

다층신경망을 이용한 모바일 자동 변환 시스템

한은정[†], 장창혁^{**}, 정기철^{***}

요 약

모바일 기술의 발전으로 오프라인 콘텐츠가 아닌 온라인 매체로 다양한 영상 콘텐츠를 제공받는 수요층이 늘어나고 있다. 그러나 모바일 단말기의 작은 화면에 맞게 수작업으로 편집/수정하기 위해서는 비용과 노력이 많이 드는 단점이 있다. 따라서 본 논문에서는 영상 콘텐츠 가운데 가장 다양한 형태를 지닌 인쇄 만화책을 모바일 단말기 환경에 맞게 자동 변환하는 Automatic Comics Conversion(ACC) 시스템을 제안한다. 모바일 단말기 화면에 적합한 형태로 기존 오프라인 만화책 각 한 면의 프레임으로 분할하기 위해 다층신경망(MLP: Multi-Layer Perceptorn)을 이용하였으며, 각 프레임은 영상의 의미 구조(Semantic Structure)의 손실을 최소화하여 적합한 크기로 분할된 영상들을 자동 변환하여 제공한다. 또한 동적인 만화 영상을 애니메이션으로 제공하기 위하여 텍스처 분석 연구를 더 했다. 이에 본 연구는 만화뿐만 아니라 프레임 단위로 되어있는 사진, 웹사이트, 다양한 영상 등을 언제 어디서나 제공받을 수 있도록 모바일 단말기에 제공함으로써 효율성을 검증한다. 또한 오프라인의 정지 영상을 분할된 영상 프레임 정보를 통해 움직이는 영상으로 제공할 수도 있다.

Mobile Automatic Conversion System using MLP

Eunjung Han[†], Changhyuk Jang^{**}, Keechul Jung^{***}

ABSTRACT

The recent mobile industry is providing of a lot of image on/off-line contents are being converted into the mobile contents for architectural design. However, it is difficult to provide users with the existing on/off-line contents without any considerations due to the small size of the mobile screen. In existing methods to overcome the problem, the comic contents on mobile devices are manually produced by computer software such as Photoshop. In this paper, I describe the Automatic Comics Conversion(ACC) system that provides the variedly form of offline comic contents into mobile device of the small screen using Multi-Layer Perceptorn(MLP). ACC produces an experience together with the comic contents fitting for the small screen, which introduces a clustering method that is useful for variety types of comic images and characters as a prerequisite as a stage for preserving semantic meaning. An application is to use the frame form of pictures, website and images in order into mobile device the availability and can bounce back the freeze images contents into dynamic images content.

Key words: Content Re-authoring(콘텐츠 재활용), Mobile Content(모바일콘텐츠), MLP(다층신경망), K-means Clustering Algorithm(K-means 분할 알고리즘)

※ 교신저자(Corresponding Author) : 한은정, 주소 : 서울시 동작구 상도5동(153-743), 전화 : 02)812-7520, FAX : 02)812-3266, E-mail : ytghbn@nate.com http://eunjung.u.to
접수일 : 2008년 5월 30일, 완료일 : 2008년 12월 15일

[†] 준회원, 숭실대학교 문화콘텐츠기술 연구소 연구원

^{**} 숭실대학교 IT대학 미디어학부 미디어공학 (석사)

(E-mail : luee@ssu.ac.kr)

^{***} 정회원, 숭실대학교 IT대학 미디어학부 교수

(E-mail : kcjung@ssu.ac.kr)

※ 본 연구 또는 저서는 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임.

“ KRF-2007-357-D00234

1. 서 론

최근 정보 기술의 빠른 발전으로 멀티미디어 데이터가 유비쿼터스 환경에 맞게 모바일 환경으로 급속히 진화하고 있으며, 최근에는 모바일 기술의 발달로 오프라인 콘텐츠를 재구성하여 다양한 정보검색, 커뮤니케이션, 오락, 영상, 만화, 교육뿐만 아니라 신문, 사진등과 같은 다양한 콘텐츠를 모바일 단말기에 제공되고 있다[1-4]. 이를 위해 대용량의 콘텐츠를 빠르고 편리하게 관리, 검색하기 위해 프레임 단위로 분할하는 연구가 많이 진행되고 있다[5-8]. 기존에 모바일 콘텐츠를 제공하기 위한 방법으로 오프라인 콘텐츠를 오프라인 콘텐츠를 수작업을 제작하여 제공하고 있지만 다양한 형태와 수많은 콘텐츠를 제공하기 위하여 많은 시간과 비용 그리고 노동력이 소요된다. 이를 해결하기 위해 오프라인 콘텐츠 데이터를 자동으로 모바일 단말기에 변환 기술이 요구하고 있다.

