Detecting Structured Image Region Using Local Features and Clustering Analysis

Huei-Yung Lin, Chin-Yu Hsu, and Yung-Yang Chiang
Department of Electrical Engineering
Advanced Institute of Manufacturing with High-Tech Innovation
National Chung Cheng University
168 University Rd., Min-Hsiung, Chiayi 621, Taiwan
hylin@ccu.edu.tw, chinyu.hsu@gmail.com, yychiang0213@gmail.com

Abstract

The structured regions in an image usually contain important clues for information understanding. Proper extraction of those regions is often a key to success for a computer vision system. In this work, we propose a method to detect the irregularly arranged region-of-interest with a fixed structure. Different from the existing techniques, our approach is able to deal with more influential factors in the images and suitable for many application scenarios. To locate a structured region, the density clustering analysis is used to summarize the intensive feature regions, followed by the iterative region selection with a specific structure. The experiments with real scene images have shown that our technique is able to provide stable results in various text extraction applications.

1 Introduction

Detecting the region of interest (ROI) in an image is an important problem which has attracted the attention of many researchers for extensive investigation for many decades. The objective is usually to identify the image region which is meaningful to the human perception or used for future processing. It is thus considered as an early stage for the automatic information extraction from the acquired images [8]. Some popular applications include face detection for person identification or camera auto-focusing, text detection for optical character recognition, or structured pattern identification for barcode or symbol reading, etc. In general, the overall performance of a machine vision system can highly depend on the correctness of the ROI detection results.

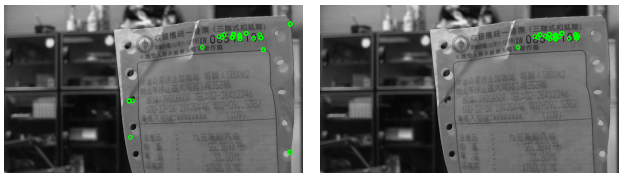
For the identification of interested regions, local feature analysis is a common technique to find the correspondence between the reference and target images [12, 10]. Viola and Jones used Haar-like features with AdaBoost machine learning methods for face recognition [14]. Dalal and Triggs used histograms of oriented gradients to derive the object gradient distribution and send to the linear SVM classifier for pedestrian detection [4]. Dorkó and Schmid used size-invariant local features to develop a vehicle detection method [5]. Cheng *et al.* used a visual attention model to decide the important region in a video sequence [3]. Although the above approaches demonstrated good region detection results, they all made an implicit assumption on the object of interest about its continuous nature appeared in the image. Consequently, those techniques cannot be directly used to detect the image region containing the object with discrete internal structures.

In this work, we are interested in detecting the structured yet discontinuous patterns in an image, and more specifically, detecting the text regions with arbitrary orientations in different scenarios. Due to a wide range of applications related to image content analysis, text detection has become an active research topic in recent years. Epshtein *et al.* converted the edge gradient information to the width of handwriting texts, and used the distribution to localize the text region [6]. Sun *et al.* used a visual attention model to simulate the immediately noticeable area by the human vision and detected the strong signal associated with texts [13]. Zhang and Kasturi adopted the HOG features to select the most likely character elements from the texts [15]. Chen *et al.* filtered out the non-text region by considering the difference between the foreground and background based on the maximally stable extremal regions [2].

This paper presents a structured region detection approach using local features and clustering analysis. Given an image, the SIFT features corresponding to the similar structures in the reference images are first extracted, followed by the ROI detection based on analyzing the clustering characteristics of the ordered feature points. The proposed method uses multiple character images in the database for feature matching, it is able to detect the text clusters even if the feature points are sparse for some characters. Once the candidate location with high feature density is identified, an iterative process to increase the ROI is carried out to derive a suitable region with structured content. The experimental results on the text detection and recognition of invoice, banknote and license plate have demonstrated the effectiveness of the proposed technique.

