2013 12th International Conference on Document Analysis and Recognition

Using Harris Corners for the Retrieval of Graphs in Historical Manuscripts

Rainer Herzog, Arved Solth, and Bernd Neumann
Department of Informatics
University of Hamburg
Germany
{herzog, solth, neumann}@informatik.uni-hamburg.de

***Abstract*—In recent years, several methods have been proposed for content-based retrieval from manuscripts, mostly based on character or word similarity. In this paper, we present a new segmentation-free method, called Harris Corner Matching (HCM), which accepts an arbitrary writing pattern as a model and allows to retrieve similar patterns from a possibly large database. Retrieval is performed in two steps. In the first step, candidate targets are determined by comparing configurations of interest points in query and data. In fact, interest points can be used as precomputed indices. In the second step, deviations between the interest point configurations of query and candidate target are used to warp the target, this way adapting it to the query. A final evaluation is obtained by template matching after binarization of query and target. The method has been evaluated for retrieval from historical Chinese and Sanskrit manuscripts and has shown good results.**

Keywords-retrieval; manuscript; interest point; graph

I. INTRODUCTION

This paper is concerned with content-based image retrieval from historical manuscripts. Such tasks may arise in many applications related to manuscript research. One may want to check whether a given (partial) manuscript image is contained in a possibly large database of manuscript images. In this case, the main challenge of a retrieval method is to provide an efficient index for targets which exactly match the query image. A more difficult task is to retrieve target patterns from a manuscript database which are in a certain sense "similar" to a query pattern. One kind of similarity, useful for many applications, is similarity by textual content. Here, the query pattern is viewed as a sequence of characters, and the goal is to retrieve the same sequence of characters from the database. This kind of retrieval, often called "word-spotting", may be quite challenging depending on what differences between the writing styles of query and target manuscripts must be tolerated. Another important application, typically requiring a tighter similarity measure, is retrieval for writer identification or verification. This task involves the search for many occurrences of the same kind of writing pattern in order to determine and compare characteristic features of unknown hands. Character or word recognition is not required as long as the retrieved patterns qualify for a valid comparison.

Our work is part of a major project[1] where scholars of the Humanities cooperate with researchers of Computer Science and Material Sciences in investigations with a special focus on material and visual properties of historical manuscripts. We are confronted with retrieval tasks for manuscripts of Asia, Africa and Europe, and hence with diverse writing systems and writing styles. The comparison of manuscripts from unknown origins is a frequent challenge.

Our approach, reported in this paper, is therefore not specialized to deal with a particular writing system and cannot be based on recognized characters or words. We aim at the retrieval of writing patterns (graphs) which are similar by appearance. We are mainly motivated by successful work in content-based image retrieval (CBIR), but also incorporate assumptions about typical manuscript retrieval tasks and properties of historical handwriting. In particular, we assume prior knowledge about the approximate orientation and scale of possible targets, a basically bimodal image content, and deviations of targets from the query pattern which are typical for handwriting.

In the first stage of our retrieval process, a configuration of local descriptors is computed for the query pattern at interest points determined by the Harris corner (HC) detector [1]. The descriptors are taken from a visual alphabet of 48 fixed types which model the typical environment of a corner point. The same computation of local descriptors must be performed for the database, albeit only once. Using the descriptors as indices, the retrieval process determines target candidates where the spatial configuration and the types of descriptors are similar to the query configuration, establishing corresponding interest points.

In the second stage of the retrieval process, query and target candidates are compared more accurately. A new similarity test has been developed where deviations between corresponding HCs are used to iteratively warp the target, thus undoing distortions between query and target. The final evaluation is obtained by template matching after binarization. In the section following this short overview, we discuss related work.

[1] Centre for the Study of Manuscript Cultures (CSMC), DFG Collaborative Research Center SFB 950.

1520-5363/13 $26.00 © 2013 IEEE
DOI 10.1109/ICDAR.2013.262

In Section III, we present our retrieval approach in more detail. Experimental results with Chinese and Sanskrit manuscripts are presented in Section IV. The paper concludes with a summary and an outlook on further work.

