

1. Working Group on Camera-Based DAS

This section summarizes the discussions of the Working Group on Camera/Video-Based Document Analysis Systems. Seven researchers from four countries participated: D. Doermann, R. Kasturi (moderator), K. Kise (scribe), Y. Nakano, P. B. Pati, H. Sako, and P. Seakins. The participants represented a mixture of both private industry and universities.

The working group shared the view that the dissemination of digital still and video cameras is now opening a new vista of Document Analysis. "Camera-based Document Analysis" (CBDA) is required because we are no longer constrained to traditional 2D images of paper documents. For example, digital cameras enable us to capture characters and documents anywhere in the 3D environment such as signs and billboards. Moreover, documents are not necessarily static; in videos and movies, there also exist scrolling text and characters on moving objects. In other words, more freedom of media enables us to open up a new chapter for document analysis systems [1,2,3].

The purpose of the working group is to clarify (1) what are domains of interest, and what are goals and targets of camera-based document analysis, (2) what makes attaining these goals difficult, (3) what have we already been able to do (proven technologies), and (4) what are open problems and future trends in this field. We started our discussion with issues on goals and targets.

1.1 Goals and targets

Compared to traditional DA for 2D static document images acquired by scanners, CBDA is characterized by a wider variety of input methods including digital still and video cameras, mobile phones and PDAs with cameras. Such freedom of input allows us to consider new goals and challenges.

An obvious and important goal is to use the technology for indexing of existing still and video images. As the amount of still and video images grows, indexing is becoming a more serious problem not only for companies but also for individuals. Because of large inexpensive storage, ordinary users have still and video imagery in quantities far greater than manual indexing will permit. A positive sign is that some images include characters which are useful for automated indexing. In particular, videos of TV programs often have captions that represent some aspects of their contents. Simpler examples are images of documents and images of slides taken with digital cameras to "record" their contents. In this case, it is natural to recognize characters for indexing purposes.

Another important example application field is interfaces. Cameras with the capability of recognition can, for example, offer functionality of reading signs on streets, which is applicable to intelligent transportation systems as well as translation systems

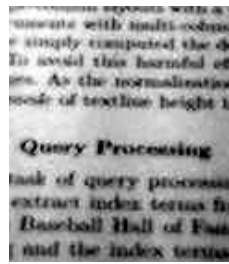
for travelers. With the same capability, we can develop a reading interface for the visually impaired.

In order to achieve such goals, document analysis systems are required to recognize either static or dynamic scene text and overlaid text in videos (i.e. moving text such as moving telops and text on moving objects). Such text is the primal target of camera-based document analysis systems.

1.2 Data quality issues

For the purpose of recognizing scene text and overlaid text successfully, problems of efficiency and accuracy of processing need to be solved. The problem about efficiency is, for example, how to achieve a level of performance so that processing does not require more than real-time. In the special case when we wish to process on the device (camera phones, PDAs, etc) there additional resource constraints. Accuracy is an issue because the data quality is much lower than that captured by scanners.

The working group highlighted the data quality issues that reveal the difficulty of camera-based DAS. Figure 1 illustrates some representative issues most of which are unrelated to ordinary scanner-based DAS.



