



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0033345
(43) 공개일자 2019년03월29일

(51) 국제특허분류(Int. Cl.)
G06F 16/00 (2019.01)

(52) CPC특허분류
G06F 16/951 (2019.01)
G06F 16/583 (2019.01)

(21) 출원번호 10-2017-0122012

(22) 출원일자 2017년09월21일
심사청구일자 2017년09월21일

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

남중호

서울특별시 용산구 한남동 810번지 한남더힐 아파트 123동 1001호

송진하

전남 순천시 연향동 호반리젠시빌 2차 203동 804호

(뒷면에 계속)

(74) 대리인

이지연

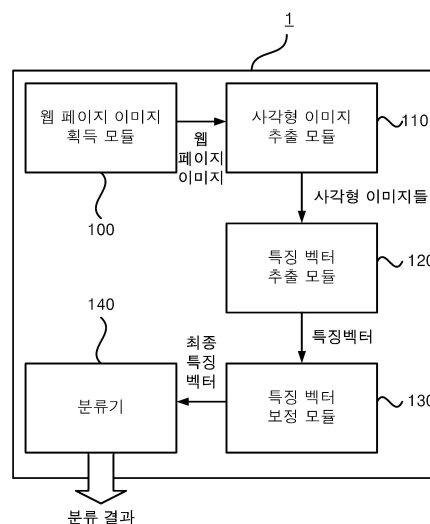
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 음란 이미지를 포함한 성인용 웹 페이지 분류 방법 및 상기 방법을 이용한 성인용 웹 페이지 검출 시스템

(57) 요약

본 발명은 성인용 웹 페이지 검출 시스템 및 방법에 관한 것이다. 상기 성인용 웹 페이지 검출 시스템은, 임의의 웹 페이지를 하나의 이미지로 획득하는 웹페이지 이미지 획득 모듈; 상기 웹 페이지 이미지 획득 모듈에 의해 획득된 웹 페이지 이미지로부터 사각형 이미지들을 추출하는 사각형 이미지 추출 모듈; 상기 사각형 이미지 추출 모듈로부터 추출된 사각형 이미지들에 대한 성인용 이미지 확률값들을 이용하여 해당 웹 페이지에 대한 특징 벡터를 생성하는 특징 벡터 추출 모듈; 상기 특징 벡터 추출 모듈에 의해 생성된 웹 페이지에 대한 특징 벡터의 차원을 보정하여 사전 설정된 차원을 갖는 최종 특징 벡터를 생성하는 특징 벡터 보정 모듈; 상기 특징 벡터 보정 모듈에 의해 생성된 최종 특징 벡터를 이용하여 해당 웹 페이지가 성인용 웹 페이지인지 여부를 판단하는 분류기;를 구비한다.

대표도 - 도1



(72) 발명자

신광수

서울특별시 관악구 신림로14길 27-, 401호

석혜경

부산광역시 동래구 충렬대로 314, 7동 806호 (낙민동, 동래한양아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711055377

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 SW중심대학 지원사업

연구과제명 SW중심대학(서강대학교)

기 여 율 1/1

주관기관 서강대학교 산학협력단

연구기간 2017.03.01 ~ 2017.12.31

명세서

청구범위

청구항 1

임의의 웹 페이지를 하나의 이미지로 획득하는 웹페이지 이미지 획득 모듈;

상기 웹 페이지 이미지 획득 모듈에 의해 획득된 웹 페이지 이미지로부터 사각형 이미지들을 추출하는 사각형 이미지 추출 모듈;

상기 사각형 이미지 추출 모듈로부터 추출된 사각형 이미지들을 이용하여 해당 웹 페이지에 대한 특징 벡터를 생성하는 특징 벡터 추출 모듈;

상기 특징 벡터 추출 모듈에 의해 생성된 웹 페이지에 대한 특징 벡터의 차원을 보정하여 사전 설정된 차원을 갖는 최종 특징 벡터를 생성하는 특징 벡터 보정 모듈;

상기 특징 벡터 보정 모듈에 의해 생성된 최종 특징 벡터를 이용하여 해당 웹 페이지가 성인용 웹 페이지인지 여부를 판단하는 분류기;

를 구비하는 것을 특징으로 하는 성인용 웹 페이지 검출 시스템.

청구항 2

제1항에 있어서, 상기 특징 벡터 추출 모듈은,

상기 이미지 추출 모듈로부터 추출된 사각형 이미지들의 각각에 대하여 사전 설정된 방법을 이용하여 성인용 이미지 확률값들을 구하고, 상기 특징 벡터는 상기 성인용 이미지 확률값들로 이루어진 것을 특징으로 하는 성인용 웹 페이지 검출 시스템.

청구항 3

제2항에 있어서, 상기 특징 벡터는 웹 페이지에 포함된 사각형 이미지의 개수에 대응되는 차원을 갖는 것을 특징으로 하는 성인용 웹 페이지 검출 시스템.

청구항 4

제1항에 있어서, 상기 특징 벡터 보정 모듈은,

상기 특징 벡터 추출 모듈에 의해 생성된 웹 페이지에 대한 특징 벡터의 차원이 사전 설정된 차원보다 작은 경우 상기 특징 벡터를 구성하는 값들을 복제하며, 상기 특징 벡터를 구성하는 값들을 분류하고, 사전 설정된 차원에 대응되는 최대값들만을 추출하여 최종 특징 벡터를 생성하는 것을 특징으로 하는 성인용 웹 페이지 검출 시스템.

