

## Optimal CCTV Installation

CUAI 5기 Panopticon팀

유나현 (응용통계학과), 방수민 (전자전기공학부), 이해연 (응용통계학과), 이윤지 (응용통계학과)

**[요약]** 김해시는 다양한 사건사고 및 사회문제로부터 시민의 안전을 확보하기 위해 매년 CCTV 설치 사업을 진행하고 있다. 이를 위하여, 112 신고 접수, 인구 관련 데이터, 시설물 데이터 등 다양한 데이터를 활용하여 시민안전 사각지대 분석을 수행하고, 시민안전을 최대한 확보할 수 있도록 최적의 CCTV 설치 위치 50개소를 제시했다.

### 1. 서 론

CCTV(Closed circuit Television)는 유동인구가 많은 지역, 범죄 우려가 있는 지역, 건물 내부 등에 방범, 감시, 화재 예방 등의 안전을 유지하기 위해 주로 설치한다. 방범용 CCTV는 범죄의 발생을 억제할 뿐만 아니라 범죄자의 검거율까지 높일 수 있다. 경찰청과 건축도시공간연구소가 공동으로 진행한 ‘범죄 예방 환경조성(CPTED·셉테드) 시설 기법 효과성 분석 연구’ 결과에 따르면 CCTV를 설치한 범위 내 인구 5대 범죄 야간 발생 건수를 분석한 결과 2016년에 41.64건에서 2018년 36.95건으로 약 11% 감소한 것으로 나타나 CCTV가 설치된 곳의 야간 범죄 발생률이 유의미하게 감소된 것을 알 수 있다. [0]

방범용 CCTV 설치 후 시민들의 안전도가 증가했다는 선행 연구도 있다. 스페인 말라가 지역의 40개 소 방범용 CCTV를 선정하여 범죄통계자료분석과 담당자 인터뷰 실시 결과, 야간에 시민들이 느끼는 안전도가 상승했다는 것을 보였다. 또한, 호주의 시드니, 멜버른, 페어필드 3개의 지역에서 CCTV 설치 후 시민들의 안전도가 61%에서 85%까지 증가하였다. [1]

안전한 도시와 시민의 안전을 위해 CCTV의 적절한 설치 위치 선정은 매우 중요하다. 그러나 설치 위치에 대한 체계적인 고려 없이 무분별하게 설치되고 있는 CCTV는 지자체의 운영·관리비용을 증가시키는

요인이 될 수 있고 관리운영상에 있어서도 다양한 문제를 야기시킬 수 있으므로 효율적으로 설치 되어야 한다. [2]

본 연구에서는 김해시 시민안전 사각지대 해소를 위한 CCTV 최적의 설치위치를 선정하고자 한다. 총 14개의 데이터를 가공, 분석하여 CCTV 설치 위치에 영향을 미치는 변수를 확인한 후, 머신 러닝 모델로 LGBM, Random Forest, Cat Boost, XGBoost를 사용하여 가장 좋은 수치를 얻는 모델을 최종적으로 선택한다. 모델의 평가 기준은 MAE로 예측 값과 실제 값의 차이에 대한 절대값에 대해 평균을 낸 값을 사용한다. 최종적으로 본 연구에서는 최적의 CCTV 설치 위치 50개소를 도출하고, 기대효과와 한계점을 제시하였다.

### 2. 본 론

#### (1) EDA - 데이터

본 연구 진행에 앞서 COMPAS에서 제공되는 데이터 들 중 상대적으로 CCTV 설치에 관련성이 높은 112신고이력, 치안 유관업종 현황, 유동인구 데이터에 관련해 유의미한 특이점 발견을 위해 EDA를 진행 한 결과 다음과 같다.

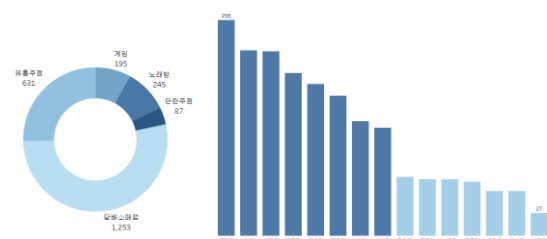


그림 1. Tableau를 이용한 치안 유관업종 데이터 시각화

치안 유관업종 데이터를 시각화 한 결과 담배 소매업이 1253개로 과반수 이상을 차지하고, 읍면동

별 치안 유관업소 개수가 많은 지역 상위 15개를 시각화 한 결과, 진영읍이 256개로 가장 많고, 이어서 외동, 삼계동 대청동에 많이 분포하는 것을 확인할 수 있었다.