2 Feature Selection and OPTICS Clustering

The first step of our structured region detection is to extract the feature points which are associated with the reference image in the database. To perform the correspondence matching, the commonly used SIFT feature is adopted in this work. Since the ROI for text detection is relatively small in general scenes, it requires a large number of database images to guarantee that the feature points are enough for region extraction. However, increasing the number of reference images for feature matching implies a high computation cost, which is not preferable in most applications. To improve the feature matching efficiency, one needs to consider not only the mismatching rate and the storage for reference



(a) Conventional method. (b) Our technique.

Figure 1. The feature extraction results.

(a) Large scale image. (b) Small scale image.

Figure 2. The clustering results.

images, but also the computation time for performing the feature matching task.

Our strategy for robust feature matching is to conduct an internal feature selection among the reference images. A correspondence matching stage is carried out for the SIFT features in all reference images, the features with good pairing results are considered as important features. Only those prominent features in the reference images are used to match the features in the target images for region detection. Moreover, we allow the one-to-many correspondences between the target and reference features to increase the matching efficiency. Figures 1(a) and 1(b) show the feature extraction results using the conventional method and our technique, respectively. It is clearly that, using the proposed method, the undesired features are removed while the important features remain.

After the feature matching process, it is reasonable to extract the feature points scattered within the structured image region. To analyze the spatial relationship among these feature points, we need to identify the dense clusters as candidate regions for ROI detection. Since there might still be some outliers in the set of image features, it is important to adopt an effective clustering method to localize the core feature points to form an initial detection region.

Almost all well-known clustering analysis algorithms require good parameter settings [9]. The parameters are not only difficult to decide but also influential to the clustering results. In this work, the OPTICS algorithm is used for hierarchical clustering [1]. A reachability plot is generated based on the ordering of reachability distances associated with the feature density. When processing a large amount of data, it is able to avoid losing important clusters due to improper parameter settings. Since there are less features in the target image for correspondence matching, the result will be more sensitive for the clustering technique. Thus, we modify the algorithm with additional constraints on the feature distribution to make it robust under the image scale change. If the feature density is less than a threshold, then the image is normalized for further hierarchical clustering. Figure 2 shows the clustering results with two different scales. The set of connecting red lines indicates the core detection region.

3 ROI Extraction

The specific area to be identified in the image usually possesses a similar type of structure in the region of interest. It is possible to extract the area by analyzing the strong relationship among the elements. As an example, a serial number of an invoice or a banknote is formed by several digits or characters. They have fixed structural properties such as the number of elements, the space between the elements, the aspect ratio of the region and individual elements, etc. This information play an important role for the selection of the region of interest.

After the feature selection and clustering analysis, a cluster in a specific area is identified. We need to select the candidate region as close to an ideal one as possible, with the information indicated by the cluster. This will then facilitate the region adjustment in the next stage. Since a serial number usually consists of multiple elements arranged along a straight line, the associated feature points should be found in a fixed direction. Thus, the line fitting can be carried out on the feature cluster, and the resulting points will scatter along the text direction. Finally, a rectangular region containing the feature points is used to represent the initial ROI and enlarged along the straight line direction to include all characters.

There are two constraints applied to remove the outliers of a cluster. First, RANSAC is used to eliminate the feature points which are further away from the initial ROI based on a line model [7]. This, however, will still have the outlier features in the line direction remain intact. The second criterion is based on the density ordering relation of the feature points derived from the OPTICS algorithm. It is found that the outliers commonly appear at the first or last point of the ordered feature string, and their corresponding distances to the connecting feature point is significant larger than others. Thus, they can be removed by thresholding the distance distribution of the feature points with a preset variation change.

In the following stage, our objective is to correctly adjust the candidate area to the region of interest correctly as long as the location is inside the region, regardless of the differences in size or the accuracy in direction. The process will be performed iteratively, until the region selection is satisfied. First, the target image is rotated using the tilt direction (with respect to the image scanline) provided by the feature point distribution of the candidate region. The objective is to facilitate the initial setting of a rectangular bounding box to fit the content of the candidate region, and decide how to perform the region expansion. By referring to the knowledge of specific structures, the bounding box region is enlarged or shifted if the horizontal and vertical projections of the connected component do not satisfy the defined area size requirement. If the boundary of the current candidate area across some characters, it indicates that the text is not included to the ROI completely, and a region expansion process will be carried out.