청구항 5

제1항에 있어서, 상기 사각형 이미지 추출 모듈은,

상기 웹 페이지 이미지 획득 모듈에 의해 획득된 웹 페이지 이미지로부터 경계선을 그릴 수 있는 사각형 이미지들을 검출하고, 상기 검출된 사각형 이미지들 중 사전 설정된 크기 조건 및 중형비 조건을 만족하는 사각형 이미지들을 추출하는 것을 특징으로 하는 성인용 웹 페이지 검출 시스템.

청구항 6

(a) 임의의 웹 페이지를 하나의 이미지로 캡처하여 획득하는 단계;

(b) 획득된 웹 페이지 이미지로부터 사각형 이미지들을 추출하는 단계;

(c) 상기 웹 페이지에 대해 추출된 사각형 이미지들을 이용하여 해당 웹 페이지에 대한 특징 벡터를 생성하는 단계;

(d) 상기 웹 페이지에 대한 특징 벡터의 차원을 보정하여 사전 설정된 차원을 갖는 최종 특징 벡터를 생성하는 단계;

(e) 상기 웹 페이지에 대한 최종 특징 벡터를 이용하여 해당 웹 페이지가 성인용 웹 페이지인지 여부를 판단하는 단계;

를 구비하는 것을 특징으로 하는 성인용 웹 페이지 검출 방법.

청구항 7

제6항에 있어서, 상기 (c) 단계는,

(c1) 상기 웹 페이지에 대해 추출된 사각형 이미지들의 각각에 대하여 사전 설정된 방법을 이용하여 성인용 이미지 확률값들을 구하는 단계; 및

(c2) 상기 사각형 이미지들의 각각에 대한 성인용 이미지 확률값들로 이루어진 특징 벡터를 생성하는 단계;를 구비하는 것을 특징으로 하는 성인용 웹 페이지 검출 방법.

청구항 8

제7항에 있어서, 상기 특징 벡터는 웹 페이지에 포함된 사각형 이미지의 개수에 대응되는 차원을 갖는 것을 특징으로 하는 성인용 웹 페이지 검출 방법.

청구항 9

제6항에 있어서, 상기 (d) 단계는,

(d1) 상기 웹 페이지에 대한 특징 벡터의 차원이 사전 설정된 차원보다 작으면, 상기 특징 벡터를 구성하는 값들을 복제하는 단계;

(d2) 상기 웹 페이지에 대한 특징 벡터를 구성하는 값들을 분류하는 단계;

(d3) 상기 분류된 값들 중 사전 설정된 차원에 대응되는 최대값들만을 추출하는 단계; 및

(d4) 상기 추출된 최대값들을 이용하여 사전 설정된 차원을 갖는 최종 특징 벡터를 생성하는 단계;

를 구비하는 것을 특징으로 하는 성인용 웹 페이지 검출 방법.

청구항 10

제6항에 있어서, 상기 (b) 단계는,

(b1) 상기 웹 페이지 이미지로부터 경계선을 그릴 수 있는 사각형 이미지들을 검출하는 단계; 및

(b2) 상기 검출된 사각형 이미지들 중 사전 설정된 크기 조건 및 종횡비 조건을 만족하는 사각형 이미지들을 추출하는 단계;

를 구비하는 것을 특징으로 하는 성인용 웹 페이지 검출 방법.

발명의 설명

기술 분야

[0001] 본 발명은 음란 이미지를 포함한 성인용 웹 페이지 분류 방법에 관한 것으로서, 더욱 구체적으로는 음란 이미지를 포함하는 웹 페이지를 분류 및 검출할 수 있도록 하는 성인용 웹 페이지 분류 방법 및 이러한 방법을 이용하여 구현된 성인용 웹 페이지 검출 시스템에 관한 것이다.

배경 기술

[0002] 최근 인터넷 기술이 발달함에 따라, 수많은 사람들이 인터넷을 통해 다양한 정보를 손쉽게 접할 수 있게 되었다. 하지만, 인터넷을 통해 유익한 정보 뿐만 아니라 유해한 정보들도 함께 접할 수 있게 될 뿐만 아니라,

성인들뿐만 아니라 청소년, 어린이들까지도 인터넷을 통해 유해한 웹사이트들에 접속할 수 있게 되었다. 그 결과, 청소년 및 어린이들이 유해한 웹 사이트, 예컨대 불법 사이트 또는 음란 이미지들을 포함하는 성인용 웹사이트 등에 무방비로 노출되는 문제점들이 발생하고 있는 실정이다.

[0003] 따라서, 청소년이나 어린이들이 이러한 유해한 인터넷 환경으로부터 보호받을 수 있도록 하기 위하여, 유해한 웹사이트들을 분류하여 검출하고, 차단할 수 있도록 하는 차단 소프트웨어들이 다양하게 제안되고 있다.

[0004] 음란 이미지들을 포함하는 성인용 웹 사이트들을 차단하기 위하여, 웹 사이트들 중 음란 이미지들을 포함하는 성인용 웹 사이트들을 분류하여 검출하여야 된다. 종래에는 성인용 웹사이트를 분류 및 검출하기 위하여, 텍스트 기반의 검색 엔진이 각 웹사이트 또는 웹페이지에 포함된 단어, 문장 등을 분석하여 유해성 여부를 판단하거나, 각 웹사이트 또는 웹페이지에 포함된 이미지들이 사전 설정된 방법을 이용하여 음란성 여부를 판단하고, 음란 이미지가 포함된 경우 성인용 웹 사이트로 판단하게 된다.

