Tracking and Retrieval of Pen Tip Positions for an Intelligent Camera Pen

Kazumasa Iwata†, Koichi Kise†, 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 method of recovering digital ink for an intelligent camera pen, which is characterized by the functions that (1) it works on ordinary paper and (2) if an electronic document is printed on the paper the recovered digital ink is associated with the document. Two technologies called paper fingerprint and document image retrieval are integrated for realizing the above functions. The key of the integration is the introduction of image mosaicing and fast retrieval of previously seen fingerprints based on hashing of SURF local features. From the experimental results of 50 handwritings, we have confirmed that the proposed method is effective to recover and locate the digital ink from the handwriting on a physical paper.

1. Introduction

Bridging the gap between paper and electronic documents is one of the important tasks in the field of document image analysis. This is because users cannot stop using both: paper documents have advantages on readability and portability, while electronic ones have different and complementary advantages such as searchability and less storage space. In this paper, we propose a solution to this problem focusing on dealing with handwriting. We propose an intelligent camera pen which allows the user to handwrite on a document printed on an ordinary paper and at the same time to keep the digitized equivalent of handwriting (digital ink) to the right position of the corresponding document.

There exist many attempts for this goal. The most important example is a camera pen called Anoto pen [2] which allows us to write on paper to obtain digital ink. Although the pen is easy to use and robust to various writing conditions, a serious limitation is the requirement of special paper called Anoto paper, which is with fine dots printed on the surface of paper for identifying locations of the pen tip. In addition, such dots spoil the appearance of paper, and it generally costs a lot, all of which result in limiting its applicability.

To solve this problem, it is necessary to have a digital pen which does not rely on the preprinted dots. A possible way is to find their substitution from documents the user is writing on. In the process of recovery of digital ink, the following requirements should be fulfilled: (1) Recovered digital ink needs to be associated with the document the user is working on. (2) Recovery should be enabled independently of the foreground: the pen should work on parts with the printed foreground as well as without it.

We have already developed camera pens aiming at satisfying the above requirements. Uchida et al. employ microscopic dimples on the surface of paper, or *paper fingerprints* for *tracking* the pen tip motion [8]. This technology is capable of satisfying the second requirement, though the first one has not been solved. Iwata et al. utilize the foreground (printed characters) of documents as a clue to find the location of the pen tip. A method of *retrieval* of document images called LLAH (Locally Likely Arrangement Hashing) [7] is employed for this purpose. Dots extracted from characters are indexed and stored in a document image database. The camera pen captures character images from which the indices of dots are calculated as well to know which dots in the database the pen is looking at. This technology fulfills the first requirement, though cannot work on blank paper.

In this paper, we integrate these two methods to satisfy both requirements simultaneously. The largest obstacle of the integration is that the imaging area required to extract features is far different with each other. In the tracking method, the camera must be placed close to the surface, since the technology requires microscopic structure of the paper surface. On the other hand, the retrieval method needs a much broader area to be captured for obtaining more "dots" for finding the location. In order to solve this problem, we employ an image mosaicing technology. We place the camera close enough to the paper surface for tracking and accumulate images for mosaicing. After an enough area is captured, the mosaicing is invoked to obtain a larger image, which is then employed for the retrieval. Another problem we have solved is to handle *reappearance* of previously seen parts in the document. It should be correctly recognized for accurate recovery of digital ink. We employ a fast hashing method of image features for this purpose. We evaluate the above method based on experimental results obtained using a prototype camera pen.

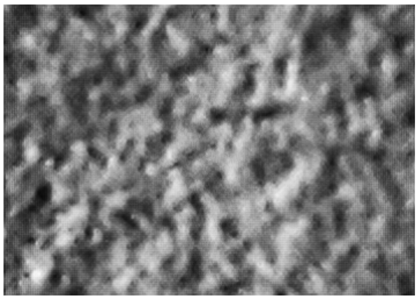


Figure 1. Paper fingerprint.

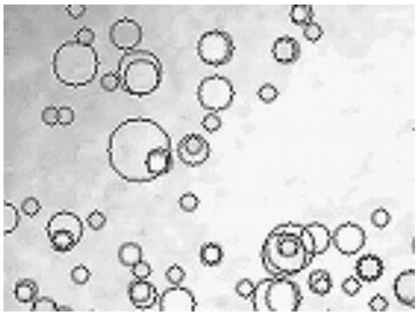


Figure 2. SURF keypoint regions extracted from the surface of blank paper (enlarged).

