

빅데이터 기반 이미지 편집 온라인 솔루션

박병창, 조인화

경희대학교 컴퓨터공학과

{ hanbyul, inhwa1025 } @khu.ac.kr

Image Editing Online Solution Based on BigData

Byungchang Park, Inhwa Jo

요 약

SNS가 활성화되고 있는 요즘, 이미지에 대한 빅데이터가 기하급수적으로 증가하고 있다. 또한 빅데이터는 몇 년 전 클라우드 컴퓨팅 등과 함께 큰 주목을 받으며 등장한 이후, 최근 국내외에서 4차 산업혁명의 핵심기술로 급부상하고 있다. 따라서 본 문서는 빅데이터를 기반으로 하며 이미지 로드, 자르기, 뒤집기, 회전, 크기 조정, 그리기 등 다양한 기능을 지원하는 이미지 편집 온라인 솔루션을 제안하고 구현한다.

1. 서 론

빅데이터는 저장되지 않았거나 저장되더라도 분석되지 못하고 버리게 되는 방대한 양의 데이터를 말한다. 빅데이터는 몇 년 전 클라우드 컴퓨팅 등과 함께 큰 주목을 받으며 등장한 후, 최근 국내외에서 4차 산업혁명의 핵심기술로 급부상하고 있다. 세계 빅데이터 시장은 현재도 성장 중이며, 시장조사기관마다 규모의 차이는 다소 있으나 공통적으로 높은 성장률을 예측한다. 실제로도 빅데이터는 페이스북, 트위터 등의 소셜 네트워크에서 많이 발생하고 있는데, 이러한 방대한 데이터들을 어떻게 효과적으로 분석하는지에 대한 관심이 높아지고 있다.

빅데이터와 인공지능은 이미지 인식기술에도 많은 영향과 발전을 주었다. 위성사진을 인공지능으로 분석해 특정 국가의 원유 저장탱크의 산출 내역을 밝혀내거나, 사용자가 촬영하거나 다운로드한 이미지와 유사한 이미지나 제품을 검색해주기도 하며, 과일의 산출량을 정렬한다거나 식물의 질병을 탐지해 낼 수도 있다. 어도비 포토샵은 2014년부터 머신러닝을 적용한 기능을 추가하였으며 현재는 이미지 합성과 영상 편집을 비롯한 다양한 분야에서 머신러닝을 적용한 소프트웨어들이 출시되고 있다.

사용자 입장에서 가벼운 이미지 편집을 위해 무거운 프로그램을 돌리는 것은 꽤나 부담되는 일이라 생각될 수 있다. 이에 본 논문에서는 빅데이터 기반의 이미지 편집 솔루션 서비스들의 동향을 살펴보고, 온라인 기반으로 사용자가 이미지 로드, 자르기, 뒤집기, 회전, 크기

조정, 그리기 등 다양한 기능을 사용할 수 있는 빅데이터 기반의 이미지 편집 솔루션의 설계를 목적으로 한다.

2. 관련 연구

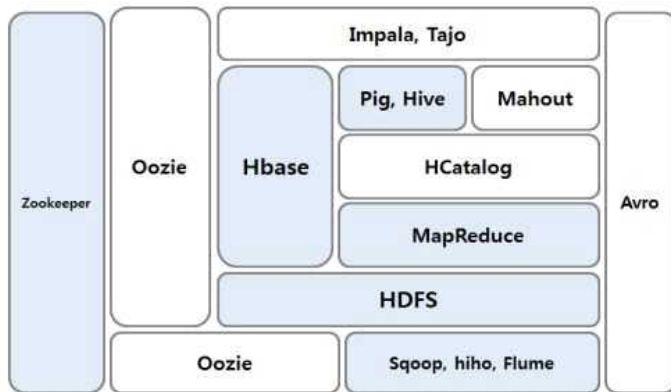
2.1 하둡(Hadoop)

Hadoop은 하나의 컴퓨터를 Scale up 하여 데이터를 처리하는 대신 적당한 성능의 범용 컴퓨터 여러 대를 Scale out한 후, 클러스터화 하여 큰 크기의 데이터를 클러스터에서 병렬로 동시에 처리한다. 이를 통해 처리 속도를 높이는 것을 목적으로 두는 오픈소스 프레임워크로 아래와 같은 모듈로 구성된다.