[0005] 이러한 종래의 성인용 웹사이트 검출 방법은 각 웹페이지에 포함된 이미지들 중 하나라도 음란 이미지로 판단되는 경우, 이에 해당하는 웹 사이트 또는 웹 페이지는 성인용 웹 사이트로 분류하게 된다.

[0006] 전술한 종래의 성인용 웹사이트 검출 방법은 해당 웹 사이트 또는 웹 페이지에 포함된 모든 이미지들에 대하여 음란 이미지인지 여부를 판단하여야 되므로, 판단 시간이 많이 소요될 뿐만 아니라, 판단의 정확성 및 신뢰도도 그리 높지 않은 문제점이 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국등록특허공보 제 10-0658799 호

발명의 내용

해결하려는 과제

[0008] 전술한 문제점을 해결하기 위한 본 발명의 목적은 웹 페이지 또는 웹 사이트가 성인용 웹 사이트인지 여부를 신속하게 판단하면서도 정확성 및 신뢰도가 높은 판단 결과를 제공하는 성인용 웹 페이지 검출 방법을 제공하는 것이다.

[0009] 본 발명의 다른 목적은 전술한 성인용 웹 페이지 검출 방법을 구현한 성인용 웹 페이지 검출 시스템을 제공하는 것이다.

과제의 해결 수단

[0010] 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징에 따른 성인용 웹 페이지 검출 시스템은, 임의의 웹 페이지를 하나의 이미지로 획득하는 웹페이지 이미지 획득 모듈; 상기 웹 페이지 이미지 획득 모듈에 의해 획득된 웹 페이지 이미지로부터 사각형 이미지들을 추출하는 사각형 이미지 추출 모듈; 상기 사각형 이미지 추출 모듈로부터 추출된 사각형 이미지들을 이용하여 해당 웹 페이지에 대한 특징 벡터를 생성하는 특징 벡터 추출 모듈; 상기 특징 벡터 추출 모듈에 의해 생성된 웹 페이지에 대한 특징 벡터의 차원을 보정하여 사전 설정된 차원을 갖는 최종 특징 벡터를 생성하는 특징 벡터 보정 모듈; 상기 특징 벡터 보정 모듈에 의해 생성된 최종 특징 벡터를 이용하여 해당 웹 페이지가 성인용 웹 페이지인지 여부를 판단하는 분류기;를 구비한다.

[0011] 전술한 제1 특징에 따른 성인용 웹 페이지 검출 시스템에 있어서, 상기 특징 벡터 추출 모듈은, 상기 이미지 추출 모듈로부터 추출된 사각형 이미지들의 각각에 대하여 사전 설정된 방법을 이용하여 성인용 이미지 확률값들을 구하고, 상기 특징 벡터는 상기 성인용 이미지 확률값들로 이루어진 것이 바람직하며, 상기 특징 벡터는 웹 페이지에 포함된 사각형 이미지의 개수에 대응되는 차원을 갖는 것이 더욱 바람직하다.

[0012] 전술한 제1 특징에 따른 성인용 웹 페이지 검출 시스템에 있어서, 상기 특징 벡터 보정 모듈은, 상기 특징 벡터 추출 모듈에 의해 생성된 웹 페이지에 대한 특징 벡터의 차원이 사전 설정된 차원보다 작은 경우 상기 특징 벡터를 구성하는 값들을 복제하며, 상기 특징 벡터를 구성하는 값들을 분류하고, 사전 설정된 차원에 대응되는 최대값들만을 추출하여 최종 특징 벡터를 생성하는 것이 바람직하다.

- [0013] 전술한 제1 특징에 따른 성인용 웹 페이지 검출 시스템에 있어서, 상기 사각형 이미지 추출 모듈은, 상기 웹 페이지 이미지 획득 모듈에 의해 획득된 웹 페이지 이미지로부터 경계선을 그릴 수 있는 사각형 이미지들을 검출하고, 상기 검출된 사각형 이미지들 중 사전 설정된 크기 조건 및 중형비 조건을 만족하는 사각형 이미지들을 추출하는 것이 바람직하다.
- [0014] 본 발명의 제2 특징에 따른 성인용 웹 페이지 검출 방법은, (a) 임의의 웹 페이지를 하나의 이미지로 캡처하여 획득하는 단계; (b) 획득된 웹 페이지 이미지로부터 사각형 이미지들을 추출하는 단계; (c) 상기 웹 페이지에 대해 추출된 사각형 이미지들을 이용하여 해당 웹 페이지에 대한 특징 벡터를 생성하는 단계; (d) 상기 웹 페이지에 대한 특징 벡터의 차원을 보정하여 사전 설정된 차원을 갖는 최종 특징 벡터를 생성하는 단계; (e) 상기 웹 페이지에 대한 최종 특징 벡터를 이용하여 해당 웹 페이지가 성인용 웹 페이지인지 여부를 판단하는 단계; 를 구비한다.
- [0015] 전술한 제2 특징에 따른 성인용 웹 페이지 검출 방법에 있어서, 상기 (c) 단계는, (c1) 상기 웹 페이지에 대해 추출된 사각형 이미지들의 각각에 대하여 사전 설정된 방법을 이용하여 성인용 이미지 확률값들을 구하는 단계; 및 (c2) 상기 사각형 이미지들의 각각에 대한 성인용 이미지 확률값들로 이루어진 특징 벡터를 생성하는 단계; 를 구비하는 것이 바람직하며,
- [0016] 상기 특징 벡터는 웹 페이지에 포함된 사각형 이미지의 개수에 대응되는 차원을 갖는 것이 더욱 바람직하다.
- [0017] 전술한 제2 특징에 따른 성인용 웹 페이지 검출 방법에 있어서, 상기 (d) 단계는, (d1) 상기 웹 페이지에 대한 특징 벡터의 차원이 사전 설정된 차원보다 작으면, 상기 특징 벡터를 구성하는 값들을 복제하는 단계; (d2) 상기 웹 페이지에 대한 특징 벡터를 구성하는 값들을 분류하는 단계; (d3) 상기 분류된 값들 중 사전 설정된 차원에 대응되는 최대값들만을 추출하는 단계; 및 (d4) 상기 추출된 최대값들을 이용하여 사전 설정된 차원을 갖는 최종 특징 벡터를 생성하는 단계; 를 구비하는 것이 바람직하다.
- [0018] 전술한 제2 특징에 따른 성인용 웹 페이지 검출 방법에 있어서, 상기 (b) 단계는, (b1) 상기 웹 페이지 이미지로부터 경계선을 그릴 수 있는 사각형 이미지들을 검출하는 단계; 및 (b2) 상기 검출된 사각형 이미지들 중 사전 설정된 크기 조건 및 중형비 조건을 만족하는 사각형 이미지들을 추출하는 단계; 를 구비하는 것이 바람직하다.

