

<https://doi.org/10.7236/JIIBC.2024.24.5.183>

JIIBC 2024-5-26

# 빅데이터를 이용한 CCTV 설치와 범죄율 간의 연관성 분석

## Relationship Analysis Between CCTV Installation and Crime Rates using Big Data

김정준\*, 황승연\*\*, 장석우\*\*\*

Jeong-Joon Kim\*, Seung-Yeon Hwang\*\*, Seok-Woo Jang\*\*\*

**요약** 최근 CCTV가 사람들이 많은 거리뿐만 아니라 사람들이 자주 안 다니는 골목길에도 설치되고 있다. 이는 CCTV로 인해 검거율을 증가시키고 있기 때문이다. 특히 검거율은 범죄율과 직접적인 영향을 미치는데, 검거율이 높은 장소일수록 잠재적 범죄자들이 그 장소를 피하기 때문이다. 따라서 본 논문에서는 공공데이터를 활용하여 CCTV 설치 지역, 각 지역의 검거율과 범죄율을 수집하고 빅데이터 시스템의 Hadoop을 이용하여 수집된 데이터를 저장하고, 처리 도구를 이용하여 데이터를 정제·처리 후, R 프로그래밍을 사용하여 CCTV 설치 개수와 범죄율 간의 연관성을 분석·시각화하는 빅데이터를 접목한 연관성 분석 기술을 사용하여 CCTV의 설치 효과를 분석하였다.

**Abstract** Recently, CCTV cameras have been installed not only in crowded streets but also in less frequently traveled alleyways. This is because CCTV increases the arrest rate. The arrest rate directly affects the crime rate, as potential criminals tend to avoid areas with high arrest rates. Therefore, this paper collects public data on CCTV installation locations, arrest rates, and crime rates in various regions, stores the collected data using the Hadoop big data system, and refines and processes the data using appropriate tools. Subsequently, R programming is employed to analyze and visualize the relationship between the number of CCTV installations and crime rates. By leveraging big data correlation analysis techniques, this study evaluates the effectiveness of CCTV installations.

**Key Words** : Big Data, HDFS, Python, R

### 1. 서론

CCTV의 용도는 대표적으로 방범, 감시, 화재 예방 등 안전을 위해 설치된다. 이런 역할을 하는 CCTV는 2014년 공공데이터를 기준으로 현재까지 설치량이 꾸준히 증가하고 있다. 이에 따라 초상권 및 사생활 침해에 따른

인권침해를 주장하며 CCTV가 실제로 범죄 예방에 도움이 되지 않는다고 생각하는 의견과 CCTV 증가에 따라서 범죄가 예방되며 검거율이 높아져 CCTV를 설치해야 한다는 의견이 나누어지고 있다. 실제로 CCTV 설치 증가율과 범죄율, 검거율을 비교해 실질적으로 어떤 상관관계를 이루고 있는지를 알아보고자 한다. 연구는 공공

\*정회원, 안양대학교 소프트웨어학과

\*\*준회원, 안양대학교 컴퓨터공학과

\*\*\*정회원, 안양대학교 소프트웨어학과

접수일자 2024년 6월 11일, 수정완료 2024년 9월 2일

게재확정일자 2024년 10월 4일

Received: 11 June, 2024 / Revised: 2 September, 2024 /

Accepted: 4 October, 2024

\*\*\*Corresponding Author: swjang@anyang.ac.kr

Dept. of Software at Anyang University, Korea

데이터 포털에서 제공하는 공공데이터를 바탕으로 2014년부터 현재까지 전국 CCTV 설치량, 전국 범죄율, 전국 검거율 데이터를 빅데이터를 활용해 진행되었다. 이 연구 결과를 토대로 어린이 납치, 아동 성범죄, 어린이 보호구역 교통사고 등 어린이 보호구역에서 보호받지 못하는 상황을 방지하기 위해 어린이 보호구역 CCTV 설치 증가에 대한 당위성을 제시하려고 한다<sup>[1-2]</sup>.

