Pattern Recognition Letters xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

# Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval

Megumi Chikano [a], Koichi Kise [a,*], Masakazu Iwamura [a], Seiichi Uchida [b], Shinichiro Omachi [c]

[a] Department of Computer and Systems Sciences, Graduate School of Engineering, Osaka Prefecture University, Sakai, Japan
[b] Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
[c] Graduate School of Engineering, Tohoku University, Miyagi, Japan

ARTICLE INFO

*Article history:*
Available online xxxx

*Keywords:*
Camera-pen
Document image retrieval
LLAH
SURF
LK tracking
Handwriting

ABSTRACT

We propose a camera-based method for digital recovery of handwritings on ordinary paper. Our method is characterized by the following two points: (1) it requires no special device such as special paper other than a camera-pen to recover handwritings, (2) if the handwriting is on a printed document, the method is capable of localizing it onto an electronic equivalent of the printed document. The above points are enabled by the following processing. The handwriting is recovered by the LK tracking to trace the move of the pen-tip. The recovered shape is localized onto the corresponding part of the electronic document with the help of document image retrieval called LLAH (locally likely arrangement hashing). A new framework for stably estimating the homography from a camera-captured image to the corresponding electronic document allows us to localize the recovered handwritings accurately. We experimentally evaluate the accuracy, processing time and memory usage of the proposed method using 30 handwritings. From the comparison to other methods that implement alternative ways for realizing the same functionality, we have confirmed that the proposed method is superior to those other methods.



## 1. Introduction

Even with a modern digital mobile environment, we still continue to use a classical method of recording information, i.e., handwriting by a pen on ordinary paper, when we write notes for recording ideas, comments on documents at a meeting, and so on. Although it is quite easy to produce handwritings by using the classical method, it is generally troublesome to convert the resultant handwritings to digital data for later use of editing and sharing; it requires at least scanning and recognizing the handwriting. Thus it is advantageous to have a mean to digitize the handwriting automatically while keeping easiness and simplicity of the classical method.

For achieving this goal, many methods and systems have been developed. A successful example is the Anoto system (http://www.anoto.com/) which enables us to digitize the handwriting by using a camera mounted on a pen and special paper with a number of fine dots on its surface. The uniqueness of local point distribution enables the system to find the absolute position, i.e., the information on which sheet of paper and where on the sheet the pen is working. The system can recover the handwriting by tracing the absolute position of the pen-tip. Although this system is in practical use, its drawback is the requirement of special paper.

The above problem can be solved by realizing the same functionality without special paper, which can be decomposed into the two processes shown in Fig. 1. One is to trace the pen-tip movement without using special paper. We call it "recovery" of handwritings. In addition, as a replacement of knowing the absolute position, it is required to relate handwritings to known documents. Suppose the case that the user is writing on a printed page of a document whose electronic version is also available. In this case the user would like to reflect the handwritings to the corresponding positions of their electronic equivalents. We call it "localization" of recovered handwritings. Needless to say, this should also be done without the use of special paper.

As a trial to realize the above functionality, we have proposed a method that employs both the paper fingerprint, i.e., microscopic structure of paper surface, and printed patterns on documents (Iwata et al., 2010). In this method, the paper fingerprints are traced by using the tracking of SURF features (Bay et al., 2008) (SURF tracking). In addition, printed patterns are used as a clue to find the corresponding electronic document as well as to find the location at which the handwriting is generated. Although this method requires no special paper there still remains a problem of accuracy; the shape of recovered handwriting is not accurate enough and sometimes fails to locate the recovered handwriting on the document.

* Corresponding author.
*E-mail address:* kise@cs.osakafu-u.ac.jp (K. Kise).

0167-8655/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2012.10.003

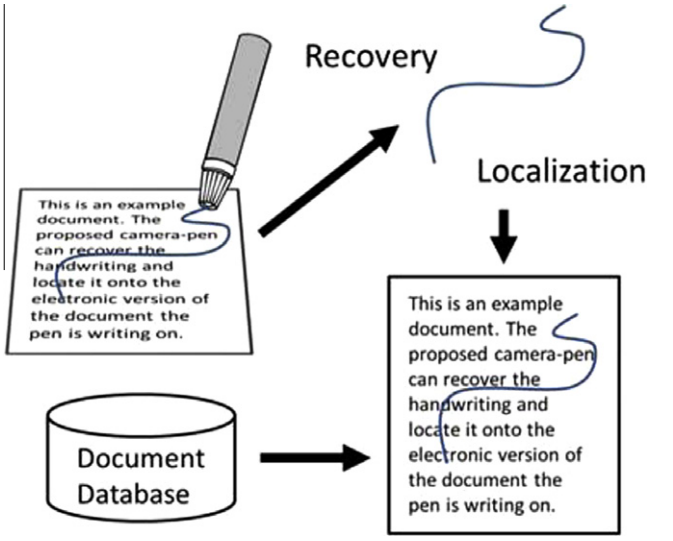



**Fig. 1.** Processes required for camera-pens.

In this paper, we propose a new method that overcomes the above problems. For the improvement of handwriting recovery, we introduce a tracking based on Lucas-Kanade Method (LK) (Lucas et al., 1981) (LK tracking) instead of the SURF tracking. This allows us to improve the stability of tracking so that the shape can be recovered more smoothly. Accuracy of document image retrieval is also improved by the query expansion (Kise et al., 2010) as well as a new framework of estimating the homography from a camera captured image to the retrieved document image. This is achieved by unifying the result of tracking and that of retrieval. Experimental results on 30 handwritings, we have confirmed that the proposed improvements are promising as compared to our previous method (Iwata et al., 2010) as well as other methods that implement possible alternatives for realizing the same functionality.

The main contribution of this paper is the proposal of an overall framework for camera-pens without special paper. In our previous trials (Kise et al., 2009, 2010; Iwata et al., 2009, 2010; Uchida et al., 2009) we have examined different technologies. This paper shows the best combination of fundamental technologies including newly introduced such as the LK tracking. More importantly, we propose a new method of stably estimating the homography which is mandatory to the localization. The technological breakthrough achieved by the above points enables us to improve the accuracy of recovered handwritings significantly. This has been confirmed by the experimental results with a much larger number of handwritings compared to the previously reported results (Kise et al., 2009, 2010; Iwata et al., 2009, 2010; Uchida et al., 2009).