- Hadoop Common: Hadoop의 다른 모듈을 지원하기 위한 공통 컴포넌트 모듈
- Hadoop MapReduce: 대용량 데이터 처리를 분산 병렬 컴퓨팅에서 처리하기 위한 목적으로 제작된 소프트웨어 프레임워크
- Hadoop HDFS: 분산 저장을 처리하기 위한 모듈, 여러 개의 서버를 하나의 서버처럼 묶어서 데이터를 저장
- Hadoop YARN: 병렬 처리를 위한 클러스터 자원 관리 및 스케줄링 담당
- Hadoop Ozone: 하둡을 위한 오브젝트 저장소

Hadoop은 시스템을 중단하지 않고 장비의 추가가 용이하며 일부 장비에 장애가 발생하더라도 전체 시스템 사용성에 영향이 적다는 장점이 있다. 그렇지만 데이터를 DISK 기반으로 처리하기 때문에 데이터 처리

시간 외에도 read/write 연산에 추가 시간이 소요된다. 또한 동일 데이터에 대해서 작업할 때, 매번 read 연산이 필요하다.



[그림 1] 하둡에코시스템 구성도
[Fig. 1] Configuration of Hadoop Ecosystem

2.2 스파크(Spark)

Spark는 빅데이터에 주로 사용되는 빅데이터에 주로 사용되는 오픈소스 클러스터 컴퓨팅 프레임워크로 Hadoop의 MapReduce와 비슷한 역할을 한다. Spark는 통합 컴퓨팅 엔진이며 클러스터 환경에서 데이터를 병렬로 처리하는 라이브러리 집합이다. RDD(Resilient Distributed Datasets)를 사용하며 인메모리 저장 및 효과적인 장애복구 지원에 기반하여 동작한다. Spark는 반복적인 기계 학습 알고리즘에 대해 Hadoop과 비교해서 10배까지 빠른 성능을 나타낸다. Hadoop의 MapReduce는 작업의 중간 결과를 디스크에 써서 IO로 인해 작업 속도에 제약이 생기는 반면 Spark는 메모리에 중간결과를 저장하여 반복 작업의 처리 효율이 높다. Spark는 SQL, Streaming, 머신러닝, 그래프 연산 등 다양한 컴포넌트를 제공해준다. 또한 다양한 클러스터 매니저를 지원하여 YARN, Mesos, Kubernetes 등 다양한 클러스터에서 작동이 가능하다.

2.3 HTML5 Canvas API

Canvas API는 HTML상에서 JavaScript를 통해 그래픽을 그릴 수 있는 API다. Canvas API는 주로 2D 그래픽에 중점을 두고 있으며 애니메이션, 게임 그래픽, 데이터 시각화, 사진 조작 및 실시간 비디오 처리에 사용된다. Canvas API를 직접 사용해 그래픽을 그리거나 WebGL의 출력으로써 사용될 수도 있다. 이를 통해 생성된 그래픽은 DOM으로 구성되어 있지 않고 이미지 그 자체로 존재하기 때문에 이미지 파일로 저장하는 등의 동작을 쉽게 구현할 수 있다는 장점이 있다.

