On a Possibility of Pen-Tip Camera for the Reconstruction of Handwritings

Seiichi Uchida
Kyushu Univ., Japan

Katsuhiro Itou
Kyushu Univ., Japan

Masakazu Iwamura
Osaka Pref. Univ., Japan

Shinichiro Omachi
Tohoku Univ., Japan

Koichi Kise
Osaka Pref. Univ., Japan

Abstract

Toward realization of "writing-life-log," a camera-based handwriting pattern acquisition system is proposed. The camera is attached around the tip of a popular pen and captures frame images around the pen tip continuously. Our problem is the reconstruction of the entire handwriting pattern by video-mosaicing of those frame images with perspective registration of consecutive frames. A key idea is to use microscopic structure of paper surface, called paper fingerprint, for the registration. Specifically, perspective transformation is estimated by using correspondence of SURF keypoints extracted on paper surface. Since the precise structure can be captured stably as the SURF keypoints from the pen-tip camera, thus it is possible to expect accurate registration of video frames.

1. Introduction

Pen and paper have been used as one of the most long-lived and familiar media for information generation and storage. In fact, we often make handwritings on papers or other materials (Fig. 1) even though we get used to using computer and keyboard. Most students still use notebook to store their knowledge. Simple pocket notebook is still a good rival of intelligent PDA.

The goal of this research is realization of "*writing-life-log*," which is a system of acquiring our daily handwritings automatically captured through a video camera attached around the tip of a popular pen (e.g., a ball-point pen) and then storing them into database. The system enables us to review our past memos on papers, even if the papers are lost. Moreover, if any character recognizer is incorporated into the system in future, it enables us to *search* our daily handwritings on papers.

For this goal, this paper tackles a novel task of reconstructing an entire handwriting (or, equivalently, estimating the motion of the pen-tip) from video frames where the handwriting is captured fragmentarily. This reconstruction process is based on so-called video-mosaicing [1], which is comprised of four steps: (i) extraction of keypoints at each frame image, (ii) determination of keypoint correspondence between every two consecutive frames, (iii) estimation of geometric (often perspective) transformation between the frames, and (iv) registration of the frames after compensating the geometric transformation.

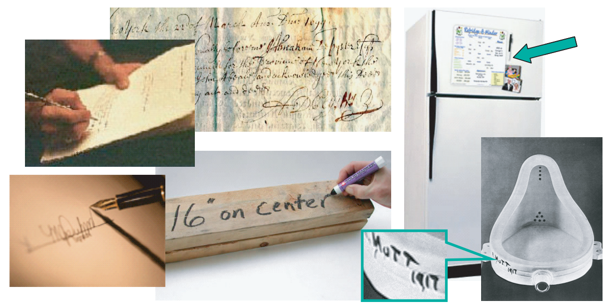

Figure 1. Handwritings everywhere.

For accurate video-mosaicing without any special requirement for paper, we will utilize the microscopic fiber structure of paper surface, called *paper fingerprint* [2] [1]. Thus, we will fully utilize background instead of foreground (i.e., handwriting patterns of black ink strokes). Recent camera technologies enable us to capture the paper fingerprint clearly even by a very small and cheap CCD camera. If it is possible to extract keypoints (e.g., corner points and edge points) from the paper fingerprint stably, they can be used as good keypoints of video-mosaicing.

The two main contributions of this paper are summarized

[1]The name of paper fingerprint ("fiber fingerprint" in [3]) is derived from the document identification system based on the microscopic structure of paper surface. For example, Fuji Xerox has developed XAYA [2] where paper fingerprint captured by a scanner is compared with stored paper fingerprints to examine their identity. More recently, Clarkson et al. [4] have developed a more sophisticated system where 3D paper surface is reconstructed from multiple scanning results and utilized for paper identification.

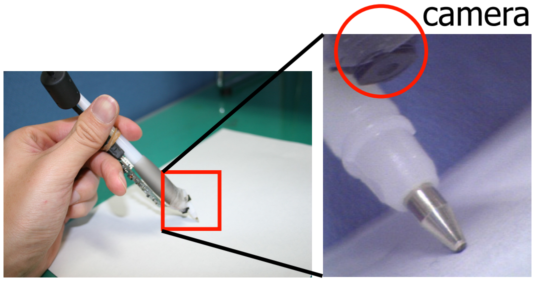

Figure 2. Pen-tip camera.