In order to make use of the proposed method as a part of a camera-pen system, it is also required to judge the state of pen up and down. In addition, as an important application, the recovered handwritings should be recognized if they are characters. However, this paper does not deal with these issues due to the following reasons. For the former point, the state of pen up and down is not necessarily determined by only using image processing. Rather it would be easier to use a micro switch to sense the touch to a paper surface by the pen-tip. For the latter point, the discussion can be safely separated since existing technologies of on-line handwriting recognition can be applied once the handwriting has been recovered.

The organization of this paper is as follows. In Section 2, we review the existing methods to clarify the problems to be solved. Section 3 is devoted to describe the goal of the proposed camera-pen system. In Section 4, we briefly review the document image retrieval employed in the proposed method. Section 5 is the main part of this paper to describe the details of the proposed method. Experimental results are shown in Section 6. In Section 7 we sum up what we have learned as well as mention the future work.

## 2. Related work

In order to bridge the gap between handwritings on paper and digital media that stores the handwritings digitally, there have been many efforts that can be divided into commercial and research developments.

As commercial systems, we focus here on the following three systems.

The first one is a tablet system that uses electromagnetic induction to capture the pen-tip movement (e.g., http://www.adesso.-com/home/tablets/158cyberpad.html). When this system is used, paper is placed on the tablet and the pen-tip movement is traced by the tablet while the user writes his/her handwriting on the paper. Although the accuracy is high, its portability and simplicity is limited.

The second is a system with a small device that emits the ultrasonic sound and/or infrared light and a special pen (http://www.pegatech.com/). The device measures the reflection from the special pen to know its position. Although it is more portable than the tablet, this has an important disadvantage: since it only measures the relative position between the device and the pen assuming that the device is fixed onto the sheet of paper, the recovery becomes impossible if the position of the device on the sheet changes.

The third is the Anoto system. It employs a camera-pen and special paper with fine dots. It captures the local distribution of fine dots to decode the global position of the pen-tip. This is the most advanced commercial system with respect to the portability and the reliability: only the camera-pen is necessary for users to carry, and no inaccuracy was caused in measuring the global position. However this still has a limitation that it requires special paper; it is not possible for users to write on ordinary paper.

As for research systems, one of the very first systems is PaperLink (Arai et al., 1997). It was proposed to establish the relation between a printed document and electronic data. Although this system works on ordinary paper, it only enables the connection from printed words; no handwriting is allowed.

As systems allowing handwriting, we introduce the following three systems. The first is a signature verification system (Yasuda et al., 2008). This system employs two cameras fixed on frontal and side positions of a sheet of paper for capturing images of the pen-tip. Tracking using the captured images enables to recover the handwriting, which is used for signature verification. The second system is proposed by Munich et al. (2002) to recover handwritings and sketches. In this system, a camera fixed in the environment captures the paper surface as well as the process of writing. Tracking allows us to recover the trajectory of the pen-tip, which is regarded as the handwriting. The third is a system proposed by Seok et al. (2008) which tracks the pen movement on printed documents similarly to the system by Munich.

The advantage shared by the above three methods is that they require no special pen nor paper. On the other hand, the cameras must be fixed in the environment. This spoils their portability. In addition, before using the system, the camera must be calibrated.

Pens equipped with cameras have also been researched.

An example is a camera-pen by tracking paper fingerprints (Uchida et al., 2009). In this method, SURF features (Bay et al., 2008) are extracted from the paper fingerprint and matched between succeeding frames so as to obtain the pen-tip movement. Its advantage comes from the reproductivity of the SURF features.

Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003

If the camera captures the same part of paper, the same features tend to be obtained. This is important, for example, for the case of writing "8"; in order to put the intersection point at the right place, it must be known where the pen goes across. The reproductivity of SURF features help us to recognize the reappearance, i.e., to find the previously seen fingerprint. However this system has a drawback that it can only know a relative pen-tip movement. This means that there is no way to put the digitized handwriting to the right place on an electronic document.

To solve the above problem of localizing the handwriting, we have already proposed a camera-pen using a document image retrieval method (Kise et al., 2009; Iwata et al., 2009). This system takes as query an image capturing a document the user is writing on and retrieves the corresponding electronic document and its captured part. This allows us to estimate the pen-tip position since the relative position between the camera and the pen is fixed. By repeating the retrieval, the pen-tip trajectory is recovered as a digitized handwriting. However, the method has two problems that prevent us from using it. The first problem is the inaccuracy of document image retrieval, which results in the failure of handwriting recovery. The second problem is that it is not possible for this system to recover the handwriting on a blank part of paper.

We have attempted to solve the first problem based on query and database expansion (Kise et al., 2010). The key idea is to make query and database images geometrically closer. The database expansion is a technique to store not only the upright images but also tilted images for better matching. The query expansion is to achieve the same goal in the opposite way. The query image that may be tilted is geometrically rotated to obtain candidates of upright images, which are employed for the retrieval.

For solving the second problem we have tried to incorporate the paper fingerprint to the camera-pen based on the retrieval (Iwata et al., 2010). Since it uses the SURF tracking, it is possible to handle handwritings on blank parts as well as to detect reappearance. In addition, this system employs image mosaicing by combining the SURF tracked sequential images. This allows us to make the query image large enough to retrieve the document image accurately. Up to now this method is the most advanced in our development and thus we call it the baseline method in this paper.

Unfortunately, however, not all problems have already been solved. We still have three major problems.

The most serious problem is inaccuracy of the SURF tracking. It affects the handwriting recovery. Recovered shape is sometimes changed due to the inaccuracy of tracking. It also causes the problem of retrieval. Recall that the query image is constructed by mosaicing based on the tracking. If it is inaccurate, the resultant query image can be far from the real image so that the retrieval fails.

