Active 카메라를 이용한 지능형 문서 영상 획득

박안진° 정기철
숭실대학교, 정보과학대학, 미디어학과
{anjin°, kcjung}@ssu.ac.kr

Intelligent Document Scanning with Active Camera

Anjin Park° Keechul Jung
School of Media, College of Information Science, Soonsil University.

요 약

문서 영상 획득(document scanning)은 문서 영상 분석과 인식을 위한 중요한 단계이다. 최근, 문서 영상 획득 장치로 스캐너(flat scanner)가 가장 많이 이용되고 있지만, 만지면 망가질 것 같은 얇은 고서, 매우 두꺼운 책과 같은 문서를 획득하기에는 어려움이 있다. 이런 어려움을 해결하기 위해, 카메라를 이용한 문서 영상 획득에 관한 많은 연구가 진행되고 있으며, 카메라의 저해상도나 잡음과 같은 문제를 해결하면, 스캐너를 대신하는 입력 장치로 이용할 수 있다. 저해상도 문제를 해결하는 방법으로 기본의 일반적인 레지스트레이션(registration) 방법은, 연결 부분(stitching position)에서 오브젝트(object: text, graphics, image)의 왜곡이 생기는 문제점이 있다. 본 논문에서는 PTZ(pan-tilt-zoom) 카메라를 이용하여 연결 부분에서 왜곡을 최소화하여 오브젝트를 획득하는 컴포넌트 기반의 영상 레지스트레이션(component-based image registration) 방법을 제안한다. 제안한 방법은 연결 부분에서 오브젝트의 수를 최소화하는데 목적이 있으며, 일반적인 레지스트레이션 방법에 비해 연결 부분에서 왜곡을 상당히 줄일 수 있으며, 상대적으로 인식률을 높일 수 있다.

1. 서 론

최근, 디지털 라이브러리(digital library) 분야의 정보 인덱싱(information indexing), 정보 검색(information retrieval)과 같은 몇몇 응용분야의 필요성으로 인하여, 문서 영상(document image)의 분석과 인식에 관한 연구가 많이 진행되고 있다[1-5]. 문서 영상 획득(document scanning)은 문서 영상의 분석과 인식을 위한 중요한 단계이다. 요즘, 문서 영상 획득 장치로 스캐너(flat scanner)가 가장 많이 사용되고 있지만, 만지면 망가질 것 같은 얇은 고서, 매우 두꺼운 책과 같은 문서를 획득하기에는 어려움이 있다. 이런 어려움을 극복하기 위해 카메라를 이용한 문서 영상 획득에 관한 연구가 진행되고 있으며[1-4], 카메라의 저해상도나 잡음과 같은 문제를 해결하면, 스캐너를 대신해 카메라를 문서 영상 획득을 위한 입력 장치로 사용할 수 있다.

카메라의 저해상도 문제를 해결하기 위해, Zappala[1] 등은 고해상도 영상을 만들기 위해 피쳐(feature) 기반의 레지스트레이션(registration) 방법을 제안하였으며, 사용자가 직접 수작업으로 문서의 모든 부분이 카메라 아래를 통과하도록 움직여야 하는 단점이 있다. Mirmdhdi[2] 등은 카메라로부터 멀리 떨어져 있는 문서를 고해상도로 획득하기 위해 광학 줌(optical zoom)을 결정하는 PTZ(pan-tilt-zoom) 카메라를 이용하여, 획득한 여러 장의 영상을 모자이싱(mosaicing)하는 방법을 제안하였다.

그러나, 앞에서 언급한 고해상도 영상을 만들기 위한 일반적인 레지스트레이션 방법은, 연결 부분(stitching position)에서 픽셀 번짐 현상으로 오브젝트(object: text, image, graphics)의 왜곡이 생기며, 이로 인해 인식률이 떨어지는 결과를 보인다. 이런 문제를 해결하기 위해 히스토그램 평활화(histogram equalization)와 같은 전역적 보간법(global interpolation)을 이용할지라도, 픽셀 번짐 현상으로 생긴 오브젝트의 왜곡을 완벽하게 제거하기는 어렵다[4,7].