그림 2. Tableau를 이용한 112 신고이력 데이터 워드 클라우드

112 신고이력 데이터를 워드 클라우드로 표현한 결과 위험방지, 교통사고, 기타 형사범 관련 신고가 많이 들어오는 것을 볼 수 있다.

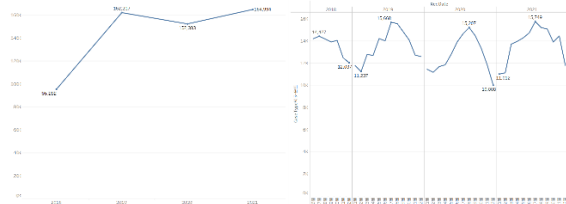


그림 3. Tableau를 이용한 112 신고 횟수 연도별 시각화

2018년부터 2021년까지 112 신고 횟수가 전반적으로 상승하고 있고, 각 연도별로 시각화 한 결과 7,8월인 여름철에 특히 신고이력이 높은 것 확인할 수 있다.

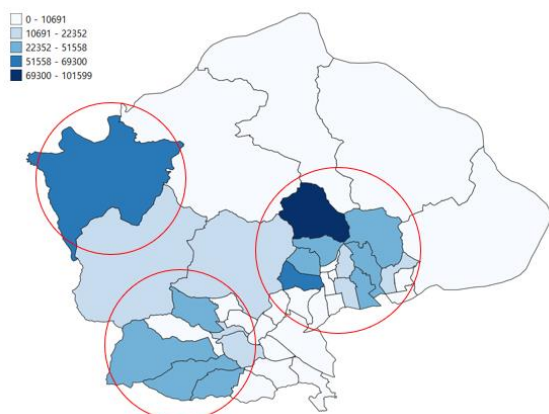


그림 4. QGIS을 이용한 유동인구 데이터 시각화

유동인구 데이터를 같은 간격의 다섯가지 범주로 시각화 한 결과, 크게 세부부분에서 유동인구가 많은

것을 볼 수 있다.

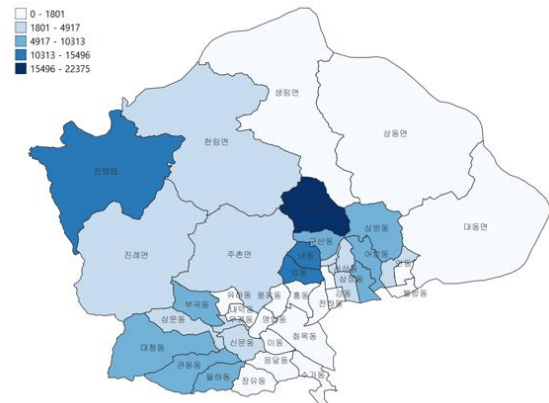


그림 5. QGIS을 이용한 여름철 유동인구 데이터 시각화

앞서 여름철에 112 신고이력이 높아, 유동인구 데이터의 6,7,8월 행을 추출하여 전체 유동인구와의 유의미한 차이를 살펴보았을 때 한림면, 내동, 삼문동, 대성동, 서상동 등에서 유동인구가 증가한 것을 확인할 수 있었다.

## (2) EDA - 중심점

본 프로젝트는 김해시를 100\*100 격자로 나누고, 이를 기준으로 데이터 프레임을 구성하고자 했다. 최종적으로, 최적의 CCTV 설치 위치는 위도, 경도로 표현해야 한다. 이를 위해, 해당 격자에 해당하는 위도와 경도를 정하는 기준을 선정했다.

결정한 기준은 격자의 각 꼭짓점 만을 추출한다. 그리고 각 꼭짓점의 중심점을 생성했고, 해당 격자에 CCTV 설치가 적합하다면, 이 중심점에 해당하는 위도, 경도 좌표를 최종 위치로 선정하기로 했다.



그림5. 중심점 도출 과정

## (3) EDA - 데이터 취합

각각의 데이터들을 100\*100 격자 데이터에 병합하고자 했다. 데이터를 병합할 때, 구분되는 데이터 프레임 유형은 총 2가지이다.

첫 번째로는 최종적으로 원하는 바와 같이 격자별로 구성된 데이터들이다. 이 데이터들은 groupby와 drop기능을 이용하여 전처리를 진행했다.

