

# 문화콘텐츠 진위 판정 지원 시스템의 고찰 및 제안

조 동 옥<sup>†</sup>

요 약

본 논문에서는 대표적인 문화콘텐츠인 게임등과 같은 프로그램, 자연어 다큐먼트 문화콘텐츠 그리고 중요한 오프라인 문화콘텐츠인 문화예술품에 대해 이를 디지털화하여 표절과 위작 여부를 판정하기 위한 기술적 방법론에 대한 고찰과 이를 지원하기 위한 새로운 방법론을 제안하고자 한다. 이를 위해 프로그램 소스 코드, 자연어 다큐먼트 문화콘텐츠에 대한 표절 검출 방법론과 이의 감정시 도움을 주기 위한 툴 등에 대해 살펴보고자 한다. 끝으로 문화예술품의 표절 및 위작 감정시 도움을 주기 위한 기술적 방법론을 작자의 드로잉 형태 분석에 의해 제안하고자 하며 실험에 의해 본 논문의 유용성을 입증하고자 한다.

## Survey and Proposal of Plagiarism Inspection Supporting System for Cultural Contents

Dong Uk Cho<sup>†</sup>

ABSTRACT

This paper describes the plagiarism inspection supporting system of typical cultural contents or properties such as game programs, cultural contents of natural language documents and art works and proposes the cultural contents the plagiarism inspection supporting system. For this, firstly, a survey of plagiarism inspection methods and tools for supporting plagiarism inspection about programs, cultural contents of natural language documents are discussed. Secondly, the technical method for supporting plagiarism inspection of cultural art works by analyzing the drawing characteristics of authors is proposed. Finally, the experimental results show the effectiveness of this paper.

키워드 : 문화 콘텐츠(Cultural Contents), 표절 감정(Plagiarism Inspection)

### 1. 서 론

최근의 IT기술은 제3세대 IT의 핵심 네트워크 분야를 넘어 네트워크상에 돌아다니는 콘텐츠가 중요한 시대가 되었고 그 중 문화콘텐츠분야, BT를 뒷받침 해주는 데이터 마이닝, 패턴인식분야등이 주된 핵심 과제가 되고 있는 실정이다[1-5].

특히 문화콘텐츠와 콘텐츠 저작자의 지적재산권 보호를 위해 표절이나 위작을 효과적으로 판정하기 위한 사회적 요구는 엄청난 사회적 요구 사항이 되고 있다[6]. 이중 문화콘텐츠와 관련된 분야는 문화예술품의 소장자, 국립중앙박물관, 문화재청등과 같은 공공기관이나 개인이나 할 것 없이 이를 어떻게 하든지 디지털화하여 표절이나 위작 감정시 기술적인 도움을 감정의 예비자료로 삼고자 하는 요구가 시급한 상황이 되었다[6]. 문화콘텐츠중 영화나 TV영상 그리고 각종 멀티미디어 저작 도구로 제작된 문화콘텐츠는

표절이나 복사를 방지하기 위해 워터마크 기법등[7,8]을 도입하고 있으나 게임같은 문화콘텐츠 프로그램, 자연어로 된 각종 다큐먼트 문화콘텐츠 그리고 문화예술품등의 경우는 실제 저작권 보호를 위한 기술적 지원 시스템이 전무했던 것이 사실이다. 특히 이같은 문화콘텐츠는 해당 분야 전문 감정인들이 표절 감정, 위작 감정등을 어떻게 하는지등에 대한 도움을 받아가며 알고리즘을 개발해야 하기 때문에 더 더욱 기술적인 표절, 위작 감정 지원 시스템의 개발이 어려웠던 것이 현 실정이다. 지금 현재 게임등과 같이 IT 기술에 기반한 문화콘텐츠에 대한 표절 감정은 프로그램심의조정위원회[9]에서 민, 형사상문제가 된 콘텐츠에 대한 표절 감정을 법원으로부터 위탁 받아 해당분야 전문가 즉, 감정인에게 의뢰를 하고 있다. 감정인은 이를 받아 고소인측의 고소이유서를 검토하고 프로그램 소스 코드 표절 감정 지원 툴을 돌려 이를 감정에 필요한 1차 예비자료로 삼는다. 추후 각종 콘텐츠에 대한 심도 있는 감정을 수행하여 이를 최종 감정 평가서로 프로그램심의조정위원회에 제출하게 되는 것이 현 감정의 흐름도이다[10]. 그러나 게임, 음

