

효과적인 연쇄 범죄 수사 지원을 위한
시공간 패턴 및 분석 기법[†]
Spatio-Temporal Patterns and Analysis Methods
for Supporting the Efficient Investigation
on Serial Crimes

홍동숙^{1*}, 서종수², 한기준³
Dong-Suk Hong^{1*}, Jong-Soo Seo², Ki-Joon Han³

¹ 연세대학교 TMS 정보기술사업단

² 연세대학교 전기전자공학부

³ 건국대학교 컴퓨터·정보통신공학과

¹ dongsukh@yonsei.ac.kr, ² jsseo@yonsei.ac.kr, ³ kjhan@db.konkuk.ac.kr

요약

연쇄 살인과 같은 강력 범죄의 심각성이 사회적 이슈가 되면서 이에 대한 효과적인 과학 수사의 필요성이 증가되고 있다. 특히, 연쇄 범죄 데이터에 대한 공간 분석을 통해 범죄자의 거점 위치를 예측하는 지리적 프로파일링과 미래에 발생할 범행 장소의 위치, 즉 기존 범행에 이어 일어날 다음 범행 위치 예측에 관한 연구가 활발하다. 그러나, 이와 관련된 기존 연구는 물리적인 거리에 대한 통계적 기법을 적용하거나 단순한 공간적 분석만을 적용하므로 낮은 예측 정확도를 보이는 문제점이 있다.

본 논문에서는 이러한 문제를 해결하고 보다 효과적인 연쇄 범죄 수사를 지원하는 방법으로 연쇄 범죄 발생에 대한 공간적·시간적 분포 특성에 따른 시공간 패턴을 기반으로 다양한 시공간 분석을 적용하는 거점 위치 예측 기법과 다음 범행 위치 예측 기법을 제안한다. 제안 기법은 중심축을 따라 나타나는 선형 분포의 연쇄 범죄에서도 정확도 높은 예측이 가능하고, 다수의 서로 다른 군집들에 대해 각 군집내 범행에 대한 지역적 예측과 대상 영역의 모든 범행에 대한 전역적 예측이 가능하다. 또한, 방향 패턴을 활용하여 다음 범행 위치 예측 정확도도 개선하였다.

1. 서론

최근 고급 GIS 및 복잡한 공간 분석 기술이 발전함에 따라 다양한 의사 결정 지원 시스템에서 보다 효율적인 지식 검색과 고급 지식의 활용을 위해 공간적 문제 해결을 위한 획기적인 기술을 필요로 하

게 되었다. 또한, 법집행 기관 및 수사 기관 등을 중심으로 효과적인 수사 및 향후 범죄 예방을 위해 과학 수사에 관한 연구의 필요성이 증대되고 있다[1,2].

특히, 범죄 발생 장소, 용의자의 거주지 등 공간적 특성을 포함하는 범죄 데이터

[†] 본 연구는 서울시 산학연 협력사업의 신기술 연구개발 지원사업의 지원으로 수행되었음.

를 분석함으로써 수사 대상 지역과 용의자를 우선순위화하기 위한 범죄 정보 분석 시스템이 활발히 연구 및 개발되고 있다[3,4,5,6].

연쇄 범죄와 관련하여 효과적인 범죄자 검거를 위해서는 범죄자의 거점 위치 예측과 다음 범행 위치 예측이 요구된다. 거점(Base)은 범죄자의 범행 장소 선택에 영향을 미칠 수 있는 거주지나 회사 등 범죄자의 근거지를 뜻한다. Criminal Geographic Targeting(CGT)[3], Canter Model(CM)[4], Journey-to-Crime(JTC)[5,6] 등 기존 연구에서 연쇄 범죄 범죄자의 거점 위치를 예측하는 방법을 제안하고 있으나 물리적인 거리에 대한 단순한 통계적 기법을 적용하며 연쇄 범죄 발생에 대한 시간 속성을 활용하지 않으므로 예측 정확도가 낮은 문제를 보인다.