두 번째로는 격자별로 구성되지 않고, 위도, 경도로

구성되어 있는 데이터들이다. QGIS를 이용하여 격자 별로 구성되도록 변환하였다.

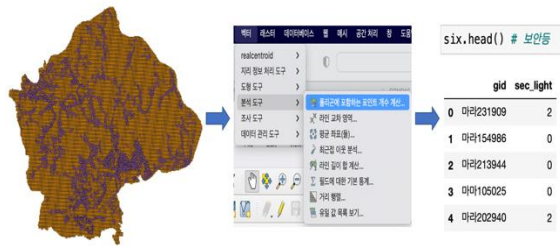


그림6. QGIS를 이용한 격자 별 데이터 변환

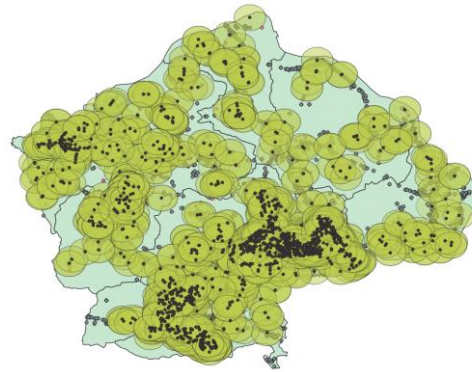


그림 8. QGIS를 이용한 CCTV 1km 버퍼 내 시설들 개수

#### (4) 변수간 상관분석

기존의 CCTV 설치 위치의 특성을 파악하기 위해 CCTV의 특정 반경 내 시설 수에 대한 상관분석을 진행하였다. 상관분석을 진행한 변수의 데이터는 김해시 CCTV 설치현황, 김해시 주차장 현황, 김해시 112 신고이력 데이터, 김해시 보안등 설치 현황, 김해시 어린이집 위치 정보, 김해시 유치원 위치 정보, 김해시 학교 위치 정보, 김해시 치안 유관업종 위치 정보, 김해시 아동안전지킴이집 위치정보이다.

112 신고이력 데이터를 제외한 위의 모든 변수 데이터에는 lon, lat 으로 표현된 위도 경도 정보에 대한 열이 포함되어 있어서 QGIS에서 이를 이용해 지도에 아래와 같이 표시 할 수 있다.

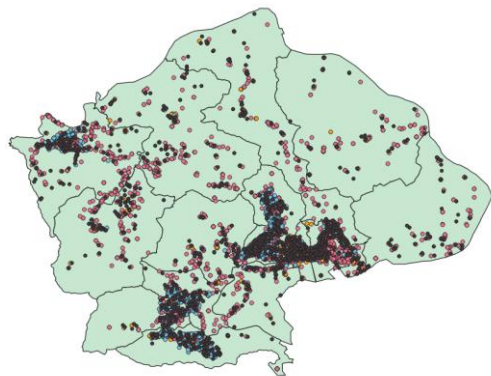


그림 7. QGIS를 이용한 112 신고이력 데이터 시각화

CCTV 설치 현황 데이터로부터 지도에 찍은 모든 점들에 대한 버퍼를 형성해서 반경 내 시설 수를 파악할 수 있도록 한다.

우선 QGIS를 이용하여 CCTV 1km 버퍼를 생성한 후 각 버퍼 내 시설들의 개수를 정리한다.

(예: 보안등)

_cctv_nm		인근 1km 내 보안등 개수
0	(동상동30)1_동상1092-31	861
1	(동상동31)1_동상1015-4	755
2	(동상동31)2_동상1015-4	755
3	(동상동31)3_동상1015-4	103
4	(동상동31)4_동상1015-4	801
...	...	...
3247	희망공원3_삼계1459-5↓	313
3248	희망공원4_삼계1459-5←	78
3249	희망공원5_삼계1459-5△	70
3250	힐하우스1_외동1244-3↑	436
3251	힐하우스2_외동1244-3↓	230

3252 rows × 2 columns

그림 9. CCTV 1km 내 보안등 개수

다른 변수들에 대해서도 동일하게 진행한 후 도출된 결과 표들을 \_cctv\_nm 열 기준으로 병합해 김해시의 CCTV 별 1km 반경 내 관련 시설 수 상관분석을 진행한다.