## II. 관련 기술

네트워크에 연결된 CCTV 기술은 빅데이터를 기반으로 하며 이를 지능형 CCTV라고 한다. 이를 통해 영상처리, 정보 수집을 하여 사람의 얼굴, 차량번호 식별, 산불과 같은 재난들을 구별 및 식별한다. 이러한 기술들을 분석해보았을 때 빅데이터는 기존 데이터베이스 관리 도구의 능력을 넘어서는 대량(수십 테라바이트)의 정형, 심지어 비정형 형태의 데이터까지 가치를 추출하고 결과를 분석하는 기술이다<sup>[3]</sup>. 이러한 빅데이터는 다양한 종류의 대규모 데이터에 대한 생성, 수집, 분석, 표현을 그 특징으로 하는 빅데이터 기술의 발전은 다변화된 현대 사회를 더욱 정확하게 예측하여 효율적으로 작동케 하고 개인화된 현대 사회 구성원마다 맞춤형 정보를 제공, 관리, 분석 가능케 하며 과거에는 불가능했던 기술을 실현시키기도 한다<sup>[4]</sup>. 또 이러한 데이터들을 분석과 처리하기 위한 기술로는 hdfs, Python, R이 있다.

hdfs는 여러 개의 저렴한 컴퓨터를 마치 하나인 것처럼 묶어 대용량 데이터를 처리하는 기술이다. 하둡은 수 천대의 분산된 x86 장비에 대용량 파일을 저장할 수 있는 기능을 제공하는 분산파일 시스템과 저장된 파일 데이터를 분산된 서버의 CPU와 메모리 자원을 이용해 쉽고 빠르게 분석할 수 있는 컴퓨팅 플랫폼인 맵리듀스로 구성돼 있다<sup>[5]</sup>.

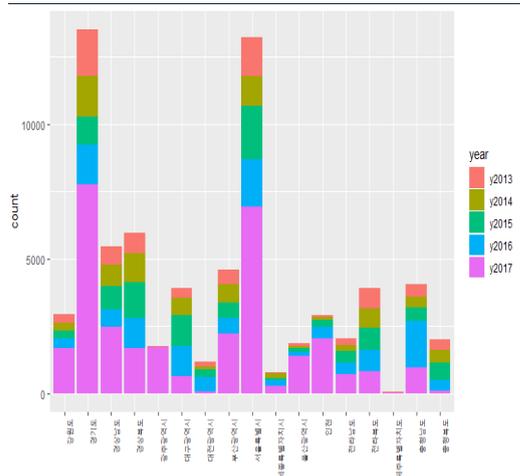
Python은 대용량 데이터 집합을 분석하는 프로그래밍 언어로 대용량 데이터를 분석하기 쉽도록 처리해준다. 다양한 라이브러리 (pandas, numpy 등)을 사용하여 데이터를 처리하고 matplotlib 라이브러리를 사용하여 데이터를 시각화하는 것도 가능하다. 첫 번째로 pandas는 Python에서 사용하는 데이터 분석 라이브러리로, 행과 열로 이루어진 데이터 객체를 만들어 다룰 수 있게 되고 좀 더 안정적으로 대용량의 데이터들을 처리하는 라이브러리입니다. 두 번째로 numpy는 행렬이나 일반적으로 대규모 다차원 배열을 쉽게 처리할 수 있고

록 지원하는 Python의 라이브러리이다<sup>[6]</sup>.

R은 통계 계산과 그래픽을 위한 프로그래밍 언어이자 소프트웨어 환경이다<sup>[7]</sup>. 뉴질랜드 오클랜드 대학의 로버트 젠들맨(Robert Gentleman)과 로스 이하카 (Ross Ihaka)에 의해 시작되어 현재는 R 코어 팀이 개발하고 있다<sup>[8]</sup>. R은 GPL 하에 배포되는 S 프로그래밍 언어의 구현으로 GNU S라고도 한다. R은 통계 소프트웨어 개발과 자료 분석에 널리 사용되고 있으며, 패키지 개발이 용이해 통계 소프트웨어 개발에 많이 쓰이고 있다<sup>[9]</sup>.

## III. 본 론

본 논문의 목적은 CCTV의 설치 비율이 범죄율과 검거율에 어떠한 영향을 끼치고 있는지를 분석하고자 한다. 연구의 목적을 달성하기 위해 다양한 공공데이터를 Python을 사용하여 데이터를 처리한 뒤 R 언어로 시각화하여 분석하였다. 전국의 CCTV 설치 비율, 범죄율, 검거율에 대한 공공데이터를 사용하였고 여기서 Python을 사용하여 공공데이터 csv 파일을 정제 및 가공하였다. 먼저 CCTV의 위치와 연도를 분석하기 쉽게 데이터를 처리하였다. 두 번째로 CCTV의 위치와 연도 별 설치에 따른 검거율, 범죄율의 영향을 분석하기 위해 어린이보호구역의 위치 데이터인 school.csv 파일을 R로 시각화하였다. 세 번째로 R 언어를 사용하여 각 데이터를 시각화하여 분석한 내용이다.



ggplot2 라이브러리에서 제공하는 ggplot 함수로 축을 만들고 geom\_bar 함수를 사용하여 그림 1과 같은 전국 각 지역의 연도별 CCTV 설치량을 막대 그래프로 시각화하였다. 전국 각 지역의 연도별 CCTV 설치량에 관한 데이터는 공공정보 포털에서 제공하는 데이터를 사용하였다. 그림 1을 통해 알 수 있는 것은 CCTV 설치량은 매년 꾸준히 증가하여 2017년에는 가장 많은 양이 설치된 것을 알 수 있다. 이것을 통해서 전국 CCTV 총량은 계속 증가하는 것을 볼 수 있다.