Correlated Walk Analysis(CWA)[6], Probability Grid Method(PGM)[7]에서 연쇄 범죄의 다음 범행 위치를 예측하는 방법을 제안하고 있으나 거리 및 방향에 대한 단순한 통계적 기법을 적용하거나 단순한 공간적 분석을 활용하므로 예측 정확도가 낮은 문제를 보인다.

따라서 본 논문에서는 이와 같은 문제점을 해결하고 효과적인 범죄 수사에 적용 가능한 정확도를 개선한 거점 위치 예측 및 다음 범행 위치 예측 기법을 제안한다. 제안 기법은 다양한 시공간 패턴을 고려하며 각 패턴에 따라 차별화된 특성을 가진다. 본 논문에서는 중심축을 따라 나타나는 선형 분포의 패턴을 거점을 중심으로 원형 분포를 보이는 패턴과 구분하여 다룬다. 또한, 범행 위치들의 다양한 공간적 분포 특성을 고려하여 다수 군집을 대상으로 하는 경우 군집 단위의 지역적 예측 기법과 모든 범행에 대한 전역적 예측 기법을 정의한다. 특히, 정확도 높은 다음 범행 위치 예측을 위하여 필요한 연쇄 범죄 데이터의 방향 패턴을 기술하고

이를 활용하는 예측 기법을 제시한다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 관련 연구에 대해 분석한다. 제 3 장에서는 연쇄 범죄 시공간 패턴에 대해 기술한다. 제 4 장에서는 시공간 분석을 기반으로 하는 거점 위치 예측과 다음 범행 위치 예측 기법을 제안한다. 제 5 장에서는 실험을 통해 제안 기법의 우수성을 보여준다. 마지막으로 제 6 장에서는 결론에 대해 언급한다.

2. 관련 연구

2.1 지리적 프로파일링

범죄 데이터의 공간적 패턴을 분석하여 거점 위치를 예측하는 지리적 프로파일링 알고리즘은 공간 분포 방법과 확률 거리 방법으로 분류된다.

공간 분포 방법은 거점으로써 모든 범죄 발생 위치들에 대한 평균 위치인 중앙점(centroid)을 찾는 방법, 모든 범죄 발생 위치들 중 중간점(median)을 찾는 방법, 모든 범죄 발생 위치들과의 거리합이 가장 작은 위치 CMD(Center Of Minimum Distance)를 찾는 방법 등이 있다. 중앙점은 범행 위치들의 산술적 평균인 위치를 의미하고, 중간점은 범행 위치들의 중간 크기에 해당하는 위치이다. CMD란 모든 범행 위치들까지의 유클리드 거리의 총합이 가장 작은 위치이다[5,6].

확률 거리 방법은 거점으로부터의 거리와 범죄 발생 수에 대한 다양한 수식을 이용하여 대상 영역내 모든 위치에 대해 거점이 될 확률값을 구하는 것으로 CGT [3], CM[4], JTC[5,6] 등이 있다.

CGT는 임의의 셀과 범죄 위치간의 맨하탄 거리를 기본으로 거리 감퇴 함수를 정의하고 있다. 주어진 버퍼 존의 크기에 따라 버퍼 존 내부에 위치한 셀과 버퍼 존 외부에 위치한 셀을 구분하고, 정의된 함수에 따라 버퍼 존 내부에 위치한 셀의

경우 범죄 위치와 멀리 있는 셀일수록 큰 확률값을 가지며, 버퍼 존 외부에 위치한 셀의 경우 범죄 위치와 멀리 있는 셀일수록 작은 확률값을 가진다.

CM은 임의의 셀과 범죄 위치간의 유클리디언 거리를 기본으로 거리 감퇴 함수를 정의하고 있다. 정의된 함수에 따라 셀과 범죄 위치간 거리에 대한 음의 지수 함수에 의하여 범죄 위치에서 멀리 있는 셀일수록 작은 확률값을 가진다.