<sup>†</sup> 정 회 원 : 충북과학대학 정보통신학과  
논문접수 : 2003년 7월 24일, 심사완료 : 2003년 12월 31일

악등과 같은 디지털소스에 대한 감정 지원 툴은 존재하고 있지만 오프라인 형태로 존재하는 문화예술품에 대한 감정 툴은 그 자체가 디지털소스가 아닌 관계로 그리고 맨 처음부터 해당분야 전문가의 전폭적 도움을 받아 알고리즘을 개발해야하는 특수성으로 말미암아 기술적 감정 지원 시스템의 개발이 전무한 실정이었다. 따라서 오프라인 형태의 문화예술품을 디지털화하여 감정시 기술적으로 도움을 줄 수 있는 위작 감정 지원 시스템의 개발은 심각한 사회적 요구 사항이 되고 있다. 이를 위해 본 논문에서는 오프라인 형태의 문화예술품에 대한 감정 요소와 방법을 해당 분야 전문가에게 도움을 받아 감정인이 위작 여부를 판정시 기술적으로 지원해 줄수 있는 방법론을 제안하고자 한다. 이는 문화재청 전문가의 도움을 받아 과제를 수행하였으며 전체 개발 시스템중 본 논문에서는 우선적으로 작자의 드로잉 특성을 분석하여 전문 감정인이 감정시 기초 자료로 활용할 수 있도록 하기 위한 방법을 제안하고자 한다. 또한 이를 위해 지금까지 있어 왔던 각종 오프라인 문화예술품 위작 판정 지원 방법들에 대해서도 살펴보고자 하며 이를 본 방법과 타 방법과의 적용대상, 기술적 방법론의 차이등에 대한 비교, 고찰도 행하고자 한다. 그리고 게임, 자연어로 된 각종 다큐멘트와 같은 문화콘텐츠에 대해 표절 감정을 행하는것이 어떤방법으로 행해지는지 그리고 이와 관련된 툴등에 대한 비교, 분석도 행하여 문화콘텐츠 표절이나 위작과 관련된 IT기술의 적용등에 대한 고찰도 행하고자 한다. 끝으로 실험에 의해 본 논문의 유용성을 입증하고자 한다.

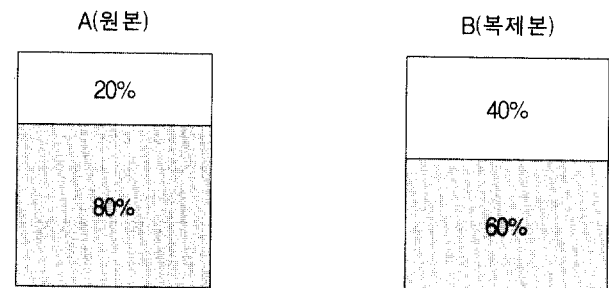
## 2. 디지털 문화콘텐츠의 표절 감정

디지털 문화콘텐츠의 대표적인 것이 게임등과 같은 경우이다. 이는 주로 프로그램 소스 코드의 비교를 통해 감정이 이루어지게 된다. 이를 위해 프로그램의 소스 코드를 감정해야 하는데 프로그램 소스 코드 감정이란 컴퓨터 프로그램 및 이와 관련된 전자적 정보등에 대하여 감정인이 전문적 지식과 기술을 활용, 비교 분석하여 복제 여부를 판정해주는 소견을 말한다[11].

통상 프로그램 소스 코드를 변형하여 표절을 행하는 방식에는 변수 이름 바꾸기, 명령문의 위치 변경, 주석문의 변경, 동의·중복 연산 사용, 동의 피연산자 사용, 인수 전개등이 있다. 이 같이 변형된 코드에 대해 표절을 감정키 위해서는 프로그램 전문가가 우선적으로 소를 제기한 쪽의 소제기의견서 그리고 이를 감정시 기술적으로 도움을 받기 위한 표절 감정 소프트웨어 지원 툴을 1차 자료로 삼고 추후 이를 기초로 전반적인 표절에 대한 감정을 수행하는 것이 현재 감정방법이다. 본 장에서는 우선적으로 소스 코드를 감정하는 일반적인 방법론에 대해 살펴보고 차후 관련 툴에 대한 소개와 분석을 행하고자 한다.

### 2.1 소스코드의 표절 감정

프로그램 소스 코드의 감정 방식은 아래 (그림 1)과 같이 우선 A라는 프로그램과 B라는 프로그램 중 어느쪽을 복제도 산출의 기준으로 하는가 이다.



(그림 1) 프로그램 복제의 산출 기준

복제도 산출의 기준은 A(예 : A의 80%)의 얼마나 많은 부분이 B에 복제되었는가 하는 원본 기준과 B의 얼마나 많은 부분이 복제한 것인지를 기준(예 : B의 60%)으로 하는가 하는 복제본 기준이 될 수 있다. 현재 통상 원본 기준 방식을 채택하고 있으며 주요 복제 대상으로 핵심(core)코드, 많은 시간/인력 소요 부분, 독창적 아이디어의 구현부분이 될 수 있다. 즉, 일반적인 감정 항목이란 프로그램의 구조, 기능, 소스코드 및 프로그램의 특성에 따른 특정 항목이 되며 감정 항목별 중요도  $W_i$ 를 결정하여 최종 복제도를 아래 식 (1)과 같이하여 구한다.

$$d = \sum_{i=1}^m W_i d_i \quad (1)$$

윗식에서  $W_i$ 는 항목별 중요도이며,  $d_i$ 는 감정 항목별 복제도 계산 결과를 뜻한다. 그러나 현재와 같은 감정 방법은 백분율로 복제도를 나타내는 것에 대한 문제 해결과 감정인이 임의로 설정하는 항목별 중요도 설정에 대한 임의성을 최소화하는 방안이 강구되어야 하는 문제를 내포하고 있다. 아울러 프로그래밍 언어별 복제도 계산방법, 프로그램 규모별 복제도 계산방법, 프로그램 기능/용도별 복제도 계산 방법등에 대한 표준화도 연구과제가 될 수 있다.

### 2.2 표절 검사 소프트웨어 툴

프로그램 소스코드의 표절을 검사하는 소프트웨어 툴은 크게 구조 연관 검사기법과 지문법이 많이 사용된다. 이중 구조 연관 검사(Structure Related Detection) 방법[12]은 프로그램 소스 코드에서 토큰을 추출하여 나열하면 선형 구조를 이루는 서열이 만들어지게 되고 프로그램 소스코드는 제어 흐름의 변경이 어렵다는 사실을 이용하여 표절을 검사하는 방법이다. 대표적인 소프트웨어 툴로는 YAP, MOSS, YAP3, JPlag등[13]이 있으며 아래 <표 1>에 이에 대한 소개를 나타내었다.