본 논문에서는 연결 부분에서 왜곡이 없는 오브젝트를 획득하고, 인식률을 높이기 위해 PTZ 카메라를 이용한 컴포넌트 기반의 영상 레지스트레이션(component-based image registration:CIR) 방법을 제안한다. CIR 방법은 컴포넌트 크기에 의해 결정된 카메라 움직임을 이용하여 영상을 획득, 각 컴포넌트 단위로 영상을 모자이싱 하며, 연결 부분에서의 오브젝트 수를 최소화하는 새로운 연결 부분을 찾음으로써, 연결 부분에서 오브젝트의 수를 줄일 수 있으며 상대적으로 인식률을 높일 수 있다.

본 논문의 구성은 다음과 같다. 제 2장에서 CIR 방법을 기술하고, 3장에서 실험결과에 대해 기술하며, 제 4장에서 향후 연구 방향을 기술한다.

2. CIR 방법

우리는 PTZ 카메라를 이용하여 연결 부분의 왜곡을 최소화하는 고해상도 영상을 획득하기 위해 CIR 방법을 제안하며, 여기서, 컴포넌트는 문자, 그래픽, 영상과 같은 오브젝트를 의미한다. CIR 방법은 5 단계로 구성되어 있다.

첫째, X-Y recursive cut 알고리즘을 이용하여 입력 받은 문서 영상을 서브(sub) 컴포넌트로 나눈다. X-Y recursive cut 알고리즘은 에지 필터(edge filter)를 이용하여 획득한 이진 영상을 입력 값으로 이용하며, 투영 윤곽(projection profile)의 골(valley)을 식별하기 위해 X, Y 방향으로 에지 픽셀을 투영한다. 그림 1은 X-Y recursive cut을 이용한 문서 영역 분할을 보여준다. 그림 1(b)는 에지 영상에서 minimum bounded rectang-

les(MBRs)를 이용하여 분할된 영상을 나타내며, 각 사각형은 컴포넌트를 나타낸다.둘째, 고해상도 영상을 만들기 위해 각 컴포넌트에서 컴포넌트 크기에 의해 결정된 카메라의 움직임(pan-tilt-zoom)을 이용하여 여러 장의 영상을 획득한다.

셋째, 두 영상의 일치점을 구한다. 본 논문에서는 Zappala[1] 등이 제안한 방법을 이용하였으며, 문서를 단어 별로 분리하고, 각 단어의 가로축 길이를 이용하여 일치점을 구한다. 그림 2(a)와 2(b)는 두 입력 영상에서의 일치점을 나타낸다.

넷째, 연결 부분에서의 오브젝트 수를 최소화하는 새로운 연결 부분을 구한다. 새로운 연결 부분은 두 영상의 중복되는 영역 중 오브젝트의 픽셀수가 가장 작은 축을 구함으로써 만들 수 있다(그림 2(c)). 오브젝트 픽셀 수는 에지 영상에서의 투영 윤곽을 이용하여 구한다.

다섯째, 식 1을 이용하여 새로운 연결 부분을 가진 영상(그림 2(c))과 그림 2(a) 두 영상을 모자이싱한다. 그림 2(c)의 (u_1,v_1)과 그림 2(a)의 (u_2,v_2) 사이의 변환은 plane-to-plane projectivity[7]에 의해 구한다:

$$\begin{bmatrix} su_2 \\ sv_2 \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \quad (1)$$