JTC는 사용자에게 의해 선택될 수 있는 몇 가지 거리 감퇴 함수를 정의하고 있다. 대표적으로 절단 음의 지수 함수는 어떤 셀이 버퍼 존의 내부에 있을 때는 1차 선형 함수에 의해 해당 셀이 거점일 확률값을 계산하고, 버퍼 존의 외부에 있을 때는 음의 지수 함수에 의해 해당 셀이 거점일 확률값을 계산한다.

이러한 기존 방법들은 실제 거점이 다수인 경우와 특정한 방향성을 갖는 연쇄 범죄의 경우 예측에 한계를 보인다.

2.2 다음 범행 위치 예측 기법

연쇄 범죄로서 다음에 발생할 수 있는 범행 위치를 예측하기 위한 알고리즘으로 CWA[6]과 PGM[7]이 있다.

CWA는 다음 범행의 위치 및 시간을 예측하기 위해 연쇄 범죄의 연속적 이동을 분석하는 기법을 제시한다. 범죄자에 의해 수행된 사건, 즉 범행들의 공간적 분포와 시간적 순서를 기반으로 다음 범행의 위치와 시간을 예측하는 것이다. 이 기법은 기본적으로 어떤 사람이 특정한 방향으로 이동한다면, 그 사람은 반대 방향 혹은 직각 방향으로 바꾸어 이동하기보다는 오히려 계속해서 원래의 방향을 유지하며 이동하기 쉽다는 이론을 근거로 하고 있다.

CWA에서는 시간, 거리, 방향에 대한 반복성을 분석하고, 각각에 대하여 반복성이 가장 높은 연속 범행들을 대상으로 평균 차이, 중간 차이 혹은 회귀식을 이용하

여 다음 위치 및 시간을 예측한다.

PGM에서는 임의 확률의 범행을 포함하는 확률 사각형 혹은 표준 편차 타원 내부, 모든 범행 위치를 포함하는 볼록 다각형 내부 혹은 마지막 범행 위치로부터 범행들의 평균 거리차만큼의 크기를 갖는 버퍼 내부 영역을 연쇄 범죄의 다음 범행 위치로써 제안한다.

CWA와 PGM는 범행들간의 평균 거리를 활용하므로 평균에서 벗어나는 범행이 많은 경우 예측 정확도가 낮다는 문제점을 내포한다. CWA에서 거리 및 방향에 대한 복잡한 패턴의 분석이 어렵고 PGM에서 기존 범행 위치들과 거리 차가 큰 다음 범행 위치 예측이 어렵다.

3. 연쇄 범죄 시공간 패턴

본 장에서는 범행 위치의 공간적 분포와 범죄 발생의 시간적 분포 특성에 따른 연쇄 범죄 데이터의 시공간 패턴에 대해 정의한다.

3.1 등방성 패턴

등방성 패턴이란 특정한 한점 혹은 점들을 중심으로 모든 방향에 유사한 선호를 보이는, 즉 원형 분포를 보이는 연쇄 범죄의 시공간 패턴을 뜻한다. 그림 1은 등방성 패턴 연쇄 범죄의 한 예로써 한점을 중심으로 8방위로 동일한 분포를 보이는 8개 범죄에 대한 발생 장소와 중심점을 보여준다.

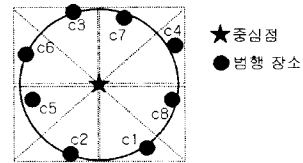


그림 1. 등방성 패턴의 예

등방성 패턴의 연쇄 범죄는 원형으로 분포된 범죄 발생 위치들로부터 중심점 식별이 가능하다. 중심점이란 등방성 패턴인 연쇄 범죄의 경우 모든 범행 위치까지

의 이동을 위한 공간적 중심인 한 점이다.