II. RELATED WORK

As indicated in the introduction, retrieval from historical manuscripts is viewed in our approach as a special kind of content-based image-retrieval, hence the large body of work on CBIR is potentially relevant [2]. As pointed out in [3], there is strong evidence that retrieval based on a "visual language" composed of local descriptors promises superior results for CBIR. Local descriptors are often determined at interest points which are expected to mark stable locations attached to the image content, invariant to different views. SIFT features [4] located at interest points computed with the Harris Corner (HC) detector [1] are in wide use for this purpose. They provide a scale- and rotation-invariant description of the neighbourhood of interest points based on gradient histograms.

In manuscripts, pronounced responses of the HC detector can be expected at line endings, line crossings or locations of strong line curvature, all of which contribute significantly to the appearance of handwriting in many writing systems. Hence there are good reasons to use HCs also for locating local descriptors in manuscripts.

Several variants of SIFT features have been proposed for character recognition [5], handwriting retrieval [6, 7] and layout analysis [8]. However, the scale and orientation invariance of these descriptors is rarely required for handwriting retrieval applications. Other local descriptors proposed for the neighbourhood of HCs are Histograms of Gradients (HOG) [9, 10] or tensor fields [11].

In our experiments with several different writing styles in historical Chinese and Sanskrit manuscripts, however, gradient-based local descriptors turned out to be often misleading in cases of degraded or distorted handwriting. Therefore our approach only exploits the spatial configuration of interest points and their qualitative orientation-specific types for an initial retrieval, and includes silhouette-based matching, as in [12], for a final test.

A large number of retrieval approaches is based on queries with words or connected components [7, 9, 12, 13, 14, 15, 16, 17, 18, 19]. A segmentation into words can offer many advantages, in particular for Latin handwriting, since features can be ordered left-to-right [7] or even assigned to a mesh of predefined subregions [14]. In our approach, a segmentation into words or characters is exploited whenever it can be reliably obtained, e.g. for Chinese manuscripts. This is, however, no prerequisite.

III. RETRIEVAL BY HARRIS CORNER MATCHING

We now present our retrieval approach HCM (Harris Corner Matching) in greater detail. In the first subsection, we describe candidate selection based on the types and the spatial configuration of interest points. If required, the matching results are further refined by comparing the silhouettes of corresponding writing patterns after warping the target pattern according to interest point deviations. This is described in Subsection 2.

A. Retrieval by Configurations of Local Descriptors

In the first stage of the retrieval process, local descriptors (LDs) are obtained for the query image. The locations (Fig. 1 left) are determined by the HC detector and associated with a strength value. The type of each LD is determined by classifying the neighbourhood according to 48 fixed 11 x 11 pixel neighbourhoods (Fig. 1 right) using an SVM. To this end, the query image has been normalized so that the line thickness of the writing pattern is in the order of the window size used by the classifier. The types have been chosen systematically to describe local appearances of line endings, bends and crossings in all orientations, discretized in 45^{o} steps. The SVM has been trained once and for all on patterns matching the templates, manually selected from manuscript databases like the ones used in our investigation (see Section IV). We believe that this approach promises more generality than a visual alphabet obtained by learning, as proposed, e.g., in [7, 13].

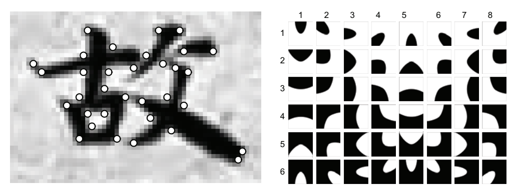


Figure 1. Interest points determined for a Chinese character using the HC detector (left). 48 templates used for classifying interest points (right).

Using the same method, LDs are computed for the manuscript database after a corresponding normalization, albeit only once as an off-line preparation for all retrieval operations, equivalent to an annotation. LD types are established as indices such that each LD location can be retrieved from the type and its eight neighbours according to the type table, this way achieving a moderate tolerance against type deviations between query and data.

To determine a possible target location for a query, a strong HC of the query is matched with a type-compatible strong HC of the data. Using this location as a reference, all other query LDs are checked for type- and location-compatible partners in the data. Acceptance of a target is controlled by a hypothesis test based on probabilities for type- and location deviations, as reported in [11].