〈표 1〉 구조 연관 검사기법 소프트웨어 툴

툴이름	방 법	웹 사이트	비 고
YAP	Longest Common Subsequence	http://www.ccsr.cam.ac.uk/~mw263/YAP.html	
YAP3	String Matching Method	http://www.ccsr.cam.ac.uk/~mw263/YAP.html	
JPlag	상 동	http://www.Jplag.de	온라인
MOSS	String Matching Method	http://www.cs.berkeley.edu/~aikem/moss.html	다양한 언어지원

이에 비해 통계적인 특징을 추출하여 표절을 검사하는 방법이 지문 검사법(Fingerprint Detection)이며, 대표적인 툴이 SIM과 Siff등이다. SIM은 소스코드를 1바이트씩 토큰들로 만들어서 토큰의 정보와 토큰의 위치들을 배열에다 할당하여 체인 해쉬 기법을 사용하여 비교하는 방식을 채택하고 있고, Siff는 50개의 대표적인 문자들을 추출하여 비교하는 방식을 사용하고 있다. 아래 〈표 2〉에 이에 대한 비교표를 나타내었다.

〈표 2〉 지문검사법을 사용하는 툴에 대한 비교

툴이름	방 법	비 고
Siff	50개의 대표 문자 추출	오프라인
Windiff	비공개	오프라인, GUI환경
SIM	토큰화 시킨후 참조 횟수 비교	오프라인

### 2.3 현재까지의 문제점 제기

현재까지 프로그램 소스코드의 표절검사 방법의 문제점을 요약하면 아래 〈표 3〉과 같다.

따라서 이같은 문제점에 대한 개선 방안이 지속적으로 연구가 수행되어야 보다 객관성 있는 표절 감정이 되리라 여겨진다.

〈표 3〉 현재까지의 문제점

문 제 점	내 용
감정방식	복제도 판단시 백분율 적용방식의 허와 실
감정기법 모델링 및 표준화	• 프로그래밍 언어별, S/W분야별 감정기법 정형화 • 복제도감정, 개발공정감정, 가치평가등의 표준기법개발
항목의 중요도 선정	감정 항목별 중요도를 감정이 설정하는데 이의 임의성을 최소화

## 3. 자연어 다큐먼트 콘텐츠에 대한 표절 감정

자연어로 된 다큐먼트 콘텐츠에 대한 표절 행위도 작게는 교육현장에서 숙제나 보고서의 표절에서부터 시작해서 크게는 기관이나 업체의 운명이 좌우될 정도의 표절까지

그 범위는 상당히 크다고 할 수 있다. 본 장에서는 이같은 자연어 다큐먼트 콘텐츠 표절에 대해 자연어 표절의 형태, 이의 표절을 검출키 위한 방법론 그리고 이의 소프트웨어 툴에 대해 살펴 보기로 한다.

### 3.1 자연어 다큐먼트 콘텐츠의 표절 형태

자연어 다큐먼트 콘텐츠의 표절은 통상 아래 〈표 4〉와 같은 형태로 이루어진다[14].

〈표 4〉 자연어 다큐먼트 콘텐츠 표절의 형태

<ul style="list-style-type: none"> <li>• 문서의 구조와 구절의 구조가 같은 경우</li> <li>• 적절한 동의어로 대체한 경우</li> <li>• 문장의 순서를 바꾼 경우</li> <li>• 추상적인 개념을 구체적으로 변경하여 서술한 경우</li> <li>• 잘못된 단어나 어구, 틀린 철자를 그대로 사용하는 경우</li> <li>• 문장 줄이기</li> <li>• 문법적 변환</li> <li>• 문장의 조합</li> </ul>
---

### 3.2 자연어 다큐먼트 콘텐츠의 표절 검출 방법

자연어 다큐먼트 콘텐츠의 표절 검출 방법은 문서의 문맥 흐름 확인, 특정 단어의 사용횟수와 위치 확인, 특정 문장을 추출하여 비교 확인과 통계적인 기법을 적용하는 방법이 있다. 아래 〈표 5〉에 통계적인 기법을 적용하는 방법에 대해 나타내었다.

〈표 5〉 자연어 다큐먼트 콘텐츠의 표절 검출을 위한 주요한 통계 자료

<ul style="list-style-type: none"> <li>• 문장의 평균 길이</li> <li>• 구나 절의 평균 길이</li> <li>• 핵심 키워드(function words)의 사용 빈도</li> <li>• 수동태와 능동태의 사용 형태</li> <li>• 전치사의 사용 빈도</li> <li>• 단어당 평균 문자수</li> <li>• 단어의 길이 분포</li> </ul>
---

### 3.3 자연어 다큐먼트 콘텐츠의 표절 검출 S/W 툴

자연어 다큐먼트 콘텐츠의 표절 검출을 행하기 위해 이의 표절을 지원해 주는 S/W 툴 등이 나와 있다. 이는 다큐먼트 콘텐츠들로부터 단어들의 출현 빈도를 추출한 뒤 이를 이용하여 가장 적합한 클러스터링을 찾아 다큐먼트 콘텐츠의 표절을 검출하는 통계적 방법을 이용한 클러스터링 방법과 키워드 집합을 이용해 클러스터링, 하고자 하는 문서가 그 키워드 집합을 내포하는 정도에 따라 클러스터링을 정하여 문서의 표절을 검출하는 지식 기반 방법 그리고 통계적 방법과 지식 기반 방법을 결합한 하이브리드 방법이 있다. 아래 〈표 6〉에 국내·외 자연어 다큐먼트 콘텐

츠 표절 검출 S/W들에 대한 소개를 나타내었다.