발명의 효과

- [0019] 본 발명에 따른 성인용 웹 페이지 검출 시스템 및 방법은, 검사하고자 하는 웹 페이지를 하나의 이미지로 캡처한 후 해당 웹 페이지 이미지에 포함된 사각형 이미지들을 추출하여 성인용 웹 페이지인지 여부를 판단함으로써, 종래의 방법들에 비해 처리 속도를 월등히 향상시킬 수 있게 된다.
- [0020] 또한, 본 발명에 따른 성인용 웹 페이지 검출 시스템 및 방법은, 웹 페이지 이미지에 포함된 사각형 이미지들로부터 추출된 특징 벡터의 차원을 사전 설정된 차원으로 일치시켜 최종 특징 벡터를 생성하고, 상기 일정한 차원을 갖는 최종 특징 벡터를 이용하여 분류함으로써, 웹 페이지 이미지에 포함된 사각형 이미지들의 위치 및 개수에 따라 특징벡터값이 바뀌는 것을 방지할 수 있게 된다.

도면의 간단한 설명

- [0021] 도 1은 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 시스템(1)을 도시한 블록도이다.
- 도 2는 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 방법을 순차적으로 도시한 흐름도이다.
- 도 3은 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 시스템에 있어서, 특징 벡터의 생성 및 보정을 설명하기 위하여 도시한 도표이다.
- 도 4 및 도 5는 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 시스템에 있어서, 사각형 이미지 추출 모듈의 동작을 설명하기 위하여 예시적으로 도시한 것으로서, 도 4는 웹 페이지 이미지이며, 도 5는 도 4의 웹 페이지 이미지로부터 추출한 사각형 이미지들이다.

발명을 실시하기 위한 구체적인 내용

- [0022] 본 발명에 따른 성인용 웹 페이지 검출 방법 및 시스템은, 임의의 웹 페이지를 하나의 이미지 파일로 캡처한 후, 웹 페이지에 대한 이미지 파일에 포함된 사각형 영역들을 잘라내어 사전 설정된 크기와 중형비를 갖는 사각

형 이미지들을 추출하며, 추출된 사각형 이미지들의 각각에 대한 성인용 이미지 확률값을 구하여 특징 벡터를 생성하고, 상기 특징 벡터를 보정하여 사전 설정된 차원을 갖는 최종 특징 벡터를 생성하고, 상기 최종 특징 벡터를 이용하여 성인용 웹 페이지인지 여부를 분류하는 것을 특징으로 한다.

[0023] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 방법 및 이를 이용한 성인용 웹 페이지 검출 시스템에 대하여 구체적으로 설명한다.

[0024] 도 1은 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 시스템(1)을 도시한 블록도이다. 도 1을 참조하면, 본 발명에 따른 웹 페이지 검출 시스템(1)은 웹 페이지 이미지 획득 모듈(100), 사각형 이미지 추출 모듈(110), 특징 벡터 추출 모듈(120), 특징 벡터 보정 모듈(130) 및 분류기(140)를 구비하여, 임의의 웹 페이지가 음란 이미지를 포함하는 성인용 웹 페이지인지 여부를 검출하는 것을 특징으로 한다. 이하, 전술한 구성 요소들에 대하여 구체적으로 설명한다.

[0025] 상기 웹 페이지 이미지 획득 모듈(100)은 임의의 웹 페이지를 하나의 이미지파일로 캡처하여 웹 페이지 이미지를 생성한다.

[0026] 상기 사각형 이미지 추출 모듈(110)은 상기 웹 페이지 이미지 획득 모듈에 의해 생성된 웹 페이지 이미지로부터 이미지내의 특징들을 이용하여 경계선을 그릴 수 있는 사각형 이미지들을 검출하여 잘라내고, 상기 잘라낸 사각형 이미지들 중 사전 설정된 크기 조건 및 중형비 조건을 만족하는 사각형 이미지들을 추출한다.

[0027] 도 4 및 도 5는 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 시스템에 있어서, 사각형 이미지 추출 모듈의 동작을 설명하기 위하여 예시적으로 도시한 것으로서, 도 4는 웹 페이지 이미지이며, 도 5는 도 4의 웹 페이지 이미지로부터 추출한 사각형 이미지들이다. 하나의 이미지에 포함된 사각형 이미지들을 검출하는 방법은 영상처리분야의 당업자들에게 널리 알려진 공지 기술이므로, 본 명세서에서는 이에 대하여 구체적인 설명은 생략한다.

