

SOM을 이용한 등록상표에 대한 내용기반 이미지 검색

이재준, 신민기, 백우진, 신문선
건국대학교 컴퓨터 · 응용과학부
{ c0mmunity, mirki, wjpaik, msshin }@kku.ac.kr

Content-based Trademark Image Retrieval System using SOM

Jaejun Lee, Minki Shin, Woojin Paik, Moonsun Shin
Dept. of Computer Science, Konkuk University

요 약

산업재산권중 하나인 상표에 대한 효율적인 이미지 검색은 상표도용 및 이로 인한 분쟁을 방지할 수 있다. 이를 위해서는 효율적인 내용기반 유사이미지 검색이 필요하다. 본 논문에서는 상표이미지검색에 있어 가시적인 특성(visual feature)을 그레이 히스토그램을 통해서 상표이미지의 특성값을 추출하여 이를 입력패턴으로 SOM(Self-Organizing Map)알고리즘을 적용한 내용기반 유사이미지 검색시스템을 제안한다.

1. 서론

현재 이미지 검색방법은 등록된 정보에 있는 키워드를 이용하는 방법이 주를 이루고 있다. 기존상표에 대한 검색기능은 이미지에 첨부된 제한된 정보에 의해서만 이미지를 검색하기 때문에 사용자가 원하는 결과를 얻지 못하는 경우가 발생한다. 이러한 문제를 해결하기 위해 이미지의 가시적인 특성(visual feature)을 이용한 내용기반 이미지검색 방안이 제시되고 있다.

산업재산권중 하나인 상표권은 자기상품을 표시하는 기호, 문자, 도형 또는 입체적 형상에 대한 보호이다. 따라서 상표에 대한 유사이미지검색은 유사이미지를 상표 등록 전에 미리 찾아 추후의 분쟁을 방지하고, 산업재산권을 보호하는 효과적인 방법이 될 것이다.

상표를 등록할 때 도형분류코드(Design Search Code)[1]를 지정할 수 있으나 실질적으로 등록된 상표들은 도형분류코드가 없는 경우가 대부분이다. 따라서 이를 이용한 검색은 실용성이 없다. 도형분류코드란 카테고리에 포함된 디자인의 상징적인 요소

(symbolic element)에 대한 체계적인 분류이다.

본 논문에서는 SOM(Self Organizing Map)[2]을 이용한 유사상표이미지 군집화를 통해 내용기반 이미지검색 방안을 제시하며, 이미지특성 데이터에 따른 군집화[3]의 정확도를 실험을 통해 효율적인 내용기반 이미지 검색방안을 제시한다.

2. 관련연구

내용기반 이미지 검색(CBIR:Content-based Image Retrieval)은 이미지로부터 컬러, 형태, 질감 등의 특징들을 추출하여 데이터베이스에 저장하는 과정과 사용자가 원하는 이미지를 검색하는 과정을 포함한다.[4]

aiPicsom 시스템의 경우 컬러, 질감 또는 그 둘을 같이 이용하여 각각의 특징을 기반으로 SOM알고리즘을 통해 검색을 한다.[5]

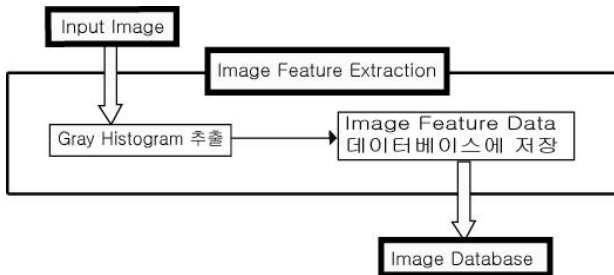
PicSOM 은 SOM 기반의 내용기반 검색시스템으로 이미지의 형태에 대한 푸리에변환(shape fourier transform)과, 컬러, 질감을 선택하여 검색을 한다.[6]

기존의 방법들은 다양한 이미지들에 대한 형태, 컬러, 질감 등의 특성으로 SOM알고리즘을 적용하였다. 그러나 상표이미지의 경우, 대상이미지들의 도메인이 한정적이어서 기존의 방법들을 그대로 적용하기에 부적합하다.

따라서 상표이미지의 가시적 특성을 반영할 수 있는 특성추출로 상표이미지 도메인에서 자기조직화 학습이 가능한 SOM알고리즘을 통해서 내용기반 유사이미지 검색시스템을 구현한다.