다수 범행 위치들을 포함하는 하나의 그룹을 군집(Cluster)이라 한다. 전체 범행 위치들이 등방성 패턴을 보일 때 모든 범행 위치들이 하나의 군집으로 표현되지만 단지 일부 범행 위치들이 등방성 패턴을 보일 때는 해당하는 일부 범행 위치들도 하나의 군집으로 표현될 수 있다.

군집이 1개인 등방성 패턴은 단일 중심점 패턴이라 하고 군집이 2개 이상인 등방성 패턴은 다중 중심점 패턴이라 한다. 모든 군집은 활성 군집이거나 활성 군집이 아닌 비활성 군집이다.

[정의 1] 활성 군집 Ca

- 1) 모든 범행 위치들을 포함하는 1개 군집 Ca
 - 2) 군집 Ca에 소속된 범행 위치들 중 최소인 범행번호가 다른 군집 Cb에 소속된 범행 위치들 중 최대인 범행번호보다 큰 군집 Ca
 - 3) 군집 Ca에 소속된 범행 위치들 중 최소인 범행번호가 다른 군집 Cb에 소속된 범행 위치들 중 최대인 범행번호보다 작고, Ca에 소속된 범행 위치들 중 최대인 범행번호가 다른 군집 Cb에 소속된 범행 위치들 중 최소인 범행번호보다 큰 군집 Ca
- (범행번호: 범행 발생 순서를 의미하는 번호)

3.2 비등방성 패턴

비등방성 패턴이란 특정한 방향 혹은 방향들에 대한 선호를 보이는, 즉 선형 분포를 보이는 연쇄 범죄의 시공간 패턴을 뜻한다.

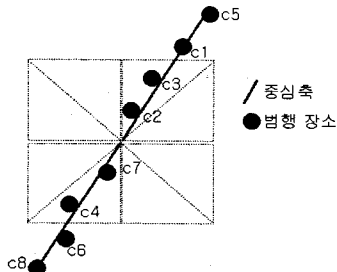


그림 2. 비등방성 패턴의 예

그림 2는 비등방성 패턴 연쇄 범죄의 한 예로써 북북동-남남서 방향으로 분포

된 8개 범죄 발생 장소와 중심축을 보여 준다.

비등방성 패턴의 연쇄 범죄는 선형으로 분포된 범죄 발생 위치들로부터 중심축 식별이 가능하다. 중심축이란 비등방성 패턴인 연쇄 범죄의 경우 범행 위치까지의 이동을 위한 공간적 중심인 선이다. 군집이 1개인 비등방성 패턴은 단일 중심축 패턴이라 하고 군집이 2개 이상인 비등방성 패턴은 다중 중심축 패턴이라 한다.

4. 연쇄 범죄 거점 및 다음 범행 위치 예측

본 장에서는 연쇄 범죄에 대하여 거점 및 다음 범행 위치 예측 기법과 이를 위한 시공간 분석 기법을 설명한다.

4.1 거점 위치 예측

‘범죄자는 다른 모든 조건이 동일하다면, 가장 적은 노력을 들여서 범행을 저지른다.’는 최소 노력 원칙[8]을 근거로 등방성 패턴의 연쇄 범죄의 경우에는 거점 위치를 예측하기 위해 중심점으로써 CMD [5,6]를 활용한다. 반면, 비등방성 패턴의 연쇄 범죄의 경우에는 중심축을 기반으로 최소 비용 지점인 CMC(Center of Minimum Cost)를 찾는 방법을 제시한다.

단일 중심점 패턴 연쇄 범죄의 거점 위치 예측을 위하여 모든 범행 위치들로부터 공간상 가장 가까운 한 점(CMD)을 찾고 그 점과 다른 모든 범행 위치들과의 평균 거리만큼의 반지름을 갖는 원 영역을 생성한다.