2. Related work

2.1. Commercial digital pens

We have already had several types of commercial products for supporting the recovery of digital ink such as electronic pen tablets, ultrasonic pens and the Anoto pen. Except for the Anoto pen, most existing technologies focus on the recovery of digital ink without the relation to the document. This spoils the usability of digital pens since manual labors are required to associate the digital ink with the document.

The Anoto pen is the first digital pen which has achieved the breakthrough to this problem. This is the main reason why it has attracted many researchers. However, for locating the digital ink onto the document, it is required to use special paper called Anoto paper, which causes new disadvantages.

2.2. Paper fingerprint

In order to solve the problem of the Anoto system, we have developed two camera pens. One is the camera pen which captures "paper fingerprint". Figure 1 shows an example of fingerprint: there is microscopic structure of an uneven surface of paper. Since it is almost impossible to have the same structure on different sheets of paper as well as in different parts of the same sheet, it can be used as a feature that represents the location in the sheet.

The concrete processing is as follows. In this method, an image feature called SURF (Speeded Up Robust Features) [1] is extracted from the surface image as shown in Fig. 2, where each circle represents a keypoint region from which a 64 dimensional feature vector is extracted. Since the feature vector has a high discrimination power, a simple nearest neighbor search allows us to obtain almost correct matching between keypoints. An example is shown in Fig. 3. By applying RANSAC to find out the homography from a set of matched keypoints, we can estimate a current pen tip position on the coordinates of the first frame.

Although this method is available even when writing on a blank part of paper, or a new sheet of paper, it is impossible for the method to associate the recovered digital ink to the document on which the pen is, unless fingerprints of all sheets of paper have been scanned and associated with documents beforehand. Since scanning fingerprints in advance is impractical, the method cannot fully solve the problem.

2.3. Document image retrieval

Another trial by us for solving the problem is to employ document image retrieval to know which document and where in it the camera captures [3, 5, 4]. As the method of retrieval, we utilize LLAH [7]. The process of capturing the digital ink is as follows. (1) Image cap-

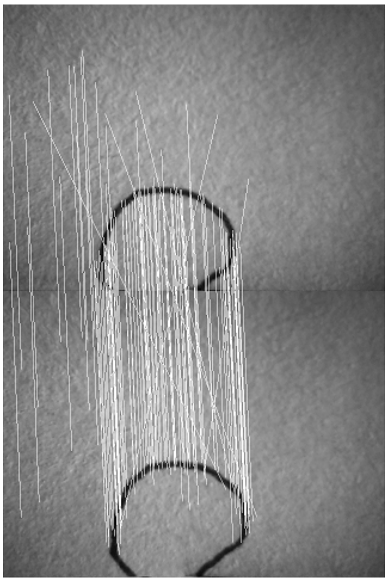


Figure 3. Matching of SURF keypoints.

ture: Capture an image of a part of the document the pen is on, (2) Retrieval: Extract centroids of connected components from the image, and use them as queries to the database. As a result, point matching between each point in the image and its corresponding point in the database is obtained. This allows us to calculate the homography. (3) Pen tip positioning: Based on the homography, calculate the coordinates of the pen tip, (4) Verification: Evaluate the coordinates and record them if there is no problem. (5) Repetition: Repeat the process of (1)-(4) and connect estimated coordinates of the pen tip to obtain the digital ink.

All centroids in the database are stored with the document ID and their positions in the document so that the recovered digital ink is located in the document. Since the centroids are extracted from the foreground objects of documents, there is no need to have special dots. However this point is also a disadvantage of the method; when the camera captures a blank part such as margin of the document, no information for the retrieval can be extracted, and thus the retrieval is failed. In order to lower the chance of the failure, a wider area is necessary to be captured. Needless to say, this method is incapable of fully solving the problem as well.

3. Proposed method

3.1. Problems for integration

We have seen that the previous two methods cannot solely solve the problem of camera pens. A simple and a reasonable solution for it is to integrate the above two methods since their pros and cons are complementary. However the solution cannot be easily obtained since the required viewing area is far different with each other: the fingerprint requires the camera position close to the paper surface, while the retrieval needs far camera position to capture a larger area of the document. This problem can be solved by using two different cameras, i.e., one is for the fingerprint and the other is for retrieval. However it is not a good idea since it increases the cost. Thus we need a different solution for the viewing area problem.