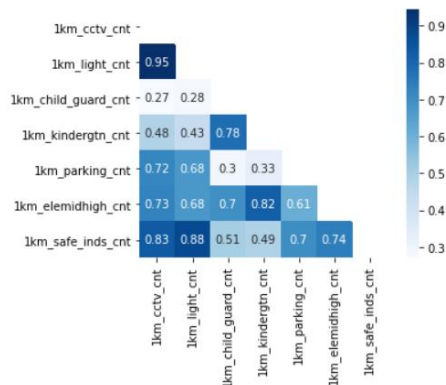


그림 10. CCTV 1km 반경 내 관련 시설 수 상관분석

상관분석 결과, CCTV 반경 1km 내 CCTV 개수 변수와 상관관계가 높은 순서대로 보안등(0.95), 치안 유관업소(0.83), 초중고 수(0.73), 주차장수(0.72), 유치원수(0.48), 아동지킴이집(0.27) 이다.

하지만 초기에 버퍼를 1km로 설정한 것은 CCTV 가 1km 반경보다 훨씬 적은 거리의 범위 까지만 커버할 수 있다는 점을 고려해서 버퍼를 500m 로 수정해 동일한 작업을 수행하였다. 또한 유관 시설 뿐 만 아니라 112 신고내역 데이터도 CCTV 설치와 상관이 있을 것으로 예상되어 112 신고내역 데이터를 추가하였다. 112 신고내역 데이터에는 다른 변수 데이터들과 달리, 위도 경도 데이터가 포함되지 않아 gid 데이터의 central point를 추출해 지도에 표시할 수 있도록 하였다.

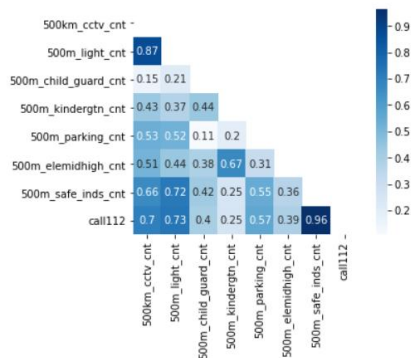


그림 11. CCTV 500m 반경 내 관련 시설 수 상관분석

버퍼를 500m로 수정하고 112 신고내역 데이터를 추가하여 상관 분석을 시행한 결과, CCTV 반경 500m 내 CCTV 개수 변수와 상관관계가 높은 순서대로 보안등(0.87), 112(0.7), 치안 유관업소(0.66), 초중고 수(0.51), 유치원수(0.43), 아동지킴이집(0.15) 이다.

이로써 기존에 설치된 CCTV 들의 위치는 보안등

설치 위치와 상관관계가 매우 크며, 치안유관 업소 또한 CCTV 설치 위치와 상관관계가 크다고 할 수 있다. 반면, 상관이 클 것으로 예상했던 유치원수가 아동지킴이 집은 CCTV 설치 위치와는 큰 상관관계가 없는 것으로 보인다.

## (5) 모델 선정

CCTV 설치 위치 선정을 위한 모델 선정 과정을 진행했다. 모델 후보는 총 4가지이며, 각각 LightGBM[3], Random Forest[4], CatBoost[5], XGBoost[6]이다. MAE을 성능 비교 척도로 선정하였고, 그림 12에서 확인이 가능하다. 모델 별 성능 비교는 그림 13에서 확인이 가능하다.

$$MAE = \frac{1}{n} \sum |\hat{y} - y|$$

그림12. MAE (Mean Absolute Error)

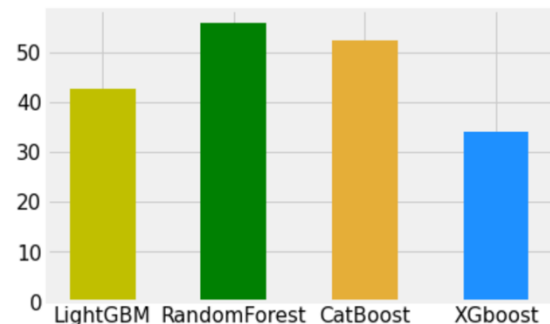


그림13. 모델 별 성능비교

## (6) CCTV 설치위치 선정 과정

모델을 통해 나온 위치가 실제로도 설치가 가능한지 여부를 판단하고 했다. 설치가 불가능한 개인 사유지, 하천을 제외한 군유지, 국유지, 시/도 지역들만을 설치 가능 지역으로 선정하였다.





그림14. CCTV 설치 가능 지역

시각화 결과 반영 요소로 판단한 치안 유관 업소 대비 CCTV가 부족한 지역인 외동, 삼계동, 어방동, 부원동, 대청동과 여름 유동인구 증가 지역인 한림면, 내동, 삼문동, 대성동, 서상동을 우선 설치 지역으로 선정하고, CCTV 설치 위치 선정에 반영하였다.