다중 중심점 패턴 연쇄 범죄의 거점 위치 예측을 위하여 활성 군집을 분석하고 각 활성 군집을 단위로 단일 중심점 패턴 연쇄 범죄의 거점 예측 기법을 적용하여 지역적 거점 영역을 생성한다. 한 거점이 모든 활성 군집의 중심 역할을 하는 경우를 고려하여 모든 활성 군집에 대한 전역적 거점 영역도 생성한다. 전역적 거점 예

측을 위하여 군집 내 범행들의 발생 빈도, 범행 발생 시간의 최근 여부에 따른 거점 가중치를 적용할 수 있다. 거점 가중치는 수식 1을 통해 구해진다.

[수식 1] 거점 가중치 = $(occRatio * (1 - recentWeight)) + (recRatio * recentWeight)$

occRatio는 전체 범행들에 대하여 해당 군집에서 발생한 범행의 발생 횟수에 대한 비율, recRatio는 최근 범행들 중에서 해당 군집에서 발생한 최근 범행의 횟수에 대한 비율을 뜻한다. recentWeight는 occRatio와 recRatio에 대한 상대적인 중요도를 결정하기 위해 부여되는 파라미터 값이다. 동일한 중요도를 가질 때 recentWeight는 기본값으로 0.5를 가진다.

그림 3은 거점 가중치를 적용하여 생성한 전역적 거점 영역의 예를 보여준다.

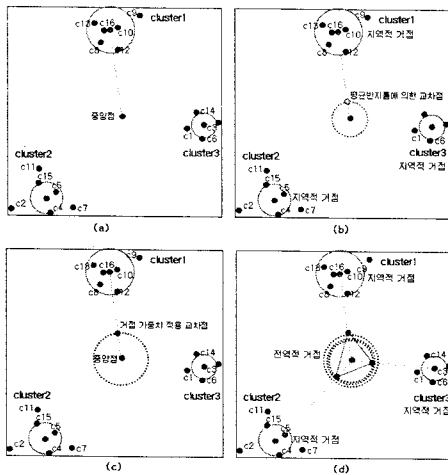


그림 3. 전역적 거점 영역 생성 예

그림 3(a)는 중앙점과 cluster1 군집에 소속된 범행 위치들의 CMD를 연결하여 생성한 선을 보여준다. 그림 3(b)는 중앙점을 중심으로 하여 평균 반지름에 의해 생성된 원과 그림 3(a)에서 생성된 선이 교차하는 점을 보여준다. 이때, 평균 반지름은 세 군집 cluster1, cluster2, cluster3에 소속된 범행 위치들의 지역적 거점 영

역인 원들의 반지름에 대한 평균이다.

그림 3(c)에서는 거점 가중치를 적용하여 교차점을 구한다. 거점 가중치를 적용한 교차점이란 그림 3(b)에서의 평균 반지름에 수식 1의 거점 가중치를 곱한 값만큼 확장된 반지름을 갖는 새로운 원을 생성하고 그것이 그림 3(a)에서의 선과 교차하는 점이다. 그림 3(d)는 전역적 거점 영역을 보여준다.

즉, 전역적 거점 위치 예측을 위해 모든 활성 군집에 포함된 범행 위치들의 중앙점을 찾고, 중앙점과 각 군집의 CMD를 연결하는 선을 생성하며 각 군집의 거점 가중치를 적용하여 얻은 원과 교차하는 점을 찾는다. 마지막으로 활성 군집의 개수만큼 얻어진 교차점들을 연결하여 생성된 다각형이 예측된 거점 영역이 된다.

단일 중심축 패턴 연쇄 범주의 거점 위치 예측을 위하여 범행들의 중심축으로써 이동선을 생성하고 이를 기반으로 최소 비용 지점을 찾는다.

연속적인 범행 장소 위치들을 연결하는 초기 이동선을 생성 후 초기 이동선을 구성하는 선분들의 중앙점들을 연결하여 갱신된 이동선을 얻을 수 있다. 이동선의 갱신은 주어지는 파라미터(중앙점 조정 비율)값에 의해 기존 이동선으로부터 반복적으로 수행된다.