The reason of the inaccuracy of SURF tracking is that the feature matching is error-prone: it is often the case that a certain number of features cannot be or erroneously matched. In particular, if the camera captures both a printed and a blank part, SURF feature points are mainly extracted from the printed part and thus little information is obtained from the blank part. The second problem is about locating the recovered handwriting. Due to inaccurately estimated homographies, the recovered handwriting is partly disconnected. The third problem is the memory consumption. In order to deal with the reappearance, a large number of SURF features must be stored. Especially, in the baseline method, two hash tables are utilized for matching SURF features. One is for short term matching of features between succeeding frame images. The other is for long term matching for reappearance. The use of two hash tables need a huge memory space.

The research described in this paper is devoted to solve the above problems for making our camera-pen system more practical.

## 3. Camera-pen system

Before describing the details of the proposed method, let us show the purpose of the development. The goal is to develop a camera-pen that works on ordinary paper by only using a camera-pen. This allows its user to recover his/her handwriting onto the document the user is working on. This functionality is mandatory since the handwriting on an existing document is strongly related to the document. Especially if the handwriting is characters, it may be OCR'ed later and put the result onto the electronic document.

Such a camera-pen system enables us to enhance the use of handwritten information. For example, if it is recognized it becomes retrievable and editable. Another objective is to enrich a lifelog. Although the purpose of lifelog is to record all activities digitally for later use, it mainly focuses on capturing what the user has seen as a video and tag it using sensory data. Needless to say but our intellectual activities are strongly related to reading and writing. Thus it is required for the lifelog to deal with such activities in a more sophisticated way. If the camera-pen is available, we can log our writing activities. We call such a functionality the "writing-life-log". The ultimate goal is to realize it which enriches our life by making our handwriting retrievable and editable.

## 4. Document image retrieval

One of the important building blocks of the proposed method is the document image retrieval method called LLAH (locally likely arrangement hashing) (Nakai et al., 2009). Thus let us start with its explanation.

### 4.1. Overview of processing

LLAH is a method of large-scale document image retrieval. A hash table is employed to deal with a large number of document image in real time. The method employs feature points called LLAH feature points extracted from printed characters for the retrieval. The index of each LLAH feature point is determined based on a feature vector calculated from a local arrangement of other LLAH feature points. Since the feature vector is highly discriminative, it is possible to obtain point-wise correspondence from each LLAH feature point of a query to that in the database. As a side effect of point-wise correspondence, we can obtain the homography which defines the perspective transformation from the camera captured query to its corresponding image in the database.

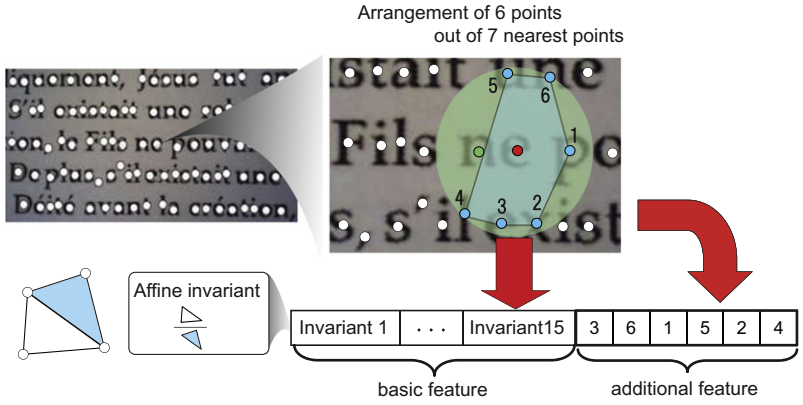
The processing of LLAH consists of two phases: storage and retrieval. In the storage process, LLAH feature points are firstly extracted from all document images. Next, for each LLAH feature point, feature vectors are extracted based on the distribution of surrounding LLAH feature points. Finally, each LLAH feature point is stored in the hash table using the indexes calculated from the feature vectors. In the retrieval process, the same process of calculating the indexes are applied. This allows us to access to the hash table to find the corresponding LLAH feature points.

In the following more details of each phase are described.

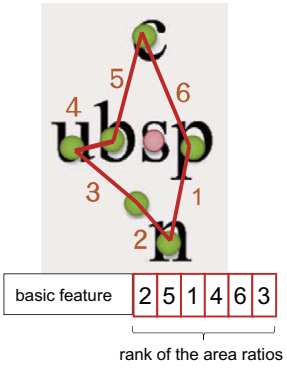
### 4.2. LLAH feature points and feature vectors

For the use of LLAH to the camera-pen it is necessary to distinguish small parts of printed documents. Thus as LLAH feature points we utilize centroids of connected components. In order to achieve high stability of the feature vectors under geometric distortion, LLAH employs a geometric invariant to define the feature.

Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003




(a) Basic feature




(b) Additional feature

**Fig. 2.** LLAH feature vector.

Fig. 2 represents the feature used for the camera-pen. The feature vector consists of two parts: the basic feature and the additional feature.

Let us first explain the basic feature. Since the perspective distortion can be approximately represented by affine transformation, we employ an affine invariant, the area ratio defined as the area of two triangles as shown in Fig. 2(a). In order to make the feature discriminative we focus on local arrangement of LLAH feature points. For each LLAH feature point, we employ its nearest six points and describe the distribution of six points by using all combination of four points out of six. This results in a 15 dimensional feature.

In order to increase the discrimination power of the feature, we employ an additional feature defined using the area ratio of two connected components as shown in Fig. 2(b). For example, the area ratio 1 represents the ratio of the area of 'p' and that of 'n'. The feature is the rank of the area ratios. For the example shown in Fig. 2(b), 2 is the largest area ratio and 3 is the smallest.

### 4.3. Storage and retrieval

In the storage process, each LLAH feature point extracted from a document image is multiply indexed using feature vectors with the information on document ID, point ID and the feature vector. The index is obtained by applying a hash function that converts a quantized feature vector to an integer. The collision of the hash table is resolved by using the chain method.

In the retrieval process, the same process is applied to obtain the hash value from a feature vector extracted from the query image. The hash table is accessed using the feature vector to obtain the chain. Then the method casts a vote to the document image which has the same quantized feature vector in the chain as the one from the query. Finally the document image with the maximum votes is regarded as the result.