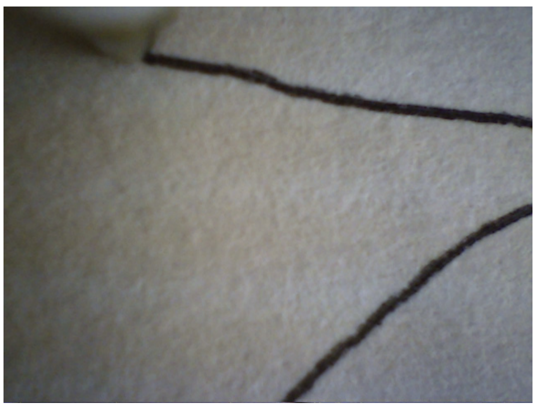

Figure 3. Image from pen-tip camera.

as follows:

- It is shown experimentally that the entire handwriting pattern on a paper can be reconstructed by mosaicing video frames from a pen-tip camera.
- It is shown that paper fingerprint is useful for the reconstruction. To the authors' best knowledge, paper fingerprint has only been utilized only for paper identification tasks [2, 4] and not for the reconstruction of handwriting patterns by pen-based interfaces.

The remaining of the paper is organized as follows. After a brief review of related work in Section 2, the principle of recovering a handwriting pattern from video frames is outlined in Section 3, and then detailed in Section 4. In Section 5, reconstruction results are shown to observe their accuracy.

2. Related Work

There are several past attempts on pen-tip camera. Anoto pen [5] will be the most famous pen-device with a pen-tip camera. It assumes a special paper where small dots are printed in advance. The dot pattern printed around a position on a paper represents not only the absolute (x, y)-position but also the ID of the paper. Thus, by capturing the dot pattern from the pen-tip camera, it is possible to estimate the motion of the pen-tip camera and thus to reconstruct handwriting patterns. A similar idea can be found in [6]. The proposed system is not necessary to assume any special paper for the reconstruction.

Arai et al. [7] have proposed PaperLink, which is a marker pen with a pen-tip camera. When marking a document, the camera captures the marked part automatically. If the marked part is captured by the camera again, the system recognizes the marked part and launches some predefined function. PaperLink is an interesting trial to link the real world (a paper document) to a cyberspace, while its purpose is not focused on reconstruction of the handwriting.

The proposed system is similar to optical mouse, which can estimate its motion by measuring the displacement of microscopic surface images of workbench. However, there are still big differences between optimal mouse and the proposed system; optical mouse is disturbed by neither various lighting environments nor perspective distortion nor (self-)occlusion. Thus, motion estimation of optical mouse is far easier than that of the pen-tip camera.

Seok et al. [8] have proposed a camera system which tracks a pen-tip in video frames from an "overlooking" camera. That is, the camera is fixed in the environment and not attached to the pen. This is a reasonable setup under the condition that target papers are always placed on the same area (such as a desk). The proposed system will relax the condition so that we can make handwritings at arbitrary locations in our daily life.

3. Overview of the Proposed Method

Figure 2 shows the pen-tip camera used in this paper. A micro-CCD USB camera (Asahi-Denshi Co., ACB-U04II) captures images around the pen-tip. The distance between the paper surface and the camera was around 1.5cm. The camera outputs non-interlace non-compressed video frames with 30fps. Figure 3 shows a frame image from the pen-tip camera. In addition to a black ink stroke, microscopic paper fingerprint can be observed here.

Video-mosaicing is employed for reconstructing an entire handwriting from the video frames. For example, by video-mosaicing, the video frames of Fig. 4 are superimposed into a single big reconstructed image of "0." As noted before, video-mosaicing is comprised of four steps. In this paper, (i) keypoints are extracted by the SURF algorithm [9], (ii) keypoint correspondence between consecutive frames is determined by the nearest neighbor matching of SURF keypoints, (iii) perspective transformation is estimated by RANSAC [10], and then (iv) the frames after

