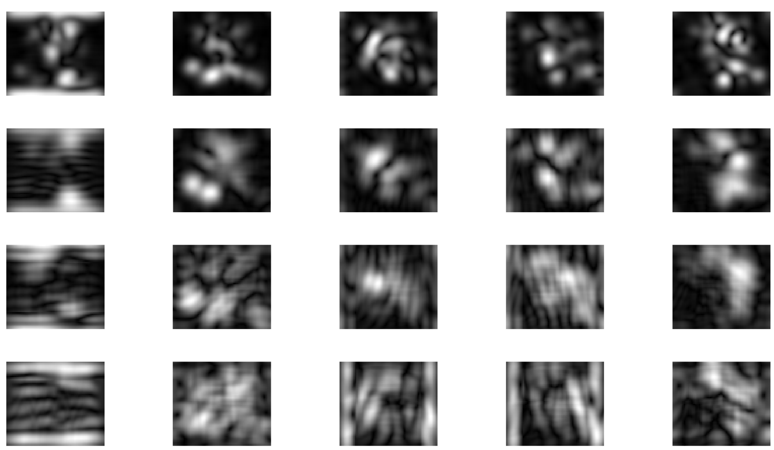
Figure 4: a) A Gabor filter bank of 4 scales and 5 orientations, b) their corresponding magnitudes.

As an example, we have a chinese character extracted from a document image shown in Figure5 and we have applied a Gabor filter bank like the one illustrated above. Figure 6 shows the resulting images in both the real part of the Gabor filter bank and their magnitudes.

Figure 5: A chinese character



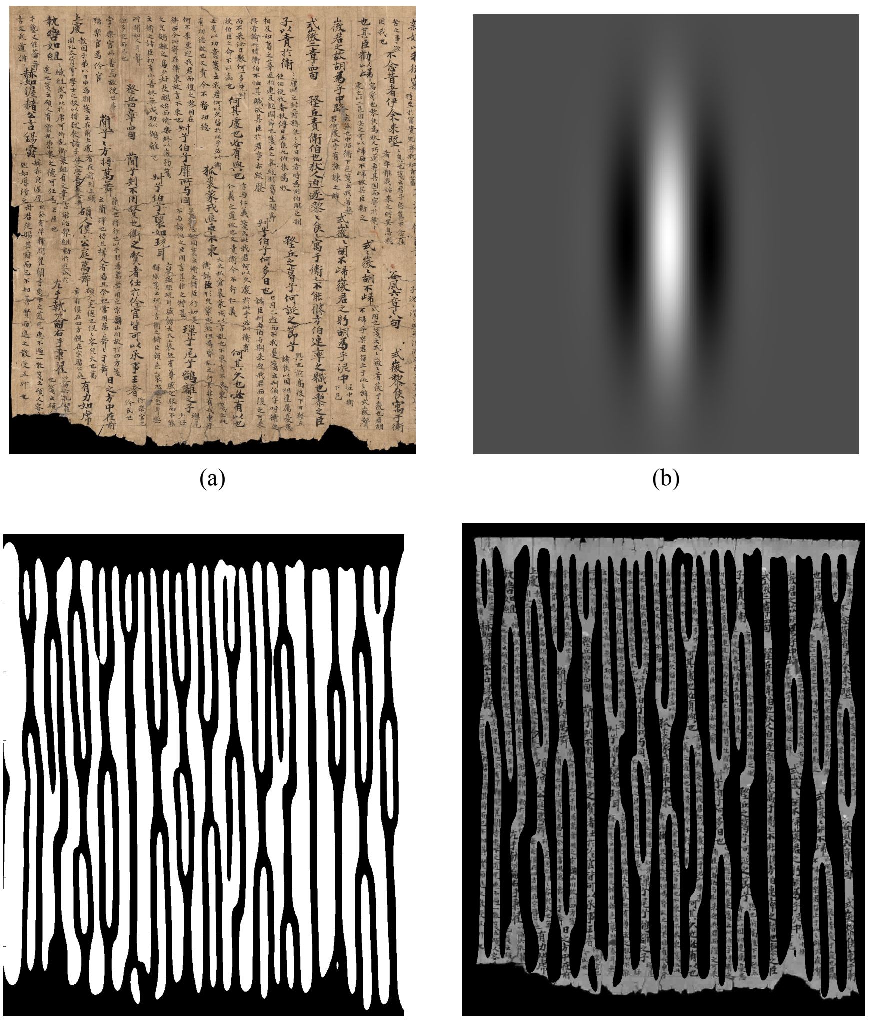
(a)



(b)

Figure 6: a) A Gabor filter bank of 4 scales and 5 orientations, b) their corresponding magnitudes.