3. SOM을 이용한 내용기반 유사이미지 검색시스템

본 논문에서는 각 상표의 시각적 특성을 추출하기 위해 그레이 스케일 변환을 하였다. 이는 유사 상표들이 색채 차이만으로 군집화 되지 못하는 현상을 방지하고 검색성능의 향상을 위해서이다. (그림 1)은 이미지의 특성추출의 과정을 나타낸다.



(그림 1) 이미지 특성 추출 흐름도

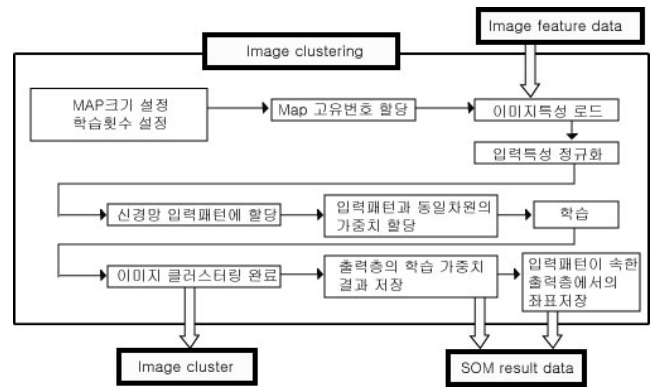
이미지를 입력하여 그레이 히스토그램을 추출된 특성값을 데이터베이스에 저장한다.

신경망알고리즘인 SOM(Self-Organizing Map)은 경쟁학습을 통해 다차원의 입력데이터를 저차원으로 시각화하며, 데이터의 군집화(data classification)에 주로 사용된다.[2]

SOM은 학습단계에서 피드백(feedback)과정이 없어 구조적으로 상당히 학습수행이 빨라 실시간 학습 처리가 가능한 모델이다. 또한 연속적인 학습이 가능하여 추가적으로 입력되는 상표이미지에 대한 군집화가 가능하다.

자기조직화를 통한 통계적 모델이기 때문에 대량의 상표이미지를 훈련데이터로 학습할 필요없이 자기조직화가 가능하여 SOM알고리즘을 상표이미지 데이터처리에 맞게 확장하였다.

내용기반 유사이미지 검색시스템에서 유사이미지 군집화를 위한 SOM기반 학습과정[7]은 (그림 2)와 같다.

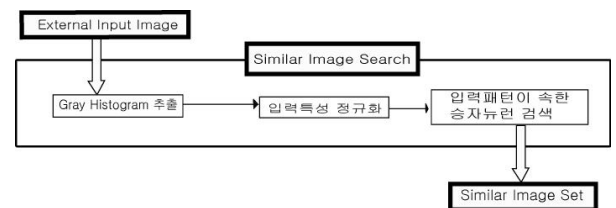


(그림 2) SOM기반 유사이미지 학습과정

먼저 이미지 특성추출을 통해 얻어진 분할히스토그램을 하나의 입력패턴으로 하여 이를 정규화 한다.

군집화하고자 하는 모든 상표에 대한 이미지특성값을 입력패턴화 하여 입력층을 생성한다. 생성된 입력층과 동일한 차원의 가중치를 할당한 출력층이 학습을 하면서 연결된 입력층의 상표를 유사한 이미지끼리 군집화 한다. 학습을 마친 후 출력층의 조직화된 가중치결과와, 군집화 된 결과를 데이터베이스에 저장하여 연속적인 학습[3]이 가능하다. 학습결과로 저장된 군집화 된 이미지는 유사이미지를 검색할 때 사용된다.

내용기반 유사이미지 검색시스템은 (그림 3)의 흐름으로 데이터베이스 내부의 상표가 아닌 사용자가 검색 하고자 하는 외부이미지를 입력으로 이와 유사한 상표를 검색하게 된다.



(그림 3) 유사 상표 이미지 검색

외부 이미지에 대한 특성추출은 로드된 SOM의 입력패턴 추출방식과 동일하게 수행된다. 추출된 외부이미지는 정규화되어 입력패턴화 된 후 출력층에서 자신이 속한 군집의 위치를 승자뉴런(Winner neuron)검색(수식 2)을 통해 확인한다. 해당좌표에 있는 군집화 결과는 사용자인터페이스를 통해 확인이 가능하다.