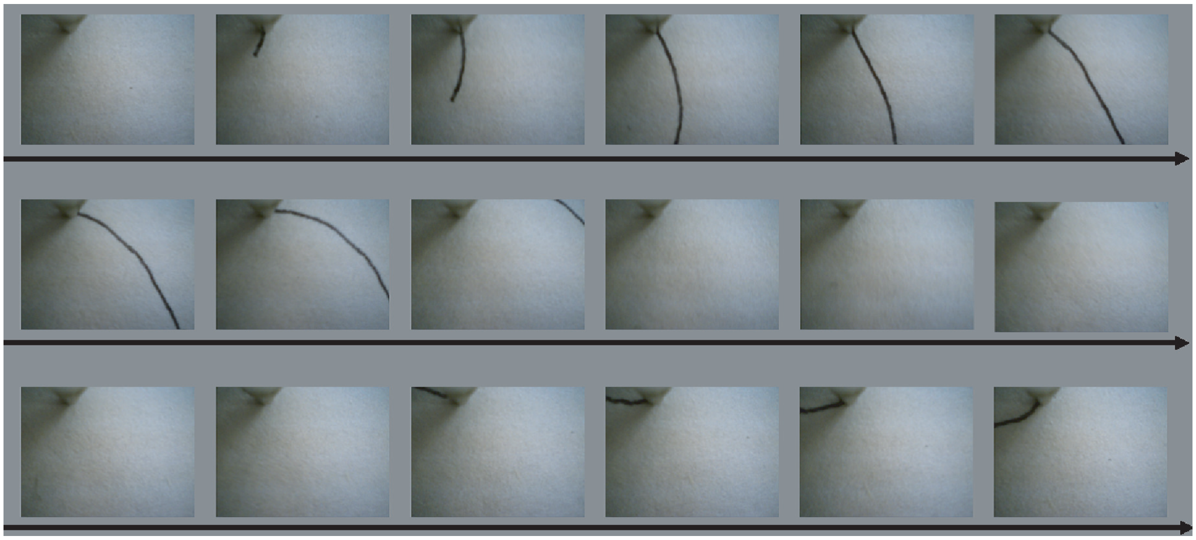

Figure 4. Frame image sequences on writing a digit pattern, “0.” (They are shown every 6 frames.) Note that at several frame images, the handwriting pattern is occluded by the pen-tip.

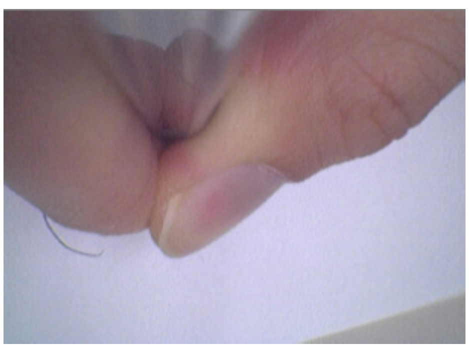

Figure 5. Image from pen-end camera.

compensating the perspective transformation are superimposed. Those steps are to be detailed in the later sections.

Note that, since the camera is fixed on the pen, the pen-tip is always seen at the same position in every frame image regardless of pen position, pen motion and pen posture (tilt and rotation). This fact can be confirmed by Fig. 4, where the pen-tip area is always seen around the left-top corner in all the frame images.

Although we will persist in using a pen-tip camera throughout this paper, we can consider to use a pen-end camera. The pen-end camera has a merit that it may capture the entire handwriting pattern within a single frame image. That is, video-mosaicing may not be necessary for the pen-end camera. In [11], a pen-end camera is used for estimating the pen posture and position by using the text contents which are printed on a paper document and stored in the system in advance.

The pen-end camera, however, also has a demerit that the handwriting pattern is heavily occluded by the hand as shown in Fig. 5. In addition, the ego-motion of the pen-end camera will be far larger than that of the pen-tip camera, and therefore larger-scale motions should be estimated with difficulties. Consequently, we will reconsider the pen-end camera in our future work as a complement of the pen-tip camera.