Besides the use of Gabor filters as a means of extracting Gabor features, they can be also used for transforming an image for better processing. An example could be the convolution of an image with a Gabor filter in order to create blobs that can be used for text localization. However, the orientation of the filter must match the orientation of the text in order to have the desired results. Figure 7 illustrates the application of a Gabor filter on a document image. Using the resulting image and a rough binirazation we are able to locate the text areas on the image.



(a) (b)

(c) (d)

Figure 7: a) Original image, b) the Gabor filter, c) rough binarization of the convolved image, d) superimposing the binary image on the original image.

3.3 Gaussian Steerable Filters

The concept behind steerable filters can be found in the work presented in [4]. The Gaussian steerable filters consist of a Gaussian kernel that is rotated at a certain angle. These filters are used for feature extraction and edge detection like the normal Gaussian filters. In the area of document analysis someone could expect the orientation of the text to follow the traditional rules, being horizontal or vertical. However, there are cases that we need to process documents with complex layouts that include a mixture of text orientations. Steerable filters are convolved with the image resulting to a certain response for each orientation. This way, we can identify the dominant orientation and apply a filter using that orientation.

3.4 Harris

The Harris corner detection algorithm [5] is a widely used algorithm that manages to detect keypoints that are located on corners. However, the algorithm fails to produce correct results in cases where image information is not very clearly separated from the background. This is the case in the chinese document images that we process. An example can be seen in Figure where there is a stain on the image causing the text to be blended into the background.

In order to determine if a keypoint is a corner or an edge, the algorithm uses the following matrix:

$$M = G \otimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{1}$$

where $G\otimes$ indicates convolution with a Gaussian kernel and I_x, I_y are the partial derivatives of the image I for the x and y directions respectively. The eigenvalues λ_1, λ_2 of matrix M are used for determining the interest points as follows:

- if $\lambda_1 \approx 0$, $\lambda_2 \approx 0$, There is not any point of interest.
- if $\lambda_1 \approx 0$, $\lambda_2 \gg 0$, there exists an edge.
- if $\lambda_1 \gg 0$, $\lambda_2 \gg 0$, there exists a corner.

As it can be seen in Figure 8, the algorithm fails to perform well in circumastance where the image information does not differ much from the background.

Although the original Harris is not scale and rotation invariant, the are versions of the algorithms that can support affine transformations. A representative work is present in [10], where a Harris based corner detection is used that can also perform well under affine transformations.