〈표 6〉 자연어 다큐먼트 콘텐츠 표절 검출 S/W 들

상품명	방법	회사명	웹 사이트
Findsame	비공개	Digital integrity	http://www.digital-integrity.com
EVE2	비공개	CaNexus	http://www.CaNexus.com
Turnitin	비공개	iParadigms	http://www.turnitin.com
Copy Catch	공통어휘 빈도수 검사	CFL s/w Developments	http://www.CopyCatch.freeseerve.co.uk
Word CHECK	단어 사용 횟수	WordCHECK Systems	http://www.WordCHECK.systems.com
교수클럽	유사 어절 트리	교수클럽	http://www.gyosuclub.com

#### 4. 문화예술품의 표절감정

고문헌 또는 예술품등에 대한 표절 및 저작자 진위 논쟁은 우리나라 뿐만 아니라 전세계적으로 많은 논란이 되어왔다. 그러나 이의 표절이나 위작 여부를 판정하기 위해서는 해당분야 전문가에게 의뢰해야 한다. 이때 감정인에게 표절 및 진위 여부 진위 판정시에 도움을 줄 수 있는 기술적 소견을 함께 제시하여 준다면 감정인의 진위 여부 판정에 큰 도움이 될 수 있으리라 여겨진다. 본 논문에서는 이를 위해 문화재청[15]의 해당 분야 전문가의 도움을 받아 문화예술품의 감정시 사용하는 요소에 대해 그리고 그 요소중 디지털 처리가 가능한 분야에 대해 조언을 받아 문화예술품의 표절 및 위작 판정시 감정인에게 도움을 줄수 있는 방법론을 제안하고자 한다. 우선 문화예술품의 감정시 사용되는 요소는 문화재청 전문가의 의견을 참조하면 아래 <표 7>과 같다.

이 중에서 사용한 재질이 중요한 판정기준이 되지만 이는 화학적 분석을 통해야만 가능하고 컴퓨터로 처리를 행할수 없기 때문에 직선이나 곡선의 처리형태, 색상특성, 낙관이나 소장자인에 대한 패턴 인식이 디지털 처리가 가능한 요소가 된다. 따라서 본 논문에서는 우선적으로 전체 진위 판정 시스템중 곡선이나 직선의 처리 형태를 분석하여 즉, 작자의 드로잉 형태 분석을 통해 표절이나 위작 여부를 진품을 판정하는데 도움을 줄수 있는 방법론에 대해 제안하고자 하며 차후 색상 특징 분석, 낙관이나 소장인에 대한 패턴 분류등과 같은 방법론을 추가로 개발하여 전체 완성된 시스템을 구축하고자 한다.

〈표 7〉 미술품이나 고문헌의 주요 특징 요소

대상	주요 특징 요소
미술품	직선이나 곡선의 처리 형태, 색상 특징, 사용한 재질, 낙관
고문헌	직선이나 곡선의 처리 형태, 서지형태, 사용한 재질, 낙관, 소장인

#### 4.1 관련 분야 연구에 대한 소개 및 본 방법과의 비교, 고찰

지금까지 문화예술품에 대한 표절, 위작 감정은 주로 문헌의 작자 진위 여부에 대해 이루어져 왔다. 이때 문헌의 작자에 대한 진위 판정은 다음과 같은 요소를 사용한다.

첫째, 기록 내용에 대한 검토와 역사적 사실에 대한 고증이다.

둘째로는 원본자료가 있을시 그 필적에 대한 감정 그리고 셋째가 사용된 종이의 지질, 잉크 혹은 먹물등의 화학적 분석에 의한 연대 추정이다. 이중 디지털 처리가 가능한 것이 필적 감정 분야이다. 지금까지 주로 해 온 방법은 문장의 길이, 단어의 길이, 단어의 문자수, 품사의 출현율, 특수한 말의 출현율등과 같은 문장의 수량적 특성을 조사하여 분석함으로써 저자의 문장스타일과 습관등을 계량적으로 파악하여 진위 여부를 판정하는 방법론이 주류를 이루고 있다. 예로서 드모르간은 캠브리지의 목사 헤랄드에게 편지를 보내 신약 성서중 바울서간 집필자에 대한 추정 문제와 관련하여 각 서한에 대한 단어당 평균 문자수를 이용[16, 17]하여 작자의 진위를 찾아 낼 수 있음을 밝혔다. 이를 바탕으로 오하이오대학의 물리학자 멘델 홀은 단어의 평균 문자수 보다는 단어 길이의 범위와 분포를 이용하는 것이 더욱 효과적이라는 제안을 하여 셰크스피어와 베이컨의 표절 여부를 판정하였다. 이외에 구약성서의 이사야서[18, 19]에 대한 진위 여부, 중국의 경우도 홍루몽의 일부분에 대한 진위 여부[20] 판정등의 연구가 진행되어 왔다. 그러나 이상의 경우는 모두 문헌의 문장 습작 특성을 분석한 방법이었다. 이를 미술품 등과 같은 문화콘텐츠에는 적용할 수 없기 때문에 작자의 드로잉 형태의 습성 분석을 통한 진위 판정 지원 시스템을 개발하는 것이 적합한 접근 방식이라는 결론에 도달하게 된다. 따라서 본 연구에서는 작자의 드로잉 형태 분석, 사용하는 색상의 분석등을 통해 진위 판정시 도움을 줄수 있는 시스템을 개발하고자 한다. 이중 본 논문은 개발하고자 하는 전체시스템중 작자의 드로잉 형태 분석에 대한 방법론을 개발하고자 한다. 이를 문헌에도 적용하여 서체의 곡선이나 직선의 처리 형태를 분석하여 진위 판정 지원 시스템을 제안하고자 한다. 아래 <표 8>에 본 방법과 타 방법과의 비교, 고찰 표를 나타내었다.

