2009 10th International Conference on Document Analysis and Recognition

Capturing Digital Ink as Retrieving Fragments of Document Images

Kazumasa Iwata†, Koichi Kise†, Tomohiro Nakai†,
Masakazu Iwamura†, Seiichi Uchida††, Shinichiro Omachi†††

† Dept. of Computer and Systems Sciences, Graduate School of Engineering,
Osaka Prefecture University, Japan
††Faculty of Information Science and Electrical Engineering, Kyushu University, Japan
†††Graduate School of Engineering, Tohoku University, Japan

Abstract

This paper presents a new method of capturing digital ink for pen-based computing. Current technologies such as tablets, ultrasonic and the Anoto pens rely on special mechanisms for locating the pen tip, which result in limiting the applicability. Our proposal is to ease this problem — a camera pen that allows us to write on ordinary paper for capturing digital ink. A document image retrieval method called LLAH is tuned to locate the pen tip efficiently and accurately on the coordinates of a document only by capturing its tiny fragment. In this paper, we report some results on captured digital ink as well as to evaluate their quality.

1. Introduction

Pen-based computing is a promising mean to bridge the gap between computers and human users because it is extremely easy to use for most people. The central and basic functionality for supporting pen-based computing is to capture the movement of the pen tip as digital ink.

In order to make pen-based computing as natural as possible, researchers have put a lot of efforts for devising new mechanisms. A fundamental issue is how to locate the pen tip on the coordinates of a document. Representative methods include tablets, ultrasonic and laser pens, and the Anoto pen. Although they have their individual strengths, a common weakness that limits their applicability is the requirement of special mechanisms which are far from ordinary writing environments. Even with the Anoto pen, which is one of the most advanced mechanisms, the special paper is required for locating the pen tip. Thus a new mechanism is still necessary to achieve the goal of pen-based computing.

This problem may be solved by looking at the pen tip for tracking. From this viewpoint, several researchers have pursued a framework with a pen and a video camera [3, 6]. By analyzing a video sequence of the pen tip movement, digital ink can be recovered. An advantage of this approach is the requirement of no special devices: ordinary pen, paper and a video camera are merely required. A drawback is, on the other hand, the lack of portability: the video camera must be fixed at a known position. If one is interested in the recovery on the coordinates of a document being annotated, the relative position between the video camera and the document must be fixed. This spoils the portability, i.e., another attractive point of pen-based computing.

The problem of portability can be solved by mounting the camera on the pen. In order to recover the digital ink on the document coordinates, the pen should be able to “perceive” on which page and where in the page the pen tip is. For this purpose, the Anoto pen employs special dots embedded on paper as landmarks for locating the pen tip. For realizing this functionality without the special dots, we need to find their substitute on documents themselves.

In this paper, we pursue the idea that printed patterns (characters, line drawings, etc.) can be landmarks for recovering digital ink. In order to realize this idea, a reliable method of locating the pen tip is required. We propose such a capability by improving a method of document image retrieval called LLAH (locally likely arrangement hashing) [4] to make it possible to find a precise location only from a tiny part of a document. In this paper we describe our prototype pen and its algorithm as well as experimental results on recovering digital ink.

2. Task and Solution

2.1. Task definition

First of all, let us define the task of digital ink detection by using foreground patterns of documents. The basic no-

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.192

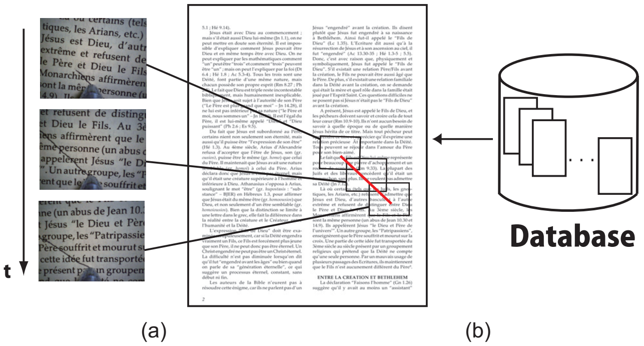


Figure 1. Task of tracking the pen tip.

