

## Neural Network을 이용한 이미지 장르 분류 시스템

안재훈<sup>0</sup>, 이한구<sup>1)</sup>, 주현호,  
건국대학교

anjh7670@konkuk.ac.kr, hlee@konkuk.ac.kr, kkabul2@konkuk.ac.kr

### Neural Network Based Image Genre Classification

Jaehoon Ahn<sup>0</sup>, Hanku Lee<sup>1)</sup>, Hyunho Ju,

School of Internet & Multimedia Engineering at Konkuk University, Seoul, Korea

#### 요 약

본 논문에서는 neural network을 이용한 이미지 장르(유형) 분류 시스템을 소개한다. 이 논문에서 제안된 시스템은 이미지를 예술(art), 사진(photo), 만화(cartoon) 이미지라는 세 가지 장르(유형) 중 하나로 분류한다. 이미지의 특성은 표준 MPEG-7 visual descriptor를 사용하여 추출된 후, neural networks를 이용하여 학습된다. 시뮬레이션 결과는 제안된 시스템이 80% 이상의 이미지들을 정확한 장르(유형)로 분류하는 것을 보여준다.

#### 1. 서론

디지털콘텐츠 시대가 도래 하면서, 우리는 매일 인터넷과 TV 상에서 전통적인 텍스트 데이터보다 훨씬 많은 디지털 데이터를 접하고 있다. 인터넷 상에서 이용자들의 정보에 대한 주된 관심은 텍스트 형식에서 디지털 멀티미디어 형식으로 급속히 전환 중이다. 디지털 콘텐츠 이미지들의 수가 급격히 늘어나고, 쉽게 접근될 수 있게 되어 짐에 따라, 우리는 새로운 문제 직면하게 된다. 예술, 사진, 만화 이미지 등과 같은 디지털 콘텐츠들을 어떻게 하면 쉽게 접근할 수 있도록 잘 분류할 수 있을까? 현재 인터넷 상에는 수많은 예술, 사진, 만화 이미지들이 있다. 디지털 멀티미디어 이미지를 이용하는 이용자들은 이러한 디지털 콘텐츠의 바다에서 그들이 좋아하는 것을 보다 효율적으로 선택해서 볼 수 있기를 원한다.

이 논문의 주목적은 MPEG-7 descriptors들을 이용하여 MPEG-7 표준을 디지털 멀티미디어 이미지에 적

용하는 것이다. 우리는 MPEG-7 descriptor를 분석하여, visual 환경에서 예술, 사진, 만화 이미지와 같은 디지털 이미지 장르(유형)를 분류하는데 이용할 수 있는 prototype 시스템을 만들었고, 실험을 통해 이미지 분류의 효과적인 방법론을 소개한다. 우리는 이 논문에서 이미지 장르(유형)의 구분을 위해 neural network을 사용하였다. neural network에 입력되는 값은 MPEG-7에 의해 추출된 시각 특성들의 여러 값들이다.

Chapter 2에서 우리는 이미지 장르(유형) 분류에 관련된 연구들을 논의하고, Chapter 3에서는 우리의 neural network 기반 이미지 장르(유형) 분류 시스템을 제안한다. 시스템의 시뮬레이션 결과는 Chapter 4에 제시되어 있으며, Chapter 5에서 결론을 맺는다.

#### 2. 관련연구

##### 2.1 MPEG-7 Descriptors and Image Classification Systems

MPEG-7은 이미지 분류 시스템에서 최근에 사용되도록 만들어진 표준이다. MPEG-7은 영상과 음성을 실제로

1) 교신저자: +82-2-2049-6082, hlee@konkuk.ac.kr

인코딩하고 디코딩하는 것에 관련된 표준이 아니라, 미디어 콘텐츠를 기술(묘사, describing)하는 표준이다. MPEG-7은 이러한 메타데이터(다른 데이터를 설명해주는 데이터)를 저장하기 위해 XML을 사용한다. MPEG-7은 시각 이미지(visual image) 콘텐츠를 기술(describe)하는 표준의 부채를 해결해준다. MPEG-7 표준의 목적과 영역 및 세부사항은 Sikora의 논문[1]에 잘 설명되어 있다.