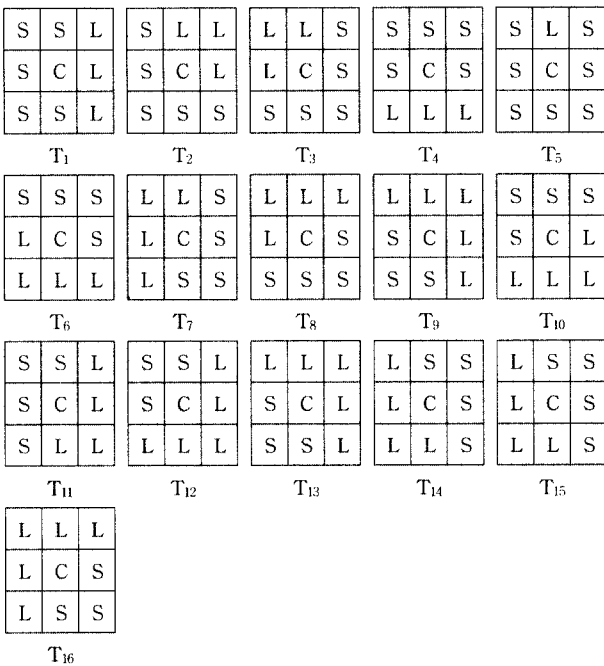
〈표 8〉 문화예술품의 진위 판정을 위한 방법론의 비교

개발자	적용 대상	주된 개발 방법
드모르간	바울서신	단어당 평균 문자수
멘델 홀	셰크스피어와 베이컨	단어 길이의 범위와 분포 이용
村上恒勝	구약 성경	무선 표집을 통해 상관관계 분석
李賢平	홍루몽	구성분분석, 정준상관 분석, 군집 분석
제안한 방법	회화, 직지	작자 고유의 드로잉 형태, 색상 분석

4.2 작자의 드로잉 형태 분석에 의한 진위 여부 판정

작자의 드로잉 형태 습성에 따라 그림이나 서체의 등근 정도나 직선 처리등이 달리 나타나므로 이를 특징 벡터로 선정해서 표절이나 위작 감정시 지원시스템의 한 요소로 삼고자 한다. 이를 위해 아래 (그림 1)과 같이 작자의 드로잉 형태 분석을 위해 영상에 있어 경계선의 형태를 16개의 유형으로 구분한다. 이때 전체 영상을 16×16의 부영상으로 나누고 이의 히스토그램 분포를 통해 특징 벡터를 추출한다.

위의 (그림 1)에서 T<sub>1</sub>~T<sub>16</sub>은 작자의 드로잉 형태에 대한 원형(Template)을 뜻한다. 또한 “S”란 중심화소 “C”(Center)를 중심으로 명암도의 변화가 작다(Small)라는 의미를, “L”은 크다(Large)는 것을 의미한다. 이때 ‘S’와 ‘L’의 계산은 식 (2), 식 (3)에 의해 구하며 히스토그램의 Y축 눈금 정규화는 식 (4)와 같이 정의한다.



(그림 1) 경계 영역의 형태 분석

$$f_{Large}(X) = \frac{X}{255} \tag{2}$$

여기서 X = |C - L|

$$f_{Small}(X) = \frac{-X+255}{255} \tag{3}$$

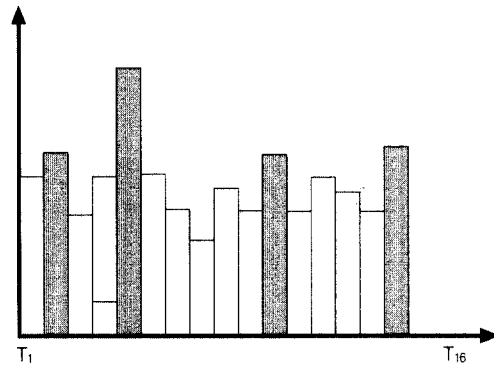
여기서 X = |C - S|

$$Y.N = \frac{\sum T_i}{14 \times 14} \tag{4}$$

여기서 i = 1, 2, ..., 16이다.

식 (2)와 식 (3)은 0~255의 명암도를 가지는 영상에서 명암도의 차를 0~1사이의 값으로 정규화를 행하는 수식이 된다. 즉, 식 (2)의 경우는 그 값이 1에 가까울수록 중앙화소를 중심으로 명암도의 차가 크다는 것을 의미하며, 역으로 식 (3)은 명암도의 차가 작다는 것을 의미하게 된다. 아울러 전체 영상을 16×16의 부영상으로 나누어 3×3의 창에 대해 적용하므로 T<sub>1</sub>~T<sub>16</sub>의 값을 히스토그램으로 나타낼시 정규화를 위해 식 (4)을 정의하였다.