As a byproduct of matching we can obtain the point-wise correspondence between LLAH feature points in the query and those in the database image. Although the point correspondence includes noise matches RANSAC (Fischler et al., 1981) is employed to obtain the homography between the query and the database images robustly. The region captured by the query is estimated by transforming the four corners of query image into the corresponding four corners int the database image.

## 5. Proposed method

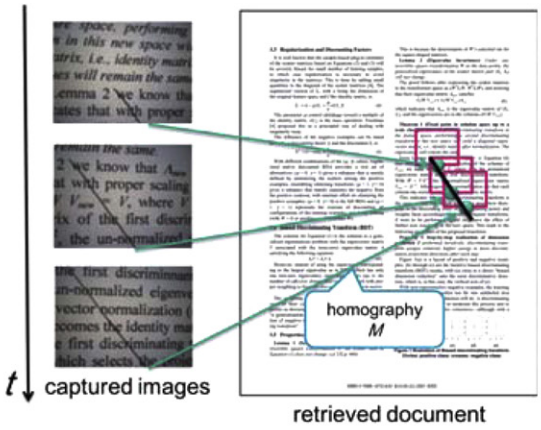
In this section, we first show our approach of designing the camera-pen for solving the problems stated in Section 2. Then the details of the proposed method are described.

### 5.1. Possible approaches

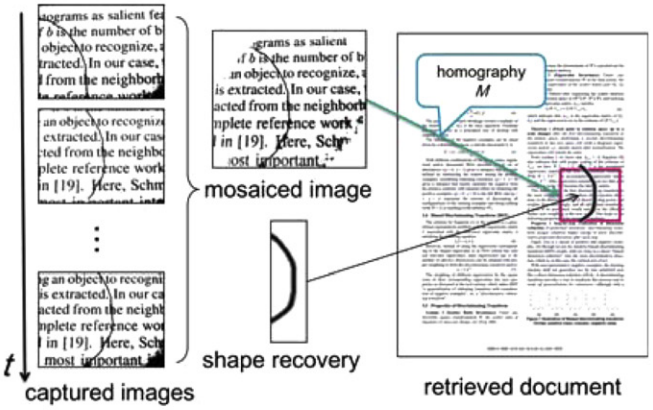
Approaches to the processing of a camera-pen can be characterized by the two aspects: handwriting recovery and handwriting localization. The former is the recovery of shape of handwriting without considering where to localize it. The latter is to find its location on the corresponding document.

The simplest approach is to realize both by solely using the document image retrieval. In this approach, as shown in Fig. 3(a), the handwriting is recovered by repeating the document image retrieval to find a sequence of pen-tip positions and connecting them to form the handwriting. The homography $M$, which represents perspective transformation from the query image to the retrieved document image is calculated every time the document retrieval is

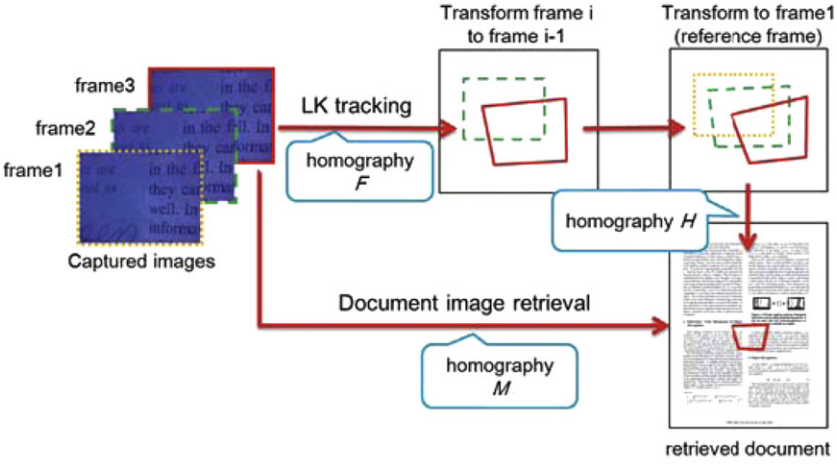
Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003




(a) Approach only using the document image retrieval




(b) Approach by using SURF tracking and mosaicing




(c) Approach by the proposed method

**Fig. 3.** Possible approaches.

applied. This approach was taken by the method described in (Iwata et al., 2009; Kise et al., 2009).

However, this approach cannot always give smoothly recovered handwritings. This is because the retrieval is error-prone: it gives us mostly correct results but sometimes includes errors, which result in deteriorating the quality of recovered handwritings: errors often cause jump or disconnection in a recovered handwriting.

Such errors were often caused by the lack of LLAH feature points due to a small region captured by the camera. Thus the problem can be solved by enlarging the captured region. A solution to the problem has been proposed in (Iwata et al., 2010). As shown in Fig. 3(b), this method does not capture a larger image at a time but to generate it from a sequence of original query images by using image mosaicing. The SURF tracking is employed in the image mosaicing to match succeeding frames. This means that the SURF tracking is responsible for the handwriting recovery, and the handwriting localization is done by the document image retrieval. To be precise, by using the SURF tracking a sequence of

Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003

frame images are transformed into the coordinates of the first frame called the reference frame. After the number of images incorporated into mosaicing becomes enough, the mosaiced image is generated and employed as a query to find the homography $M$ from the mosaiced image to the retrieved document image.

This allows us to locate the recovered handwriting onto the retrieved document image. As described earlier this method is called the baseline to the current proposal.

Unfortunately, however, this improvement is still not enough to recover handwritings smoothly. Although the probability of correct retrieval is improved, it is still imperfect. If an error of retrieval occurs, the situation is even worse than the previous approach since the error means the loss of all frames that are incorporated into the mosaicing. In addition, low stability of SURF features makes the situation more difficult.

The proposed method is designed for solving the same problem in a different way. The baseline method is to improve the accuracy of a single document image retrieval by sacrificing the number of applications of document image retrieval. On the contrary, the proposed method attempts to use the document image retrieval as many as possible to increase the chance of obtaining correct results. This allows us to estimate more reliably the homography to be used for the localization.

