학습용 시각 정보 인식 시스템의 설계 및 구현

신현경
가천대학교 수학정보학과

요 약

모바일 기반의 스마트 기기의 보급이 확대됨에 따라 교육 현장에서 이를 활용하는 사례가 증가하고 있는 추세이며, 가까운 장내에는 매우 중요한 교육용 기자재로서의 위치를 차지할 것으로 예측된다. 이러한 추세에 맞춰 교육과학기술부는 스마트 교육에 대한 중장기 추진 계획을 발표하였고 현재 추진을 준비 중에 있으며, 다양한 산업계 학계 연구 기관에서 관련 연구 결과물과 시제품들을 활발히 발표하고 있는 현실이다.

본 논문에서는 모바일 스마트 기기에 장착된 비디오카메라를 이용하여 촬영된 영상 내부에 포함된 문자를 인식하는 모듈을 구현하고 이를 응용하여, 교육환경에서 현실적으로 적용 가능한 학습용 시각 정보 인식 시스템에 관련한 설계 및 구현 방안을 제안하였다.

본 논문에서 제안한 학습용 시각 정보 인식 시스템은 비디오 영상취득, 영상 처리, 정보 추출, 지식 표현 등 4개의 모듈로 구성되었으며, 실제적인 예제를 통해 각 모듈을 설명 하였다.

키워드: 비디오 문자 인식, 지식 표현, 학습용 시각정보 인식 시스템

Design and Implementation of Visual Information Extraction System for Education

Hyunkyung Shin
Gachon University, Department of Mathematics & Information

ABSTRACT

As propagation of mobile smart devices is widespread, it is an observable trend that the cases of utilizing them are increasing in the school programs, and it is also anticipated that they will be very important part of the educational equipment in near future. For this reason the department of education and science technology has announced a medium and long term project on the education with smart device, which is undergoing the preparation stage, and the various academic and industrial institutes have actively produced the related research results and the application prototypes.

In this paper we propose a framework on design and implementation of a visual context recognition system for educational purpose usable in the school program by utilizing a module for recognition of the texts embedded in the image captured by video camera from mobile smart device. The system proposed in this paper is consisted of the four modules, such as, image acquisition, image processing, information extraction, and knowledge representation, which are explained in details with the practical examples.

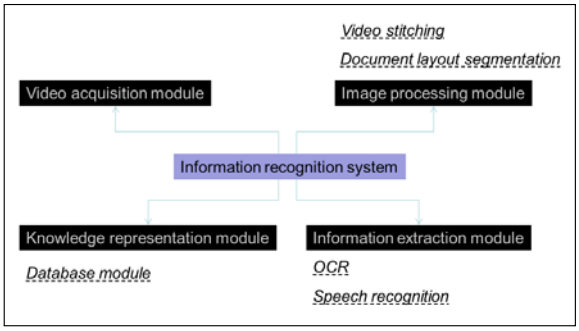
Keywords: video character recognition, knowledge representation, visual information extraction system for education

본 논문은 2012년 가천대학교 교내연구비 지원을 받아 시행된 연구임.
논문투고 : 2012-12-04
논문심사 : 2012-12-04
논문완료 : 2012-12-21