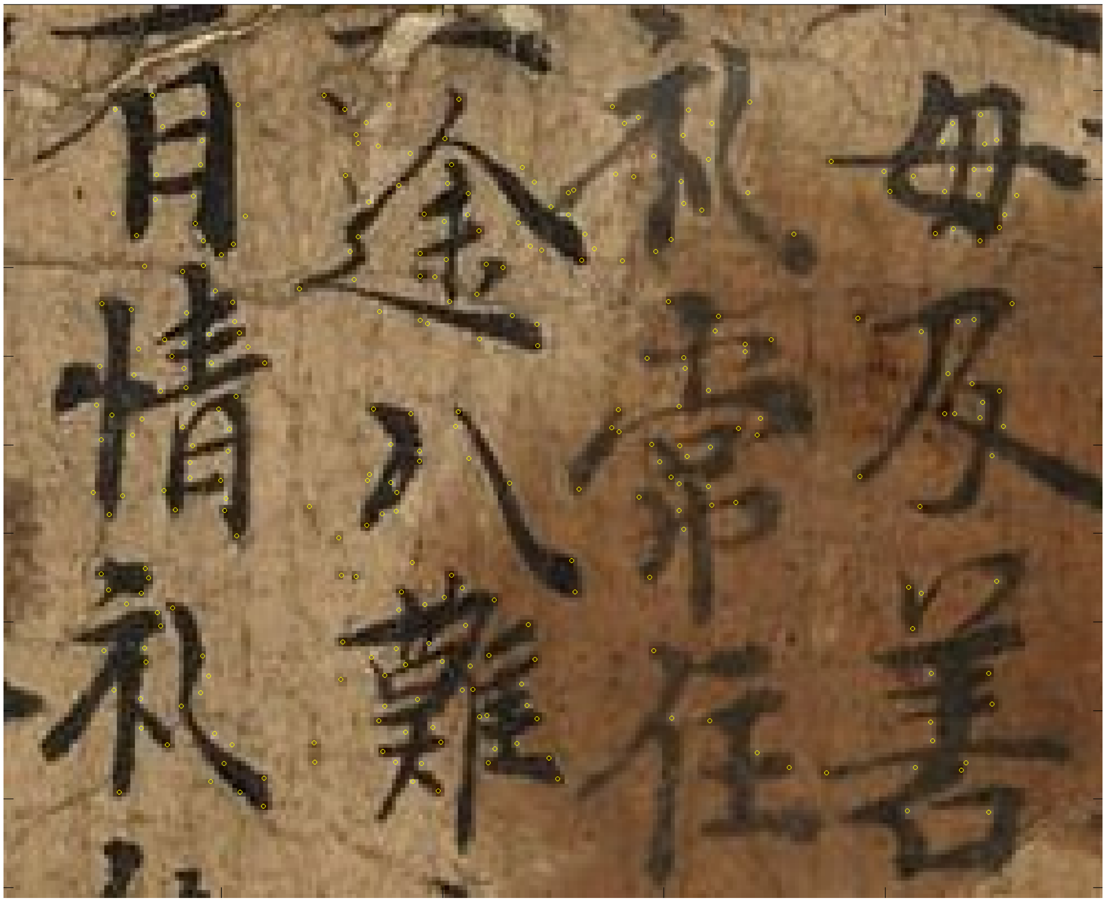


Figure 8: Harris Keypoints.

Although, the original Harris algorithm is not affine invariant, there are proposed methods that extend the functionality of the algorithm so as to be able to cope with scale and rotation changes. A representative method is presented in [10].

3.5 Shi-Tomasi

The Shi-Tomasi algorithm [12] follows the same principle as the Harris corners algorithms with one main difference. The algorithm uses the $min(\lambda_1, \lambda_2)$ in order to find the interest points. The Shi-Tomasi algorithm comes to overcome the weakness of the Harris algorithm in cases where background and image information do not pose significant differences. Figure 9 illustrates the Shi-Tomasi algorithm in a Chinese manuscript image as above. It is clear that the algorithm can perform better in cases where that Harris algorithm fails.