Due to many reasons, the LD of the query which has been chosen as reference may not have a corresponding partner in a valid target. Therefore, the search is repeated a few times with other LDs as reference points. In our experiments, three cycles sufficed to locate all targets.

The retrieval performance obtained by this matching procedure is satisfactory for queries with a sufficiently large number of LDs and limited deviations between query and target, as illustrated in Fig. 2. For many manuscripts, however, valid targets may be more severely distorted with respect to the query, and HCs may be displaced or missing. We have therefore developed a second stage for the retrieval process, described in the following subsection.

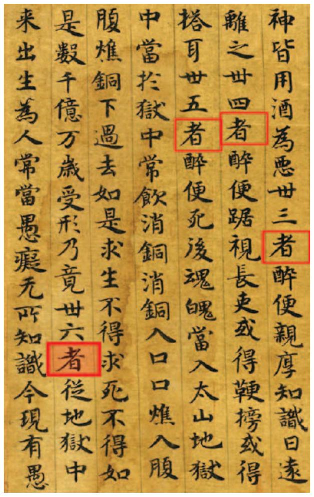


Figure 2. Section of the Fo shuo Tiwei jing (© The British Library Board, Or.8210/S.2051). The Chinese character in the left-most box has been used for a query. The retrieval results, all correct, are shown in the other boxes.

B. Matching Silhouettes after Warping the Target Pattern

We are interested in retrieving target patterns which differ from the query by larger distortions and whose LDs may be so noisy that a reliable identification is not possible by solely comparing LD configurations. In our approach, the first retrieval stage, described above, is used for a rough selection of candidates, hopefully including all valid targets. This can be achieved by setting the acceptance threshold sufficiently low. In the second stage, the candidate patterns are more elaborately compared with the query. The main idea is to use the discrepancies of corresponding LD locations as indicators for distortions and warp the targets accordingly to undo these distortions. This enables template matching of binarized versions of query and target as a final test. Silhouette-based warping has also been proposed for word recognition by [12].

Several steps of this procedure require careful treatment. First, the warping method should reflect typical distortions of writing patterns. We take this into account by selecting warping with Gaussian radial basis functions which have the desired local effect while preserving global smoothness requirements. Second, not all LD correspondences determined in the first stage are reliable. Hence warping the target according to a bad correspondence may corrupt the result. We therefore ascertain that a warp improves the match and ignore the corresponding LD pair otherwise. Third, when comparing the binarized query pattern with the binarized target, one may want to ignore differences due to additional components in the target window. We therefore offer the choice of two similarity measures, S_{WINDOW} based on all different pixels in the target window, and S_{QUERY} only based on differences within the query silhouette. Thus S_{QUERY} tests to which degree the query pattern is *contained* in the warped target pattern. Let Q and T be the pixels of the query and target silhouette, respectively, and Q^c and T^c be the respective complement areas in the window. Then the similarity measures are:

$$S_{WINDOW} = 1 - (| Q^c \cap T | + | Q \cap T^c |) / (| Q | + | Q^c |)$$

$$S_{QUERY} = 1 - (| Q \cap T^c |) / | Q |$$

In the experiments described in the following section, S_{WINDOW} is used.

The effect of warping is illustrated in Fig. 3 for the retrieval of a Chinese character from Database B (see below). Mismatching line patterns (white areas) are considerably reduced by warping.

Figure 3. Retrieval of a Chinese character, query and target left, matching results before and after warping right. White areas show mismatching line patterns, reduced by warping.

IV. Experimental Evaluation

In this section, we present evaluation results for manuscript databases with historical Chinese and Sanskrit handwritings of diverse origin and different degree of variability. To facilitate evaluation, we use characters here as retrieval patterns.

Database A is a section of the Fo shuo Tiwei jing with ca. 2000 Chinese characters, part of which is shown in Fig. 2.

Database B is a section of Or.8210/S.2717-Recto, also with ca. 2000 Chinese characters, partly shown in Fig. 4.

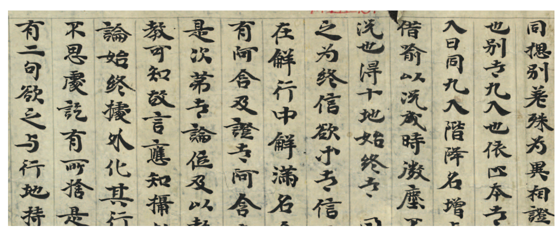


Figure 4. Section of the manuscript used for Database B with Chinese Characters (© The British Library Board, Or.8210/S.2717).