[0028] 상기 특징 벡터 추출 모듈(120)은 상기 사각형 이미지 추출 모듈로부터 추출된 복수 개의 사각형 이미지들을 이용하여 해당 웹 페이지에 대한 특징 벡터를 구한다. 본 발명에 따른 상기 특징 벡터 추출 모듈이 웹 페이지에 대한 특징 벡터를 구하는 과정을 보다 구체적으로 설명하면, 상기 특징 벡터 추출 모듈은, 상기 이미지 추출 모듈로부터 추출된 복수 개의 사각형 이미지들의 각각에 대하여 사전 설정된 방법을 이용하여 성인용 이미지 확률값들을 구하고, 상기 성인용 이미지 확률값들로 이루어진 특징 벡터를 생성한다. 이때, 상기 사각형 이미지에 대한 성인용 이미지 확률값은 해당 사각형 이미지가 얼마나 포르노 영상에 가까운지를 나타내는 값으로서 0 ~ 1 사이의 숫자로 표현된다. 또한, 상기 특징 벡터는 웹 페이지에 포함된 사각형 이미지의 개수에 대응되는 차원을 갖는 것을 특징으로 한다.

[0029] 한편, 사각형 이미지에 대한 성인용 이미지 확률값을 구하고 이를 이용하여 분류하는 방법은 당업계에서 공지된 기술로서, NSFW(Not Suitable for work) 등이 있다. NSFW(Not Suitable for work)는 폭력적/포르노 이미지 등을 구별할 수 있도록 CNN(Convolutional Neural Network)모델을 미리 학습시켜 놓은 분류기로서, 이는 야후(Yahoo)에서 공익 목적으로 인터넷 상의 다수의 폭력적/포르노 이미지를 수집하여 학습시킨 CNN기반 분류기이다. 본 발명에 따른 시스템 및 방법은 전술한 NSFW 뿐만 아니라 다른 종류의 분류기에도 적용될 수 있다.

[0030] 예를 들면, 하나의 웹 페이지 이미지에 10개의 사각형 이미지들이 있고, 그 중 3개의 사각형 이미지가 사전 설정된 크기와 중형비를 만족하는 경우, 사전 설정된 크기와 중형비를 만족하는 3개의 사각형 이미지를 추출하여 이들 각각에 대하여 성인용 이미지 확률값들을 구하고, 3개의 성인용 이미지 확률값들로 구성된 3차원의 특징 벡터를 생성하게 된다. 여기서, 3개의 성인용 이미지 확률값들이 0.57, 0.9, 0 이라고 가정하면 상기 특징 벡터는 $< 0.57, 0.9, 0 >$ 이 된다.

[0031] 상기 특징 벡터 보정 모듈(130)은, 각 웹 페이지에 포함된 사각형 이미지들의 개수가 서로 상이하므로 이를 보완하기 위하여, 상기 특징 벡터 추출 모듈에 의해 구한 웹 페이지에 대한 특징 벡터의 차원을 보정하여 사전 설정된 차원을 갖는 최종 특징 벡터를 구한다. 본 발명에 따른 상기 특징 벡터 보정 모듈이 최종 특징 벡터를 구하는 과정을 보다 구체적으로 설명하면, 상기 특징 벡터 추출 모듈에 의해 구한 웹 페이지에 대한 특징 벡터가 사전 설정된 차원보다 작은 경우 상기 특징 벡터를 구성하는 값들을 복제하여 사전 설정된 차원과 같거나 크게 만들어, 웹 페이지에 대한 특징 벡터의 차원이 사전 설정된 차원과 같거나 그보다 크게 만든다. 다음, 상기 웹 페이지에 대한 특징 벡터를 구성하는 값들을 분류하고, 상기 특징 벡터로부터 사전 설정된 차원에 대응되는 최대값들만을 추출하여 최종 특징 벡터를 생성한다.