2.4 기존 이미지 편집에 관련된 솔루션 연구

어도비 포토샵은 2014년부터 머신러닝을 적용한 기능을 추가하였으며 현재는 이미지 합성과 영상 편집을

비롯한 다양한 분야에서 머신러닝을 적용한 소프트웨어들이 출시되고 있다. 이미지 합성 분야에서 어도비 포토샵은 AI기능을 사용한 여러 소프트웨어의 기능을 대부분 포함하고 있었으나 영상 편집 분야에서 프리미어는 아직 다른 AI기반 영상편집 프로그램들과는 전혀 다른 기능을 볼 수 있다. 이는 여러 AI기반 영상편집 프로그램들이 텍스트나 음악에 맞추어 영상을 선별하여 일반인들이 쉽게 만드는 것을 목적으로 하는 것과 달리 프리미어는 범용 툴로서 전문적인 영상 편집과 이미지 보정이 목적이기 때문이다.

2.5 기존 연구의 문제점 및 해결 방안

포토샵 등의 기존 이미지 편집기들은 사용자의 컴퓨터에 용량이 큰 프로그램을 설치해야 할 뿐만 아니라 인터페이스가 복잡해 사용자가 손쉽게 사용하기 어렵다. 적은 시간을 들이고도 이미지를 쉽게 편집하려면 한 눈에 기능의 종류와 사용법을 알 수 있어야 한다. 그러기 위해서는 단순하고 명확한 사용자 인터페이스를 제공해야 한다. 또한, 인터페이스를 토대로 사용자가 쉽게 이미지를 조작할 수 있도록 해야 한다.

따라서 본 연구를 통해 HTML5 Canvas API와 Fabric.js를 이용하여 사용자가 프로그램 설치 없이 온라인 기반으로 간편하게 이미지를 편집할 수 있도록 웹 애플리케이션 서비스를 구현하고자 한다.

이미지 편집기 기능을 온라인 기반으로 간편하게 사용할 수 있도록 HTML5 Canvas API를 활용하여 웹 서비스를 제공한다. HTML5 Canvas API를 활용하면 HTML 상에서 JavaScript를 통해 그래픽을 그릴 수 있다. HTML5 Canvas는 웹을 통해 실현할 수 있는 그래픽 수준을 크게 끌어올린 반면, 이를 구현하기 위해 사용하는 API는 하위 수준에 머물러 있어서 복잡하고 어려운 편이다. 이러한 측면에서 Fabric.js 라이브러리는 동일한 그래픽 결과물을 구현함에 있어 방법이 단순하고, 객체 집합이나 사용자 인터랙션을 지원한다.

3. 연구 내용

3.1 시나리오

이미지 편집 온라인 솔루션의 개발을 목표로 한다. 사용자 요구사항은 다음과 같다. 사용자는 온라인 어플리케이션에서 편집할 이미지를 Canvas에 불러올 수 있다. 어플리케이션은 이미지 회전, 크기 변경, 자르기, 도형 추가, 텍스트 추가 등의 이미지 편집의 기능을 제공한다. 수행한 작업을 실행 취소 / 다시 실행 가능하도록 한다. 해당 어플리케이션을 통해 편리하게 이미지를 편집할 수 있다.

3.2 웹 서버

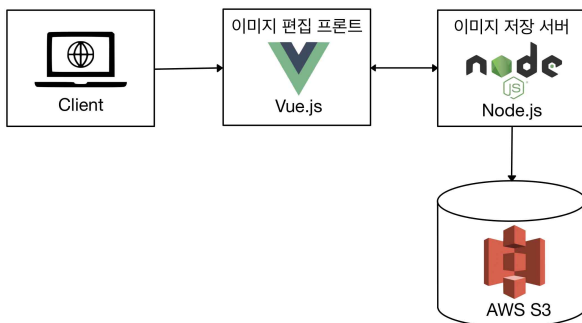
웹 서버는 이미지 편집 애플리케이션으로부터 사용자의 이미지 데이터를 받아 서버에 저장하는 역할을 담당한다. 웹 서버는 Node.js 기반으로 개발되고 있는

며, AWS S3와 연동되어 이미지를 저장하고 있다. Docker를 이용해 쉽게 배포 가능하도록 구현하고 있다.