1. 서론

주변 시각 정보를 기반으로 고급 수준의 반응 형 에이전트를 구현하는 것은 유아 및 청소년들의 학습 흥미 유발을 일으키는 응용소프트웨어 개발의 핵심이다. 주변 시각 정보를 인식하고 적절히 반응하는 시스템의 구현은 아직 매우 단순한 수준에 머물러 있다. 예를 들어 유아들의 문자 학습 도움을 위한 응용소프트웨어 등이 있다[1] 인공지능 기술을 활용한 교육용 어플리케이션 개발은 모바일 기기의 CPU 및 RAM 성능의 향상[7], 클라우딩 서비스 등 서버 이론의 발전[16], Hadoop 에코 시스템 등으로 대표되는 대용량 정보 처리를 위한 확장 가능한 분산 컴퓨팅 프레임워크(scalable distributed computing framework)의 성공적 구축[15], 개방형 소스 개발을 위한 Apache 소프트웨어 그룹(apache software foundation) [17] 등 소프트웨어 인프라의 활성화와 맞물려 그 실효성이 증가되고 있고 분야별 산학 공조의 요구가 높아졌다. 비디오 카메라를 통해 취득되는 시각 정보의 처리 문제는 매우 어려우면서도 광범위한 응용 가능성을 내포하고 있다. 움직이는 객체 인식을 통한 감시 카메라 자동화의 문제[8], 안면 인식 등 생체 인식을 통한 보안 자동화의 문제[10], 제조 상품 정보 인식(product scanning)을 통한 물류 유통 현황 파악[3] 등을 그 예로 들 수 있다. 실시간 프로그래밍의 측면에서는 영상 처리 모듈의 비디오 스티칭(video stitching)과 정보 추출 모듈의 OCR을 위한 하부 모듈 구현이 대표적 난제이다. 첫째, 비디오 스티칭을 위한 일반적인 방법론으로 사용되는 크기와 방향에 무관한 탐색자(detector) 및 기술자(descriptor)기반의 SIFT(Scale Invariant Feature transform) 또는 SURF(Speeded Up Robust Features)의 실행 속도는 일반적으로 10분을 초과하는 실정으로 실시간 처리를 위해 새로운 알고리즘의 개발이 절실한 상황이다. 둘째, 비디오 영상의 낮은 해상도와 영상 내에 문자 이외의 다양하고 복잡한 배경 패턴이 존재하는 문제로 인해 OCR의 정확도가 현실적으로 사용하기에 현저히 부족한 실정이다.

본 논문에서는 스마트 폰 등 모바일 기기의 비디오카메라를 통해 취득한 영상 내부의 문자 정보를 활용해 사용자 대화형 학습용 어플리케이션 개발의 문제점을 분석하고 전체적 시스템의 설계 및 구현 방안을 제안 하였다. (그림 1)와 같은 제안된 시스템의 구성은 네 분야의 독립적 모듈과 각 모듈을 구현하는 하부 모듈로 구성된다.



(그림 1) 비디오 문자 정보 인지 및 반응 시스템 모듈 디자인

본 논문은 제 2장에서 기존 연구 현황을 분석하였으며, 제 3 장에서 본 시스템의 설계 및 구현을 위한 4 가지의 모듈과 그 하부 모듈에 대하여 구체적으로 기술한 후 기존 시스템과의 비교 분석을 실시하였다. 마지막으로 제4장 결론으로 이루어져 있다.

2. 관련연구

비디오 스트림에서 파노라마 영상을 만들어내는 소프트웨어는 몇 가지 존재한다[18][19]. 그러나 카메라의 위치가 고정되어 있어야 하는 등 제약 조건이 심각하다. 자유로이 움직이는 비디오카메라에서 파노라마 영상을 만들어내기 위한 연구 노력으로 SIFT/SURF 등 고전적 방법론[6][11], Perspective-n-point(PnP) 모델[9][13], pure visual SLAM[4][16] 등이 있다. 고전적 방법론들은 알고리즘의 복잡도에 의한 실행 시간 초과로 실시간 처리에 부적합하고, PnP 모델은 수학적으로 고급스러운 이론이나 안정적이지 못한 결점이 있다.

비디오 문자 인식의 연구 분야는 크게 두 가지로 구분된다. 하나는 배경 화면에 존재하는 복잡한 패턴들로 인한 문자 추출의 성능 저하에 관한 연구[2][5]이고 다른 하나는 낮은 해상도 및 움직임에 기인한

번짐(blurring)현상에 따른 성능 저하에 관한 연구 [3][12] 이다.