승자뉴런 검색에서 입력이미지와 이미지군집간의 유사도측정은 유클리드 거리함수(수식 1)를 이용한다.

$$D_{ij} = |X^l - W_{ij}| = \sqrt{(x_1 - w_{ij1})^2 + \dots + (x_n - w_{ijn})^2}$$

(수식 1) 유클리디언 거리함수

$$D(k_1, k_2) = \min(D_{i,j})$$

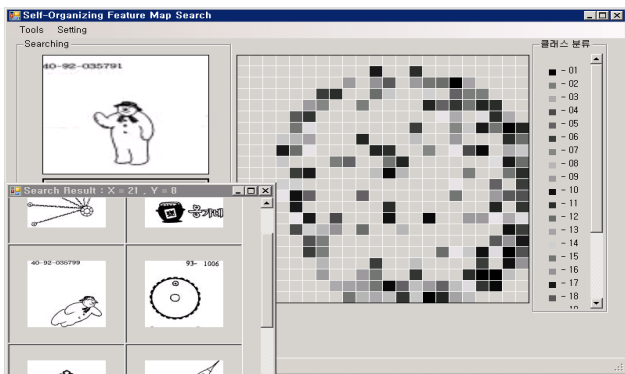
(수식 2) 승자뉴런 검색

i 와 j 는 출력층의 뉴런에 대한 좌표이며, 입력층의 입력패턴과 출력층의 가중치의 거리계산을 통해 가장 가까운 거리에 있는 뉴런이 승자가 된다. k_1, k_2 는 승자뉴런의 출력층에 대한 좌표가 된다. 이 승자뉴런의 가중치만이 학습이 되며, 승자뉴런 검색 방법은 외부 이미지를 입력패턴화 후 유사이미지 검색결과로 출력층의 k_1, k_2 좌표에 있는 이미지를 확인 가능하다.

4. 실험 및 결과

시스템 구현은 .Net framework 2.0 기반으로 개발 언어로는 Visual Studio .NET 2005를 이용하여 C#을 사용하였다. 데이터베이스로는 Oracle Database 9.2.0.4를 활용하였다. 실험데이터는 1995년에서 2002년 특허등록상표[8] 100만개 중 도형분류코드를 가지는 23.03%의 상표이미지를 대상으로 샘플링 하였다. 이는 실험데이터에 다양한 종류의 이미지를 포함시킬 수 있기 때문이다.

(그림 4)는 유사이미지 내용기반 검색시스템을 구현하여 실행한 화면이다.



(그림 4) 학습을 마친 결과 화면

구현된 내용기반 유사이미지 검색시스템은 군집화된 결과가 사용자 인터페이스를 통해 확인가능하며 좌표를 클릭할 때 해당 좌표의 이미지 군집결과를 확인할 수 있다. 또한 외부이미지를 검색할 때 해당 이미지가 속한 좌표와 유사이미지도 확인할 수 있다.

Input Image	Similar Image Set

(그림 5) 유사이미지 검색 결과

입력이미지가 저장된 이미지와 유사하거나 같은 경우 그 결과를 (그림 5)에서처럼 확인할 수 있다.

5. 결론 및 향후 연구 계획

본 논문에서는 상표이미지의 효율적인 검색을 위해 신경망 알고리즘인 SOM을 이용한 내용기반 유사이미지 검색시스템을 설계하고 구현하였다. 구현된 시스템

은 이미지의 특성추출을 위하여 그레이 히스토그램을 이용하였으며 실험결과 단순한 특성추출로 인하여 군집이미지의 응집력이 낮다는 것을 확인할 수 있었다.

향후 군집이미지의 유사도를 높이기 위해 상표이미지의 에지특성을 추가하고, 이미지 분할기법을 이용하는 것이 필요하며 이에 대한 연구를 진행 중이다.

참고문헌

[1] USPTO Design Search Code manual , <http://tess2.uspto.tmdb/dscm/index.htm>
 [2] Teuvo Kohonen, "Self-Organizing Maps", third edition Springer Series in Information Science, 2001
 [3] Kyung Ah Han, Jong Chan Lee, Chi Jung Hwang, ChungNam National University, Taejun, KOREA , "Image Clustering using Self-Organizing feature map with Refinement", IEEE, 1995

[4] 반종오, 강문주, 최형진, “칼라 공간과 형태 정보를 이용한 내용기반 이미지검색 시스템 구현”, 정보처리학회논문지, 2003, 10

[5] Self Organizing Map AI for Pictures By Casey Chesnut, 2004

[6] PicSOM information browsing and retrieval,
<http://www.cis.hut.fi/picsom>

[7] Juha Vesanto , Esa Alhoniemi, “Clustering of The Self-Organizing Map”, IEEE TRANSACTION NEURAL NETWORKS, VOL.11, NO.3 MAY 2000

[8] 한국특허정보원 ,
<http://www.kipris.or.kr>