- [0032] 도 3은 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 시스템에 있어서, 특징 벡터의 생성 및 보정을 설명하기 위하여 도시한 도표이다. 예를 들면, 도 3에 표시된 바와 같이, 상기 특징 벡터가 $< 0.57, 0.9, 0 >$ 인 경우, 이를 4번 복제하여 $< 0.57, 0.57, 0.57, 0.57, 0.9, 0.9, 0.9, 0.9, 0, 0, 0, 0 >$ 의 12 차원이 되고, 이를 순차적으로 sorting 하여 $< 0.9, 0.9, 0.9, 0.9, 0.57, 0.57, 0.57, 0.57, 0, 0, 0, 0 >$ 이 된다. 다음, 사전 설정된 차원이 10 차원인 경우, 순서대로 분류된 12차원의 특징 벡터로부터 10개의 최대값들을 선택하여 사전 설정된 10차원의 최종 특징 벡터인 $< 0.9, 0.9, 0.9, 0.9, 0.57, 0.57, 0.57, 0.57, 0, 0 >$ 을 생성하게 된다.
- [0033] 상기 분류기(140)는 상기 특징 벡터 보정 모듈에 의해 구한 사전 설정된 차원을 갖는 최종 특징 벡터를 이용하여 해당 웹 페이지가 성인용 웹 페이지인지 여부를 판단하게 된다.
- [0034] 본 명세서에서는, 설명의 편의상, '블랙 페이지(Black page)'는 성인용 이미지를 포함하거나 성인 콘텐츠 또는 사행 콘텐츠 등을 포함하는 성인용 웹 페이지들을 지칭하는 것으로 사용하며, '화이트 페이지(White page)'는 성인용 이미지, 성인 콘텐츠, 사행 콘텐츠가 없는 일반 웹 페이지들을 지칭하는 것으로 사용한다. 상기 분류기는 다수개의 블랙 페이지들과 다수 개의 화이트 페이지들을 포함하는 데이터셋을 이용하여 사전 학습하여 완성된 분류기이다. 따라서, 상기 분류기는 입력된 최종 특징 벡터를 이용하여 해당 웹 페이지를 블랙 페이지 또는 화이트 페이지로 분류하게 되며, 블랙 페이지는 성인용 웹 페이지로 판단하게 된다.
- [0035] 이하, 도 2를 참조하여 전술한 성인용 웹 페이지 검출 시스템에 적용된 성인용 웹 페이지 검출 방법에 대하여 구체적으로 설명한다. 본 발명에 따른 성인용 웹 페이지 검출 방법은 검출 시스템을 구성하는 마이크로프로세서 등의 제어부에 의해 구현된다.
- [0036] 도 2는 본 발명의 바람직한 실시예에 따른 성인용 웹 페이지 검출 방법을 순차적으로 도시한 흐름도이다. 도 2를 참조하면, 본 발명에 따른 성인용 웹 페이지 검출 방법은, 먼저 임의의 웹 페이지를 하나의 이미지로 캡처하여 웹 페이지 이미지를 생성한다(단계 200).
- [0037] 다음, 상기 획득된 웹 페이지 이미지로부터 이미지내의 특징들을 이용하여 경계선을 그릴 수 있는 사각형 이미지들을 검출하여 잘라내고, 상기 잘라낸 사각형 이미지들 중 사전 설정된 크기 조건 및 중형비 조건을 만족하는 복수 개의 사각형 이미지들을 추출한다(단계 210).
- [0038] 다음, 상기 웹 페이지에 대해 추출된 복수 개의 사각형 이미지들의 각각에 대하여 사전 설정된 방법을 이용하여 성인용 이미지 확률값들을 구하고, 상기 성인용 이미지 확률값들로 이루어진 특징 벡터를 생성한다(단계 220).
- [0039] 다음, 상기 웹 페이지에 대한 특징 벡터의 차원을 보정하여 사전 설정된 차원을 갖는 최종 특징 벡터를 구한다(단계 230). 상기 특징 벡터의 차원을 보정하는 과정을 보다 구체적으로 설명하면, 상기 특징 벡터 추출 모듈에 의해 구한 웹 페이지에 대한 특징 벡터가 사전 설정된 차원보다 작은 경우 상기 특징 벡터를 구성하는 값들을 복제하여 사전 설정된 차원과 같거나 크게 만든 후, 상기 웹 페이지에 대한 특징 벡터를 구성하는 값들을 분류하고, 사전 설정된 차원에 대응되는 최대값들만을 추출하여 최종 특징 벡터를 생성한다.
- [0040] 다음, 상기 웹 페이지에 대한 최종 특징 벡터를 이용하여 해당 웹 페이지가 성인용 웹 페이지인지 여부를 판단한다(단계 240).
- [0041] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