The overview of the processing is shown in Fig. 3(c). In the proposed method the recovery is also done by using the tracking. The difference is that it employs a more stable tracking called LK tracking by giving up the recognition of reappearance. Using the LK tracking we can estimate the homography $F$ between the succeeding frames. The estimated $F$'s enable us to map the current frame back to the reference frame to obtain the shape. The localization is also applied to query images as many times as possible if time permits. This allows us to obtain many homographies $M$ from queries to the document image. The key of this approach is the validation step. We employ the fact that the transferred frame by using the homographies $F$ is the same as the frame transferred by the homography $M$. This means that the homographies $H$ between the reference frame and the retrieved document image, which are calculated every time the retrieval is applied, must be identical. As described later, we employ this fact to estimate robustly the homography for localizing the handwriting.

### 5.2. Overview of the processing

The overall processing consists of handwriting recovery and handwriting localization.

The handwriting recovery is based on a sequence of images captured during writing. The LK tracking enables us to obtain the trajectory of the pen-tip, which is regarded as recovered handwriting. Through this process, the shape of handwriting is recovered. The handwriting localization is applied once in every $m(\geqslant 1)$ frames. In order to achieve the localization, we need to know which page of a document and where in the page the handwriting is on. LLAH used in the localization enables us to obtain this information. As a result, the handwriting is localized in the retrieved document image.

In the following, each processing is described in more details.

### 5.3. Handwriting recovery

The process of handwriting recovery is as follows. First, the first frame image is selected as the reference frame. Next, a corner detection is employed to extract feature points from paper fingerprints as well as printed foreground. These feature points are used to calculate the optical flow between succeeding frames. By computing the optical flow, correspondence of feature points between succeeding frames is obtained. Based on this correspondence, the homography between succeeding frames is calculated.

Let $F_i$ be the homography calculated between the frames $i$ and $i+1$. The pen-tip position $(x_p, y_p)$ in the current frame $i+1$ is transformed back to the coordinates $(x'_p, y'_p)$ in the reference frame using $F_i$ as follows:

$$(x'_p, y'_p) = (x_p, y_p)\prod_{k=1}^{i} F_k \tag{1}$$

By using this transformation, all pen-tip positions are represented in the reference frame. Finally, pen-tip positions are connected in the chronological order to recover the handwriting.

### 5.4. Handwriting localization

For the handwriting localization, we employ document image retrieval with the query expansion. After explaining the query expansion, we describe the localization process.

#### 5.4.1. Query expansion

First, the homography $M$ is calculated between the captured image and the corresponding document image. As an expanded query, the query image is converted onto the coordinates of the document image in the database by using the homography $M$. An example is shown as the expanded query 1 in Fig. 4(a).

This converted image may not be upright due to the error included in the homography. In order to cope with it, we further rotate the expanded query 1 into three different ways shown as the expanded queries 2–4 in Fig. 4(a). The generation of the query image, which is a modified version of the method in (Kise et al., 2010), is done as follows. First, define the axes for rotation as shown in the expanded query 1 of Fig. 4(a): one is parallel and the other is perpendicular to the base of the expanded image in the expanded query 1. Note that both intersect at the centroid of the rectangle of the original query. Since the hand of the user is at the base side of the captured image in Fig. 4(a), we select three frequent rotations as shown in expanded queries 2–4 of Fig. 4(a).

An important point of the query expansion is the computational burden. If we directly apply the homography $M$ to convert the captured query image to obtain the expanded query, it requires a lot of time. Since LLAH only needs the centroids and areas of connected components, we calculate them with much lower cost as follows. The centroid of a connected component is transferred by applying the homography $M$. In order to calculate the area in an efficient way we employ an approximate method. As shown in Fig. 4(b), we first obtain a circumscribing rectangle. Then four apexes of the rectangle are transformed by using the homography between the captured query and the expanded query. The area ratio of these rectangles is multiplied to the area of the connected component for the conversion.

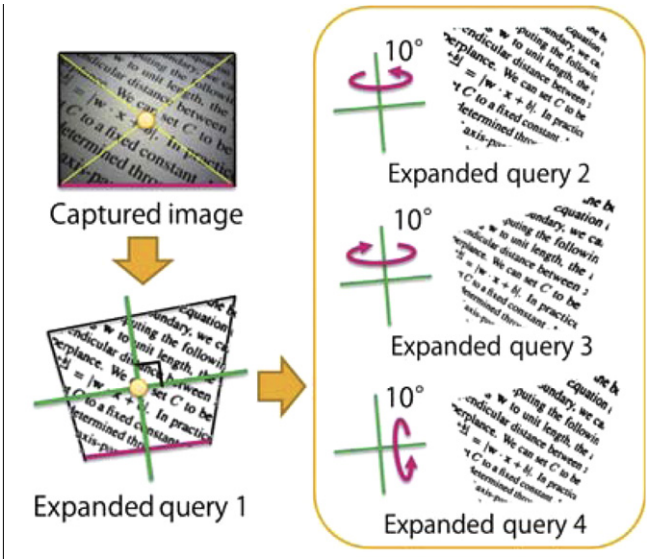
#### 5.4.2. Use of expanded queries for estimating the homography

Every time a query is fed to the document image retrieval, it is expanded and utilized for retrieval. The retrieval result is determined by the query that has the maximum number of corresponding points to a document image in the database.

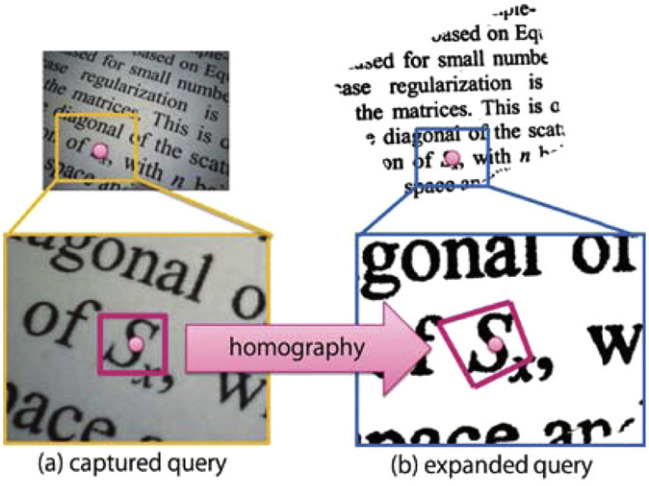
By using the point correspondence we can estimate the homography $M$ that converts the query image to the coordinates of the retrieved document image as shown in Fig. 3(c). In addition, we also know the location of the same query image in the reference frame. Based on the fact that these two images are the same, we can also calculate the homography $H$ in Fig. 3(c), which allows us to locate the recovered handwriting onto the document image.