오프라인 영상을 모바일 단말기에 적합하게 변환하는 연구로써, Chen[1] 등은 웹 페이지를 무선 단말기 상에 보여주기 위해 썸네일(thumbnail)방식으로 큰 영상을 작은 영상으로 변환하였고, Lin[2] 등은 웹 문서들을 자동으로 변환하는 시스템으로 온라인에서 웹 콘텐츠 마이닝을 통해 콘텐츠 탐색을 향상시켜 주는 기법을 제공하였으며, 김원철[9] 등은 방송사 웹 페이지 내에서 실시간으로 사용자가 선호하는 방송사 웹 페이지의 섹션을 추출하고, 모바일 환경에 적합하도록 각 섹션의 순서를 재구성하여 모바일 단말기에 제공하였으며, 강신상[10] 등은 모바일 단말기에서 영상 및 폰트를 효율적으로 처리할 수 있게 비트맵으로 처리하는 기법으로 영상을 제공하였다. Liu[3] 등은 큰 영상들을 모바일 단말기에 자동으로 변환하는 연구로써 영상의 특정 중요 영역 부분만 순차적으로 보여주는 기법을 연구하였으나, 특정 영역 부분을 각 장마다 추출하기에는 시간이 걸릴 뿐만 아니라, 특정 영역, 배경 그리고 특정 오브젝트 외에 다양한 영상 정보의 형태를 가지고 있는 영상을 제공하기 위한 방법으로는 적합하지 않다.

본 논문에서는 다양한 형태의 만화 영상을 모바일 단말기에 맞게 자동 변환하는 Automatic Comics Conversion(ACC) 시스템을 제안한다. 제안한 실험은 오프라인 콘텐츠 중 가장 많은 영상 데이터 정보

를 가진 콘텐츠 만화를 이용하였다. 만화는 하나의 문화 장르로써 자리 매김하고 있으며 다양한 매체와 폭넓은 수요층을 형성하고 있음에도 불구하고, 국내의 모바일 만화 콘텐츠 이용환경은 아직까지 콘텐츠 제작에 관한 연구에 대해 사회가 궁극적으로 지향하는 수준에는 도달하지 못한 상태에 있다. 그의 예로 국내 모바일 만화 콘텐츠를 제공하는 업체 중 시스튜디오에서는 인쇄 만화 콘텐츠의 특정 영역에 애니메이션을 적용하여 모바일로 서비스를 제공하고 있다. 또한 Wireless-watch. Community의 일본에서 열린 시사회에서는 다양한 콘텐츠 서비스를 제공하는데 있어서 “최근에는 3G 네트워크의 경제나 기술적인 이유 때문에 2G의 생존력이 줄어들며, 이동하기 쉬운 새로운 형태의 모바일 콘텐츠 서비스를 제공하는 것이 효율적이다.”라고 발표하였으며, 그 중 휴대폰에 만화책을 제공하기 위해 모바일 만화 콘텐츠를 제작하였다. 이와 같이 모바일 단말기에 제공된 만화 콘텐츠의 대부분은 스캔한 인쇄 만화의 임의 영역을 지정해 모바일 단말기 화면 크기에 맞게 수작업으로 편집/수정하거나 모바일용 만화를 제작하여 제공하기 때문에 많은 비용과 시간이 소요된다.

따라서 ACC 시스템은 다양한 형식의 인쇄 만화 콘텐츠의 스캔 영상을 모바일 단말기 화면에 맞게 프레임 단위로 분할함에 있어서 다중신경망(MLP: Multi-Layout Proceptorn)을 기반으로 하는 분할점 선택 기법을 사용한 X-Y 재귀분할 알고리즘을 사용함으로써 영상의 노이즈, 다양한 프레임 외곽 형식 등에 적응성을 기한다. 또한 영상처리나 패턴인식은 입력양이 많은 신경망 응용 분야 중 하나이다. 예를 들어 신경망을 이용한 컨벌루션(convolution)을 영상 전체 영역에 수행할 때는 상당한 수행 시간을 필요로 한다[11]. 본 논문에서는 모바일 화면보다 큰 프레임 영상에서 만화내용을 이해하기에는 불필요한 배경을 자르고 중요한 부분(배경이 아닌 모든 부분: 사람, 중요한 오브젝트)만을 제공한다. 프레임 단위로 되어 있는 만화 영상을 분할하기 위해 k-means 알고리즘을 이용한 텍스처 분석 기법을 사용한다. 이와 같이 특정 만화 영상을 프레임 단위로, 또한 의미 손실을 줄이는 긴 프레임 분할 방법을 사용함으로써 다양한 종류의 만화 영상을 모바일 영상으로 변환 시스템을 구성할 수 있다.

본 논문의 구성은 다음과 같다. 제 2장에서는 다중

신경망을 이용한 프레임 분할 방법, 긴 프레임 분석 후 자르는 방법에 대해 설명한다. 제 3장에서는 제안된 방법을 이용한 실험 결과와 분석을 보이고, 제 4장에서는 결론과 향후 연구 분야를 기술한다.

2. MLP를 이용한 프레임 분할

인쇄 만화에는 다양한 프레임 형식이 있다(그림 1). 여러 형식들 중 정형화된 프레임 추출을 위한 일반적인 방법인 X-Y 재귀분할 만을 이용한 방법[12]은 구현이 쉬운 반면, 프레임 크기, 프레임 간의 거리, 프레임 주변의 노이즈 등에 많은 영향을 받으며, 정밀한 영상 분할 알고리즘이 필요하기 때문에 인쇄 만화 영상과 같은 잡음이 많은 영상이나 저해상도 영상 또는 다양한 모양과 크기의 영상에는 적합하지 않다.