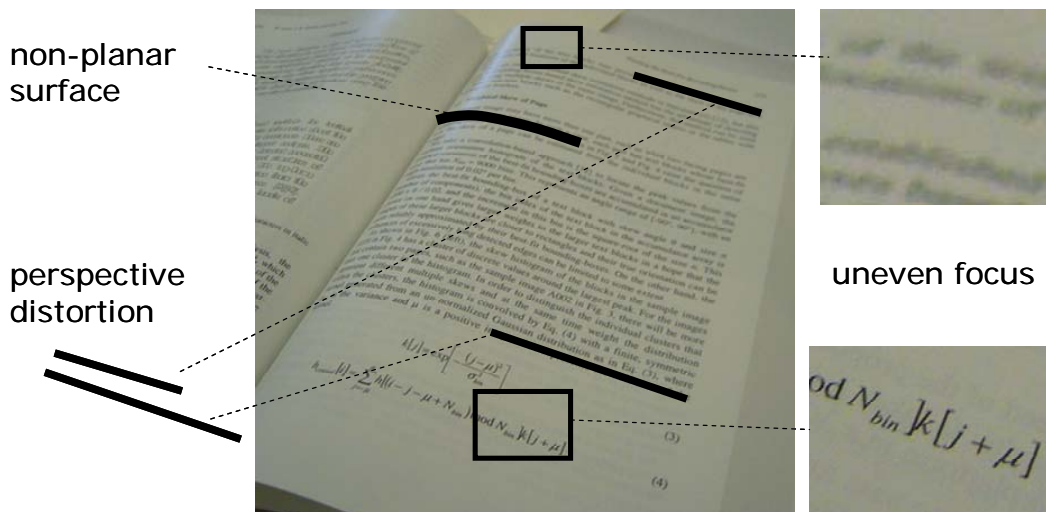
(a) low resolution



(b) uneven lighting & reflection



(c) motion blur



(d) uneven focus, perspective distortion and non-planar surfaces

Fig. 1 Issues of data quality. (a) The resolution is limited especially with cameras on mobile devices as well as video cameras. (b) The lighting environment and the reflection of a surface make character images difficult to read. (c) For videos and movies motion of objects and/or cameras blurs images. (d) Uneven focus caused by a narrow depth of field. Non-planar surfaces as well as perspective distortion make textline extraction more difficult.

1.3 Proven technologies

Next, the discussion transitioned to what are the proven and cutting-edge technologies in this relatively new field of research. The following is the list of technologies pointed out during the discussion.

- mosaicing & super-resolution,
- perspective correction,
- detection of overlaid still text on videos,
- constrained scene text recognition (on presentations, license plates, etc.)

The first two technologies are fundamental low-level image processing. Most of these technologies work under assumptions about the shape of document surfaces and the layout of text. For example, document surfaces are often assumed to be planar, and text lines linear, parallel and justified. The last two technologies focus on segmentation and recognition of text, which are also characterized by assumptions or constraints on text. In other words, current technologies rely on these assumptions to be useful.

1.3 Open problems

As complements to what has been done, the group discussed open problems, which are twofold: problems with data quality, and those of segmentation and recognition.

(1) Data quality

Problems falling into this category are fundamentally the same as shown in Figure 1: low resolution, uneven illumination and focus, motion blur, non-planar surfaces and sensor noise. Some of them have been addressed by proven technologies, but under some strict assumptions. For example, the problem of low resolution can be solved by super-resolution but it fundamentally holds under the assumption that the surface is planar. Thus some of the open problems here are to remove such assumptions.

(2) Segmentation and recognition

The same holds for some of the open problems in segmentation and recognition. Segmentation and recognition of scene text must work with fewer constraints. For example, it is worth pursuing segmentation and recognition of text on non-planar surfaces, non-rigid and moving objects. In such an unconstrained environment another important problem is how to cope with occlusion. For the overlaid text on videos, scrolling text, designed text (stylized, textured and dithered text), text changing its shape are still hard to segment and recognize.

Another important set of open problems centers around the use of these technologies. Since there is a continuum between the types of processing that needed for camera captured documents and scanned documents, we need techniques for triage. The

community has become very good at developing techniques for specific problems, but they only work if the document has those problems! A classic example is enhancement. If we try to enhance a document which has a specific type of degradation, we can do very well. But if the document does NOT, we can actually make things worse. As we encounter a wider variety of documents that we need to process, this will become a bigger and bigger issue.

1.4 Future trends

Finally the group discussed possible future trends in research which develop camera-based DAS. As mentioned earlier as one of the goals, indexing and searching of still and video images are important functions to be pursued in the future. For video images another important area is the analysis of embedded text. The current major topic of CBDA is still imagery, but video continues to gain attention, because they are the primary documents in the future. Even devices which we now use to take static snapshots of documents, such as mobile phones, will ultimately be use to scan the documents dynamically.

In order to boost the research activities of CBDA, it was pointed out that we need some killer applications of the technologies. Helping disabled people and assisting tourists with the technologies are important areas, but the size of the market is not necessarily large enough to be drivers of research. If document capturing with still and video cameras including mobile devices will be common ways of recording documents, a huge number of imaged documents will be produced and stored, and thus the size of application field as well as the market will become large. In such a situation, technologies of indexing and recognition are keys for extracting fruitful information from such large archives.

1.5 Summary of the working group

Since the research field of camera-based DAS is new, there is no clear definition of the boundaries of the field. However the members of the working group felt that camera-based DAS would be one of the major research fields because of the rapid dissemination of digital still and video cameras. Addition of dimensions from 2D to 3D or 3D plus time may change not only technologies but also the notion of “documents” themselves.

Rangachar Kasturi and Koichi Kise

References

- [1] S.Antani, R.Kasturi, R.Jain: “A survey on the use of pattern recognition methods

for abstraction, indexing and retrieval of images and video”, Pattern Recognition, vol.35, pp.945-965, 2002.

- [2] J.Liang, D.Doermann and H.Li: “Camera-based analysis of text and documents: a survey”, International Journal of Document Analysis and Recognition, vol.7, pp.84-104, 2005.
- [3] Proceedings of the first International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2005), available at <http://www.m.cs.osakafu-u.ac.jp/cbdar/> .