Database C is one part of the Sanskrit manuscripts of Raghuvamsa (Kathmandu) with ca. 1800 Newari characters. A section is shown below in Fig. 5.

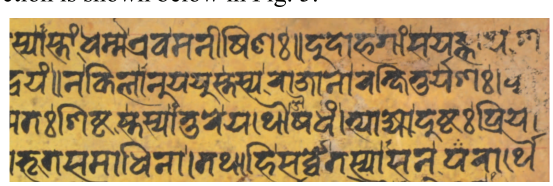


Figure 5. Section of the manuscripts of Raghuvamsa with Newari characters.

We first present a quantitative appraisal of the variability of characters in the manuscripts. We then report performance results for our retrieval procedure.

A. Quantitative Evaluation of Character Variability

In order to assess the variability of characters within each manuscript and the distortions which must be removed by warping, we determined the displacements of corresponding HCs for sample characters. Corresponding HCs were selected manually to exclude displacements due to noisy correspondences rather than character distortions.

Figure 6. Two occurrences of the same Chinese character in Database A (left and center), displacement vectors of corresponding HCs relative to the character width, normalized to zero mean. (right).

As shown in Fig. 6, the displacements, normalized to zero mean, are typically smaller than 3% of the character width. They may be much larger, however, for fairly isolated endpoints of strokes, such as the right endpoint of the bottom stroke in Fig. 6. Hence warping should have a local effect, adapted to the individual displacement of each HC pair.

Fig. 7 shows analogous results for Database B. By subjective impression, the handwritings are less regular, and this is also reflected in the quantitative displacements of HCs which are in the order of 5% of the character width for this database.

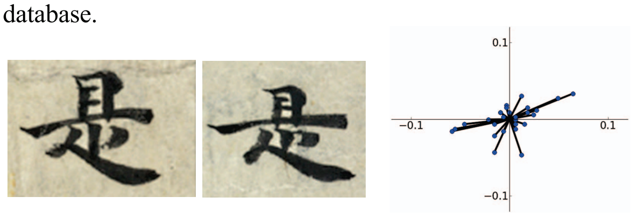


Figure 7. Displacement vectors for characters of Database B.

Fig. 8 shows analogous results for two occurrences of the same Newari character in Database C with displacements up to 8% of the character width. The example also illustrates the smaller number of HCs to be expected for a Newari character.

Figure 8. Displacement vectors for characters of Database C.

The results in Figs. 6 - 8 illustrate the variability of characters within the same manuscripts, quite likely written by the same hand. If query and data belong to different manuscripts, much larger displacements of HCs have to be expected, of course. This is shown in Fig. 9 for a query character taken from Database A and a target in Database B.

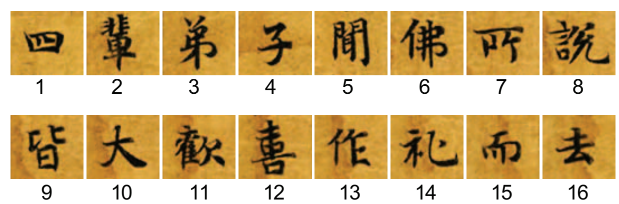
Based on this quantitative appraisal of possible character distortions, we have set the threshold for acceptable correspondences in our HCM procedure to the maximally expected distortions. This may give rise to a considerable number of false positives, as will become evident in the following subsection, but seems necessary to minimize false negatives.

Figure 9. Displacement vectors occurrences of the same Chinese character in Database A (left) and Database B (center).

B. Retrieval Performance

The retrieval performance was evaluated with characters from a single column or line of each of the manuscripts to provide a random selection. The number of characters which could be tested was limited because the ground truth had to be provided manually: For each test character, the complete database had to be scrutinized for true positives. As the number of occurrences of each test character varies considerably, we present results for individual characters.