4. Detail of Video Mosaicing

4.1. Keypoint extraction from paper fingerprint

There are two possible “clues” for mosaicing video frames from the pen-tip camera: the foreground pattern (i.e., the handwriting pattern drawn as a black ink stroke) and the background pattern (i.e., paper fingerprint). As noted before, we will use paper fingerprint as the clue of the proposed method. This is because of the following two advantages of paper fingerprint over handwriting pattern:

- Avoidance of occlusion. If the pen-tip moves down from the current position in Fig. 3, the resulting black ink stroke will be totally occluded by the pen in the succeeding frame images. This fact also can be confirmed by the images of latter frames in Fig. 4. Thus, by observing the black ink stroke, the estimated pen-tip motion may become erroneous. In contrast, paper

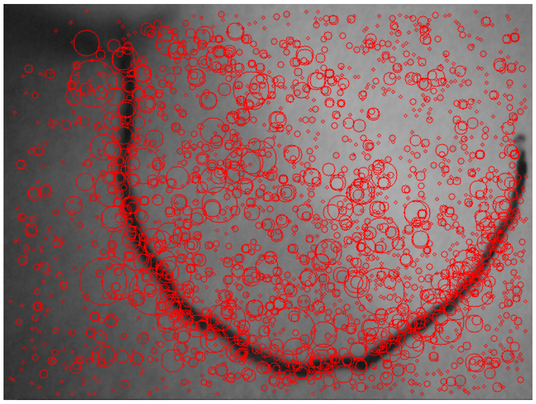

Figure 6. SURF keypoints on a frame image.

fingerprint can be always observed (except for a small part occluded by the pen-tip) and thus free from this occlusion problem.

- Avoidance of aperture problem. If the pen moves linearly to draw a long straight line, frame images are misrecognized as "still" frames showing the same straight handwriting pattern. Thus, it is difficult to judge whether the pen moves or stops. This is so-called the aperture problem on motion estimation. In contrast, it is possible to judge it correctly by observing paper fingerprint. That is, if the pen moves, the paper fingerprint also moves.

From paper fingerprint, keypoints are determined and represented by a certain feature vector. The feature should be rotation and scale invariant to have stable keypoint correspondence under pen rotation and pen tilt. The feature should also be robust to lighting condition change by, for example, the shades of the pen-tip or the hand.

Among various keypoint detection and representation algorithms, we choose the SURF (Speed-up Robust Features[9]) algorithm, which extracts keypoints and describes them as rotation, scale, and intensity invariant feature vectors. SURF is a fast version of SIFT (Scale-Invariant Feature Transform). SURF extracts keypoints by using efficient Haar wavelet instead of the DoG operator. Note that we can choose other detection and representation algorithms, such as FAST [12] and DAISY [13]. Keypoints specialized for paper fingerprint will be an interesting research topic.

Figure 6 shows the extracted SURF keypoints on a frame image. Each keypoint is depicted as a circle whose radius represents the estimated scale of the keypoint. Many keypoints appear on paper fingerprint, while they also appear around the black ink stroke.

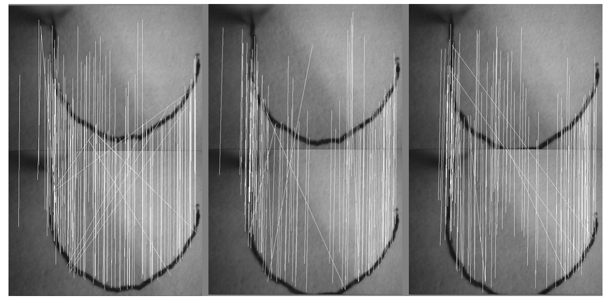

Figure 7. Keypoint correspondence between two consecutive frames.

4.2. Determination of keypoint correspondence

The keypoint correspondence between consecutive frames (say, the $(t-1)$th frame and the tth frame) can be determined by searching for pairs of SURF keypoints with very similar feature vectors. This search is simply realized by some nearest neighbor search and a thresholding operation on the difference of feature vectors; if the feature difference between a keypoint on the tth frame and its nearest neighbor keypoint on the $(t-1)$th frame is smaller than a fixed threshold, those keypoints are considered as a corresponding pair.

Figure 7 shows an example of keypoint correspondence between two frames. Each corresponding pair is connected by a line. Keypoints of the same position are often connected correctly and therefore it is shown that SURF-based matching of paper fingerprint is quite accurate. It is also shown that there are several erroneous corresponding pairs. In the next section, those erroneous pairs are excluded as possible.