먼저 CCTV 총량 증가와 범죄율 감소의 상관관계가 있는지 확인하기 위하여 전국 지역별 범죄율을 시각화하였다. 전국 지역별 범죄율에 관한 데이터는 공공정보 포털사이트에서 제공하는 데이터를 사용하였다.

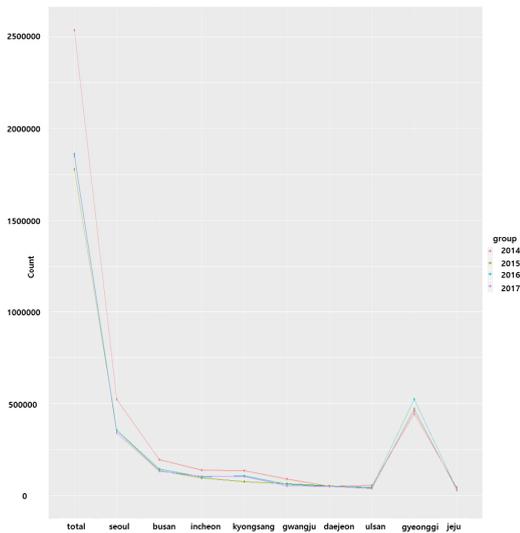


그림 2. 지역별 범죄율  
 Fig. 2. Crime Rate by Region

ggplot2 라이브러리에서 제공하는 ggplot 함수로 축을 만들고 geom\_line 함수를 사용하여 그림 2와 같이 표현하였다. 그림 2를 보면 연도별로 전국 전체의 범죄율이 소폭 감소하는 것을 볼 수 있다. 2014년에서 2015년 사이에서 유의미한 감소 폭을 보여주고 있다. 반면 2015년에서 2016년과 2017년까지의 변동은 소폭 증가와 소폭감소를 나타내고 있는 것을 볼 수 있다. 그림 2를 통해서 CCTV 총량 증가에 따라서 범죄율이 감소한다고 볼 수 있다.

다음으로는 CCTV 총량 증가에 따른 전국 연도별 검거율을 알아보기 위해 전국 범죄별 검거율을 시각화하였다. 전국 범죄별 검거율에 관한 데이터는 공공정보 포털

사이트에서 제공하는 데이터를 사용하였다. ggplot2 라이브러리에서 제공하는 ggplot 함수로 축을 만들고 geom\_line 함수를 사용하여 그림3과 같이 표현하였다. 그림 3을 보면 연도별로 전국 범죄별 검거율이 선형적으로 증가하고 있는 것을 볼 수 있다. 그림 3을 통해 CCTV 총량 증가와 검거율 증가에 비례하는 상관관계를 알 수 있다. 즉 CCTV 총량이 증가할수록 검거율이 상승하는 것을 알 수 있다.