The problem here is that $H$ is not always stable due to the inaccuracy of document image retrieval. To solve this problem we record the homography every time the retrieval is applied and at the end of handwriting we estimate the plausible homography

Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003




(a) Expansion by rotation




(b) Calculation of the transformed area

**Fig. 4.** Query expansion.

based on all recorded $H$'s. Although they are identical in the ideal case, some outliers are included in the estimated homographies. Thus we exclude such outliers and average the remaining homographies to determine the final plausible homography for localizing the recovered handwriting.

As a method for excluding outliers, we employ the following simple method. For each homography, compute the distance to other homographies in the parameter space of perspective transformation and count how many homographies are included in the hypersphere with a fixed radius from the homography. The homography with the maximum number of other homographies in its hypersphere is employed for the calculation of average: the homographies within the hypersphere are averaged to obtain the final result.

### 5.5. Client–server model

The proposed method is implemented based on the client–server model. Fig. 5 illustrates the timing of each processing in the proposed method. The server plays a role of database retrieval, and the client is responsible for the remaining processing including image capture. At the client side, after capturing a $i$-th frame image, LLAH feature points are extracted from the image and send them to the server for the retrieval. At the server side, the received LLAH feature points are fed to the query expansion and then employed for the retrieval. Then, based on the result of retrieval, the homography $M$ is estimated at the step of homography estimation. The final retrieval result is sent back to the client.

As shown in Fig. 5, the whole processing is parallelized by the client–server model to speed up the processing. In general, the processing at the server side is faster than that at the client side, we insert the waiting at the server side.

## 6. Experiments

### 6.1. Methods for comparison

In order to evaluate the effect of different tracking methods and query expansion, we employed the above four possible combinations. The details are as follows.

1. Method 1 for comparison (baseline method):
Recovery: SURF tracking
localization: image mosaicing + retrieval without query expansion
2. Method 2 for comparison:
Recovery: SURF tracking
Localization: retrieval with query expansion
3. Method 3 for comparison:
recovery: LK tracing
localization: retrieval without query expansion

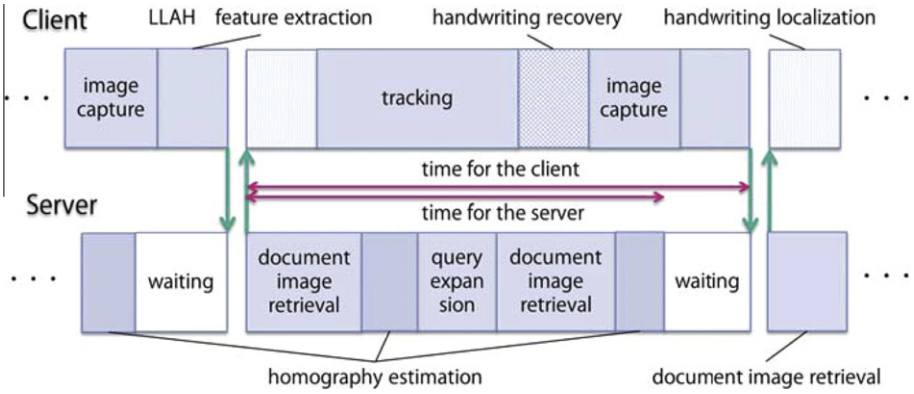



**Fig. 5.** Client–server model.

Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003

4. Proposed method:
Recovery: LK tracking
Localization: retrieval with query expansion

The method 1, which is the baseline method, consists of handwriting recovery with SURF tracking and handwriting localization with image mosaicing. Since the mosaicing of method 1 is troublesome, the method 2 employs handwriting localization with query expansion instead of the mosaicing. In addition, we modified the SURF tracking of method 2 to improve the stability as follows. In order to extract an enough numbers of SURF features even from the mixture of printed and blank parts, the captured image was divided into $48(=8\times6)$ small regions from which SURF features were independently extracted. In addition, we utilized only one hash table so as to reduce the memory consumption. For the purpose of not reducing the accuracy of correspondence of LLAH feature points, we employed a new matching method that takes into account not only the distance of feature vectors itself but also the location of the feature point in the document image. Moreover, in order to make the recovered handwriting smooth, the estimation of the homography $H$ using the average, which was described in Section 5.4.2, was also introduced. In the method 3, the handwriting recovery was with the LK tracking. The localization for this method was kept simple without the query expansion. The estimation of $H$ was also introduced. The proposed method is a modified version of method 3 by introducing the query expansion. It can also be said that the method 4 is a modified method 2 with LK tracking.

In the query expansion of the methods 2 and 4, the expanded query 1 shown in Fig. 4(a) as well as the expanded queries 2–4 in Fig. 4(a) with 10 degrees were employed. In the method 1, the SURF tracking was employed every five frames, the image mosaicing was applied every 25 frames, and document image retrieval every 50 frames where video data was taken 30 frames per second. In other methods, tracking and retrieval were applied every five frames.

### 6.2. Experimental conditions

The database used for the experiments contained 1000 English pages whose image size was $5100\times6600$. In order to obtain the groundtruth while writing, we put a camera onto a pen for a tablet. Several pages were printed out on sheets of A4 paper and used for handwriting on them. The query images were taken during writing on the pages. Note that we did not include neither the case of handwriting on blank sheets nor the case of handwriting without images of printed parts. The size of query images was $640\times480$, the frame rate of the camera was 30 [fps]. During the experiments one writer wrote 30 handwritings. The written objects were English words or simple shapes as shown in Fig. 6. Their sizes were from $5\,\text{mm}\times5\,\text{mm}$ to and $70\,\text{mm}\times15\,\text{mm}$. Since the current implementation is still not able to deal with pen up and down, all shapes and words were of a single stroke. The size of the image used for handwriting recovery was $1700\times2200$.