여러 가지 MPEG-7 descriptor를 이용하는 많은 연구들이 진행되었다. Ro와 그의 연구진[2]은 texture 기반의 이미지 descriptor와 MPEG-7의 동일한 texture descriptor의 개선된 버전을 사용한 검색 기법을 연구하였다. 이미지 분류의 다른 연구들은 contour-based shape descriptor[3], histogram descriptor[4], 그리고 color structure와 homogeneous의 조합[5]과 같은 descriptor를 사용한다. EU aceMedia 프로젝트 연구의 일환으로써, Spyrou와 그의 연구진은 여러 종류의 하위 수준 MPEG-7 영상 descriptor들의 연동에 기반한 세 가지 이미지 분류 기법을 제안하였다[6]. descriptor를 직접 삽입하는 것은 부적절하고 호환성이 없기 때문에, 목표의 의미 분류와 하위 수준 descriptor와의 semantic gap을 이어주기 위해 융합이 요구된다.

CBIRS는 neural network와 MPEG-7 표준을 결합한 것이다. Helsinki University of Technology의 연구진은 그들의 콘텐츠에 기반한 이미지들을 검색하는 스스로 조직되는(self-organizing) 시스템인 PicSOM[7]을 개발했다. 이 기법은 그림샘플과 관련 피드백에 기초를 둔다. PicSOM 시스템은 트리 구조의 SOM을 이용하여 실행된다. PicSOM 시스템에는 MPEG-7 콘텐츠 descriptor가 제공된다. 논문 [7]에서 PicSOM 인덱싱 기법을 벡터 양자화(vector quantization, VQ)에 기반한 대조 시스템과 비교한다. 그들의 연구 결과는 유클리드(Euclidean) 거리 계산법이 모든 경우 최적이지 않음에도 불구하고, MPEG-7 콘텐츠 descriptor가 PicSOM 시스템에서 사용될 수 있음을 보여준다.

Kim과 Lee [8, 9]는 MPEG-7을 사용하는 neural network 기반의 분류 모듈을 제안했다. 이 모델에서, neural network의 입력은 이미지에서 추출된 MPEG-7

descriptor의 특성 값이다. 여러 가지의 descriptor들은 주어진 이미지의 특정 성분을 나타낼 수 있기 때문에, 성인(adult) 이미지 분류를 위해 최적의 방식을 선택하기 위한 적절한 평가 과정이 요구된다. 성인 이미지 분류 시스템에서 color와 관련된(Color Layout, Color Structure) descriptor들이 매우 효과적인 결과를 보여준다.

## 2.2 Image Genre Classification

사진 혹은 만화를 포함한 그래픽과 같은 WWW 이미지의 자동 분류는 잘 알려진 콘텐츠 이미지 분류다. 이것의 예로는 WebSeek 검색엔진 [10]이 있으며, 논문 [11]은 WebSeek에 관해 기술하고 있다. 유감스럽게도, WebSeek은 만화 이미지에는 존재하지 않는 aspect ratio와 이미지 URL에서 나타나는 단어와 같이 두드러지는 웹 이미지의 특성을 이용한다. 논문 [12]에서는 논문 [10, 11]과 같은 방식으로, 만화의 기본적인 특성과 1200개 샘플 데이터베이스 상의 사진과 그래픽 자료를 분류하기 위해 구현된 9가지 color descriptor를 강조한다.