이제 각 히스토그램 중에서 히스토그램 누적 분포가 가장 큰 4개의 부분들을 예로서 아래 그림과 같이 선정하여 다음과 같은 다항식으로 나타낸다.



(그림 2) Dominant 점들의 선정

(n+1)개의 점에 대하여 n차식의 다항식은

$$(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n) \tag{5}$$

$$P_n(x) = C_0 + C_1X + \dots + C_n X^n \tag{6}$$

$$P_n(x_k) = y_k : k=0, 1, \dots, n \tag{7}$$

이다. 차수 n에 대한 Lagrange 다항식은 식 (8)을 만족한다.

$$\begin{aligned} \text{if } i \neq k \quad L_k(x_i) &= 0 \\ \text{else } L_k(x_i) &= 1 \end{aligned} \tag{8}$$

따라서 n차다항식은 식 (9)와 같이 나타내는 것이 가능하게 된다.

$$P_n(x) = \sum_{k=0}^n L_k(x)y_k = \sum_{k=0}^n \left[ \prod_{i=0, i \neq k}^n \frac{x - x_i}{x_k - x_i} \right] y_k \tag{9}$$

비매개변수 Hermite 곡선에 대해서는 식 (10)과 같이 표현할 수 있다.

$$P_{0,3}(x) = y_0L_0(x) + y_1L_1(x) + y_2L_2(x) + y_3L_3(x) \tag{10}$$

5. 실험 및 고찰

본 논문에서의 시험은 IMB-PC상에서 행하였다. 아래 (그림 3)과 (그림 9)이 원본이고 (그림 4), (그림 5) 그리고



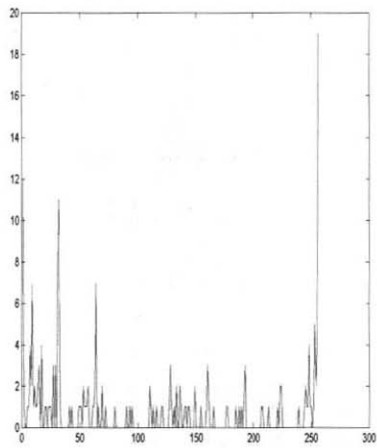
(그림 3) 실험미술품(원본)



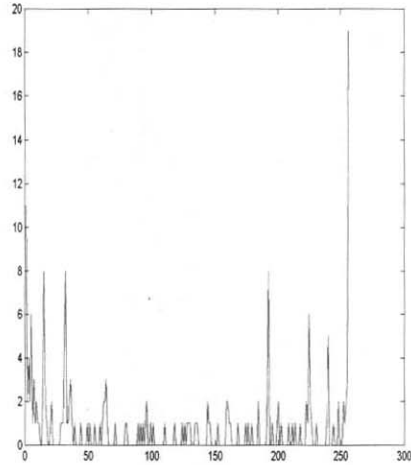
(그림 4) 실험미술품(표절)



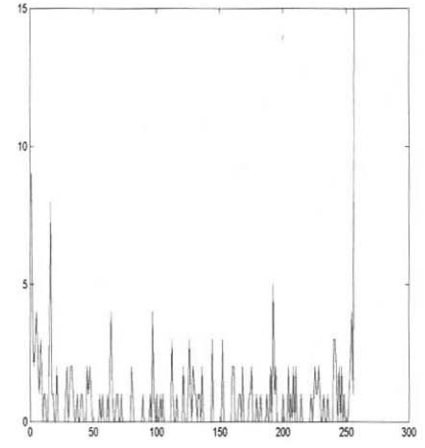
(그림 5) 실험미술품(표절)



(그림 6) 처리형태분석(원본)



(그림 7) 처리형태분석(표절)



(그림 8) 처리형태분석(표절)



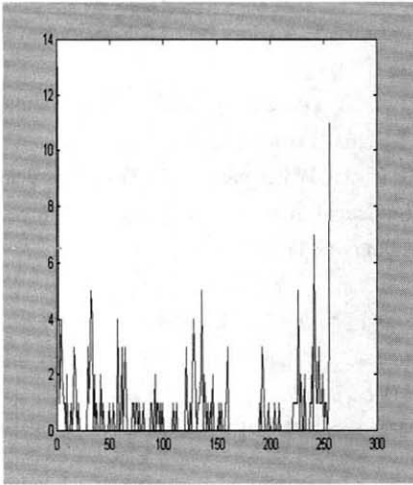
(그림 9) 실험미술품(원본)



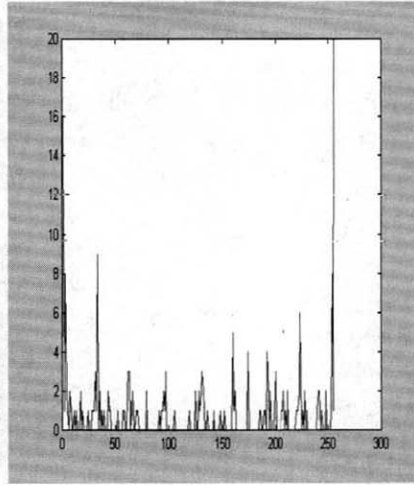
(그림 10) 실험미술품(표절)