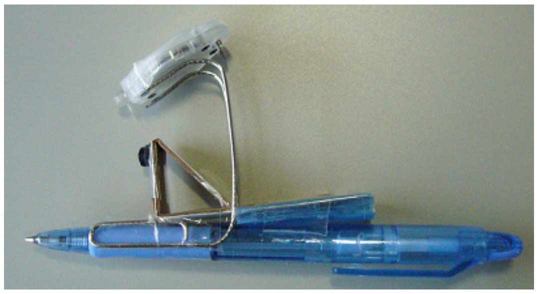


Figure 4. Camera pen.

Another problem we encounter for the integration is that errors of the position estimation by the fingerprint method are easily accumulated. Thus, the digital ink becomes different from the handwriting on the paper. For example, when a circle is drawn, the beginning part and the ending part must be connected. However it is not always possible just only by following the recovered digital ink. This problem is called *reappearance*, since the method cannot recognize the previously seen same part.

3.2. Solution for the camera position

The proposed method is to solve the above two problems for integration as follows.

For the first problem we employ image mosaicing. It is not necessary for us to retrieve the pen tip position at each image frame. This allows us to postpone the retrieval until an enough area of the document has been captured. The details are as follows.

First, the camera is mounted close enough for the paper fingerprint. The real position of the camera is shown in Fig. 4, where the black part at the corner of the triangle is the camera module and the white part above the camera is an ultraviolet light for observing the paper fingerprint.

Next, an image mosaicing technique is applied to combine frame images into one large image, which is

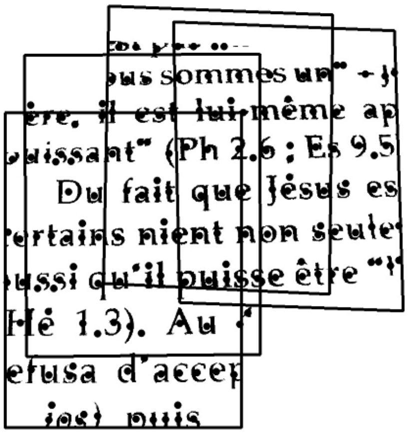


Figure 5. LLAH feature points extracted from a mosaic image.

then utilized for the retrieval. Figure 5 illustrates an image mosaicing result obtained by combining four frame images. Dots in the figure represent centroids of connected components.

3.3. Solution for the reappearance

In order to solve the problem of reappearance, it is necessary for the method to recognize the same part of the document even if it is not in the consecutive frames. This is realized by recording all SURF keypoint vectors in a database and applying the matching of each keypoint to the previously seen keypoints in the database. By using this function, the method is capable of detecting the known keypoints to cancel the accumulated errors of the estimation of homography. A typical example is the character “8” when the pen crosses the previously written stroke.

The next question is how to do such an extensive matching efficiently. Fortunately we can apply a method based on hashing local features [6]. It has been proved in [6] that the method enables us to match 200 feature vectors of 36 dimensions to the database of 2.6 billion feature vectors in 60 ms. Since the feature vectors extracted from a single document image is far less than the above case, there would be no problem.

The details of the processing is as follows. First, feature vectors obtained by SURF is converted into binary vectors of 16 dimensions. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_{64})$ be an original SURF feature vector. The corresponding binary vector $\boldsymbol{u}= (u_1, \ldots, u_{16})$ is obtained as follows:

$$u_j = \begin{cases} 1 & \text{if } x_j - \mu_j \geq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad (1)$$

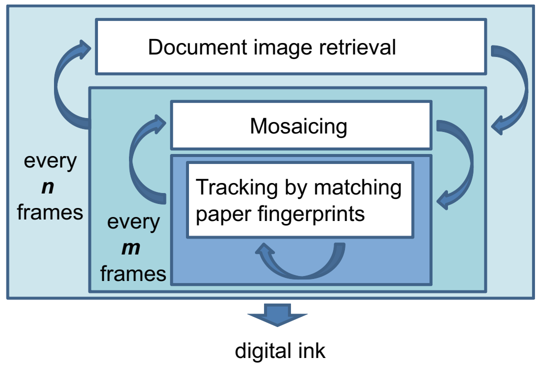


Figure 6. Processing flow.

where μ_j is the median of values of the dimension j, which is predetermined using learning samples. Then the hash index to store the feature vector is obtained by

$$H_{\text{index}} = \left(\sum_{i=1}^{16} u_i 2^{i-1} \right) \bmod H_{\text{size}} \qquad (2)$$

where H_{size} is the size of the hash table. In case that multiple feature vectors have the same hash index, they are stored by the chaining method.