논문 [13]은 실시간으로 연속된 프레임의 특정 색, 구성, 움직임 등을 분석하여 cartoon 혹은 non-cartoon로서 mpeg-2 비디오 시퀀스를 분류하는 접근 방식을 제안한다. 이는 만화, 광고, 음악, 뉴스, 스포츠와 같은 TV 방송 장르(유형)들이 연구되는 곳에 잘 알려진 영상 장르(유형) 분류 문제이다. 이 시스템에서 visual descriptor에서 추출된 특성들은 S형 함수로 가중치된 비선형(non-linear weighed with sigmoid function)이며, 후에 안정적인 인지를 할 수 있도록 multi layer perceptron을 사용하여 결합된다.

논문 [14]는 만화 혹은 사진 이미지로써의 개별 비디오 프레임들을 분류하는 접근 방식을 제안한다. 이 방식은 포물선 분포의 패턴 스펙트럼과 이미지 압축률에 의해 근사되는 이미지 신호의 복잡성에 기반한 descriptor를 소개한다. 이 방식은 데이터 수집을 위해 논문 [12]에서 제안된 the farthest neighbor histogram descriptor를 사용한다. 하지만 이 방식은 오차율이 개선되지 않으면서, 계산 비용이 많이 든다.

논문 [12]의 사진/그래픽 분류 시스템은 이미 Acoi 시스템 [15]의 일환으로 개발되었다. Decision rule classifier는 [16] 다양한 수량의 학습 데이터 상에서 논문 [12]에서 주어진 특성들을 학습하였다. 실험결과는 www 상에서 수집된 14,040개의 사진과 9,512개의 그래픽 데이터 집합 상에서 0:9라는 분류 결과를 보여 주었다.

논문 [17]은 motion만을 사용하는 만화 같은 video fragment들의 분류를 위한 접근 방식을 제시한다. 이 연구의 데이터베이스는 단지 8개의 만화와 20개의 비 만화(non-cartoon) 시퀀스로 구성되어있어서, 이것이 어떻게 동작할지 예측하는 것은 어렵다. 그리고 이 연구의 데이터 집합이 공개적으로 이용 가능한 것도 아니다. 논문 [18]은 비디오를 7개의 종류로 분류하는 또 다른 시도를 제시하였다. 논문 [13]과 2개의 특성은 유사하지만, 접근 방식이 다르고, 실험도 비슷하지가 않다.

### 3. The Proposed Image Genre Classification System

#### 3.1 The Proposed Architecture

예술, 만화, 사진의 예시 이미지는 Figure 1에 제시되어 있다. 제안된 시스템은 이러한 이미지들을 3가지 종류 중 하나로 분류한다. 이미지 종류는 (a), (b), (c)와 같이 미리 정의되어 있다.

Figure 2는 이미지 장르(유형) 분류 시스템의 전체적인 구조를 보여주고 있다. 제안된 구조는 특성 추출(feature extraction) 모듈과 분류(classification) 모듈 두 가지 구성되어 있다. 주어진 query image의 MPEG-7 descriptor에 정의된 특성들은 추출되고, 이렇게 추출된 특성은 이후 분류 모듈의 입력으로 사용된다.