The test characters for Database A are shown in Fig. 10. The 3rd, 11th and 14th character did not occur in the rest of the manuscript, the others occurred between once and 28 times. The high regularity of the characters in this manuscript (see Fig. 6) allowed 99% recall at an average of 80% precision already after the first stage of the HCM procedure, without any warping in a second stage. The individual results are shown in the table.



Character	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
True Positives	13	2	0	5	12	10	28	2	16	2	0	7	4	0	5	1
False Positives	5	0	0	0	5	0	1	0	10	0	0	4	0	0	1	0
False Negatives	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Figure 10. Test characters used for retrieval from Database A (upper section) and retrieval results (lower section). Only the first stage of HCM has been used, with the acceptance threshold set to minimize false negatives.

Retrieval from Database B was tested with 8 query characters chosen from a single column of the manuscript. As illustrated in Fig. 7, the variability of characters is much higher than in Database A, hence the effectiveness of the warping stage of HCM could be evaluated. The 8 plots in Fig. 11 show precision-recall curves for each test character (above the plot), in red for retrieval without warping, in green with warping. The curves without warping have been obtained by varying the acceptance threshold in the hypothesis test of Stage 1. For the curves with warping, this threshold has been fixed at a tolerant level, while a threshold for the similarity measure S_{WINDOW} has been varied.

The results in Fig. 11 show that the precision is increased considerably by warping in Stage 2 of HCM. This means that a user, interested in complete recall, will retrieve much fewer false positives than without warping. In our experiments, complete recall was not achieved for two cases due to the fixed threshold in Stage 1. The jerkiness of some of the curves reflects the low number of targets, less than 30, including false positives, for the corresponding queries.

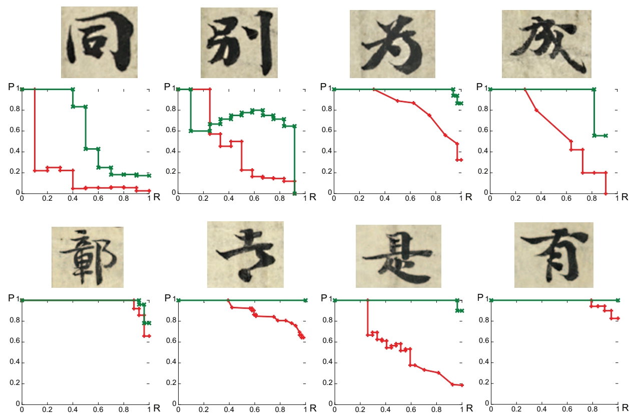


Figure 11. Precision-recall plots for retrieval of test characters (above the plots) from Database B without warping (red plots) and with warping (green plots).

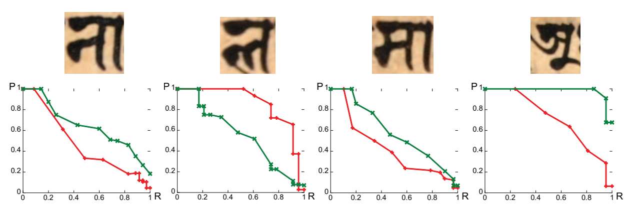


Figure 12. Precision-recall plots for retrieval of test characters (above the plots) from Database C without warping (red plots) and with warping (green plots).

Fig. 12 shows the results for 4 Newari characters of Database C. Except for the second character, where HC locations are noise-prone due to soft bends, warping improves the retrieval performance notably. All examples give rise to much fewer HCs than Chinese characters, degrading the performance. The fourth character shows the strongest advantage of warping due to its complexity.

V. Conclusions and Outlook

In this paper, we have explored a new method for retrieving writing patterns from historical manuscripts. In a 2-stage procedure, called Harris Corner Matching (HCM), the spatial configuration of HCs is first used to locate possible targets for a given query, exploiting HC types as indices. In the second stage, the displacements between corresponding HCs is used to warp the target, this way achieving a much higher precision than using the point configurations alone. Results obtained for two different Chinese and a Sanskrit manuscript indicate that HCM can be applied to diverse handwritings, provided that HC locations at endpoints, crossings and sharp bends of strokes provide sufficiently stable characteristic patterns. In the future, we will explore the applicability of HCM to further writing systems (Hebrew, Arabic, Latin) and develop a user interface for Computer Vision laypersons.