At the retrieval phase, the hash index is first calculated in the same way stated above for each SURF feature vector from the captured image. Then the index is utilized to access to the hash table. In order to cope with variations of feature vectors, a multi-probe strategy is employed. Multiple bit vectors close to the original one are generated as indices for a single feature vector and employed for the access to the hash table. Finally, a set of feature vectors are obtained as the union of sets of feature vectors obtained by multiple bit vectors, and within the union the feature vector nearest to that from the captured image is determined simply by calculating the Euclidean distance.

3.4. Processing flow

The above functions are put together to make the proposed method. Figure 6 illustrates the overall processing in the method. For each frame, the tracking process is applied to obtain the digital ink. At every $m(> 1)$ frames, mosaicing is employed for combining frame images. The process of document image retrieval is then executed at every $n(> m)$ frames on condition that the number of centroids exceeds a threshold. If the retrieval is successful, the recovered digital ink is placed onto the document by using the estimated homography. Mosaicing is initialized after the application of retrieval for avoiding the accumulation of errors.

4. Experiments

4.1. Conditions

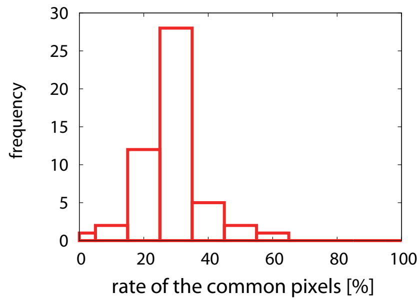
The proposed method was evaluated using 50 handwritings on documents printed on recycled paper. The maximum frame rate of the camera is 30 [fps]. The mosaicing is applied at every 10 frames ($m = 10$), and the retrieval is executed at every 30 frames ($n = 30$). Since the frame rate of the camera is not high, the speed of writing needs to be slow for avoiding the motion blur. To be precise, a character is written in 5-10 seconds.

The quality of the recovered digital ink was evaluated using the groundtruth obtained by a pen-tablet. While handwriting with the proposed camera pen, the trajectory of the pen tip was also recorded by the pen-tablet. The recovered digital ink is evaluated by the rate of common pixels to the pixels of groundtruth. Since the groundtruth includes a displacement due to the nature of the pen-tablet, we matched the pixels with the tolerance of 3 mm. This means that the recovered digital ink was moved within 3mm on the coordinates of groundtruth, for finding the best match before measuring the rate.

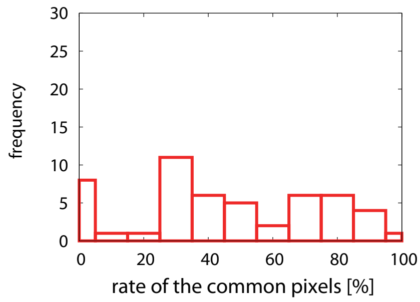
4.2. Results and discussions

Figure 7 shows the results of evaluation. Figure 7(a) is the result obtained by applying solely the tracking. We were capable of obtaining the rates about 30%, because the tracking only enables us to find the coordinates not of the groundtruth but of the first frame. Thus if the first frame is not upright, the digital ink is deformed as shown in Fig. 8. Figure 7(b) shows the result by both tracking and retrieval. The rate was improved since the pen tip was located on the coordinates of the document as a result of retrieval.

An example of successful cases is shown in Fig. 9. This is the case that the pen tip moved from a part with document foreground to a blank part. Thus with the document image retrieval alone it is difficult to recover the digital ink correctly. In the proposed method, the functions of tracking and retrieval enabled us a correct recovery. Another example of successful cases is in Fig. 10, where only a slight deformation is observed. On the other hand, we also have various erroneous cases. The rate of 0 % was due to the highlight clipping (over exposure). If the image is under this condition, (1) tracking is impossible and (2) retrieval is failed since no centroids are obtained from the image. Another typical example is the lack of captured area. An example is shown in Fig. 11. Since the handwriting is horizontally long, only a limited number of centroids are obtained in the vertical direction. This deteriorates the precision of location estimation by the retrieval. In order to solve this problem it is necessary to improve the discrimination power of the retrieval.