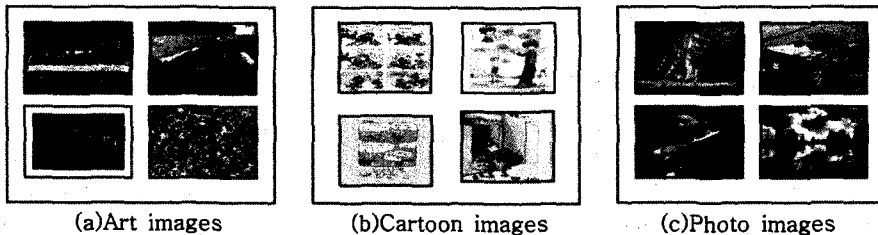


Figure 1. Example images from the three genres used for simulation

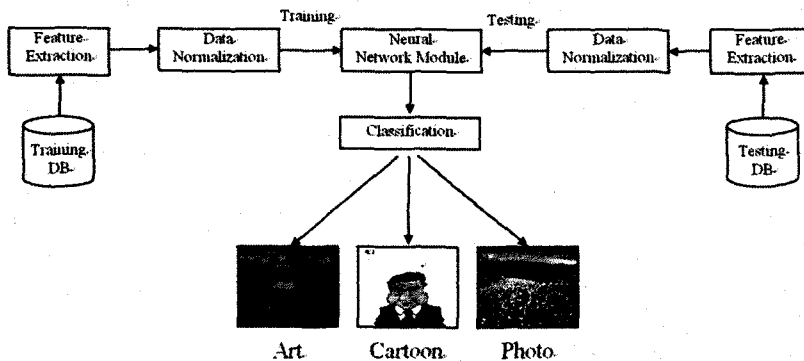


Figure 2. The overview of the proposed Image Genre Classification System

### 3.2 Feature Extraction

MPEG-7 XM 프로그램을 실행시켜서 학습 이미지의 특성들이 XML format으로 추출된다. XML 포맷으로 된 특성 정보는 다음 단계에서 parse 되며, 각각의 descriptor에 의해 생성된 값을 기준으로 0과 1 사이의 값으로 표준화(normalize) 된다. 이렇게 표준화된 값은 neural network classifier의 입력 값으로 사용된다. 분류 정보(class information)가 있는 original 값은 표준화된 값으로 바뀐다. 특성 값에 부착되어 있는 분류 정보는 직교(orthogonal) 벡터 값이다. 예를 들자면, 예술 이미지는 (1, 0, 0)으로 표현되고, 만화 이미지는 (0, 1, 0), 사진 이미지는 (0, 0, 1)로 표현된다.

### 3.3 Classification Module

분류 모듈은 neural network을 사용한다. neural network classifier는 노드 간의 가중치(weight value)를 수정함으로써 특성값(feature value)과 이에 상응하는 class 사이의 관계를 학습한다. Neural network을 학습시키기 위해 backpropagation 알고리즘을 사용한다. Neural network는 입력층(input layer), 출력층(output layer), 다수의 은닉층(hidden layer)으로 구성되어 있다. 입력 노드의 개수는 개별 descriptor의 크기에 의해 결정되지만, 출력 노드의 개수는 만화-사진 이미지 분류용에서는 2개, 예술-만화-사진 이미지 분류용에서는 3개이다. 출력 노드를 위한 분류 정보는 위에 설명한대로 분류에 따라 직교적으로 제시된다. 학습 과정과 유사한 testing 과정에서, 시스템은 MPEG-7 descriptor를 사용하여 query(질의) 이미지로부터 특성들을 추출하고, 학습과정을 거쳐 생성된 neural network을 이용하여 이미지를 분류한다.

## 4. Simulation

### 4.1 Environments

만화-사진 분류 시뮬레이션에서는 학습과정을 위해 1600개의 이미지(각 장르나 유형별로 800개씩)가 사용되었고 testing 과정을 위해 720개의 이미지(각 장르나 유형별로 360개씩)가 사용되었다. 예술-만화-사진 분류 시뮬레이션에서는, 학습과정을 위해 2400개의 이미지

(각 장르나 유형별로 800개씩)가 사용되었고 testing 과정을 위해 1080개의 이미지(각 장르나 유형별로 360개씩)가 사용되었다. 두 시뮬레이션 모두 특성 값을 위해 Color Layout, Color Structure, Homogeneous Texture, Region Shape, Edge Histogram 와 같은 5가지 descriptor를 사용하였다. 만화-사진 분류 시뮬레이션에서, 분류 모듈은 각각 50개의 노드를 가진 2개의 은닉층(hidden layer)을 갖췄으며, 100,000번 반복 학습을 하였다. 예술-만화-사진 분류 시뮬레이션에서, 분류 모듈은 각각 30개의 노드를 가진 2개의 은닉층(hidden layer)을 갖췄으며, 100,000번 반복 학습을 하였다.