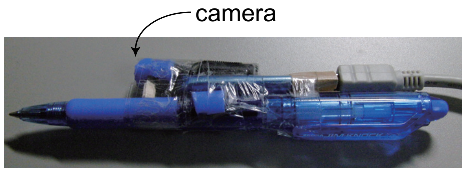
tion is quite simple. We assume that a camera is fixed on the pen and thus the pen tip is always at the fixed position of captured images. This means that, as shown in Fig. 1 (a), the task of finding the position of the pen tip is equivalent to locating images captured by the camera on the document coordinates.

We assume that documents to be annotated have already been stored in the database. Since most documents are printed from their electronic version in recent years, we think this assumption is not unrealistic. For instance, many scientific papers can be accessed via the Internet. Under this assumption, the above mentioned task of locating captured images is equivalent to retrieving their corresponding document and its parts from the database of documents as shown in Fig. 1(b). Note that the pen is permitted to use an actual ink; the retrieval method needs to be robust to handwritten patterns newly added to the document. Another problem is to cope with various image distortions including illumination change and geometric distortions. Thus the retrieval method must be robust to such distortions. The majority of document image retrieval technologies that have been proposed so far can work only on scanned documents; only a few methods [4, 5, 2] are capable of handling severe distortions caused by the camera.

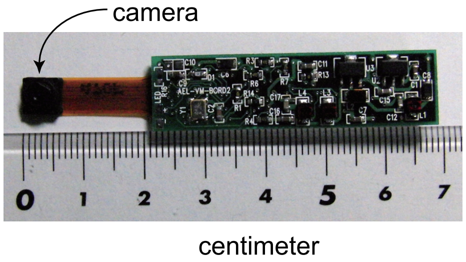
In order to focus first on the above issues of retrieval, the following additional assumptions are made in this paper.

First, we exclude from our task the case that the user writes on blank paper or wide margins of documents. Although this case needs further development, we think that it is not a serious problem; we can apply technologies that complement the current camera pen. For example, a mechanism similar to the optical mouse can be applied to obtain the relative position of the pen tip. The absolute position on the coordinates of a document can be recovered when the camera captures its printed foreground patterns.

Secondly, we do not deal with the issue of recognizing pen-up and pen-down states. It is a non-trivial task when they are detected only by analyzing images [3, 6]. However, this issue has another solution by attaching a micro switch for detecting pen-down states based on the pressure to the pen tip.



(a) Prototype camera pen



(b) Its pen camera

Figure 2. Prototype camera pen and its camera module.

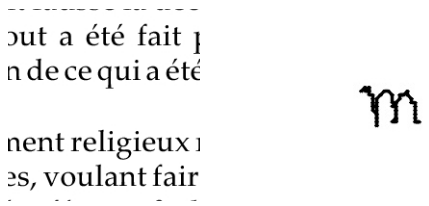
2.2. System configuration

The proposed system is based on the client-server model: the camera pen is connected to the client for processing of captured images. The result of processing is sent to the server for finding the corresponding part from the stored document images. The result of retrieval at the server side is returned to the client for keeping track of the pen tip movement on the coordinates of the retrieved document.

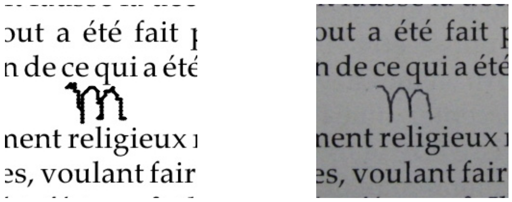
Figure 2(a) shows the prototype of our camera pen. The camera module mounted on the pen (called "pen camera" in the following) is with the VGA resolution and its viewing angle is 52°. The pen camera and the client computer are currently connected with USB but it can be replaced by a wireless connection. The pen camera is shown in Fig. 2(b) in which the size is shown in [cm]. The black square part on the left is the camera module; because it is tiny, we can mount it on the pen.