4.3. Estimation of perspective transformation

For flat documents, two consecutive frames captured by the pen-tip camera can be registered accurately with a perspective transformation. Thus, it is necessary to estimate the most reliable perspective transformation by using the corresponding keypoint pairs. If this estimation is performed for every consecutive frame pair, it is possible to superimpose all the frames to have the final mosaicing result, that is, the entire handwriting pattern.

As noted in Section 4.2, there are erroneous corresponding pairs. Estimation results by LMS-like techniques are heavily degraded by those erroneous pairs, i.e., outliers. Unfortunately, those erroneous pairs are almost inevitable

Figure 8. (Upper) Numeral written on paper and (lower) its reconstruction result.

because paper fingerprint is a very delicate texture and easily distorted by pen-tilt, pen-rotation, motion blur, etc.

RANSAC [10] is employed in order to suppress the effect of the erroneous pairs during the estimation of the perspective transformation. RANSAC is a robust estimation method where perspective transformation is initially estimated from a small number of corresponding pairs and then evaluated by the "consensus" of the remaining corresponding pairs. After repeating this evaluation while changing corresponding pairs for the initial estimation, the best perspective transformation (i.e., the transformation with the widest consensus) is selected.

In addition to the use of RANSAC, the following three considerations are made to improve the estimation accuracy and the reconstruction result:

Keypoint removal on pen-tip area: The keypoints on the pen-tip area are removed in advance. Since the pen-tip area is always seen at the same position regardless of pen motion, thus the corresponding keypoint pairs on the pen-tip area are erroneous. (They wrongly indicate that there is no pen motion between the frames.) By simply removing the keypoints of the (known) pen-tip area, their effect can be suppressed easily.

Keypoint removal around pen-tip: The keypoints around the pen-tip area are removed in advance. Our mosaicing task deal with a "dynamic" target where a handwriting pattern grows frame by frame. Thus, the handwriting pattern of the tth frame has a new-born part around the pen-tip area and the part has no corresponding part on the handwriting pattern of the $(t-1)$th frame. By removing the keypoints around the pen-tip area, it is possible to avoid erroneous corresponding pairs on this part.

Skipping less reliable frames: The frame images with less reliable perspective transformation are *skipped* when superimposing frame images. Even though above two considerations, we still have erroneous estimation of perspective transformation. This erroneous estimation at the tth frame can be detected by checking how the estimation differs from that at the $(t-1)$th frame; since the pen motion is generally smooth, this difference should be small. If a large difference is found, the estimation at the tth frame is discarded and the perspective transformation is then estimated between the $(t-1)$th and $(t+1)$th frame.

5. Experimental Results

5.1. Reconstruction accuracy

In order to observe the accuracy of reconstructed handwriting patterns, video frames of writing a digit ("0"∼"9") were captured from the pen-tip camera of Fig. 2. A popular recycle paper was used in the experiment. The size of a written digit was about 2.0cm ×1.5cm on average. Writing speed will be discussed later.

Figure 8 shows the mosaicing results and their ground-truth, that is, the handwriting pattern on the paper. For showing the reconstructed handwriting pattern clearly, the small white circles representing the pen-tip position on each frame image are plotted on the mosaicing results. While the reconstructed handwriting patterns become slightly jaggy, they are accurate enough to show the original handwriting pattern. These results, consequently, indicate the validity of the paper fingerprint on our mosaicing task.

It was also shown that the occlusion problem noted in Section 4.1 could be avoided by the use of paper fingerprint. The video frames of Fig. 4 are the digit pattern "0" of Fig. 8. As shown in Fig. 4, the black ink stroke is often occluded by the pen-tip area and thus it is impossible to reconstruct the handwriting pattern "0" by using the black ink stroke as the clue of video-mosaicing. In contrast, "0" was reconstructed successfully by the paper fingerprint as shown in Fig. 8. It is interesting to note that the start and the end points of "0" were located very closely like the original handwriting pattern. This fact also indicates good accuracy of the proposed system.