The determination of moving direction is based on both the distribution of the connected components and

Table 1. The ROI extraction algorithm.

Algorithm : ROI extraction

comment : extract ROI from candidate

```
ROI ← RotateByFirstOrientation(candidateROI)
iterative ← True
WHILE iterative
    iterative ← False
    Binarization(ROI)
    (midAreaCC, quantity) ←
        ConnectedComponentAnalysis(ROI)
    IF heightAccError > heightRatio
        Rotate(ROI)
        Truncate(ROI)
        iterative ← True
        continue
    IF quantity < totalAmount
        IF margin is confused
            trend ← ComputePointDistributed()
            ExpandByTrend(ROI) or
            ShiftByTrend(ROI)
        ELSE
            ExpandByMargin(ROI) or
            ShiftByMargin(ROI)
        iterative ← True
END WHILE
return ROI
```

the regions formed by the horizontal and vertical projections of the area. Less blank areas implies there might be more components in that direction. It is possible that erroneous results exist due to the inaccurate initial assessment of the area. We need to have a timely response since it could cause a serious problem due to the accumulation of the improper expansion of the detected region. Whether rotate the ROI or not is based on the accumulation error derived from the elements and reference to a medium size element in the region. Because there might be noise present in the image, the assessment of the accumulation error is the difference of heights between the element and the reference object in the region. In case the error is over a proportion of the threshold compared to the reference object, the ROI is rotated with an angle derived based on the accumulation error. Since the size of the elements does not differ too much in general, the selected reference object is not only for the assessment of the height error and used for rotation, but also serves as a basis of a noise filter. The algorithm for ROI extraction, including the iterative process and region rotation is shown in Table 1.

4 Experimental Result

This section presents the experimental results of the proposed technique for structured image region detection. We have investigated three applications– text detection for invoice, serial number identification for banknote, and vehicle license plate detection. The experiments contain feature point matching, feature point clustering, and specific region selection. All these stages are closely related, and the results are tabulated for illustration. To demonstrate the effectiveness of our region detection results, we also perform the optical character recognition using a neural network based approach [11].

Table 2. The region detection results.

	Invoice	Banknote	License Plate
# of Samples	113	109	108
ROI	116	114	110

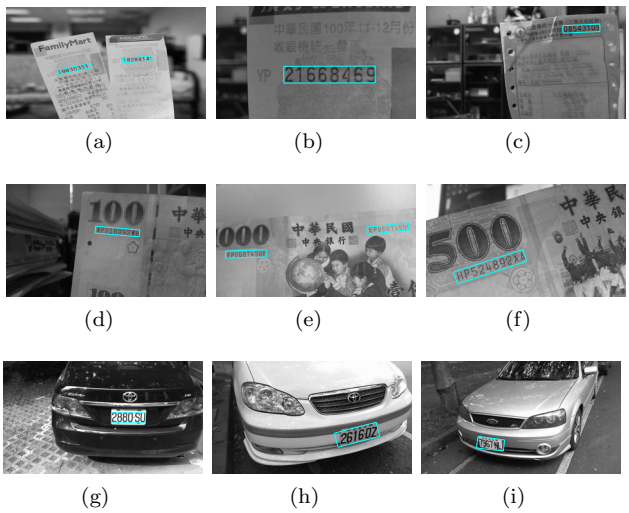
Table 3. The clustering results.

	Invoice	Banknote	License Plate
# of ROI	116	114	110
Detected	107	106	70
Accuracy	92 %	93 %	63 %

The proposed technique is developed with the idea of simple and easy to use. For the consideration of practical applications, the test samples in the experiments are complicated in order to illustrate the robustness of our method. Table 2 tabulates the number of samples and the detected region of interest for three different applications. We do not enforce the constraint on the number of detected region, so each sample image can contain more than one region of interest as indicated in the table. The performance evaluation on the correctness of the detection is based on the observation of regions of interest.