2.3. Overview of the processing

As discussed in **2.1**, the key processing of our method is to locate captured images on the coordinates of the annotated document. A simple solution is the template matching of a camera-captured image to all the document images in the database. However it is strictly prohibitive due to its computational cost. Therefore, a fast and accurate method



(a) Retrieved fragment (b) Recovered digital ink



(c) Superimposed ink on the electronic document (d) Handwriting on the paper document

Figure 3. Example of recovered digital ink.

of image matching is mandatory for our framework. We have solved this problem by adopting a method of document image retrieval called LLAH (Locally Likely Arrangement Hashing) with some improvements. This will be described in the next section.

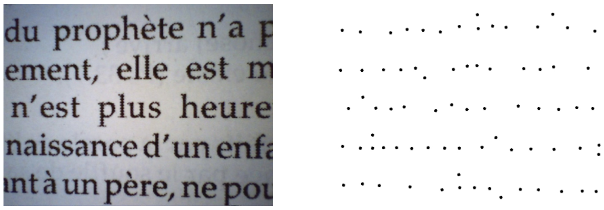
Once the captured image is matched to its corresponding page, the recovery of digital ink is straightforward. LLAH tells us not only the matched part, but also values of parameters of projective transformation. Using these values, we can locate the pen tip precisely on the document coordinates. Figure 3 illustrates an example of recovered digital ink. By connecting the coordinates of the pen tip using retrieved fragments (Fig. 3(a)), we can recover the digital ink (Fig. 3(b)). The recovered digital ink can be recorded on the original document (such as PDF) as shown in Fig. 3(c). Although there is still some jagged noise on the recovered digital ink, it looks similar to the annotation written on the paper document shown in Fig. 3(d).

3. Improvement of LLAH

This section describes the improvement of LLAH that enables us to track the pen tip from a sequence of camera-captured images.

3.1. Problems of the use of LLAH

LLAH is a method of fast and accurate retrieval of document images taken as queries camera captured portions of printed documents. It was first proposed for documents written in Latin scripts [4] and recently extended to be applicable to many other scripts [5]. Since it is intended to retrieve documents, the size of regions acceptable as queries are from whole pages down to about 1/8 pages.



(a) Captured image (b) Centroids of connected components

Figure 4. Feature points extracted from a captured image.

As discussed later in detail, the size of regions captured by the pen camera is quite limited due to the restriction of its position on the pen. Typically only 4-6 text lines can be captured by the pen camera. This is clearly much smaller than the size required by LLAH. The application of LLAH to the camera pen, therefore, needs its improvement.

3.2. Feature points

LLAH indexes documents using feature points extracted from their images. For documents written in Latin scripts, centroids of word regions are employed as feature points for the original retrieval purpose. Their local arrangements defined with nearby 6–10 points are unique enough to distinguish a feature point among those whose number is more than 10 million. The indexing and matching of feature points are not perfect and thus not all feature points are correctly matched. Nevertheless documents can be retrieved with high accuracy because of the effect of voting by matching of many points. However, images captured by the pen camera includes word regions (feature points) whose number is about 10–20, which is too small to be beneficial from the voting effect. In other words, we need more feature points for a successful retrieval.

In order to solve this problem, we employ not word regions but connected components from which feature points are extracted as their centroids. An example is shown in Fig. 4. As shown in this figure, connected components allow us to obtain many feature points required for the retrieval.

3.3. Indexing

In LLAH, each feature point is indexed using multiple feature vectors calculated from surrounding feature points.