(a) Digital ink obtained only by tracking



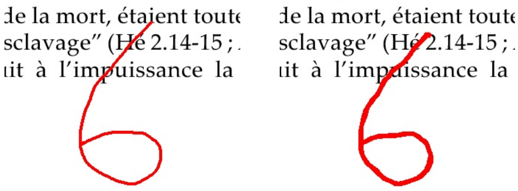
(b) Digital ink obtained by tracking and retrieving

Figure 7. Evaluation of digital ink.

(a) Groundtruth

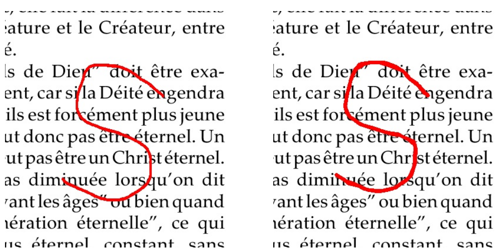
(b) Recovered digital ink

Figure 8. Digital ink recovered without the retrieval and its corresponding ground truth.



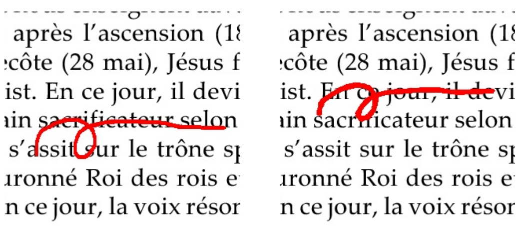
(a) Groundtruth (b) Recovered digital ink

Figure 9. Handwriting on both printed and blank regions.



(a) Ground truth (b) Recovered digital ink

Figure 10. An example of successful cases.



(a) Ground truth (b) Recovered digital ink

Figure 11. An example of erroneous cases.

5. Conclusion

In this paper, we have proposed a method of recovery of digital ink for an intelligent camera pen which can deal with the handwriting on ordinary paper. By integrating two fundamental components of tracking and retrieval, we have solved the problems of locating the recovered digital ink onto the coordinates of the document and of supporting the handwriting on a blank part of the document. The key for the integration is twofold: mosaicing and matching keypoints. From the results of fundamental experiments, we have shown that the camera pen works properly both on printed and blank parts.

The future work to be explored includes the improvement of the accuracy of retrieval as well as to increase the speed of acceptable handwriting. The latter may be enabled by using a higher speed camera.

Acknowledgment

This work was supported in part by the Grant-in-Aid for Scientific Research (B)(19300062, 20300049, 22300062) and the Grant-in-Aid for Challenging Exploratory Research (21650026) from Japan Society for the Promotion of Science(JSPS).

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, 2008.

[2] http://www.anoto.co.jp/.

[3] K. Iwata, K. Kise, T. Nakai, M. Iwamura, S. Uchida, and S. Omachi. Capturing digital ink as retrieving fragments of document images. *Proceedings of the 10th International Conference on Document Analysis and Recognition*, pages 1236–1240, 2009.

[4] K. Kise, M. Chikano, K. Iwata, M. Iwamura, S. Uchida, and S. Omachi. Expansion of queries and databases for improving the retrieval accuracy of document portions: An application to a camera-pen system. *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, pages 309–316, 2010.

[5] K. Kise, K. Iwata, T. Nakai, M. Iwamura, S. Uchida, and S. Omachi. Document-level positioning of a pen tip by retrieval of image fragments. *Proceedings of the Third International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2009)*, pages 61–68, July 2009.

[6] K. Kise, K. Noguchi, and M. Iwamura. Robust and efficient recognition of low-quality images by cascaded recognizers with massive local features. *Proceedings of the 1st International Workshop on Emergent Issues in Large Amount of Visual Data (WS-LAVD2009)*, pages 2125–2132, Oct. 2009.

[7] T. Nakai, K. Kise, and M. Iwamura. Real-time retrieval for images of documents in various languages using a web camera. *Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR2009)*, pages 146–150, July 2009.

[8] S. Uchida, K. Itou, M. Iwamura, S. Omachi, and K. Kise. On a possibility of pen-tip camera for the reconstruction of handwritings. *Proceedings of The Third International Workshop on Camera-Based Document Analyais and Recognition*, pages 119–126, 2009.