<표 1> 비디오 문자인식 관련연구

명칭	관계 연구	특징
비디오 스티칭	상용 소프트웨어[8][9]	제한적 카메라 궤적 요구 사항
	SIFT/SURF [10][11]	영상 registration 기술. 계산 복잡도.
	PnP[12][18]	영상에서 카메라 포즈 매트릭스 추정
	Visual SLAM [13][19]	비디오에서 카메라 궤적 추정
복잡한 배경에서 문자추출	배경 문자추출 [14]	특징 선택을 통한 배경과 전방의 구별에 대한 방법론
저 해상도 영상에서 패턴 인식	저해상도 처리[15][16]	카메라 렌즈의 저해상 능력과 손 떨림에 따른 저해상도 해결 방법론

3. 시스템 설계 및 구현

본 논문에서 제안한 학습용 시각정보 인식 시스템은 비디오 영상 내부에 포함된 문자의 인식을 수행하며, 지식 표현된 데이터 베이스와의 연동을 통해 사용자 질의에 답변하는 것을 목적으로 한다. 제안된 시스템은 크게 비디오 영상 취득(video acquisition module), 영상 처리(image processing), 정보 추출(information extraction), 지식 표현(knowledge representation) 등 4 가지 모듈로 구성하였다. 각 모듈에 대한 설명은 아래와 같다.

3. 1 비디오 영상 취득 모듈

뒷면이 보이지 않는 실린더 형태 또는 카메라로 한 번에 촬영할 수 없는 넓은 형태의 물체에 대한 영상 정보를 취득하기 위해 카메라 영상 대신 비디오 영상을 사용하였다. 구체적으로 모듈 구현을 위해 공개 소프트웨어 라이브러리인 OpenCV의 mpeg codec 등을 사용할 수 있다. (그림 2)에 시계방향으로 회전하며 촬영한 비디오에서 추출한 몇 개의 프레임을 보여준다.

(그림 2) 넓은 물체를 시계방향으로 비디오 스캔 한 프레임들

3.2 영상 처리 모듈

본 시스템의 영상 처리 모듈은 다음 단계인 문자 인식 모듈의 성능 향상을 주요 목적으로 하며 크게 비디오 스티칭을 사용한 영상 내 문자들의 문장 인식 향상을 위한 문서 레이아웃 분할 모듈로 이루어진다.

3.2.1 비디오 스티칭 모듈

SIFT, SURF 등 크기와 방향에 무관한 탐색자와 기술자들 적용한 이미지 등록(image registration) 방법론 기반의 스티칭은 알고리즘의 복잡성으로 인해 실시간 처리가 요구되는 어플리케이션 구현에 현실성이 없다. 대안으로서, 본 논문에서는 순수시각정보 SLAM(Simultaneous Localization and Mapping)에서 효과적으로 적용되고 있는 시각 흐름(optical flow)을 사용해 프레임 간의 카메라 이동 경로를 예측하고, 이를 바탕으로 프레임 간의 호모그래피를 측정해 카메라 포즈를 예측하는 방식을 사용한다.

3.2.2 문서 레이아웃 분할 모듈

스트칭을 통해 만들어진 영상은 문서 레이아웃 분할을 통해 문자 추출 및 추출된 문자들 간의 연관 관계를 사용해 문장 형태로 정비된다. 문자 추출은 비문자 제거를 목적으로 수행되며 방법론으로는 기계학습 기반의 CART(Classification And Regression

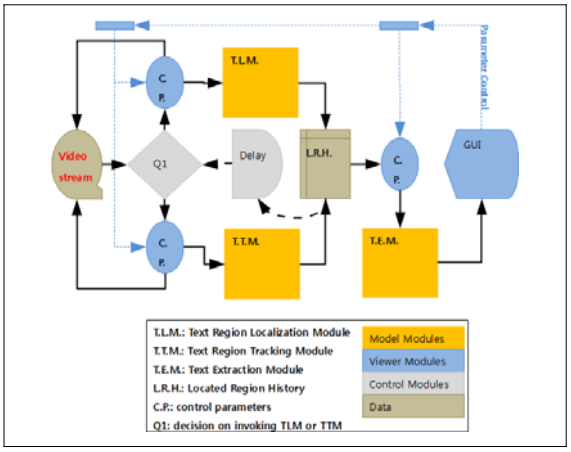
Tree)를 구현한다. 학습을 위한 특성벡터는 2차원 탐색자를 위한 central momentum과 각 객체간의 가장 가까운 이웃(nearest neighbor)와의 거리를 원소로 13차원 벡터로 구성된다. 레이아웃 분할을 위한 구체적인 방법론은 문자 객체의 컨투어(contour)를 그래프 노드로 하는 네트워크상에서 그래프를 적용한 클러스터링을 재귀적으로 적용하는 알고리즘을 구현하였다. (그림 3)는 서류 레이아웃의 결과를 표현한 것이다. 추출된 문자들은 단어와 문자열 단위로 정비되어 다음 단계인 문자 인식의 성능을 크게 향상 시키는 역할을 하게 된다.