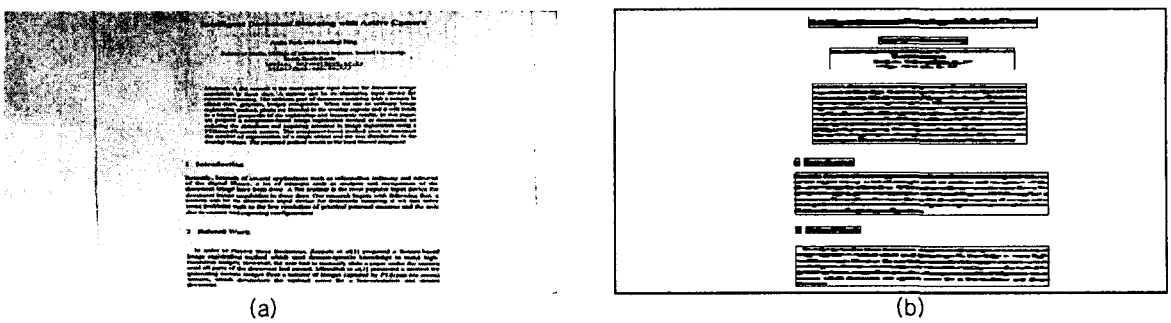
projectivity의 8개 파라미터(parameter)는 일치점의 4개 쌍으로부터 구할 수 있다.

처음 두 단계는 각각의 컴포넌트를 구하는 단계이며, 나머지 세 단계는 각 컴포넌트 별로 영상을 모자이싱하는 단계이다. 그림 2(d)는 일반적인 영상 레지스트레이션 방법을 이용하여 획득한 영상이며, 그림 2(e)는 제안된 방법을 이용한 결과영상이다. 연결 부분은 그림 2(d), 그림 2(e)에서 화살표로 표시되어 있으며, 그림 2(d)는 모자이싱된 영상의 연결 부분에서 오브젝트의 픽셀이 번지는 결과를 보이며, 반대로, 그림 2(e)는 연결 부분에서 오브젝트의 수가 적기 때문에, 그림 2(d)에 비해 상대적으로 오브젝트의 픽셀이 덜 번지는 결과를 보여준다.

3. 실험 결과

본 논문의 실험 환경은 Wellner의 디지털 데스크와 비슷하게 구성되어 있으며[3,5], 영상을 획득하기 위한 AXIS 2130 PTZ 카메라, 물리적 데스크(physical desk), 2.66GHz PC로 구성되어 있다. 우리는 문서는 이미 배경과 분리, 기울기 보정되어 있으며, 방사상 왜곡(radial distortion) 에 대해서 생각하지 않는다는 가정을 가진다.

그림 3은 Zappala[1] 등이 제안한 방법과 본 논문에서 제안



(a) (b)

그림 1. 문서 영역 분할 결과: (a) 문서 입력 영상, (b)에지 영상에서 문서 영역 분할 결과 영상.