References

[1] C. Harris and M. Stephens "A combined corner and edge detector," Proc. 4th Alvey Vision Conference, 1988, pp. 147–151.

[2] R. Datta, D. Joshi, J. Li, and J.Z. Wang "Image Retrieval: Ideas, influences, and trends of the New Age," ACM Computing Surveys, Vol. 40, No. 2, 2008, pp. 1-60.

[3] J. Sivic and A. Zisserman "Efficient visual search for objects in videos," IEEE Proceedings, Special Issue on Multimedia Information Retrieval, 96 (4), 2008, pp. 548-566.

[4] D.G. Lowe "Distinctive image features from scale-invariant keypoints," Int. Journal of Computer Vision, 60.2, 2004, pp. 91-110.

[5] Z. Zhang, L. Jin, K. Ding, and X. Gao "Character-SIFT: A novel feature for offline handwritten Chinese character recognition," ICDAR '09 Proc. 10th Int. Conf. Document Analysis and Recognition, 2009, pp. 763-767.

[6] M. Rusiñol, D. Aldavert, R. Toledo, and J. Lladós "Browsing heterogeneous document collections by a segmentation-free word spotting method," Proc. ICDAR 2011, Int. Conf. on Document Analysis and Recognition, 2011, pp. 63-67.

[7] I.Z. Yalniz and R. Manmatha "An efficient framework for searching text in noisy document images," 10th IAPR Int. Workshop on Document Analysis Systems (DAS'12), 2012, pp. 48-52.

[8] A. Garz, R. Sablatnig, and M. Diem "Layout analysis for historical manuscripts using Sift features," Int. Conf. Document Analysis and Recognition (ICDAR), 2011, pp. 508-512.

[9] J.A. Rodrıguez, F. Perronnin "Local gradient histogram features for word spotting in unconstrained handwritten documents," in Proc. 1st Int. Conf. Frontiers in Handwriting Recognition (ICFHR'08), 2008, pp. 7-12.

[10] K.Velmurugan and S. Santhosh Baboo "Image retrieval using Harris Corners and Histogram of Oriented Gradients," Int. Journal of Computer Applications, Volume 24, No.7, 2011, p. 6.

[11] R. Herzog, A. Solth, O. Bestmann, J. Scheel, and B. Neumann "Retrieving writing patterns from historical manuscripts using local descriptors," in Digital Humanities 2012, Hamburg University Press, 2012, pp. 288-291.

[12] T. Adamek, N.E. O'Connor, N. Murphy, and A.F. Smeaton "Word matching using single closed contours for indexing handwritten historical documents," Int. Journal on Document Analysis and Recognition, 9 (2-4), 2007, pp. 153-165.

[13] T.M. Rath, R. Manmatha, and V. Lavrenko "A search engine for historical manuscript images," Proc. SIGIR '04 27th Annual Int. ACM SIGIR Conf. Research and Development in Information Retrieval, 2004, pp. 369-376.

[14] S. Srihari, C. Huang, and H. Srinivasan "A search engine for handwritten documents," in Proc. of SPIE-IS&T Electronic Imaging, 2005, pp. 66-75.

[15] S.A. Shabab, W.G. Al-Khatib, and S.A. Mahmoud "Computer aided indexing of historical manuscripts," Int. Conf. Computer Graphics, Imaging and Visualisation, 2006, pp. 287-295.

[16] T.v.d. Zant, L.R.B. Schomaker, and K. Haak "Handwritten-word spotting using biologically inspired features," IEEE PAMI, Vol. 30, No. 11, 2008, pp. 1945-1957.

[17] Y. Leydier, A. Ouji, F. LeBourgeois, and H. Emptoz "Towards an omnilingual word retrieval system for ancient manuscripts," Pattern Recognition 42, 2009, pp. 2089-2105.

[18] E.F. Can and P. Duygulu "A line-based representation for matching words in historical manuscripts," Pattern Recognition Letters 32, 2011, pp. 1126-1138.

[19] M. Baechler, A. Fischer, N. Naji, R. Ingold, H. Bunke, and J. Savoy "HisDoc: Historical document analysis, recognition, and retrieval," in Digital Humanities 2012, Hamburg University Press, 2012, pp. 94-97