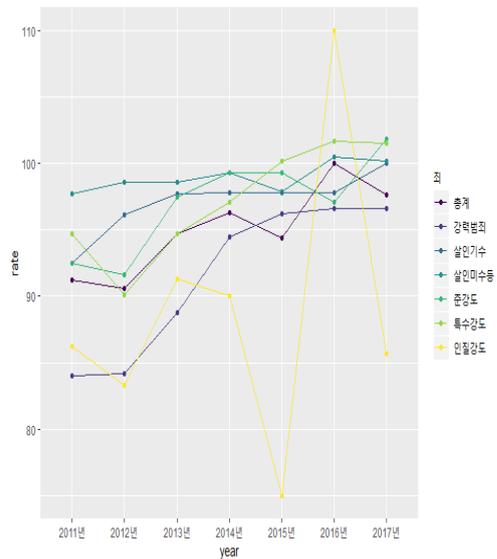


그림 3. 연도별 범죄별 검거율  
 Fig. 3. Annual Arrest Rate by Crime

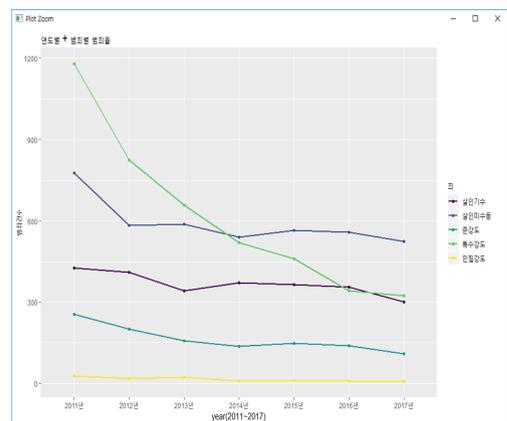


그림 4. 연도별 및 범죄별 범죄율  
 Fig. 4. Crime Rate by Year and Crime

그림 1, 그림 2, 그림 3을 보면 CCTV 총량의 증가에 검거율은 비례하고 범죄율은 반비례하는 것을 볼 수 있

다. 다시 말해서 CCTV 총량이 증가하면 검거율은 증가하고 범죄율은 감소하는 효과가 있다고 말할 수 있다. 하지만 범죄 요소마다 CCTV가 많은 효과를 발휘하여 큰 감소가 되는 것이 있고 CCTV가 설치되어도 많은 감소가 되지 않는 범죄 요소가 있다는 것을 확인하였다. 감소비율이 큰 5가지 범죄요소를 따로 R로 시각화했을 때 특히 특수강도, 준강도가 비중이 큰 것을 확인하였다.

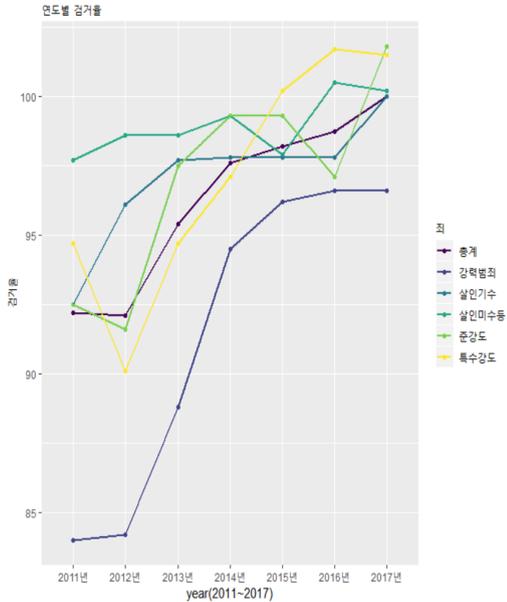


그림 5. 연도별 검거율  
Fig. 5. Annual Arrest Rate

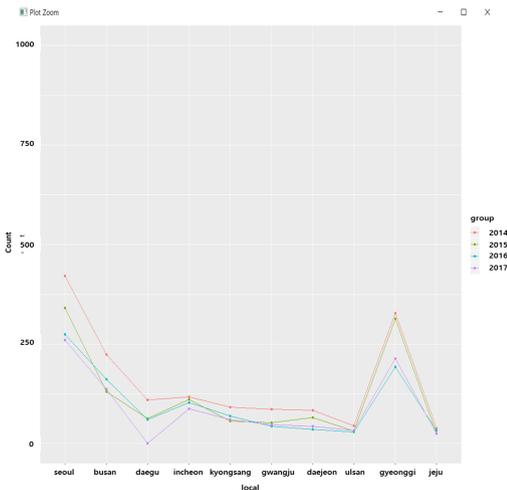


그림 6. 연도별 및 지역별 강도 범죄율  
Fig. 6. Annual and Regional Robbery Rates

검거율 또한 특수강도, 준강도가 큰 폭으로 증가하는 것을 확인하였고 이는 CCTV가 강도 범죄율에서 많은 효과가 있다는 결론을 내릴 수 있었다.

강도 범죄율만 연도별로 시각화를 하였을 때 경상도, 광주, 울산 지역이 범죄율이 지속적으로 감소하지 않았고 이 지역들은 CCTV를 추가적인 설치가 필요하다.



그림 7. 범죄 키워드 워드 클라우드  
Fig. 7. Crime keyword Word Cloud

이제 전국의 어린이 보호구역 CCTV 설치 비율과 전국 지역별 인원수 비율을 살펴볼 것이다. 먼저 그림 7의 워드 클라우드를 통해 현재 기준 범죄에 대한 가장 키워드 이슈가 무엇인지 보여준다. 아동, 성범죄자, 조두순이 대표적이다. 이에 따라 어린이 보호구역 CCTV 설치 비율에 대해 분석해 볼 것이다.