(그림 3) document layout segmentation의 예

3.3 정보 추출 모듈

정보 추출 모듈은 문자 인식과 TTS(text-to-speech) 두 가지 모듈로 이루어진다. 앞 단락의 레이아웃 분할의 결과는 복잡한 배경의 비 문자를 제거해 문자 인식의 성능과 속도를 향상시키고, 더 나아가 인식된 문자들을 단어와 문자열 단위로 조직 하는데 결정적인 역할을 한다. 문자 인식은 공개 소프트웨어인 google tesseract를 사용해 수행하고, 인식된 문자의 발성을 위해 공개 소프트웨어인 e-speak기반의 TTS를 사용하였다.

(그림 4) 비디오 OCR 흐름도



3.4 지식 표현 모듈

인공 지능 요소를 구현하기 위한 핵심적인 모듈로 구현에 가장 어려운 부분이다. 공개 소프트웨어로서 UIMA (Unstructured Information Management Architecture)가 있으며 아직 실제적 문제를 해결하기에는 초기 단계에 머물러있는 수준이다, 본 논문에서는 apache 공개 소프트웨어인 openNLP를 기반으로 문장 분석 모듈을 구현하였다. 문제 정의된 도메인에 특화된 데이터를 학습시켜 참조테이블(look-up table)을 템플렛으로 하는 데이터베이스를 구축하고, 예를 들어, 알츠하이머 관련 자료 또는 영수증 관련 자료 등, 사용자의 키워드 탐색에 반응하는 웹 기반 시스템을 구현하였다.

3.5 성능평가

본 논문에 제안한 스마트 기기 상에서의 학습용 시각 정보 인식 어플리케이션의 설계 및 구현 방안은 새로운 분야로서 비교 대상이 풍부하지 않다. <표 2>는 본 개발 시스템의 특징을 몇 가지 관점에서 기존 시스템들과 비교 분석 내용을 요약한 것이다.

<표 2> 개발 시스템특징 비교

	개발 시스템	기존 시스템
문자 추출 방법론	문단 모델을 적용한 전처리기능을 통해 보다 많은 문자를 획득	서류 영상에 특화된 문자 인식 엔진을 적용
시스템 설계	카메라를 통한 시각적 정보를 담당 처리 하는 고위 모듈 구현을 통해 지식 기반 데이터베이스의 연속적 업데이트 활성화	지식 기반 (KB) 데이터베이스에 저장한 지시사항의 탐색 및 실행을 통한 반응형 에이전트 구현
적용 알고리즘	영상 취득 및 처리, 문자 추출, 문자 인식, 정보 추출, 자연어 처리, 지식 표현 알고리즘 등을 네 단계별 모듈로 구분하여 적용	지식 기반 데이터 베이스 구축과 그 접근 및 저장을 위한 알고리즘 위주

4. 결론

모바일 기반 스마트 기기의 보급 확대로 교육 환경에서 이를 활용하는 사례가 증가하고 있으며, 그 추세는 지속적으로 증가할 것으로 예측된다. 이에 따라 본 논문에서는 모바일 기반의 스마트 기기를 활용하여 이미지를 촬영한 후 이미지에 포함 된 문자 정보를 인식 하여 관련 교육 정보를 확인할 수 있는 시각 정보 인식 시스템을 설계하고 구현 방안을 제안 하였다. 본 논문에서 제안한 시스템은 현재 새롭게 관심 받고 있는 분야의 내용으로 관련 연구가 풍부하지는 않으나 유사한 시스템들과의 비교 분석을 통하여 요소 기술 측면에서 몇 가지 우수성을 입증하였다. 향수 보완 연구가 필요한 사항 등을 정리 하면 다음과 같다. 첫째, 스마트 폰 등 휴대 기기 상에서 미리 기계 학습 과정을 거쳐 체계화된 지식들을 데이터베이스로 구축하고 키워드 탐색 기능을 활용해 학습을 위한 지능형 어플리케이션을 구현할 필요가 있다. 둘째, 무엇보다도 지식 표현 모듈의 한계성이 두드러지며 이 문제의 해결책으로 프레임 시스템의 도입을 고려하고 있으며 중간 단계로 의미론적 온톨로지의 기법을 활용한 방안에 대한 연구가 필요할 것이다.