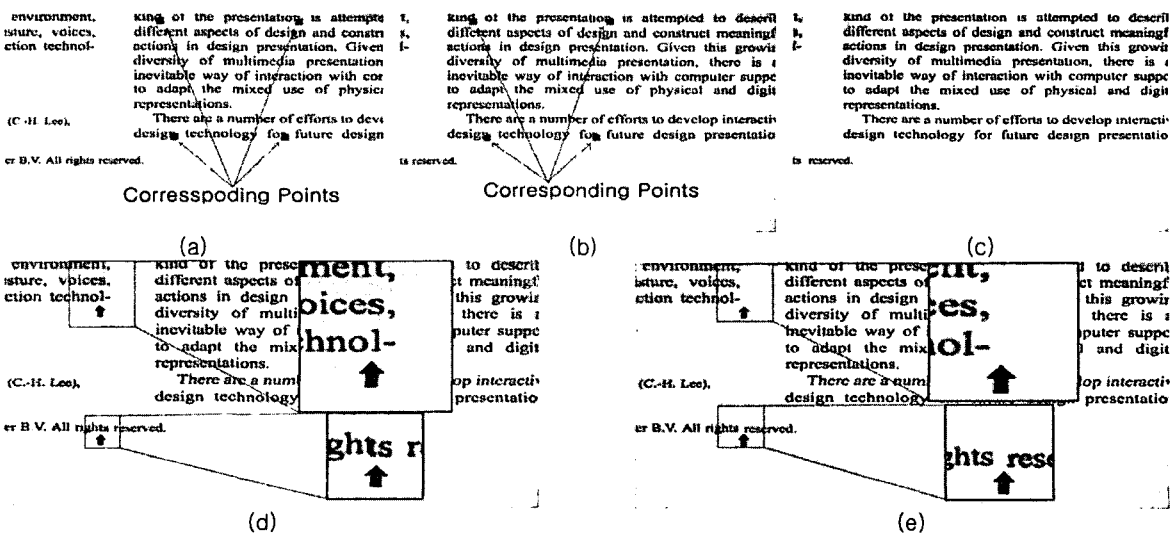


그림. 2. 두 연속 영상을 이용한 모자이싱 결과: (a) n번째 입력 영상, (b) (n+1)번째 입력 영상, (c) (b)의 새로운 연결 부분, (d) (a)와 (b)를 이용한 입력적인 모자이싱 결과, (e) (a)와 (c)를 이용한 모자이싱 결과.

한 방법의 비교를 위해, 모자이싱 결과를 이진 영상으로 만들고, 연결 부분을 표시한 영상이며, 약 220 dpi(dot per inch) 해상도를 가진다. 그림 3(a), 3(c), 3(e)는 Zapplar[1] 등이 제안한 방법을 이용한 결과 영상이며, 그림 3(b), 3(d), 3(e)는 본 논문에서 제안한 방법을 이용한 결과 영상이다. 표 1은 그림 3 영상에서의 연결 부분 상의 문자 수와 상업적인 OCR 시스템을 이용한 평균 OCR 인식률을 비교한 결과이며, CIR 방법은 연결 성분 상의 문자 수가 적으며, 상대적으로 높은 OCR 인식률을 보인다.

표 1. 연결 부분 상의 문자 수와 OCR 인식률 비교.

	연결 성분 상의 문자 수	OCR 인식률(%)
Zappala[1] 등이 제안한 방법	33	92.84
본 논문에서 제안한 방법	16	94.76

4. 결론

실험 결과에서 보는 바와 같이, 본 논문에서 사용한 CIR 방법은 일반적인 영상 레지스트레이션 방법보다 연결 부분에서 오브젝트의 수가 적으며, 인식률이 높은 장점을 가지고 있다. 그러나 카메라 기반의 문서 영상 획득이 갖은 이점을 가지고 있을지 라도, 우리는 여전히 일치점계산, 정교한 모자이싱 방법, 고르지 않은 빛, 원근 왜곡, 복합 문서, 렌즈 왜곡 등의 환경 문제와 같은 다양한 문제를 가지고 있다[4].

영상(image) 컴포넌트는 문자 컴포넌트보다 인식을 위해 더 높은 해상도를 요구한다. 앞으로, 우리는 문서 영상에서 문자와 영상 컴포넌트에 차이를 두고 레지스트레이션 할 수 있는 방법을 연구할 것 이다.

4. 참고 문헌

[1] Anthony Zappala, Andrew Gee and Michael Tayor, "Document Mosaicing," Image and Vision Computing, Vol. 17, pp. 589-595, 1999.
[2] Majid Mirmehdi, Paul Clark and Justin Lam, "Extracting Low Resolution Text with an Active Camera for OCR," Proceedings of Pattern Recognition and Image Processing, pp. 43-48, May 2001.
[3] Pierre Wellner, "The DigitalDesk Calculator: Tactile Manipulation on a Desktop Display," Proceedings of ACM Symposium on User Interface Software and Technology, pp. 27-33, 1991.
[4] David Doermann, Jian Liang and Huiping Li, "Progress in Camera-based Documents Image Analysis," Proceedings of International Conference on Document Analysis and Recognition, Vol. 1, pp. 606-616, Aug. 2003.
[5] Kwangjin Hong and Keechul Jung, "Advanced Paper Document in a Projection Display," Proceedings of Pacific-Rim Conference on Multimedia, To be published.
[6] Barbara Zitova and Jan Flusser, "Image Registration methods: a survey," Image and Vision Computing, pp. 977-1000. Oct. 2003.
[7] M. J. Taylor and J. Lee, "Interactive text segmentation of gray scale images of documents," Technical Report EPC-19970194, Xerox Research Centre, Apr. 1997.