먼저 학교 어린이 보호구역 CCTV 설치량에 관한 데이터시각화하기 위해 공공정보 포털사이트에서 제공하는 데이터를 사용하였다. 그리고 plotrix 라이브러리의 pie 함수를 사용하여 지역별 어린이 보호구역 CCTV 현황을 시각화하여 그림 8과 같이 나타내었다.

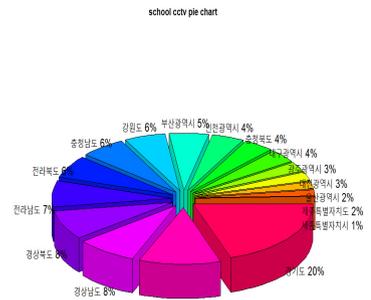


그림 8. 지역별 어린이 보호구역 CCTV 현황  
Fig. 8. Current Status of CCTV in Child Protection Zones by Region



## References

- [1] Hyungjin Lim, Julak Lee., "The Seasonal Crime Prevention Effects of Open-street CCTV", The Korean Association of Police Science Review (KAPS), Vol. 16, No. 2, pp. 99-120, 2014.
- [2] Eung-Ryul Choi, Yeonsoo Kim, "The Effectiveness of the Closed Circuit TVs (CCTVs) on the Crime Prevention", Korean Journal of Public Safety and Criminal Justice, Vol. 16, No. 1, pp. 143-186, 2007.
- [3] Joo, J. H., "The Improvement Plans of Legal Framework and Regulations of IoT Healthcare Service", Kangwon Law Review, Vol. 50, pp. 801-837, 2017.  
DOI: <http://doi.org/10.18215/kwlr.2017.50..801>
- [4] Yoon Su Jeong, Kun-Hee Han, "Service Management Scheme Using Security Identification Information Adopt to Big Data Environment", Journal of Digital Convergence, Vol. 11, No. 12, pp. 393-399, 2013.  
DOI: <http://doi.org/10.14400/JDPM.2013.11.12.393>
- [5] Kwang-Min Ahn, Jong-Yoon Lee, DongMin Yang, Bong-Hwan Lee, "Design and Implementation of a Hadoop-based Efficient Security Log Analysis System", Journal of the Korea Institute of Information and Communication Engineering, Vol. 19, No. 8, pp. 1797-1804, 2015.  
DOI: <http://doi.org/10.6109/jkiice.2015.19.8.1797>
- [6] Sang-Yong Choi, Eun-Young Cheon, Dae-Sik Ko, "Analysis of the Network Traffic Collection Performance Using Suricata", Journal of Korean Institute of Information Technology, Vol. 17, No. 8, pp. 59-66, 2019.  
DOI: <http://doi.org/10.14801/jkiit.2019.17.8.59>
- [7] Yang-Ui Lim, Kang-Chul Kim, "Methods to Propel Tourism of Yeosu City Using Big Data", The Journal of the Korea Institute of Electronic Communication Sciences, Vol. 15, No. 4, pp. 739-746, 2020.
- [8] Min-Kwan Kim, Yong Lee, Chang Hee Han, "Analysis of Consulting Research Trends Using Topic Modeling", Journal of Korean Society of Industrial and Systems Engineering, Vol. 40, No. 4, pp. 46-54, 2017.
- [9] Jong-Ki Lee., "A Case Study of R Analysis of Big Data in Seoul Special City Expenditure Settlement Data", The Journal of Business Education, Vol. 29, No. 4, pp. 57-74, 2015.  
DOI: <http://doi.org/10.34274/krabe.2015.29.4.003>

## 저 자 소 개

## 김 정 준(정회원)

• Jeong Joon Kim received his BS and MS in Computer Science at Konkuk University in 2003 and 2005, respectively. In 2010, he received his PhD in at Konkuk University. He is currently a professor at the department of Computer Science at Anyang University. His research interests include Database Systems, Big Data, Semantic Web, Geographic Information Systems (GIS) and Ubiquitous Sensor Network (USN), etc.

## 황 승 연(준회원)

• Seung-Yeon Hwang received his BS in Department of Computer Science at Korea Polytechnic University in 2019. He is currently studying MS in Department of Computer Science at Anyang University. His research interests include Database System, Big Data, Data Analysis, Machine Learning, etc.

## 장 석 우(정회원)

• Seok-Woo Jang received his BS and MS in computer science at Soongsil University in 1995 and 1997, respectively. In 2000, he received his PhD at Soongsil University. He is currently a professor in the department of software at Anyang University. His research interests include Artificial Intelligence (AI), Big Data, Video Processing and Block Chain.