참 고 문 헌

[1] Alvaro, A. K. S. (2010), Basic handwriting instructor for kids using OCR as an Evaluator, Networking and Information Technology (ICNIT), 2010 International Conference on. IEEE.

[2] B. Epshtein, E. Ofek, and Y. Wexler (2010), Detecting text in natural scenes with stroke width transform, CVPR, 2963 - 2970.

[3] Beller, William E., and Ynjiun P. Wang. (1997), Bar code dataform scanning and labeling apparatus and method, U.S. Patent No. 5,602,377. 1-16

[4] B. Williams, G. Klein, and I. Reid (2007), Real-time SLAM relocalisation. In Proc. 11thIEEEInt'lConf. ComputerVision, 1-8

[5] C. Mancas-Thillou and B. Gosselin (2007), Natural Scene Text Understanding, Vision Systems: Segmentation and Pattern Recognition.

[6] D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, IJCV 04.

[7] Economides, Anastasios A., and Nick Nikolaou (2008), Evaluation of handheld devices for mobile learning. International Journal of Engineering Educatiorn 24-1, 3-11

[8] Foresti, Gian Luca, et al (2005), Active video-based surveillance system: the low-level image and video processing techniques needed for implementation, Signal Processing Magazine, IEEE 22-2, 25-37.

[9] G. Schweighofer and A. Pinz (2008), Globally Optimal O(n) Solution to the PnP Problem for General Camera Models," British Machine Vision Conference 2008 (BMVC'08).

[10] Gorodnichy (2005), Dimitry. Video-based framework for face recognition in video.

[11] H. Bay, A. Ess, T. Tuytelaars, L. Gool (2008), SURF: Speeded Up Robust Features, Computer Vision and Image Understanding (CVIU), 110-3, 346-359.

[12] K.K. Kim and Y.K. Chung (2004), Scene Text Extraction in Natural Scene Images Using Hierarchical Feature Combining and Verification, Proc. Int. Conf. on CVPR, 2, 679-682.

[13] L. Quan and Z.D. Lan (1999), Linear n-point camera pose determination, IEEE Transactions on PAMI, 21-8, 774-780 .

[14] M. Zheng, X. Chen and L. Guo (2008), Stitching Video from Webcams, Lecture Notes In Computer Science; Vol. 5359 archive Proceedings of the 4th International Symposium on Advances in Visual Computing, Part II. 429-429

[15] Park, J, Leung, C.M., Wang, C. Shon, T. (2012), "Future Information Technology, Application, and Service", FutureTech 2012 Vol. 1, Lecture Notes in EE 164

[16] Rimal, Bhaskar Prasad, Eunmi Choi, and Ian Lumb (2009), A taxonomy and survey of cloud computing systems. INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on. IEEE. 44-51

[17] Roberts, Jeffrey A., Il-Horn Hann, and Sandra A. Slaughter. (2006), Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects, Management science 5-7, 984-999.

[18] Kolor autopano (2011), http://www.autopano.net/blog-en/tag/videostitching/.

[19] MindTree, (2011), http://www.slideshare.net/MindTreeLtd/mindtreevideo-analytics-suite-real-time-image-stitching-1135870.

신 현 경

2002년 8월 State University of New York at Stony Brook. 대학원 응용수학과(Ph.D.)

2007년 8월 현재 가천대학교 수학정보학과 조교수

관심분야: Image Processing, Neural Network, Machine Learning.

e-mail: hyunkyung@gachon.ac.kr