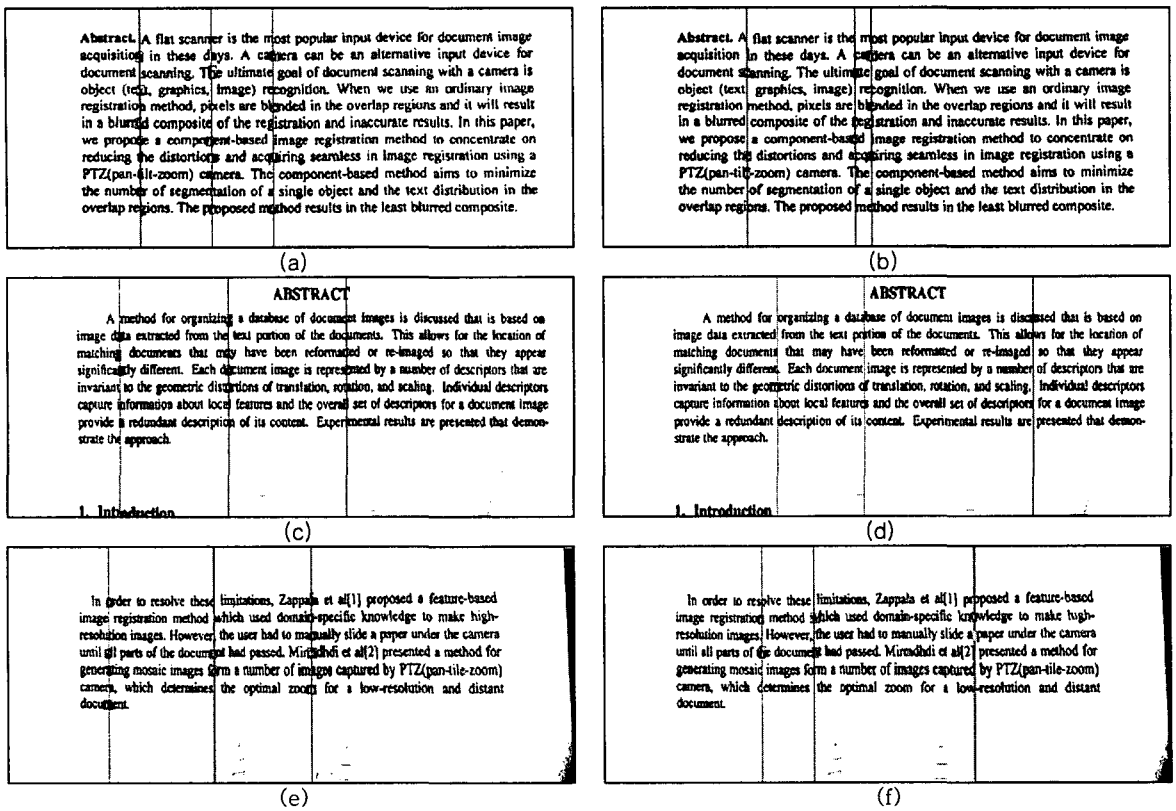


그림 3. 연결 부분을 표시한 레지스트레이션 결과의 이진 영상: (a),(c),(e) Zappala[1] 등이 제안한 방법을 이용한 결과 영상, (b),(d),(f) 본 논문에서 제안한 방법을 이용한 결과 영상.