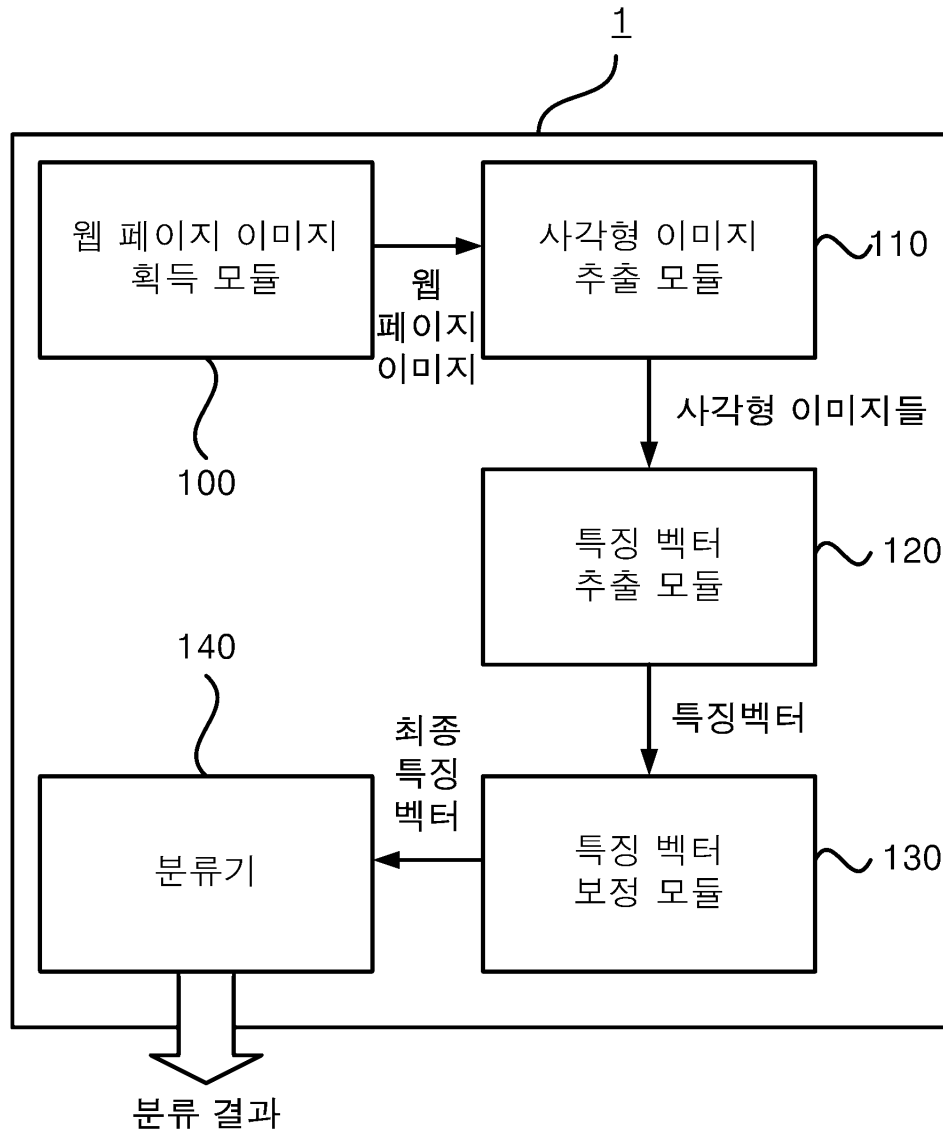
부호의 설명

- [0042] 1 : 웹 페이지 검출 시스템
100 : 웹 페이지 이미지 획득 모듈
110 : 사각형 이미지 추출 모듈
120 : 특징 벡터 추출 모듈
130 : 특징 벡터 보정 모듈