The results were evaluated using the following three criteria: (1) accuracy, (2) processing speed, and (3) memory usage.

#### 6.2.1. Accuracy

Examples of recovered handwritings are shown in Fig. 6. As shown in this figure, we evaluated the accuracy with two viewpoints: the evaluation solely on the recovered shape, and the evaluation including localization. The former can be done only with the step of handwriting recovery so that the ability of tracking was evaluated. On the other hand, the latter requires the handwriting localized on the correct place of the corresponding page.

The latter is always more difficult than the former and thus the accuracy can be lower. However, even with the failure of localization, there still remains a value of recovered handwriting, since, for example, it can be manually localized to the corresponding document. Thus we evaluated the results based both on these points.

The accuracy was evaluated based on pixel wise comparison between the groundtruth and the recovered shape. For this purpose we first applied the thinning to both of them. Examples of thinned images are also shown in Fig. 6.

Once the scale, rotation and translation of the recovered shape on the coordinates of groundtruth are fixed, we can compare them as follows. From each pixel of the thinned ground truth, find the nearest pixel on the thinned recovered shape. If the Euclidean distance between these two pixels is less than or equal to the tolerance $d$, it is regarded as correct. The accuracy is defined as the rate of the number of correct pixels to the total.

For the case of evaluation of shape recovery, since there was no fixed position, rotation and scale, we explored the possible space and find the parameters that maximizes the accuracy.

For the case of evaluation including the localization, it is not necessary to explore the parameter space. However, since the groundtruth obtained by the tablet tends to be displaced due to the problem of device, we applied a simple compensation for finding the displacement before the evaluation: the recovered handwriting was moved up to $\pm20$ pixels to find the best transformation that gives the highest accuracy. The accuracy was evaluated after this process.

#### 6.2.2. Processing time and memory usage

As the processing time, we measured the time for processing one frame excluding the waiting time shown in Fig. 5. As the memory usage, we measured the maximum amount of memory used for the processing. It changed depending on handwritings so that we show their average as the result. The computers used for the experiments were as follows: The server was with the CPU Opteron 8378(2.4 GHz) and 128 GB memory. The client was with the CPU Intel Core i7-920 and 3 GB of memory.

### 6.3. Experimental results

#### 6.3.1. Examples of recovered handwritings

Examples of recovered handwritings are shown in Fig. 6, where the numbers on the left column indicate individual handwritings and columns show the employed methods. For each handwriting, the upper row shows the results using only the handwriting recovery. Some results are tilted since it only relies on the relative motion of the pen. In the process of tracking all results were adjusted to the pose of the first frame. Thus if it is not upright, results were deformed. The lower row shows the results by the whole process including the localization. "No results" in the figure indicates the failure of localization.

As shown in Fig. 6, methods with LK tracking were capable of smoother recovery of handwritings than that with SURF tracking. This indicates that the LK tracking is superior to the SURF tracking. In addition when we compare the results for handwriting 1 and 5 by the method 3 and the proposed method, which were without and with the query expansion, respectively, it can be said that the query expansion is effective to find positions of handwritings more accurately. On the other hand, for the handwriting 2 by the method 3 and the proposed method, the query expansion had a negative effect. This was caused by the inaccurate homography caused by the erroneous matching of LLAH feature points between query and database images.

An example of erroneous recovery by the LK tracking is shown as the handwriting 6 in Fig. 6. At the end of the handwriting recovered by both the method 3 and the proposed method, the stroke

Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003


**Fig. 6.** Examples of recovered handwriting.

was corrupted. This is because of the drift of the estimated position. The SURF tracking worked good for this case since it prevents the drift by checking the reappearance.

*6.3.2. Accuracy*

Fig. 7 shows the average accuracy as a function of the tolerance $d$, where 1 pixel corresponds to 0.12 mm in the images used for the experiments.

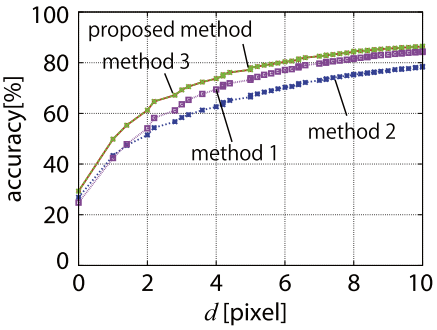
Fig. 7(a) represents the average accuracy obtained by only using the handwriting recovery. As shown in this figure, the LK tracking used in the methods 3 and 4 was superior to the SURF tracking in the methods 1 and 2. This was simply because the recovered handwriting was tilted if the reference frame to which the handwriting was recovered was not upright.

The accuracy after handwriting localization is shown in Fig. 7(b) and (c). Fig. 7(b) shows the accuracy averaged over all 30 handwritings, while Fig. 7(c) was calculated from the handwritings excluding the ones that failed in localization.

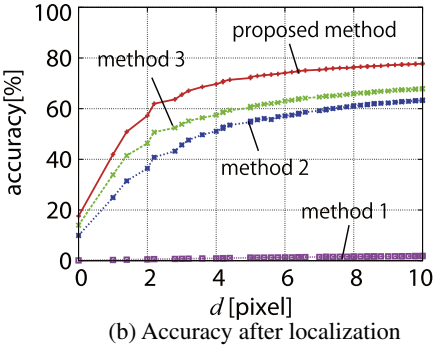
As shown in these two figures, about 10% difference was observed at $d = 2$. This was caused by the handwritings that were not correctly localized. For example, the method 1 successfully localized only 2 handwritings out of 30. The number of successfully localized handwritings by the methods 2 and 3, and the proposed method was 22, 23, 25, respectively.

The method with the best performance was the proposed method in both cases of including and excluding the erroneously localized handwritings. In the case of Fig. 7(c), for example, the accuracy of 93% was obtained at $d = 10$, which means 1.2 mm.