이동선을 기반으로 범의자가 동일한 경로로 이동한 경우 모든 범행에 대하여 이동선 진입 비용, 이동선상 이동 비용, 이동선에서 범행 위치까지의 이동 비용의 총합이 최소가 되는 지점을 CMC라 한다. 대상 영역의 정규화를 통해 전체 후보 CMC를 선정하고 각 후보 CMC에 대해 모든 범행 위치들까지의 이동 비용을 합한 총비용을 계산하여 최소값을 갖는 후보 CMC를 선택한다. 그림 4는 임의의 후보 CMC에서 범행 장소(c3)까지의 이동 경로를 보여준다. CMC를 찾기 위해 이동선상의 반복 이동에 대한 해당 경로의 누적

방문 횟수와 범죈자 학습률에 따른 가중치를 적용할 수 있다.

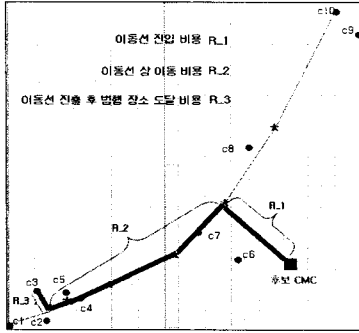


그림 4. 이동선 기반 이동 경로
가중치를 적용한 이동선상 이동 비용은 수식 2를 통해 구해진다. 진출점은 범죈 위치까지 최소 거리에 있는 이동선상 한 점이다.

[수식 2] 가중치 적용된 이동선상 이동비용 = 이동선상 이동거리 - {이동선상 이동거리 * (해당 진출점에 대한 누적 방문 횟수/진출점 총개수) * 학습률}

다중 중심축 패턴 연쇄 범죈의 거점 위치 예측을 위하여 활성 군집을 분석하고 각 활성 군집을 단위로 단일 중심축 패턴 연쇄 범죈의 거점 예측 기법을 적용하여 지역적 거점 영역을 생성하고 모든 활성 군집에 대한 전역적 거점 영역도 생성한다. 군집별 이동선과 모든 범죈 위치들로부터 생성된 이동선이 요구된다.

4.2 다음 범죈 위치 예측

중심점이 범죈시 이동을 위한 공간적 중심이 되는 등방성 패턴의 연쇄 범죈의 경우에는 다음 범죈 위치를 예측하기 위해 중심점과 범죈 장소간 방향 패턴을 활용한다. 반면, 중심축이 이동을 위한 공간적 중심이 되는 비등방성 패턴의 연쇄 범죈의 경우에는 중심축상의 연속 정점간 방향 패턴을 활용한다. 따라서, 중심점과 중심축의 결정이 예측에 큰 영향을 끼치게 된다. 등방성 패턴의 경우 중심점은 CMD를 활용하여 결정하고 비등방성 패턴

의 경우 중심축은 앞서 언급한 이동선 생성 방식에 따라 결정된다.

단일 중심점 패턴 연쇄 범죈의 다음 범죈 위치 예측을 위하여 모든 범죈 위치들로부터 CMD를 찾고 CMD와 모든 범죈 위치들간의 방향 패턴을 분석하여 다음 범죈의 방향을 예측한다. 그리고, 예측 방향에 존재하는 범죈 위치들의 중앙점을 구해 이를 원점으로 하고 원점과 CMD의 거리 크기만큼을 반지름으로 하는 원 영역을 생성한다.

다음 방향 예측을 위하여 CMD를 기준으로 전체 영역을 8방향으로 표현되는 영역으로 정규화한다. 따라서, 모든 범죈 위치들의 CMD에 대한 방향이 8방향중 하나로 결정된다. 연속적인 3건의 범죈을 단위로 모든 범죈 장소의 내부/외부방향 패턴과 연속 방향차 패턴을 분석한다. 내부/외부방향 패턴은 연속적인 3건의 범죈이 있을 때 첫 번째 범죈의 방향을 기준으로 세 번째 범죈의 방향이 두 번째 범죈의 방향보다 가까이 있는지(내부) 혹은 멀리 있는지(외부)에 대한 패턴이다.