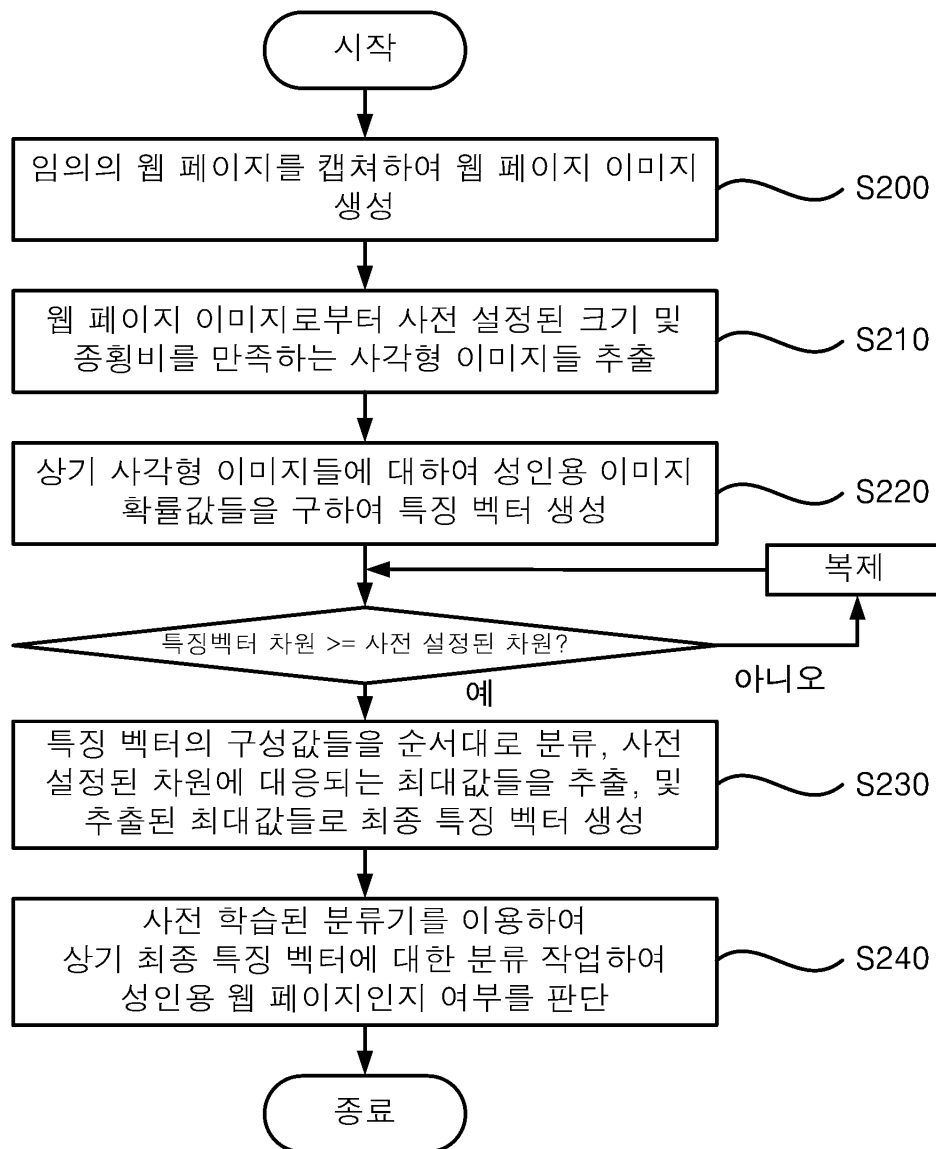
140 : 분류기

도면

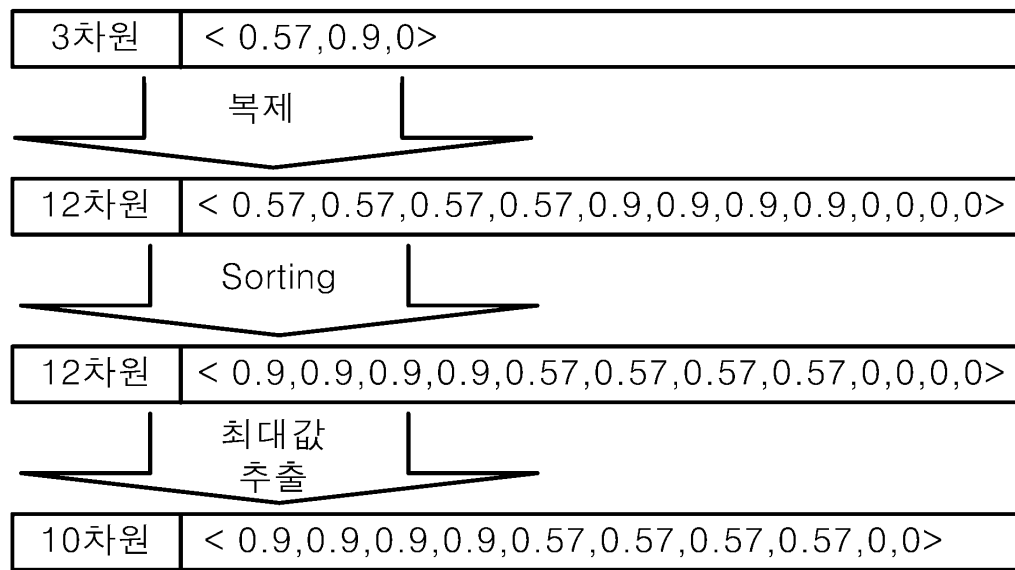
도면1



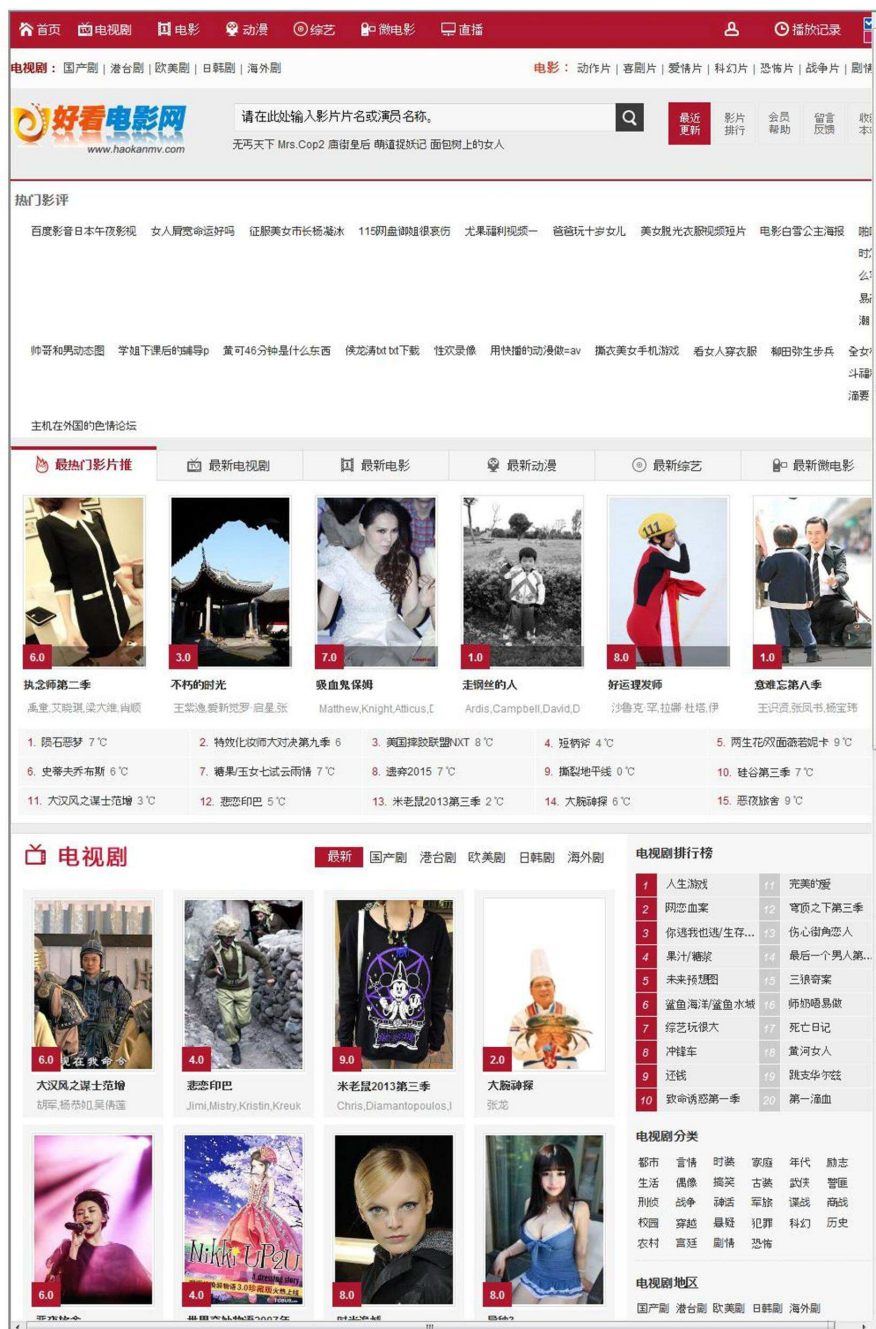
도면2



도면3



도면4



도면5