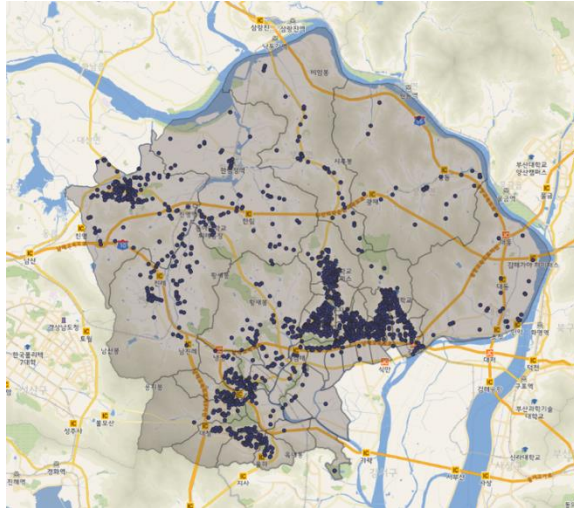


그림15. 치안 유관업소 대비 CCTV 부족 지역

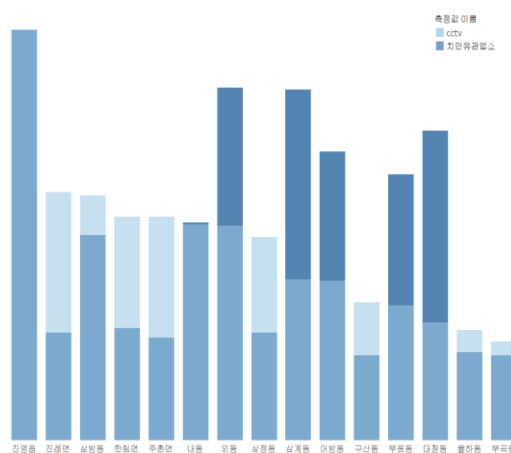


그림16. 여름 유동인구 증가 지역

### (7) CCTV 위치 좌표 선정 결과

모델을 통해 나온 결과와 설치 가능 지역, 그리고 시각화 결과를 반영한 최종 결과는 다음과 같다. 시각화한 자료는 그림 17을 통해, 최종적으로 도출한 50곳의 좌표들은 그림 18을 통해 확인이 가능하다.



그림17. CCTV 위치 좌표 시각화

gid	순번	x좌표(경도)	y좌표(위도)
마라258935	1.0	128.882980	35.229451
마라243939	2.0	128.866661	35.233244
마라247971	3.0	128.871441	35.262040
마라243938	4.0	128.866645	35.232342
마라243943	5.0	128.866621	35.236850
마라184895	6.0	128.801104	35.194293
마라247972	7.0	128.871456	35.262942
마라246970	8.0	128.870327	35.261151
마라241926	9.0	128.864167	35.221550
마라183894	10.0	128.799991	35.193404
마라276943	11.0	128.902879	35.236435
마라259933	12.0	128.884048	35.227635
마라185893	13.0	128.802173	35.192479
마라247970	14.0	128.871426	35.261139
마라118020	15.0	128.730324	35.307738
마라118019	16.0	128.730310	35.306836
마라242940	17.0	128.865477	35.234158
마라243940	18.0	128.866576	35.234145
마라243941	19.0	128.866591	35.235047
마라242938	20.0	128.865447	35.232355
마라275943	21.0	128.901780	35.236447
마라278953	22.0	128.905232	35.245424
마라277943	23.0	128.903978	35.236422
마라258937	24.0	128.883010	35.231254
마라258936	25.0	128.882995	35.230352
마라256934	26.0	128.880767	35.228574
마라276937	27.0	128.892898	35.231140
마라117020	28.0	128.729224	35.307749
마라245971	29.0	128.869243	35.262065
마라185895	30.0	128.802202	35.194281

gid	순번	x좌표(경도)	y좌표(위도)
마라185896	31.0	128.802216	35.195183
마라247968	32.0	128.871395	35.259336
마라259934	33.0	128.884063	35.228537
마라244942	34.0	128.867705	35.235936
마라282951	35.0	128.909596	35.243570
마라129014	36.0	128.742338	35.302205
마라246971	37.0	128.870342	35.262053
마라257933	38.0	128.881850	35.227660
마라256938	39.0	128.880828	35.232180
마라277954	40.0	128.904149	35.246338
마라258934	41.0	128.882964	35.228549
마라190905	42.0	128.807838	35.203237
마라281953	43.0	128.908528	35.245385
마라186894	44.0	128.803286	35.193368
마라244927	45.0	128.867478	35.222414
마라245970	46.0	128.869228	35.261164
마라184893	47.0	128.801075	35.192490
마라281951	48.0	128.908497	35.243582
마라277942	49.0	128.903962	35.235521
마라189876	50.0	128.806321	35.177106