중심점과 범죈 장소간 방향에 대한 연속 방향차 패턴은 연속적인 3건의 범죈이 있을 때 첫 번째 범죈의 방향을 기준으로 두 번째 범죈의 방향과의 크기차와 세 번째 범죈의 방향과의 크기차에 대한 패턴이다. 그림 5는 등방성 패턴의 16개 범죈중 연속적인 3건의 범죈으로 그림 5(a)c1-c2-c3과 그림 5(b)c2-c3-c4의 경우 연속 방향차를 보여준다. 시계 방향을 (+)로 할 때 그림 5(a)와 그림 5(b) 모두 (+3,-3)의 연속 방향차를 보인다.

모든 범죈 장소의 내부/외부방향 패턴과 연속 방향차 패턴을 근거로 다음 방향을 예측한다. 즉, 모든 연속적 이동에 대하여 중복 발생이 가장 큰 패턴에 따라 다음 범죈 위치에 대한 내부/외부방향과 연속 방향차를 예측하여 마지막 범죈을 기준으로 다음 방향을 결정한다.

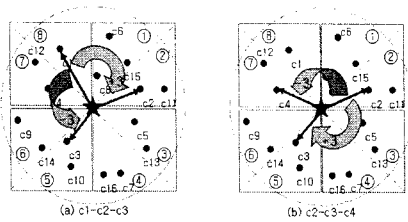


그림 5. 연속 방향차 예

다중 중심점 패턴 연쇄 범주의 다음 범행 위치 예측을 위하여 활성 군집을 분석하고 활성 군집 중 다음 범주가 수행될 군집을 선택한 후 단일 중심점 패턴 연쇄 범주에 대한 예측 기법을 적용하여 지역적 예측을 수행한다. 모든 범행에 대하여 전역적 예측도 수행한다. 다음 군집 예측을 위해 군집 간의 연속적 이동에 대한 중복 발생이 가장 큰 패턴을 활용한다.

단일 중심축 패턴 연쇄 범주의 다음 범행 위치 예측을 위하여 모든 범행 위치들로부터 이동선을 생성하고 이동선상의 연속 정점간 방향 패턴을 분석하여 다음 범행의 방향을 예측한다. 그리고, 이동선의 마지막 정점으로부터 예측 방향으로 이동선을 확장한다. 예측 영역은 확장된 이동선상의 정점들로 구성되는 영역이다.

이동선상의 연속적인 3개 정점을 단위로 모든 정점의 순/역방향 패턴과 연속적인 4개 정점을 단위로 모든 정점의 연속 방향차 패턴을 분석한다. 순/역방향 패턴은 연속적인 3개 정점이 있을때 두 번째 정점과 첫 번째 정점간 방향과 세 번째 정점과 두 번째 정점간 방향의 유사성에 대한 패턴이다. 그 차이가 +/-1 이하이면 다음 방향이 이전 방향에 대하여 순방향, 그렇지 않으면 역방향이라 한다.

중심축상의 연속 정점간 방향에 대한 연속 방향차 패턴은 연속적인 4개 정점이 있을때 두 번째 정점의 첫 번째 정점에 대한 방향(이전 방향)과 세 번째 정점의 두 번째 정점에 대한 방향(다음 방향)의 크기차와 세 번째 정점의 두 번째 정점에

대한 방향(이전 방향)과 네 번째 정점의 세 번째 정점에 대한 방향(다음 방향)의 크기차에 대한 패턴이다.