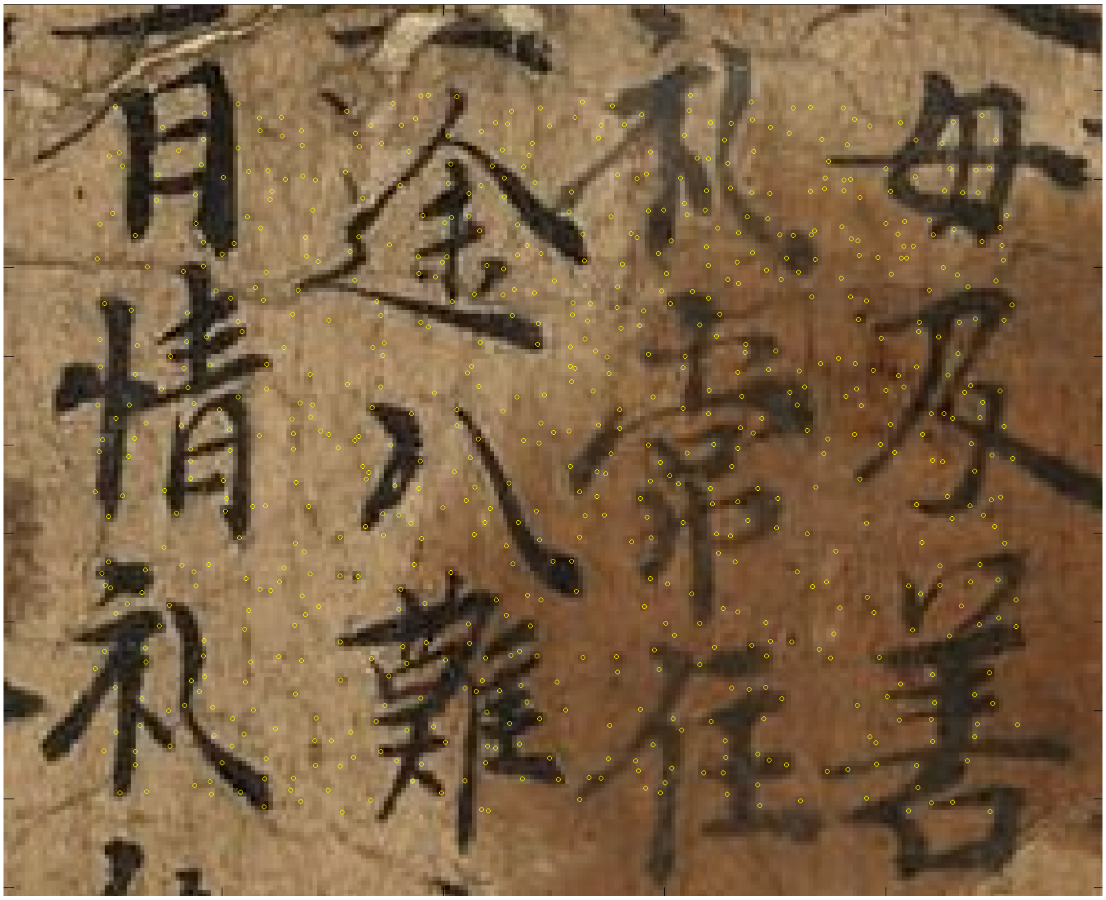


Figure 9: Shi-Tomasi Keypoints.

3.6 SIFT

The Scale Invariant Feature Transform (SIFT) algorithm, developed by David Lowe [8], is an algorithm originally targeted for object detection. The key aspects of the algorithm focus on its abillity to handle both scale and rotation changes. The algorithm consists of two distinct parts, namely the keypoint detector and the keypoint descriptor, respectively.

The keypoints are detected by searching for local extrema on images. From the original image a series of images are created using different gaussian filters rerpesenting this way a scale space. Applying Difference of Gaussians (DoG), the images used for the detection of the keypoints are extracted.

The descriptor part of the algorithm involves the creation of a histogram for a predetermined area around each keypoint. The histogram is repeated for many orientations so as to ensure the rotationally invariant character of the algorithm. The size of the descriptor is 128.

The algorithm performs well in situations where the information that we need to find is not repeated. However, in cases where we have the desired information located multiple times on an image the algorithms fails to perform well out of the box. That is the case where document images cannot be segmented and are treated as a single entity. Furthermore, the algorithm extract a large number of keypoints, many of which are located to positions different than the actual information, such as the background. An example of SIFT keypoints on a sample document image is shown in Figure 10. Indeed, we need to smooth the image in order to eliminate as much as possible unwanted keypoints.