본 논문에서는 정형화된 프레임을 가진 만화 콘텐츠의 분할을 위해서 기존의 일반적인 X-Y 재귀분할 분할 알고리즘을 사용하면서, 다양한 환경에 적응력

을 제공하기 위해서 학습 기반의 신경망을 이용하여 영상을 분할할 지점(cutting point candidate)를 생성한다. 그림 3과 같이 프레임 추출의 전처리 단계(pre-process step)에서는 영상 윤곽의 특정 영역을 MLP를 적용하여 분할 지점을 찾으며, 투영 프로파일을 통해 미리 학습된 사용자가 마우스로 표시한 분할점 데이터를 찾는다. 분할된 영상이 모바일 단말기 화면에 맞게 제공하기 위해서 분할된 긴 영상의 프레임을 의미영역(semantic regions)을 찾아 특정 영역만을 제공한다(그림 2).

2.1 전처리

프레임 추출의 전처리 단계로써, 임계값(threshold)으로 영상을 이진화한 후 X-Y 재귀분할 알고리즘의 분할점(cutting point)을 추정하기 위해서 투영 프로파일의 특정값 이상인 영역에 대해서만

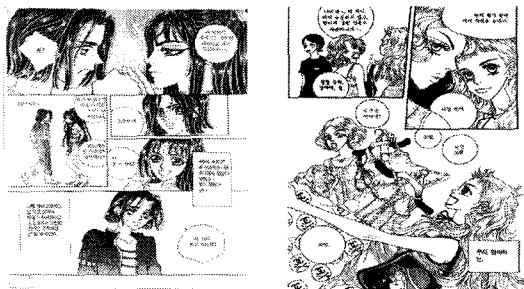


그림 1. 일정하지 않은 프레임 영상: (a)프레임과 프레임 사이의 배경에 말풍선이 있는 영상, (b)정사각형이 아닌 프레임 영상, (c)프레임이 없이 배경으로 만화의 내용이 있는 영상, (d)프레임과 프레임 사이에 문자가 있는 영상

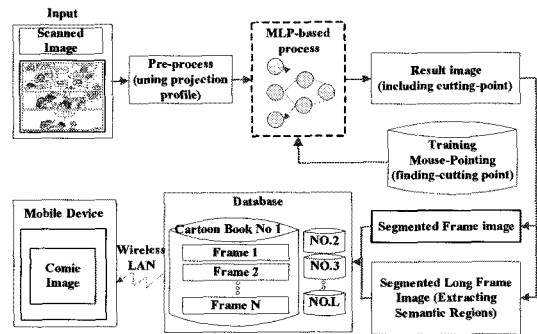


그림 2. ACC시스템의 구조

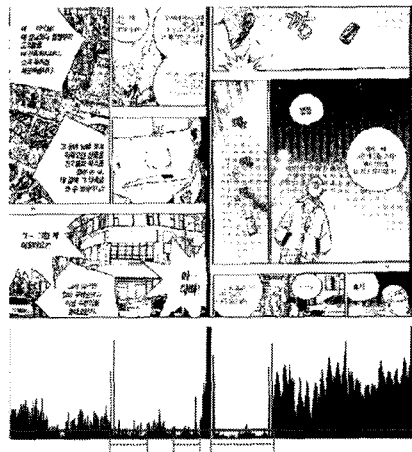


그림 3. 실험 과정의 입력 영역

프레임 분할 신경망을 적용함으로써 정확도의 향상과 수행시간의 효율성을 기한다. X-Y 재귀분할 알고리즘은 기본적으로 탑-다운(Top-down) 방식으로 픽셀을 읽어 들여 반복적으로 영상을 사각형으로 블록화해 여러 조각으로 분할하는 기술로써, 각 과정에서 픽셀로 잡힌 윤곽들이 계산된다[13-15]. 본 논문에서는 실험 과정의 입력을 생성하기 위해서 만화 영상에 대한 투영 프로파일 방법을 사용하였다. 그림 3과 같이 느슨한 임계값을 사용하여 후보 분할점을 찾기 위해 영상의 히스토그램을 사용하였으며, X-Y 투영 프로파일에서의 위치는 실험과정에서의 입력값을 나타낸다. 이러한 방법은 효율적인 처리 시간을 위해 영상 전체보다 적은 부분을 입력값으로 사용하게 된다.

2.2 신경망 구조

신경망이라고 불리는 인공신경망(artificial neural network)은 인간 두뇌의 동작 방식을 따라 구현되었다. 다양한 모양의 신경망이 사용되고 있지만 대표적으로 multiayer perceptron, ling vector quantization, radial basis function, Hopfield, kohonen등을 주로 많이 사용한다[16,17].