Figure 3 presents the detection results of specific areas for the invoice, banknote and license plate images. If the clustering result of the feature points in an image is located at the correct region of interest, then it is considered as the correct clustering and the follow-up setting of the candidate will also be correct. As shown in Table 3, the test samples of banknote and invoice achieve good clustering results, but not for the samples of license plate. The main reason is that there are not enough representative feature points extracted from the reference images in the database. Consequently, many samples can only obtain a very limited number of correspondence in the local feature matching, which is not able to provide the points dense enough for the clustering stage. Currently, we are not able to derive a general parameter setting for license plate detection using our clustering method. The clustering result is not correct if there are only vey few feature correspondences identified. This will be investigated in the future, especially for the cases with non-ideal cases such as the long distance capture or the license plate appeared with a large tilt angle.

The correctness evaluation of specific regions is determined by the final selection of regions of interest, only those with all contents fully covered are regarded as correct results. Since the specific region selection depends on the candidate regions derived from successful clustering, it is more reasonable to to discuss the success rate without considering the clustering results. As shown in Table 4, the statistics of the ROI detection rate is over 90% for the indoor scene. The accuracy of optical character recognition using the ROI detection results are illustrated in Table 5.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 3. The region detection results for three different applications.

Table 4. The ROI detection results.

# of ROI	Invoice	Banknote	License Plate
ROI	107	106	70
Correct	100	102	61
Accuracy	93 %	96 %	87 %

5 Conclusion

This paper presents a novel method to detect the region of interest by local feature correspondence matching and clustering analysis. The objective is to search a target area for specific region detection and identification. The technique is designed for general situations and with a high degree of fault tolerance. It is able to deal with the problems such as the target area in any direction, the region size change, and the perspective distortion. We applied SIFT features with scale and rotation invariant for local correspondence matching to increase the flexibility. By using the OPTICS algorithm with density analysis to derive the clustering characteristics of the corresponding points, the feature distribution of the structured region associated with the original area is adjusted to obtain the proper location of the region of interest. The future work will focus on the enhancement of tolerance on special circumstances, such as enhancing the region detection with severe perspective distortion or defocus blur.

Acknowledgment

The support of this work in part by the National Science Council of Taiwan, R.O.C, under Grant NSC-99-2221-E-194-005-MY3 is gratefully acknowledged.

References

[1] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. Optics: ordering points to identify the clustering structure. In *Proceedings of the 1999 ACM SIGMOD international conference on Management of data*, SIGMOD '99, pages 49–60, 1999.

[2] H. Chen, S. Tsai, G. Schroth, D. Chen, R. Grzeszczuk, and B. Girod. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 2609 –2612, sept. 2011.

[3] W.-H. Cheng, W.-T. Chu, and J.-L. Wu. A visual attention based region-of-interest determination framework for video sequences*. *IEICE - Trans. Inf. Syst.*, E88-D(7):1578–1586, July 2005.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, pages 886–893, 2005.

[5] G. Dorkó and C. Schmid. Selection of scale-invariant parts for object class recognition. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 634–, 2003.

[6] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2963 –2970, june 2010.

[7] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[8] N.-C. Huang and H.-Y. Lin. A multi-stage processing technique for character recognition. In *Advanced Intelligent Mechatronics (AIM), 2012 IEEE/ASME International Conference on*, pages 1081–1085, July.

[9] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323, Sept. 1999.

[10] H.-Y. Lin and W.-C. Fan-Chiang. Reconstruction of shredded document based on image feature matching. *Expert Syst. Appl.*, 39(3):3324–3332, Feb. 2012.

[11] H.-Y. Lin and C.-Y. Hsu. Optical character recognition with fast training neural network. In *International Conference on Image Processing*, pages 793–796, 2012.

[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[13] Q. Sun, Y. Lu, and S. Sun. A visual attention based approach to text extraction. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3991 –3995, aug. 2010.

[14] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, May 2004.

[15] J. Zhang and R. Kasturi. Text detection using edge gradient and graph spectrum. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3979 –3982, aug. 2010.

Table 5. The OCR results.

	Invoice	Banknote	License Plate
# of Char.	808	949	332
Correct	673	722	260
Accuracy	83 %	76 %	78 %