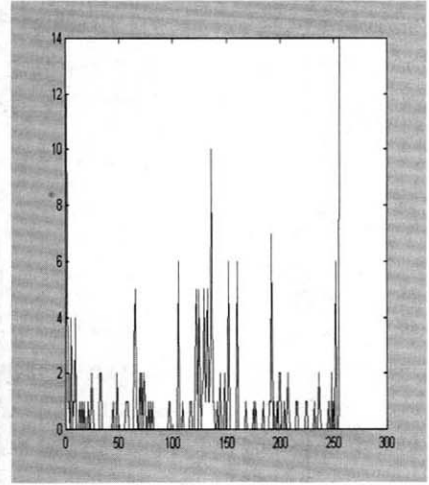
(그림 11) 실험미술품(표절)



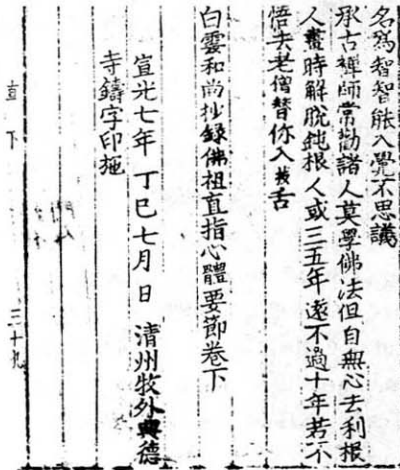
(그림 12) 처리형태분석(원본)



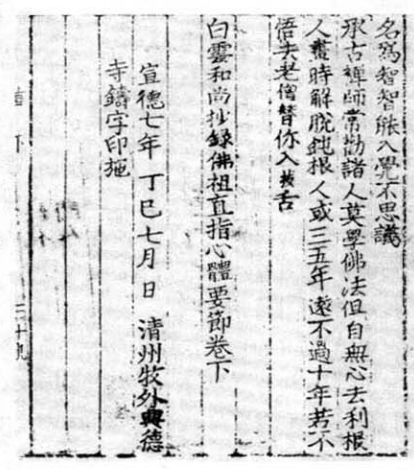
(그림 13) 처리형태 분석(표절)



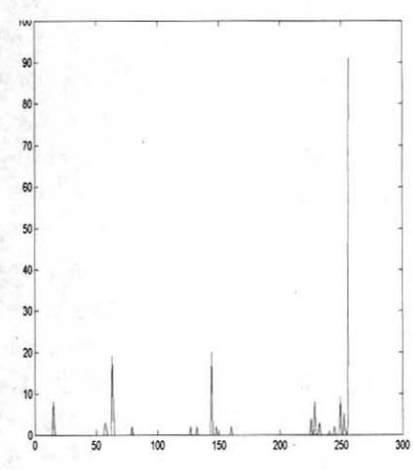
(그림 14) 처리 형태 분석(표절)



(그림 15) 고문헌직지(진품)



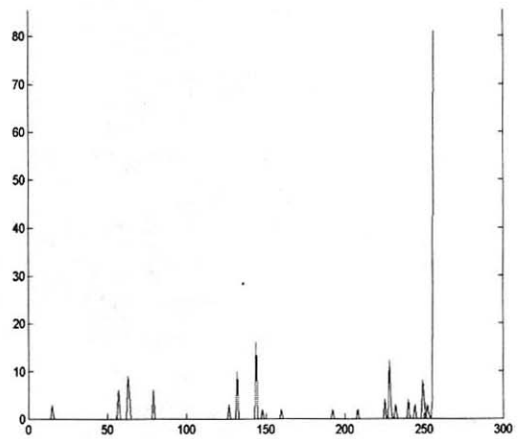
(그림 16) 고문헌직지(위작)



(그림 17) 처리형태분석(진품)

(그림 10)과 (그림 11)은 표절작품이다. 그리고 이에 대해 직선이나 곡선의 처리 형태를 분석한 결과 즉, 드로잉 형태를 분석한 결과를 히스토그램으로 나타낸 것이 (그림 6), (그림 7), (그림 8) 그리고 (그림 12), (그림 13), (그림 14)이다. 실험결과에서 알수 있듯이 미술품을 눈으로 보았을시는 구분이 제대로 안되지만 이에 대해 드로잉 형태를 분석한 결과는 원본과 표절 작품이 구분이 됨을 확인할 수 있다. 또한 고문헌에 대한 실험예를 (그림 15)에 보인다. 이는 청주 고인쇄 박물관에 소장되어 있는 직지인데 이를 연대를 조작하여 모작한 것이 (그림 16)이다. 이에 대한 패턴 처리의 결과 예를 (그림 17)과 (그림 18)에 보인다. 고문헌도 미술품과 마찬가지로 히스토그램의 분석 결과가 차이가 발생함을 확인할 수 있었다. 따라서 본 논문에서 특징 벡터로 선정된 곡선이나 직선의 처리 형태 분석은 대단히 유용한 특징 벡터가 됨을 확인할 수 있었으며 이는 저작의 습작 형태를 분석할 수 있는 효과적인 방법임을 알 수 있었다. 또한 본 방법은 문화재청 전문가의 의견과 도움을 받아

구현한 방법이기 때문에 즉, 해당 분야 전문 감정인의 의견과 식견을 디지털로 처리가능한 부분을 선택하여 처리한 것이기 때문에 감정시 충분히 감정인을 지원해 줄 수 있는



(그림 18) 처리형태분석(위작)



지원시스템이 될 수 있으리라 여겨진다. 향후는 작자마다 고유한 형태로 주로 사용하고 있는 색상 정보의 분석등도 행하여 중요한 지적재산권인 미술품이나 문헌등의 표절형태를 보다 판별력있게 처리할수 있는 시스템이 되기 위한 확장을 꾀해야 하리라 여겨진다.