본 논문에서 구현하고자 하는 신경망은 다층 퍼셉트론이다(그림 4). MLP는 48개, 은닉층(hidden layer) 40개를 가지며 1개의 출력 노드로 구성되어 있다. MLP는 인접한 층(layer)의 노드들이 완전연결(fully-connected)되어 있고, 층의 개수나 각 층의 노드수 등에서 변화가 있을 수 있지만 기본적으로 각

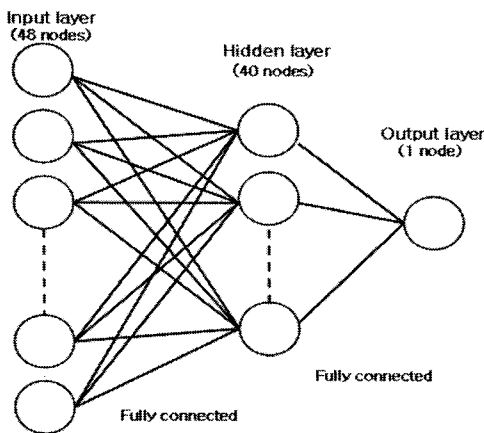


그림 4. 두 단계 신경망 구성

노드는 웨이트벡터와 해당 노드의 입력벡터의 내적 연산을 수행한 후 활성화함수 연산을 수행한다. 그림 4는 2단계 신경망의 구성을 나타내며, BP(Back Propagation) 학습 알고리즘을 사용하였다. 신경망의 입력으로 30×40 크기로 정규화된 이진 영상으로부터 추출된 48차원 벡터가 사용되며, 48개의 정수값은 정규화된 이진영상에서 각 5×5 윈도우의 픽셀 개수로부터 얻어진 25개의 휘도(intensity)는 0과 1사이로 정규화 된다. 결국 48개의 값은 열 우선 순위(column major order)에 의해 신경망의 입력값으로 사용된다. 만화 영상의 위치를 마우스로 표시한 점을 MLP를 적용하여 그 점을 영상을 분할하기 위한 프레임 경계로 인식을 한다. MLP의 첫번째 은닉층의 노드들부터 수행하여 차례로 출력층까지 계산한다. MLP 노드들과 유사한 연산을 수행하기 때문에, 본 논문의 연구 결과는 쉽게 다른 신경망에 적용될 수 있다. 이와 같이 원하는 분할점을 수작업으로 결정하여 48차원 벡터와 함께 저장된다. 그림 5는 신경망의 입력값을 구하는 과정을 나타낸다.

MLP의 출력값은 0 혹은 1이 되며, 전처리 단계에서 48개의 노드를 가진 30×40 입력 픽셀로 분할되는 입력 영상은 MLP의 출력값을 가진다. 만약 30×40 영역의 MLP 출력값이 1이면 해당 영역에 분할점을 표시한다. 또한 MLP의 출력값이 0일 경우 프레임 해당 영역은 프레임 경계가 아님을 나타낸다. 프레임 경계선은 분할 지점을 나타내며, 분할된 영역은 MLP 결과로 알아볼 수 있다. 그림 6과 같이 인위적으로 오프라인 만화 영상을 학습 영상으로 분할할 수 있다. 비록 프레임들이 얼마간의 잡음을 포함하고는 있지만, MLP는 프레임 경계를 알아낼 수 있다. 이와 같이 경계선에 분할 지점을 찾을 수 있지만 MLP는 아직 정확하게 찾을 수 없다. 만화 영상은

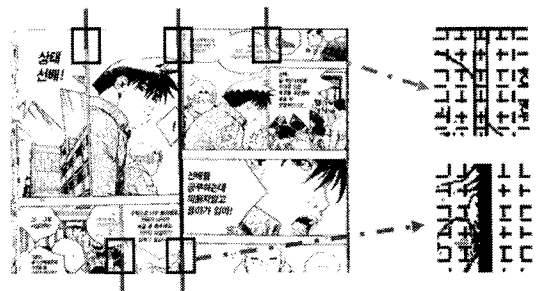


그림 5. 만화영상의 MLP입력 데이터

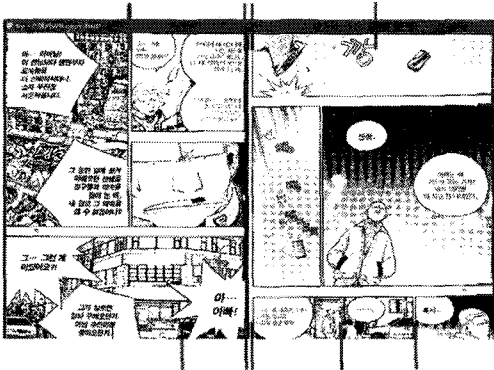


그림 6. 분할 지점을 찾은 결과 영상

프레임 경계선에 가까이 다른 오브젝트나 노이즈가 있거나, 프레임의 선과 비슷한 영상의 특성이 프레임 안에 있게 되면 프레임 경계와 비슷한 영상 영역으로 찾게 되며, MLP만을 이용하게 되면 완전하게 프레임 분할할 영역을 찾지 못하는 단점이 있다(그림 7).