그림18. 최종 선정 지역 50곳

## 3. 결 론

본 연구에서는 CCTV 설치현황, 치안 유관업종, 유동인구 및 보안등 설치 등 다양한 데이터들과 112 신고 데이터 간의 EDA와 상관분석 등을 통해 CCTV 설치우선 지역 50곳을 제안하였다. 본 연구를 진행하기 위해 앞서 CCTV 설치와 112신고 사이의 유의미한 관계가 있다고 예측하여 실제로 CCTV 500m 반경 내 관련 시설 수 상관분석을 한 결과, 필터링한 112신고 이력 데이터와 0.7이라는 높은 양의 상관관계를 보임을 알 수 있었다. 위치를 나타낼 때는 100\*100 격자 gid를 기준으로

사용하여 데이터들을 정리하여 나타내고 이후 격자에 해당하는 위도, 경도를 파악하였다. 예측모델의 train set은 CCTV가 없는 격자가 많은 것을 고려하여 CCTV 개수가 1개 이상인 데이터로만 학습을 진행하였다. MAE 지표를 통해 LGBM, Random Forest, Cat Boost, XGBoost로 예측모델을 평가하였고 그 결과, XGBoost가 34.13으로 가장 좋은 수치를 얻어 최종 모델로 선정하였다. 모델을 통해 선정한 위치에 지역적 안배문제 등(여름철 유동인구, 토지소유정보, 하천지역)을 고려하여 실제로 CCTV 설치가 가능한지 여부를 판단하기 위해 상위 100곳을 추출한 뒤 최종적으로 50곳을 정하였다. CCTV와 112 신고 사이의 높은 양의 상관관계를 보인 것에 따라 최종적으로 CCTV 설치 장소를 예측했을 때도 112신고가 다수 발생했던 지역 위주로 할당되었다. 총 14개의 데이터를 취합하여 CCTV 설치 고려 요소를 정하였으며, CCTV 입지선정과 상관계수가 높아 관련성이 높은 데이터들을 활용하여 데이터 분석을 이루었기 때문에 정확도가 좋다고 할 수 있다. 본 연구를 진행하면서 사용한 모든 패키지는 오픈소스였기 때문에 활용성도 높다고 생각한다. 또한, 실제로 설치가 가능한 지역들로만 위치를 선정했기 때문에 현실적으로 적용이 가능하다는 장점이 있다. 데이터가 이후에 추가되거나 변경사항이 있는 경우 이에 해당하는 데이터만 수정하면 기존 모델을 그대로 사용하는 것이 가능하기 때문에 경제성도 좋다고 할 수 있다.

본 연구에서는 CCTV 커버리지 각도를 360°로 설정해 모든 구역이 감시 가능하다고 가정하여 분석하였다. 그러나 360°를 모두 감지하는 것은 현실적으로 어렵고, 특정 각도에 건물이나 장애물이 위치한 부분은 제한적으로 감시가 가능하다는 한계점이 있다. 또한, 위도 경도 좌표를 파악하기 위해 격자들을 이차원으로 구성된 점으로 가정하여 연구를 진행하였으나 높이와 같이 더 다양한 요소를 추가하여 실제 가시권 범위를 고려한다면 보다 더 정확한 결과를 도출할 수 있을 것으로 기대된다.

## 참고 문헌

[0] 도내 5대 강력범죄 발생률, 김해가 가장 높아 (<http://www.knnews.co.kr/news/articleView.php?idxno=1291826>)

[1] 방법용 CCTV의 계절별 범죄예방효과

[2] 도시민의 일상활동 패턴에 따른

방법용 CCTV의 입지적 개선방안에 관한 연구

[3] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree", Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017

[4] Andy Liaw and Matthew Wiener, "Classification and Regression by randomForest", R news, 2002

[5] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, Andrey Gulin, "CatBoost: unbiased boosting with categorical features", arXiv:1706.09516v5, 2019

[6] Tianqi Chen, and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System", Proceedings of the 22<sup>nd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016