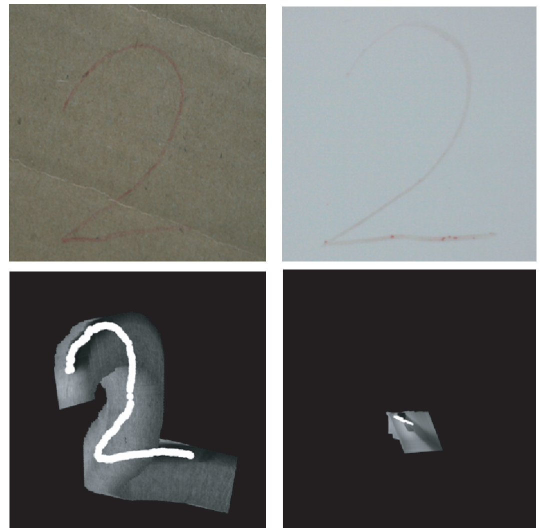

Figure 9. Reconstruction results on cardboard paper (left) and glossy photo-print paper (right).

5.2. Effect of paper type

Figure 9 shows reconstruction results under various paper types. A good reconstruction result was obtained on the cardboard paper because the cardboard paper has a rich paper fingerprint. In contrast, a poor reconstruction result was obtained on the glossy photo-print paper. This result was due to heavy reflection and less textured paper surface.

5.3. Writing speed

In this experiment, writing speed was kept rather slow. In fact, the number of frames was about 120. Since the frame rate was 30fps, it took 4s for writing a digit on the 2.0cm ×1.5cm area. This slow writing was necessary for avoiding motion blur, which destroys paper fingerprint. Figure 10 shows corresponding keypoint pairs under motion blur.

It is important to note that the weakness against motion blur does *not* prevent the possibility of the pen-tip camera for capturing paper fingerprint at all. Nowadays, high-speed cameras, which are free from motion blur, are available easily ($< 1,000$USD) and thus we can expect that its micro

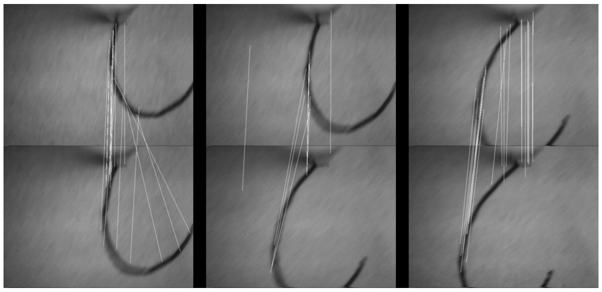

Figure 10. Corresponding keypoint pairs under motion blur.

version will also be available soon. In fact, recent optical mouse devices are already furnished with a very high-speed camera (often over 1500fps!).

5.4. Skipping less reliable frames

Figure 11 shows the number of keypoint pairs and skipped frames on writing a star symbol. On the graph, the vertical green lines indicate skipped frames, that is, less reliable frames. Roughly speaking, when keypoint pairs are not many, less reliable perspective transformation is estimated. From the principle of RANSAC, this relation is quite possible.

A small number of the skipped frames will not affect reconstruction results because the remaining frames (which will have large overlaps with the skipped frames) will complete the reconstruction. A larger number of skipped frames, however, are not negligible. A remedy to reduce the skipped frames is to increase keypoint pairs by removing motion blur (as noted in Section 5.3), enhancing paper fingerprint by some image processing, and so on.

6. Conclusion

It was examined to reconstruct handwriting patterns from video frames captured by a pen-tip camera. Video-mosaicing was performed for the reconstruction, while utilizing paper fingerprint, that is, a microscopic fiber structure of paper surface. Specifically, as the clue for the registration of consecutive frame images, SURF keypoints of paper fingerprint were utilized instead of ink stroke. It was showed experimentally that we could have good reconstruction results without any serious degradation by the proposed system.

This is the first trial on examining possibility of pen-tip camera and paper fingerprint and therefore there are many future work.