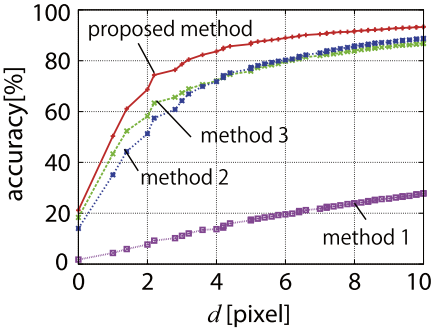
Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003




(a) Accuracy only by handwriting recovery




(b) Accuracy after localization




(c) Accuracy for samples successfully localized

**Fig. 7.** Average accuracy.

The proposed method performed 20% better than the method 2 at $d = 2$ in Fig. 7(c). This indicates the superiority of the LK tracking which is capable of generating smoother handwritings without losing any parts. The proposed method improved 10% of the performance of the method 3 at $d = 2$ in Fig. 7(c). The method 3 sometimes failed to retrieve the document as well as to localize recovered handwritings to incorrect positions as shown in the handwritings 2 and 5 of Fig. 6. On the other hand, the proposed method was with less failures thanks to the expanded queries.

Reasons of the errors occurred in the proposed method are as follows. One is the failure of document image retrieval because of the limited region of the captured image that caused the lack of LLAH feature points. This happened especially when the image included a larger blank area. We cannot place the camera apart from the paper surface in order to obtain the paper fingerprint, methods such as using multiple cameras need to be attempted to solve this problem. Another reason is that handwritings themselves changed LLAH feature points because of the overlap. This problem could be solved by estimating possible changes of LLAH feature points.

*6.3.3. Processing time and memory usage*

Table 1 lists the processing time and the amount of memory used by each method. At the client side, processing time of the methods 3 the proposed method with the LK tracking was shorter than that of the methods 1 and 2 with the SURF tracking, because the SURF tracking required longer processing time. At the server side, processing time of the methods 2 and the proposed method with the query expansion was longer than that of the methods 1 and 3 without it. However, since the longer processing time is still shorter than the processing time of the client, it has no negative influence.

**Table 1**
Processing time and memory usage.

| | Time (ms) | | Memory (MB) | |
|---|---|---|---|---|
| | Server | Client | Server | Client |
| Method 1 (baseline method) | 7.4 | 131.0 | 3140 | 414 |
| Method 2 | 24.5 | 163.7 | 3140 | 39 |
| Method 3 | 8.6 | 57.3 | 3140 | 22 |
| Proposed method | 24.7 | 60.3 | 3140 | 21 |

The query expansion has no influence on the memory usage at the server side. An important difference was observed at the client side. Since the methods 1 and 2 with the SURF tracking need to record all SURF features for recognizing reappearance, they used a larger amount of memory. In particular, the method 1 used less efficient way to record the SURF features, it required the largest amount. On the other hand, it is not necessary for the method 3 and the proposed method to record the features for tracking, they required less amount of memory.

*6.3.4. Discussions*

From the results shown above, we can conclude that the LK tracking is superior to the SURF tracking in terms of accuracy, processing time and memory. The query expansion was effective to improve the accuracy. Its longer processing time have no negative impact since even longer processing time was needed at the client side. The proposed method allowed us the best accuracy among the methods, it is still necessary to further improve the accuracy especially of the document image retrieval. One of the reasons of errors was the limited captured area by the camera. Thus a possible improvement is to increase the number of cameras to achieve a wider capturing area. Another important point is that it is necessary for the proposed method to recognize the reappearance for dealing with handwritings on blank parts.

## 7. Conclusion

Our activity of writing on paper with a pen is an important source of information to be utilized by computers. In order to achieve a natural way of recording handwritings as well as localizing them to the corresponding electronic document, we have developed a method that works only with a camera-pen. The best accuracy was achieved by combining the LK tracking at the client side and the query expansion at the server side, though there is still a room for the improvement.

The future work include further improvement of accuracy by modifying the document image retrieval as well as to deal with the reappearance.

## Acknowledgement

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

Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003

## References

Arai, T., Aust, D., Hudson, S.E., 1997. Paperlink: A technique for hyperlinking from real paper to electronic content. Proc. of the SIGCHI conference on Human factors in computing systems, 327–334.

Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., 2008. Surf: Speeded up robust features. CVIU 110 (3), 346–359.

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

http://www.adesso.com/home/tablets/158cyberpad.html.

http://www.anoto.com/.

http://www.pegatech.com/.

Iwata, K., Kise, K., Nakai, T., Iwamura, M., Uchida, S., Omachi, S., 2009. Capturing digital ink as retrieving fragments of document images. Proc. ICDAR2009, 1236–1240.

Iwata, K., Kise, K., Iwamura, M., Uchida, S., Omachi, S., 2010. Tracking and retrieval of pen tip positions for an intelligent camera pen. Proc. ICFHR2010, 277–282.

Kise, K., Iwata, K., Nakai, T., Iwamura, M., Uchida, S., Omachi, S., 2009. Document-level positioning of a pen tip by retrieval of image fragments. Proc. CBDAR2009, 61–68.

Kise, K., Chikano, M., Iwata, K., Iwamura, M., Uchida, S., Omachi, S., 2010. Expansion of queries and databases for improving the retrieval accuracy of document portions. Proc. DAS2010, 309–316.

Lucas, B.D., Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. Proc. IJCAI, 674–679.

Munich, M.E., Perona, P., 2002. Visual input for pen-based computers. IEEE Trans. PAMI 24 (3), 313–328.

Nakai, T., Kise, K., Iwamura, M., 2009. Real-time retrieval for images of documents in various languages using a web camera. Proc. ICDAR2009, 146–150.

Seok, J.H., Levasseur, S., Kim, K., Kim, J.H., 2008. Tracing handwriting on paper document under video camera. In: Proc. of ICFHR2008.

Uchida, S., Itou, K., Iwamura, M., Omachi, S., Kise, K., 2009. On a possibility of pen-tip camera for the reconstruction of handwritings. Proc. CBDAR2009, 119–126.

Yasuda, K., Muramatsu, D., Matsumoto, T., 2008. Visual-based online signature verification by pen tip tracking, Intfl Conf. on Computational Intelligence for Modelling. Control and Automation, 175–180.

Please cite this article in press as: Chikano, M., et al. Recovery and localization of handwritings by a camera-pen based on tracking and document image retrieval. Pattern Recognition Lett. (2012), http://dx.doi.org/10.1016/j.patrec.2012.10.003