### 4.2 Result

Table 1은 만화-사진 분류의 시뮬레이션 결과를 보여주며, Table 2는 예술-만화-사진 분류결과를 보여준다. 만화-사진 분류 작업은 Table 1에서처럼 Color Layout, Color Structure, Edge Histogram에서 좋은 결과를 나타내었다. 제안된 neural network는 Color Structure와 Edge Histogram descriptor 시뮬레이션에서, 만화 이미지를 각각 97.79 %와 97.24 %로 성공적으로 분류했다. 또한 사진 이미지 분류에서 Homogeneous Texture는 93.63 %라는 좋은 수행결과를 보여줬다.

예술-만화-사진 이미지 분류 시뮬레이션은 세 개의 장르(유형)를 다루기 때문에, 만화-사진 이미지 분류에 비해 낮은 결과를 보여준다. 그러나 시스템은 만화 이미지 분류에서 훌륭한 결과를 보여준다. 왜냐하면 사진 이미지는 아주 독특한 이미지 특성을 지니기 때문이다. 일반적으로 시스템은 Color Structure descriptor에서 좋은 결과를 보이고, Region Shape descriptor에서 나쁜 결과를 보인다. 시뮬레이션 결과는 매우 고무적이며, 여러 다른 이미지 처리 영역에서 사용될 수 있을 것이다. 본 시스템은 이미지 검색 엔진이나 이미지 수집 엔진의 중심부로 사용되거나 개발될 수 있을 것이다. 시스템은 대형 이미지 데이터베이스에서 이미지 검색 시스템을 위한 아주 유용한 도구이다.

Table 1. Classification Results of Cartoon-Photo Images (%)

Color Layout(CL)			Color Structure(CS)		
	Cartoon	Photo		Cartoon	Photo
Cartoon	92.82	7.18	Cartoon	97.79	2.21
Photo	3.87	96.13	Photo	4.14	95.86

Region Shape(RS)			Edge Histogram(EH)		
	Cartoon	Photo		Cartoon	Photo
Cartoon	78.18	21.82	Cartoon	97.24	2.76
Photo	15.51	84.49	Photo	6.65	93.35

Homogeneous Texture(HT)			Average		
	Cartoon	Photo		Cartoon	Photo
Cartoon	84.53	15.47	Cartoon	90.12	9.89
Photo	6.37	93.63	Photo	7.31	92.70

Table 2. Classification Results of Art-Cartoon-Photo Images (%)

		art	cartoon	photo
art	CL	69.71	8.61	22.22
	EH	79.17	6.94	13.89
	HT	75.00	10.28	14.72
	RS	50.56	17.78	31.67
	CS	82.50	1.39	16.11
	average	70.99	9.00	19.73
cartoon	CL	5.83	86.94	7.22
	EH	11.39	84.44	4.17
	HT	8.33	86.94	4.72
	RS	24.72	59.72	15.56
	CS	1.11	98.06	0.83
	average	10.28	83.22	6.50
photo	CL	25.00	2.50	72.50
	EH	17.22	5.56	77.22
	HT	20.00	4.72	75.28
	RS	32.50	15.28	52.22
	CS	19.72	1.67	78.61
	average	22.89	5.95	71.17

5. Conclusion

이 논문에서 우리는 이미지 장르(유형) 분류 시스템을 제안했다. 제안된 시스템은 이미지를 예술, 사진, 만화라는 세 가지 장르(유형) 중 하나로 분류한다. 시스템의 분류 모듈은 MPEG-7 descriptor로부터 추출된 특성 값에 따라, 그에 상응하는 분류 작업을 학습한다. 시뮬레이션 결과는, 제안된 시스템이 사용된 특성 내에서 이미지를 80% 이상 올바른 장르(유형)로 성공적으로 분류하는 것을 보여준다.