이를 해결하기 위하여 Sung과 Poggio[18]에 의해 제시된 부트스트랩 방법을 사용한다. 이 방법은 얼굴

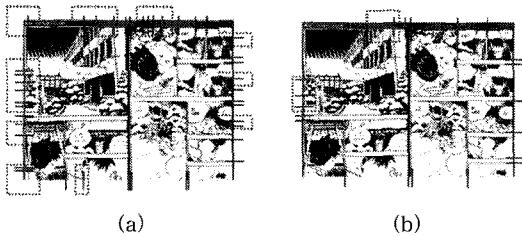


그림 7. 프레임 분할의 오류 영상: (a) MLP만을 이용한 영상, (b) MLP방법에 부트스트랩 방법을 이용한 결과 영상

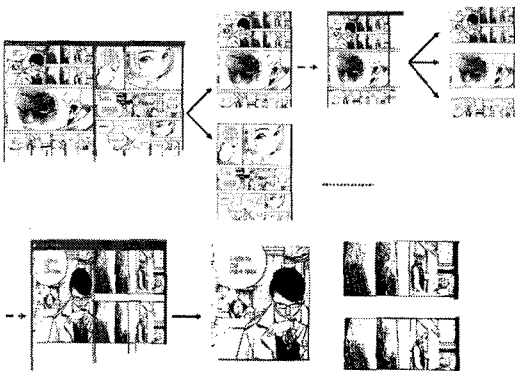


그림 8. 프레임 분할 영상 결과

신경선 탐지를 위해 예전부터 전개되던 방법으로써, 프레임이 없는 샘플들을 선택하여 실험에 사용한다. 또한 부분적으로 학습된 MLP는 좀 더 완벽한 분할을 위해 여러 이미지들에 대해 반복적으로 적용되며 정확한 출력을 가진 패턴들은 프레임 단위로 되어 있지 않는 영상의 예제 파일들은 학습데이터에 추가된다. 만화의 영상의 순서는 영상의 첫 프레임부터 데이터가 시작되며 첫 프레임에 분할하는 영역이 있으면 그 영상을 두 번째 데이터로 순서를 정한다(그림 8).

3. 의미구조(Semantic Structure)를 고려한 긴 프레임 분할

일반적으로 모바일 단말기의 화면은 가로의 길이보다 세로의 길이가 긴 반면에, 만화 영상은 가로의 길이가 긴 영상보다 세로의 길이가 긴 영상이 더 많다. 즉 모바일 단말기 화면에 영상이 보여질 때에 과도한 횡적 축소가 일어나게 되는데, 본 논문에서는 이를 방지하기 위하여 프레임이 다음과 같은 조건을 만족할 경우 프레임을 자르는 과정을 수행한다.

$$\frac{I_w}{I_h} \geq \frac{M_w}{M_h} \times 1.3 \quad (1)$$

I_w 는 만화 영상의 가로, I_h 는 만화 영상의 세로, M_w 는 모바일 단말기 화면의 가로, M_h 는 모바일 단말기 화면의 세로 길이를 의미한다. 즉, 만화 영상의 세로 길이를 모바일 단말기 화면의 세로 길이에 맞추어 영상을 확대 또는 축소 시에 변환된 영상의 가로 길이가 화면의 가로 길이보다 130% 이상일 경우 프레임을 자르는 과정을 수행한다. 130% 이하의 영상의 경우 모바일 단말기 화면에 제공하였을 시에 영상의 내용을 깨뜨리지 않는 정도의 축소가 발생한다. 그리고 130% 이상의 영상의 경우 모바일 단말기 화면에 적용하기 위한 과도한 축소가 발생한다. 이를 극복하기 위해 영상의 의미 정보가 아닌 외곽 부분을 모바일 단말기 화면에 맞게 제공하기 위하여 영상의 정보를 분리한다.

본 논문에서의 만화 영상은 주로 흑백 영상이며, 만화 영상의 각 세로줄의 정보량은 각 줄의 서로 다른 질감(texture) 수와 비례한다고 가정한다. 그림8은 영상에서 같은 질감을 갖는 블록들의 군집화에 적용한 K-means 클러스터링 알고리즘이다. n 은 각

블록의 질감(v)의 개수, k 는 군집의 개수를 나타내며, 무작위 v 중 k 개를 선택하여 군집의 초기 평균(mean)값으로 설정하여 거리를 계산하여 가장 작은 군집에 속하게 한다. 각 만화 영상에 대한 텍스처 분할 완료 후에 의 영상 내의 각 화소들은 각 군집에 속해진 v 값들로 군집의 평균값을 새롭게 정해준다(그림 9)[19].

그림 10에서는 긴 프레임 만화 영상 I 를 $M \times N$ 개의 블록으로 나누고, 각 블록의 크기는 (i, j) 로 표현한다. K-means 클러스터링에 사용되는 입력값은 그림 10의 (a)와 같이 블록의 질감 v 는 블록의 중심을 기준으로 가로, 세로, 대각선에 위에 존재하는 $2 \times (i+j) - 4$ 개의 픽셀이며, 이들 픽셀들의 명암(Intensity)값을 사용한다.

- Begin initialize n, k
divide the image to n components
initialize textures of n components using pixels values
randomly choose k means among n components
- Do
classify each n component to the cluster which has the nearest mean from the component
recalculate mean of cluster
until mean equals to previous mean of cluster
end

그림 9. K-means 클러스터링 알고리즘

긴 프레임 만화 영상 I

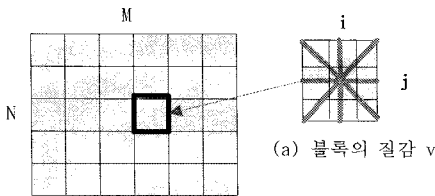


그림 10. 긴 프레임 만화 영상의 질감 영상 분석

$$precision(\%) = \frac{\# \text{ of correctly detected cutting points}}{\# \text{ of detected cutting points}} \times 100 \quad (2)$$

$$recall(\%) = \frac{\# \text{ of correctly detected cutting points}}{\# \text{ of cutting points}} \times 100 \quad (3)$$

K-means 클러스터링 알고리즘을 통해 파악된 만화 영상의 의미구조를 바탕으로 모바일 단말기에 적절한 만화 콘텐츠를 생성한다. 그림 10와 같이 같은 정보의

질감을 가진 컴포넌트들이 모여 군집화된다. 따라서 우리는 다른 질감을 가지는 군집들이 많이 모여있는 영역을 보존하고 그 외의 영역을 삭제하는 방법을 사용한다.