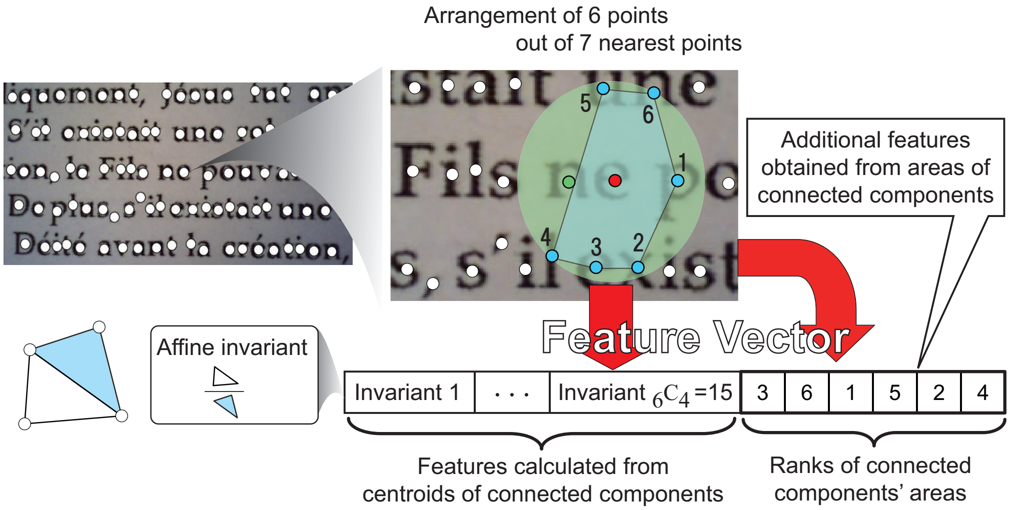


Figure 5. Calculation of feature vectors.

These feature vectors represent local arrangement of points. Figure 5 shows the overview of the calculation.

The central point surrounded by the points 1–6 is a point of interest to which a feature vector is now calculated. First, note that the point to immediate left of the point of interest is not included in the feature calculation. This is to cope with disappearance of points: in the case of Fig. 5, we take 6 points out of 7 points nearest to the point of interest. This enables us to have a feature vector robust to disappearance of one point. Since we have 7 ways of selecting 6 points, the point of interest is indexed with 7 feature vectors.

In general, for each point of interest p, we first take n points nearest to p and then combine $m(\leq n)$ points out of n points to extract m feature vectors for indexing p.

Details of the feature vectors are as follows. Let us go back to the example shown in Fig. 5. Feature vectors for indexing should be invariant to geometric distortion caused by the pen camera. In LLAH, this is fulfilled by using a geometric invariant called "affine invariant". Although the geometric distortion by the camera is beyond the class of affine transformation, the distortion can be well approximated by it. As an affine invariant, we employ the area ratio of two triangles shown on the lower right corner of Fig. 5.

With the affine invariant, the feature vector is defined as follows. For each combination of m points, we take all combinations of 4 points out of m points. From each of 4 point combinations, we calculate the affine invariant. Thus the dimension of the feature vector is $_mC_4$ (no. of combinations of 4 out of m). In the case of Fig. 5, a 15 dimensional vector is the feature vector for $m = 6$. Although values of the affine invariant are real numbers, they are quantized to define the feature vector. This is the fundamental way of indexing by LLAH.

The above features are, however, not sufficient for distinguishing a tiny fragment image. Thus for the purpose of this research, we need more discriminative features. In order to complement the feature vector, we add new features at some dimensions as shown in Fig. 5. Additional features are obtained by taking into account the areas of connected components. To be precise, ranks of the areas of connected components are newly added features, since the ranks are fairly stable with the variation of images.

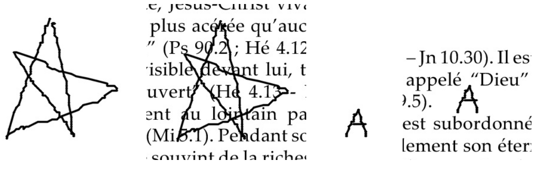
The total number of dimensions of a feature vector is 21 $(= 15+6)$ for the case of $n = 7$ and $m = 6$. The number of feature points is about 3K–4K / page, and thus the number of feature vectors for indexing one page is about 20K – 30K.

3.4. Retrieval

The fundamental computation of the retrieval is shared by the indexing. Each captured image is employed as a query to the database. First, feature vectors are calculated for all feature points extracted from the captured image. Each feature vector is utilized to find its nearest neighbor (NN) from the database. Because the number of feature vectors stored in the database is so large that the sequential matching is prohibitive. In order to solve this problem we employ a hash table for storage and retrieval of feature vectors. Although the retrieved feature vector may not be the "real" nearest neighbor, the retrieval with the hash table is extremely fast.