References

1. Sikora, T.: The MPEG-7 visual standard for content description - an overview. IEEE Transactions on Circuit and Systems for Video Technology, Vol. 11, No. 6 (2001) 696-702
2. Ro, Y., Kim, M., Kang, H., Manjunath, B., Kim,

- J.: MPEG-7 homogeneous texture descriptor. ETRI Journal, Vol. 23, No. 2 (2001) 41-51
3. Bober, M.: The MPEG-7 visual shape descriptors. IEEE Transactions on Circuit and Systems for Video Technology, Vol. 11, No. 6 (2001) 716-719
4. Won, C., Park, D., Park, S.: Efficient use of MPEG-7 edge histogram descriptor. ETRI Journal, Vol. 24, No. 1 (2002) 23-30
5. Pakkanen, J., Ilvesmaki, A., Iivarinen, J.: Defect image classification and retrieval with MPEG-7 descriptors. Lecture Notes in Computer Science, Vol. 2749. Springer-Verlag, Berlin Heidelberg, New York (2003) 349-355
6. Spyrou, E., Borgne, H., Mailis, T., Cooke, E., Arvrithis, Y., O'Connor H.: Fusing MPEG-7 visual

- descriptors for image classification. Lecture Notes in Computer Science, Vol. 3697. Springer-Verlag, Berlin Heidelberg, New York (2005) 847-852
7. Laaksonen, J., Koskela, M., Oja, E.: PicSOM - Self-organizing image retrieval with MPEG-7 content descriptor. IEEE Transactions on Neural Networks: Special Issue on Intelligent Multimedia Processing, Vol. 13, No. 4 (2002) 841-853
8. Wonil Kim, Hanku Lee, S. Yoo, S. Baik.: Neural Network Based Adult Image Classification. Lecture Notes in Computer Science, Vol. 3696, pp 481-486, Springer-Verlag, Berlin Heidelberg, New York in 2005
9. Wonil Kim, Hanku Lee, J. Park, Kyoungro Yoon: Multi Class Adult Image Classification using Neural Networks. Lecture Notes in Computer Science, Vol. 3501, pp 222-226, Springer-Verlag, Berlin Heidelberg, New York in 2005
10. J. R. Smith and S.-F. Chang, "Searching for images and videos on the world wide web," Tech. Rep. 459-96-25, Center for Communications Research, Columbia University, 1996.
11. N. C. Rowe and B. Frew, "Automatic caption localization for photographs on word wide web pages," Tech. Rep., Department of Computer Science, Naval Postgraduate School, 1997.
12. V. Athitsos, M. J. Swain, and C. Frankel, "Distinguishing photographs and graphics on the world wide web," in IEEE Workshop on Content-Based Access of Image and Video Libraries, Puerto Rico, June 1997.
13. R. Glasberg, K. Elazouzi and T. Sikora, "Cartoon-Recognition Using Visual-Descriptors and a Multilayer-Perceptron," in WIAMIS, Montreux, April 13-15, 2005.
14. T. Ianeva, A. de Vries, and H. Rohrig, "Detecting cartoons: A case study in automatic video-genre classification," in *IEEE International Conference on Multimedia and Expo*, 1, pp. 449-452, 2003.
15. M. A. Windhouwer, A. R. Schmidt, and M. L. Kersten, "Acoi: A System for Indexing Multimedia Objects," in International Workshop on Information Integration and Web-based Applications & Services, Yogyakarta, Indonesia, November 1999.
16. J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, 1993.
17. M. Roach, J. S. Mason, and M. Pawlewski, "Motion-based classification of cartoons," in Int. Symposium on Intelligent Multimedia, 2001.
18. B. T. Truong, C. Dorai, and S. Venkatesh, "Automatic genre identification for content-based video categorization," in Proc. 15th International Conference on Pattern Recognition, Barcelona, Spain, September 2000, vol. II, pp. 230-233.