그림 11에서는 긴 프레임을 모바일 단말기에 적용하기 위해 영상 상의 컴포넌트로 분할된 세로줄(Y)을 가로줄(X)로 분석하여, 유사한 질감을 가진 블록들끼리 모여 몇 개의 군집으로 이루어졌는지 분석한다. t 는 세로줄에 존재하는 재 군집한 서로 다른 군집의 총 개수이며, 세로줄의 t 를 각각 줄마다 비교하여, t 의 개수가 많을 때 복잡도가 높다고 인식한다. 영상 외곽에서부터 영상에 정보가 작은 군집의 개수가 나온 곳을 S (시작점)과 E (끝점)이라 하며 그 부분만 분할하여 모바일 화면에 제공한다.

그림 11은 긴 프레임 분할을 실험한 영상으로 그림 12에서의 방법을 직접 실험한 결과로써, 긴 프레임의 주요한 정보가 아닌 부분을 분할한 것을 그래프로 제시한다.

4. 실험 및 결과

4.1 프레임 분할 MLP 성능 분석

MLP의 학습을 위하여 만화 영상 30장을 사용하였

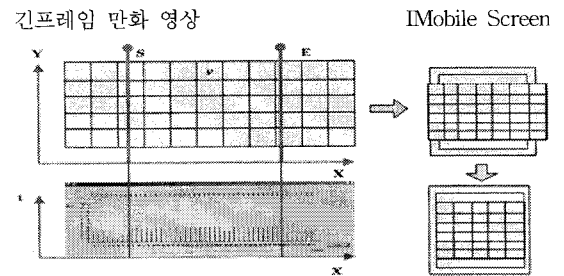


그림 11. 긴 프레임 영상 모바일 단말기에 자동

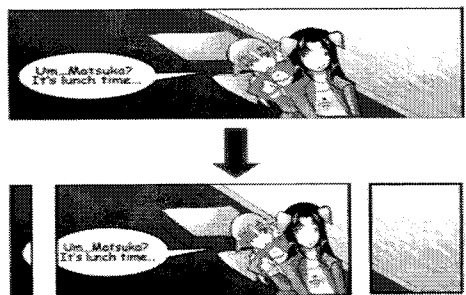


그림 12. 긴 프레임 자른 예

으며, 많은 연구 결과에서 프레임 분할 결과의 정확도를 비교하기 위해, X-Y 재귀분할 방법을 이용한 방법과 MLP를 이용한 방법에서의 정확도를 아래의 재현률(precision rate)과 정확률(recall rate)를 이용하여 측정하였다. 재현률은 찾아내야 할 총 분할점의 개수 대 시스템이 찾아낸 정확한 분할점의 개수의 비율로서, 시스템의 이해력을 나타낸다. 정확률은 시스템이 찾아낸 총 분할점의 개수 중 정확하게 분할점의 개수 비율로써 시스템의 정확도를 나타낸다.

그림 13은 MLP를 방법을 통한 프레임 분할과 X-Y 재귀분할 알고리즘을 이용한 방법에 대한 실험 결과 영상이다. 각 그림에서 13(a)는 X-Y 재귀분할 결과 영상이고 13(b)는 MLP 방법을 이용한 프레임 분할 영상이다. 기존의 X-Y 재귀분할을 이용한 방법은 입력영상에 노이즈가 많거나 수평/수직선이 아닌 프레임은 정확하게 분할하지 못한다. 이런 이유로 그림 14와 같이 MLP를 사용하지 않은 방법은 분할 프레임 아닌 영역에도 분할이 되어 재현율보다 정확도 비율이 낮은 것 같이 보이지만, MLP를 사용한 학습 과정을 통해 불필요한 분할점을 최소화함으로써, 프레임을 분할하는데 있어서 더욱더 효율적인 것을 알 수 있다.

실험 결과 만화 영상이 120장이 각 프레임으로 분할되었으며, 나머지 30장은 아래 그림 14와 같은 이유로 분할되지 않았다. 그림 14(a)(b)에서는 잘려질 프레임에 문자가 겹쳐지는 영상, 큰 프레임의 가로 세로 비율이 맞지 않는 경우 프레임이 분할되지 않았으며, 그림 14(c)는 긴 프레임으로 잘려진 프레임 중에서 프레임 중간에 문자가 있는 경우이다. 또한 모바일 화면에 맞게 제공하기 위해 긴 프레임 분할 하지만, 배경에 글자 있는 경우는 적은 정보를 갖고 있기 때문에 분할된다.