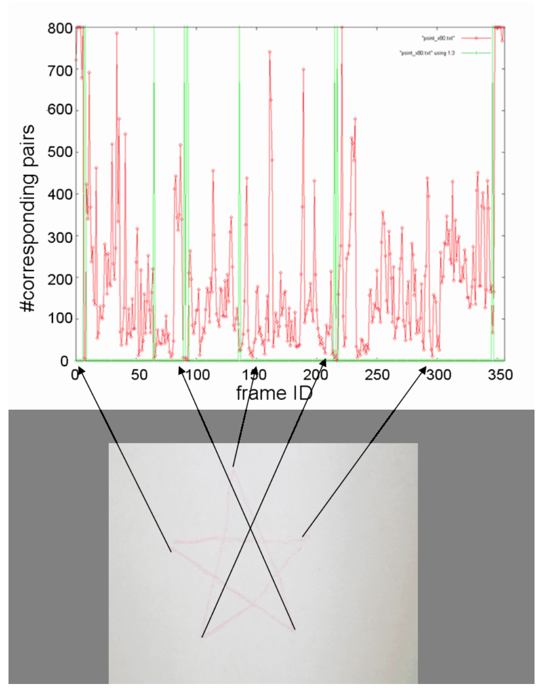

Figure 11. The number of keypoint pairs and skipped frames on writing a star symbol.

1. Image enhancement should be examined to have more reliable keypoints from paper fingerprint. Especially, it will be useful to detect stable keypoints on/around black ink strokes by the enhancement.

2. The strategy of video-mosaicing can be improved. In this paper, we used a frame-by-frame strategy; thus, accumulated registration errors degrades reconstruction accuracy in latter frames. Bundle-adjustment for improving perspective transformations among multiple frames simultaneously will be a possible remedy; by searching for corresponding SURF keypoint from not only the latest frame but also past frames [14, 15], it is possible to grasp the correct geometric relations among multiple frames.

3. Removal of motion blur is an important future work. The most solid remedy will be deblurring. It may be possible by using ink stroke [16]. A more simple and straightforward remedy will be to use a high-speed camera. As emphasized in Section 5.3, we can expect that commercial micro high-speed cameras will be available soon.

4. The extension to deal with multiple-stroke handwriting patterns is necessary. Bundle-adjustment will be useful again. The combination with a pen-end camera may

be also be useful to capture global structure of multi-stroke patterns.

Acknowledgment

This work was supported in part by the Research Grant (No.20300049) of The Ministry of Education, Culture, Sports, Science and Technology in Japan.

References

[1] M. Irani and P. Anandan, "Video indexing based on mosaic representations," Proc. IEEE, vol. 86, no. 5, pp. 905-921, 1998.

[2] http://www.fujixerox.com/eng/company/technology/xaya/

[3] E. Metois, P. Yarin, N. Salzman, and J. R. Smith, "FiberFingerprint identification," Proc. 3rd Workshop on Automatic Identification, pp. 147-154. 2002,

[4] W. Clarkson T. Weyrich, A. Finkelstein, N. Heninger, J. A. Halderman, and E. W. Felten, "Fingerprinting blank paper using commodity scanners," Proc. IEEE Symposium on Security and Privacy, 2009.

[5] http://www.anoto.com/

[6] D. A. Jared, et al., "Methods and apparatus for robust decoding of glyph address carpets," US Patent, no. 6208771, 2001.

[7] T. Arai, D. Aust, S. E. Hudson, "PaperLink: a technique for hyperlinking from real paper to electronic content," Proc. ACM Conf. Human Factors in Computing Systems (CHI'97), pp. 327-334, 1997.

[8] J. -H. Seok, S. Levasseur, K. -E. Kim, and J. H. Kim, "Tracing handwriting on paper document under video camera," Proc. ICFHR, 2008.

[9] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," Proc. ECCV, 2006.

[10] 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, vol. 24, no .6, pp. 381-395, 1981.

[11] 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," Proc. CBDAR, 2009.

[12] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," Proc. ECCV, 2006.

[13] E. Tola, V. Lepetit, P. Fua, "A fast local descriptor for dense matching," Proc. CVPR, 2008.

[14] T. Sato, S. Ikeda, M. Kanbara, A. Iketani, N. Nakajima, N. Yokoya, and K. Yamada, "High-resolution video mosaicing for documents and photos by estimating camera motion," Proc. SPIE Electronic Imaging, vol. 5299, pp. 246-253, 2004.

[15] N. Nakajima, A. Iketani and T. Sato, "Video Mosaicing for Document Imaging," Proc. CBDAR, Demo Session, 2007.

[16] X. Y. Qi, L. Zhang, C. L. Tan, "Motion deblurring for optical character recognition," Proc. ICDAR, pp. 389-393, 2005.