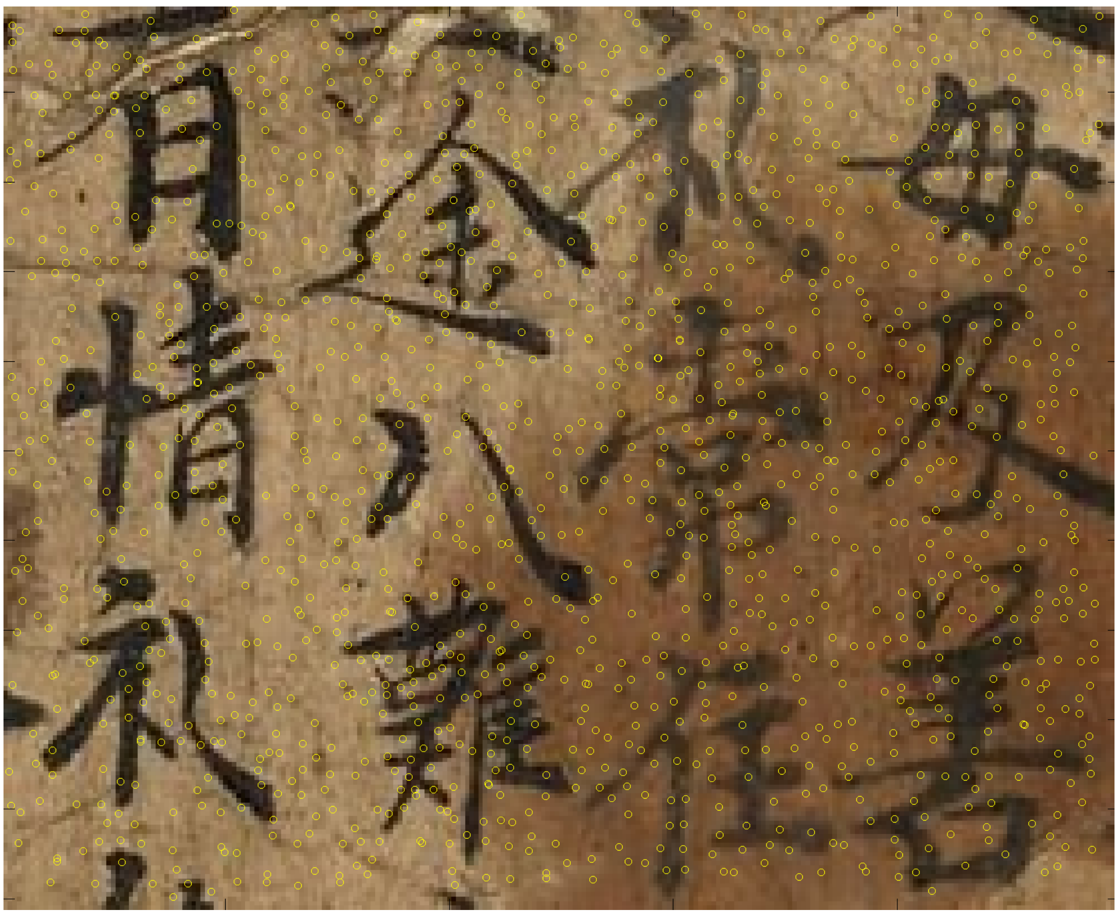


Figure 10: SIFT Keypoints.

In the same path as the SIFT algorithm, Bay et. al. [1] proposed an affine invariant algorithm for keypoint extraction. The algorithm instead of using DoGs, it used the Determinant of the Hessian (DoH) matrix on integral images. This causes a significant speed up compared to SIFT. Furthermore, it is claimed to produced more robust features since it does not rely on DoGs.

4 Conclusion

The use of features in the processing of document collections is a crucial part for the development of robust methods. There are features that can be used in both segmented and non-segmented documents. The later seems to gain more and more popularity since it avoids the segmentation step which can produce errors due to the bad quality of historical document images. In this report we have used different features on ancient handwritten chinese documents. The aim was to check the different characteristics and how they effect the performance on various applications. From our point of view, there are no good or bad features. There are features that can perform better than others under certain requirements.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 -- 359, 2008. Similarity Matching in Computer Vision and Multimedia.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*, pages 886--893, 2005.

[3] N. Feddaoui, H. Mahersia, and K. Hamrouni. Iris recognition method based on gabor filters and uniform local binary patterns. *International Journal on Image Graphics*, 12(2), 2012.

[4] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:891--906, 1991.

[5] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147-151, 1988.

[6] Q. Huo, Y. Ge, and Z. Feng. High performance chinese OCR based on gabor features, discriminative feature extraction and model training. In *IEEE International Conference on Acoustics, Speech, and*

Signal Processing, ICASSP 2001, 7-11 May, 2001, Salt Palace Convention Center, Salt Lake City, Utah, USA, Proceedings, pages 1517--1520, 2001.

[7] J. A. Laborda, L. Salgado, and M. Camplani. Image-based on-road vehicle detection using cost-effective histograms of oriented gradients. *J. Visual Communication and Image Representation*, 24(7):1182--1190, 2013.

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

[9] H. B. Maguluri, Q. Tian, and B. Li. Detecting text in floor maps using histogram of oriented gradients. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26-31, 2013*, pages 1932--1936, 2013.

[10] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *In Proceedings of the 7th European Conference on Computer Vision*, pages 0--7, 2002.

[11] F. Riaz, A. Hassan, S. Rehman, and U. Qamar. Texture classification using rotation- and scale-invariant gabor texture features. *IEEE Signal Processing Letters*, 20(6):607--610, 2013.

[12] J. Shi and C. Tomasi. Good features to track. In *9th IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[13] S. Tian, S. Lu, B. Su, and C. L. Tan. Scene text recognition using co-occurrence of histogram of oriented gradients. In *2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, August 25-28, 2013*, pages 912--916, 2013.

[14] S. Zhang and X. Wang. Human detection and object tracking based on histograms of oriented gradients. In *Ninth International Conference on Natural Computation, ICNC 2013, Shenyang, China, July 23-25, 2013*, pages 1349--1353, 2013.

[15] A. G. Zuñiga, J. B. Florindo, and O. M. Bruno. Gabor wavelets combined with volumetric fractal dimension applied to texture analysis. *Pattern Recognition Letters*, 36:135--143, 2014.