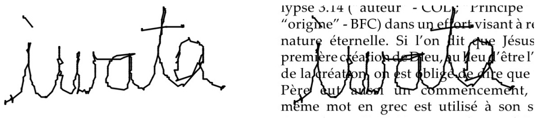
As a result of NN search, each feature vector from the query corresponds to a feature vector in the database. This point-wise correspondence is considered as a vote for a document. The document which has the maximum votes is regarded as the result of retrieval.

Once the retrieved document is determined, the next step is to find the region to which the query image corresponds in the document. This is achieved by using again the point-wise correspondence to the document. The goal here is to



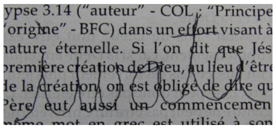
(a) (b) (c) (d)

Figure 6. Examples of simple annotations.



(a) Digital ink

(b) Superimposed annotation



(c) Written annotation

Figure 7. Captured signature.

estimate the homography using the point correspondence. We apply RANSAC [1] to make the estimation reliable.

As a result of estimation, the location of the pen tip is estimated as a fixed position in the region projected to the document.

4. Experiments

We have implemented the prototype system with the camera pen and the improved LLAH which enables us to capture digital ink.

In order to evaluate whether or not the system proposed in this paper works properly, we used it for annotating some documents. In the database of documents, we stored 100 printed pages. The camera pen captured images with the rate of 20–25 fps.

As the annotations, simple line drawings including straight lines, triangles, rectangles, circle, pentagrams, etc., as well as capital alphabets were employed. Examples are shown in Fig. 6, where (a) and (c) are recovered digital ink and (b) and (d) are document images with digital ink. We also tried to write a simple signature as shown in Fig. 7.

As shown in these figures, the system works well not only for the annotation on printed regions but also for the annotation on margins of documents, if they are not so wide. An obvious drawback, on the other hand, is the jagged on the recovered digital ink. This is simply because the retrieval is done independently on each frame of images. Thus this problem would be solved by taking into account the continuity between adjacent frames. The matching of digital ink to camera captured annotation written on a document would also provide a good clue to make the digital ink smooth.

5. Conclusion

We have presented a prototype system of a camera pen which works on ordinary paper to recover the digital ink. The characteristic point of our method is the use of foreground patterns on documents as landmarks for locating the pen tip. The novel algorithm called LLAH is effective to find the location in real-time. From the preliminary experiments, we have confirmed the effectiveness of our framework, though we still have room for further improvement. One is to make the digital ink smoother. Another open problem is to extend the method to work on blank paper. They will be solved in the future work.

Acknowledgment

This work was supported in part by the Grant-in-Aid for Scientific Research (B)(19300062, 20300049) and the Grant-in-Aid for Challenging Exploratory Research (21650026) from Japan Society for the Promotion of Science(JSPS).

References

[1] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–395, 1981.

[2] X. Liu and D. Doermann. Mobile retriever — finding document with a snapshot. In *Proc. of the 2nd CBDAR*, pages 29–34, 2007.

[3] M. E. Munich and P. Perona. Visual input for pen-based computers. *IEEE Trans. PAMI*, 24:313–238, 2002.

[4] T. Nakai, K. Kise, and M. Iwamura. Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. In *Lecture Notes in Computer Science (7th International Workshop DAS2006)*, volume 3872, pages 541–552, 2006.

[5] T. Nakai, K. Kise, and M. Iwamura. Real-time retrieval for images of documents in various languages. In *Handout of 8th IAPR International Workshop on Document Analysis Systems (DAS2008)*, page 37, 2008.

[6] J.-H. Seok, S. Levasseur, K.-E. Kim, and J.-H. Kim. Tracing handwriting on paper document under video camera. In *Proc. of Int'l Conf. on Frontiers in Handwriting Recognition*, 2008.