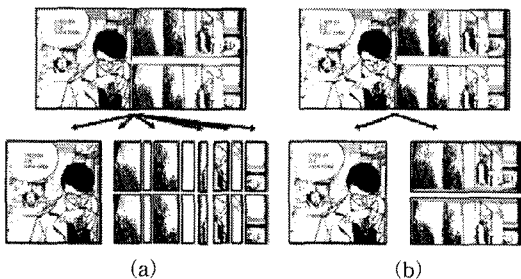


그림 13. 프레임 분할 비교 방법 예: (a) X-Y 재귀분할 결과, (b) MLP-based 분할 결과

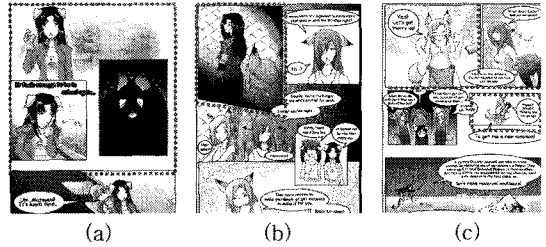


그림 14. 페이지 프레임 자르기 힘든 실험예제: (a) 배경 프레임 일정하지 않게 배치되어 있는 프레임, (b) 정사각형이 아닌 프레임, (c) 말 풍선이 프레임과 겹치는 영상

4.2 실험 분석

본 시스템에 사용한 PDA는 Pocket-PC 2003기반의 카메라가 장착된 POZ x301을 사용하였다. POZ x301 모델은 400MHz 인텔 프로세서를 사용하며, 62MB SDRAM /160MB Flash ROM을 사용한다. 그림 15는 긴 프레임 만화 영상을 모바일 단말기 화면에 자동 변환한 결과 화면이다. 그림 15(a)는 원본 영상이며, 그림 15(b)는 MLP 방법을 통해 프레임 분할한 영상이며, 그림 15(c)는 추출한 프레임과 추출한 문자를 모바일 단말기에 제공한 결과영상이다. ACC 시스템은 프레임 분할을 한 후 문자 추출을 하여 자동으로 변환하는 시스템이라면, 기존 방법은 수동으로 포토샵 작업을 하여 모든 만화를 수동으로 제공한다(그림 15).

기존 영상의 방법에 대한 편의를 위하여 30장을 수작업 했을 경우의 수행 시간과 제공된 자동 변환 시스템을 사용한 수행 시간을 측정 한 결과이다. 표 1에서는 만화 콘텐츠 공급에 있어서 가장 큰 걸림돌이었던 오프라인 콘텐츠의 디지털화에 필요한 시간의 비효율성과 알기 위한 실험으로써, 기존 방법은 모바일 단말기에서 만화를 볼 수 있게 하는 시간이 더 많이 소요 된다는 것을 알 수 있었다.

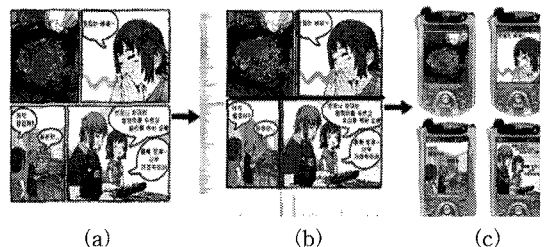


그림 15. 결과 화면: (a) 원본 영상, (b) 프레임 분석 (프레임 분할), (c) ACC를 적용한 모바일 단말기 화면상에 제공된 영상

표 1. 수작업과 자동변환 시스템의 수행시간 결과

	수작업	자동변환시스템
수행시간	10 (min)	1.5 (sec)

5. 결 론

기존의 오프라인 영상 콘텐츠를 자동으로 제공하기 위해서는 많은 시간과 노력이 필요할 뿐만 아니라 콘텐츠마다 다양한 형태와 모양으로 제작되었기 때문에 자동으로 변환하는 시스템을 연구하기에는 많은 어려움이 있다. 본 논문에서는 MLP를 사용함으로써 다양한 환경의 만화 영상에 대한 프레임 분석기를 별도의 특징값 추출 없이 자동으로 구현하였으며, K-means 알고리즘을 긴 프레임 영상을 분석하여 불필요한 영상의 외곽 부분을 분할하는 방법으로 실효성을 검증하였다.

향후 연구 과제로는 고정되어 있지 않은 영상을 모바일 단말기에 제공하는 연구를 통해 모든 인쇄영상을 자동으로 모바일 단말기에 제공할 것이며, 문자 부분이 잘리는 단점은 문자 영역의 추출을 통해 문자의 이동/변환 방법을 적용시킴으로써 이러한 문제점을 해결할 것이다. 이를 확장하여, 만화 콘텐츠 자동 변환 시스템의 지속적인 개발로 보다 효과적인 콘텐츠와 저렴한 비용의 만화 콘텐츠 영상의 공급에 관한 연구를 할 예정이다. 뿐만 아니라 프레임으로 분할할 수 있는 콘텐츠(신문, 문서 그리고 사진) 등에도 활용이 가능하다.