3.3 이미지 처리

웹 프론트엔드 부분인 이미지 편집 어플리케이션은 TUI Image Editor (<https://github.com/nhn/tui.image-editor>)를 참고하여 개발하였다. 편집 어플리케이션은 프론트엔드 라이브러리인 Vue.js를 이용하여 개발되었으며, 기능상의 버그를 수정하는 작업을 주로 진행하고 있다.

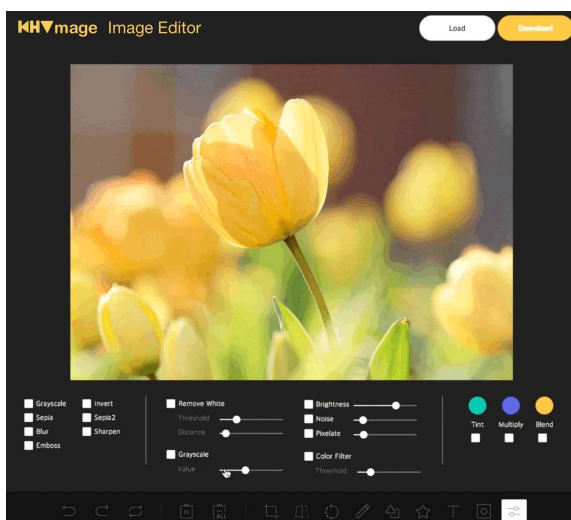
3.4 프로젝트 아키텍처



[그림 2] 프로젝트 아키텍처
[Fig. 2] Project Architecture

현재 어플리케이션, 서버로 나누어 작업중이며, 각 파트가 완성됨에 따라 프로젝트 아키텍처와 같이 통합하는 작업을 진행하고자 한다.

최종적으로 Image Editor UI와 이미지 편집 기능을 제공한다



[그림 3] 사용자 화면
[Fig. 3] User Interface

4. 결론 및 향후연구

본 연구에서 제안하는 솔루션은 사용자가 프로그램 설치 없이 온라인 기반으로 간편하게 이미지를 편집할 수 있도록 웹 어플리케이션을 통해 서비스를 구현하는 연구를 진행할 예정이다. 해당 솔루션을 통해 사용자는 편리하고 직관적인 UI를 제공함으로써 적은 시간을 들이고도 손쉽게 편리하게 이미지를 편집할 수 있을 것이다. SNS가 활성화되고 있는 요즘, 이미지 편집에 대한 수요 또한 급증하고 있기 때문에 이를 필요로 하는 사용자에게 유용한 서비스를 제시할 예정이다.

5. 참고 문헌

- [1] Zaharia, Matei, 「Spark: Cluster computing with working sets」, 『HotCloud 10.10-10』, 2010.
- [2] 박지훈, 「빅데이터 시스템의 데이터 수집 및 저장에 관한 연구」, 『2017년 추계학술발표대회 논문집 제24권 제2호』, 2017.
- [3] 인포매티카, 「빅 데이터(Big Data)의 폭발적 증가 - 빅 데이터를 큰 비즈니스 기회로 연결시키는 Informatica 9.1 플랫폼」, 2011.
- [4] 송재민, 「이미지 인식 기술의 산업 적용 동향 연구」, 『한국콘텐츠학회논문지 제20권 제7호』, 2020.
- [5] hs_seo 저, 빅데이터 - 하둡, 하이브로 시작하기.
- [6] hs_seo 저, 빅데이터 - 스칼라, 스파크로 시작하기.
- [7] HTML5 Canvas API Reference: https://developer.mozilla.org/en-US/docs/Web/API/Canvas_API
- [8] Fabric.js Reference: <http://fabricjs.com/>
- [9] 임철완, 「인공지능에 대한 사용자 관점의 이해와 이를 응용한 이미지 효과 및 영상 편집 프로그램에 대한 고찰」, 『만화애니메이션연구 통권 제60호』, 2020.
- [10] 안광민, 「하둡 기반의 효율적인 보안로그 분석시스템 설계 및 구현」, 『한국정보통신학회논문지 제19권 제8호』, 2015.