## 6. 결 론

본 논문에서는 문화콘텐츠의 표절과 위작을 판정하는데 도움을 주기 위한 현재까지의 기술적 방법론에 대한 고찰을 행하였으며, 주요한 문화콘텐츠 분야인 문화예술품의 위작과 표절 행위 감정 지원 시스템에 대해 제안하였다. 크게 게임등과 같은 문화콘텐츠인 프로그램 소스코드, 자연어 다큐먼트 문화콘텐츠의 표절등에 대해 현재까지의 기술 방법, 현황, 소프트웨어 틀등에 대한 고찰을 행하였으며 문화예술품의 경우 저작자의 직선이나 곡선 형태 처리를 분석하여 즉, 드로잉 형태를 분석하여 원본과 표절 작품을 감정하는 지원시스템에 대해 제안하였다. 향후 연구과제로는 문화콘텐츠 관련 프로그램 소스코드의 표절을 검출하는 각종의 틀에 대해 여러 각도에서 비교·분석하여 각 도구의 유용성, 제안성, 주요 적용 환경 및 분야, 사용 방법들을 제시하고자 하며 궁극적으로는 표절 검출 도구를 자체 개발하는 것까지 연구를 수행하고자 한다. 또한, 문화예술품의 표절 검출에 대해서는 색상 정보에 대한 분석을 행하고 낙관이나 소장인에 대한 패턴인식을 행하여 제안한 방법의 유용성을 보완하고 더욱 효율성을 높이기 위한 알고리즘의 추가 개발과 실험을 수행하여 해당 분야 전문가들이 진위 판정을 행할시 보다 많은 객관적, 기술적 자료를 제공해 주는 전체시스템으로 확장하기 위한 노력이 지속적으로 이루어져야 하리라 여겨진다. 아울러 문화예술품은 진품이 하나이므로 이를 소지하고 있는 소장자의 협조를 받아 작가의 고유한 통계적 특성을 추출하기 위한 노력도 행해져야 하며 이를 통해 보다 효율적인 감정지원시스템이 되기 위한 종합적 작업이 이루어져야 하리라 여겨진다.

## 참 고 문 헌

[1] 서병문, 이제는 문화콘텐츠 산업이다, 한국정보처리학회 추계종합학술대회 초청강연, Nov., 2002.  
 [2] 서병문, 문화콘텐츠는 미래 경쟁력이다, 한국콘텐츠학회 춘계종합학술대회 초청강연, May, 2003.  
 [3] 조환규, "Genomic Sequence Aligment and its Application to Computing Linear Structure Similarity," 한국생물정보학회, 2002.  
 [4] Bioinformatics, 기술/시장보고서, 한국전자통신연구원, Oct., 2001.

[5] Bioinformatics특집, 한국정보과학회지, Aug., 2000.  
 [6] 중앙일보, 다산 정약용 글씨 '가짜'논란, 문화면, p.29, Oct., 2003.  
 [7] M. Barnietal, "A DCT-Domain System for Robust Image Watermarking," Signal Processing 66, 1998.  
 [8] Gerit C. Langella et al., "Watermarking by DCT Coefficient Removal : A Statistical Approach to Optimal Parameter Setting," Proceedings SPIE, Vol.3657, 1999.  
 [9] <http://www.pdmc.or.kr>.  
 [10] 프로그램심의조정위원회간 월간지, enter, 2003.  
 [11] 이중구, "컴퓨터 프로그램 저작권 보호 제도", 한국멀티미디어학회 워크샵, Dec., 2000.  
 [12] 황대준, 한국소프트웨어감정평가학회 춘계종합학술대회 논문집, June, 2003.  
 [13] <http://www.cs.usyd.edu.au/~michaelw/YAP.html>.  
 [14] 조동욱, 조맹섭, 전병태, "컴퓨터소프트웨어 감정 관련 국내·외 동향 조사 및 분석", 로그램심의조정위원회 최종연구보고서, 2002.  
 [15] <http://www.ocp.go.kr>.  
 [16] R. D. Lord, Studies in the History of Probaility and Statistics VIII. De Morgan and the Statistical Study of Literature Style, Biometrike, 45, 1958.  
 [17] S. E De Morgan, Memoir of Augutus de Morgan by his wife Sophie Elizabeth DeMorgan with selection from his letters. Longman, Gree and Co., 1982.  
 [18] L. L. Adams and A. C. Rencher, The popular critical view of the Isaiah problem in light of ststistical style analysis, Computer Studies in the Humanities and Verbal Behavior, 1973.  
 [19] 村上征勝, 眞偽の 科學, 朝倉書店, 1994  
 [20] 李賢平, 紅樓夢成書新說, 夏旦大學, 第5期, 1987.



## 조 동 욱

e-mail : [ducho@ctech.ac.kr](mailto:ducho@ctech.ac.kr)

1983년 한양대학교 공대 전자공학과 (공학사)

1985년 한양대학교 대학원 전자공학과 (공학석사)

1989년 한양대학교 대학원 전자통신공학과 (공학박사)

1991년~2000년 서원대학교 정보통신공학과 부교수

2001년 한국정보처리학회 우수논문상

2002년 한국콘텐츠학회 학술상

2003년 한국정보처리학회 우수논문상

2000년~현재 충북과학대학 정보통신학과 교수

관심분야 : 문화콘텐츠의 지적재산권 보호, 영상 콘텐츠 공학, 소프트웨어 프로그램 표절 감정, 영상 처리 및 인식