참 고 문 헌

[1] Y.Chen, W.Y.Ma and H.J.Zhang, "Detecting Web Page Structure for Adaptive Viewing on Small Form Factor Devices," International WWW Conference, pp. 225-233, 2003.
 [2] Baldonado Lizhen Lin, Junjie Chen and Hantao Song, "The research of Web mining," World Congress on Intelligent Control and Automation, Vol.3, No.4, pp. 2333-23337, 2002.
 [3] Hao Liu, Xing Xie, Wei-Ying Ma and Hong-Jiang Zhang, "Automatic Browsing of Large Pictures on Mobile Devices," Eleventh ACM International Conference on Multimedia,

pp. 148-153, 2003.
 [4] J. Casares, K. Cross and A. Klein, "Redesigning Traditional media and Assessing Entertainment value With Online Comics," Conference on Human Factors in Computing systems CHI'01, Vol.34, No.10, pp. 427-428, 2001.
 [5] D.Wang and S.N.Srihari, "Classification of newspaper image blocks using texture analysis," Computer Vision Graphics, and Image Processing, Vol.47, pp. 327-352, 1989.
 [6] K. etemad, D.S. Doermann, and R. Chellappa, "Multiscale segmentation of unstructured document pages using soft decision integration," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, pp. 92-96, 1997.
 [7] N. Amamoto, S. Torigoe, and Y. Hirogaki, "Block Segmentation and Text Area Extraction of Vertically/Horizontally Written Document," Document Analysis and Recognition, pp. 739-742, 1993.
 [8] J.Li and R.M.gray, "Text and picture segmentation by distribution analysis of wavelet coefficients," International conference on Image Processing, (Chicago, Illinois), pp. 790-794, 1998.
 [9] 김원철, 이수철, 황인준, 변광준, "모바일 환경에서 사용자 검색 성향을 반영한 웹 방송 정보 재구성 기법," 정보처리학회논문지 D, 제11-D권, 제5호, pp. 149-1158, 2004.
 [10] 강신상, 옥경달, 이상범, "모바일 단말기 상에서의 효율적인 영상 및 폰트 처리," 한국정보과학회지, 제31권, 제1호, pp. 0685-0687, 2004.
 [11] Haykin, "Neural Networks: a comprehensive foundation," Prentice Hall, 2003.
 [12] H. Fujisawa and Y. Nakano, "A Top-Down Approach for the Analysis of Documents," Pattern Recognition, No.10, pp. 113-122, 1990.
 [13] Y. Chenevoy and A. Belaid, "Hypothesis Management for Structured Document Recognition," Document Analysis and

Recognition, pp. 121-129, 1991.

- [14] R. Ingold and D. Armangil, "A Top-Down Document Analysis Method for Logical Structure Recognition," Document Analysis and Recognition, pp. 41-49, 1991.
- [15] Eunjung H., Sungkuk J., Anjin P. and Keechul J, "Automatic Conversion System for Mobile Cartoon Contents," International Conference on Asian Digital Libraries, Vol. 3815, pp. 416-423, 2005.
- [16] J. Ha, R. Haralick, and I. Phillips, "Document Page Decomposition by the Bounding-Box Projection Technique," Document Analysis and Recognition, pp. 1,119-1,122, 1995.
- [17] K. Kise, O. Yanagida and S. Takamatsu, "Page Segmentation Based on Thinning of Background," The 13th Int'l Conf. Pattern Recognition, pp. 788-792, 1996.
- [18] K. K. Sung and T. Poggio, "Example-based Learning for View-based Human Face Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.20, No.1, pp. 39-51, 1998.
- [19] Anil K. Jain and Kalle Karu, "Learning Texture Discrimination Masks," IEEE Transactions on pattern analysis and machine intelligence, Vol.18, No.2, pp. 195-205, 1996.



장 창 혁

2005년 숭실대학교 미디어학과 학사(공학사)
 2009년 숭실대학교 미디어학과 석사(졸업예정)
 2009년 숭실대학교 미디어학과 박사과정(입학예정)

관심분야 : 영상처리, 컴퓨터 비전, HCI

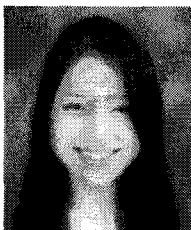


정 기 철

1994년 경북대학교 컴퓨터공학과(공학사)
 1996년 경북대학교 대학원 컴퓨터 공학과(공학석사)
 1999년 Intelligent User Interfaces group at DFKI (The German Research

Center for Artificial Intelligence GmbH), Germany, 방문연구원

2000년 Machine Understanding Division, Electro Technical Laboratory in Japan, 방문연구원
 2000년 경북대학교 대학원 컴퓨터공학과(공학박사)
 2000년~2002년 PRIP lab, Michigan State University Postdoc
 2003년~현재 숭실대학교 IT대학 미디어학부 교수
 관심분야: HCI, 콘텐츠공학, 인터랙티브 게임, 영상처리 / 컴퓨터 비전, 증강현실, 인공지능



한 은 정

2003년 신라대학교 국어국문학과(인문학사)
 2005년 동서대학교 소프트웨어대학원 소프트웨어학과(공학석사)
 2007년 숭실대학교 IT대학 미디어학부 콘텐츠 공학(공학박사)

2007년~2008년 University of British Columbia MAGIC (Post-Doctor)
 2009년 숭실대학교 IT대학 문화콘텐츠연구소 연구원
 관심분야: 콘텐츠공학, E-Learning, M-Learning, HCI, 영상처리, 증강현실, 인터랙티브 아트