이동선의 모든 정점들에 대한 순/역방향 패턴과 연속 방향차 패턴을 근거로 다음 방향을 예측한다. 예측된 방향이 순방향인지 역방향인지에 따라 예측 방향으로 확장되는 이동선의 크기가 해당 순/역방향의 연속점들의 평균 길이로 결정된다.

다중 중심축 패턴 연쇄 범주의 다음 범행 위치 예측을 위하여 활성 군집을 분석하고 활성 군집 중 다음 범주가 수행될 군집을 선택한 후 단일 중심축 패턴 연쇄 범주에 대한 예측 기법을 적용하여 지역적 예측을 수행한다. 모든 범행에 대하여 전역적 예측도 수행한다.

5. 실험 및 평가

살인, 강간, 강도 등 다양한 유형의 16건 연쇄 범주 데이터를 이용하여 예측 정확도 실험을 하였다. 실험 평가를 위해 다음의 평가 지표를 정의한다. 첫째, ErrorDistance-TopPoint는 예측 거점과 실제 거점간의 최소 거리로써 예측 결과에 대한 정량적 오차 분석을 위해 사용된다. 둘째, ErrorDistance-TopArea는 예측 거점 영역(혹은 다음 범행 영역)과 실제 거점(혹은 다음 범행 위치)와의 최소 거리이다. ErrorDistance-TopArea는 모든 후보 CMC들 중에서 비용이 적은 0.01% CMC를 대상으로 한다. 두 가지 평가 지표의 값은 작을수록 예측 정확도가 높음을 의미한다.

그림 6은 CGT, CM, JTC, CMD와 제안 기법에 대한 평균 ErrorDistance-TopPoint 및 평균 ErrorDistance-TopArea를 비교한 것을 보여준다. CGT, CM, JTC에서는 거리 감퇴 함수의 입력 파라미터들에 대해 기본값을 적용하였고, 제안 기법을 위해 중앙점 조정 비율 0.4, 학습률 0.4를 적용하였다.

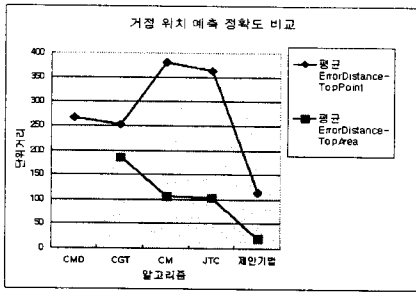


그림 6. 거점 예측 오차 비교

거점 위치 예측시 제안 기법은 평균 ErrorDistance-TopPoint와 평균 ErrorDistance-TopArea 모두 가장 작은 예측 오차를 가진다. 기존 알고리즘에 비하여 평균 6배 정도 높은 정확도를 보인다. 이는 거점이 범행 위치들의 중앙에 위치하지 않을 경우나 거점이 2개인 경우에도 기존 방법들은 단순히 거리를 기반으로 모든 범행들의 무게 중심을 거점으로써 높은 확률을 주기 때문이다.

그림 7은 CWA, PGM과 제안 기법에 대한 평균 ErrorDistance-TopArea를 비교한 것을 보여준다. 제안 기법을 위해 중앙점 조정 비율 0.4를 적용하였다.

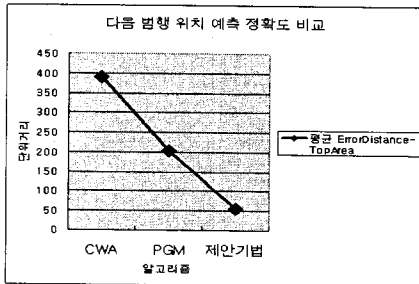


그림 7. 다음 범행 예측 오차 비교

다음 범행 위치 예측시 제안 기법은 기존 알고리즘에 비하여 평균 5배 정도 높은 정확도를 보인다. 이는 기존 방법들이 일정 간격동안의 연속적 이동에 규칙성이 뚜렷하지 않거나 이동 거리의 표준 편차가 큰 경우에도 단순히 평균 거리에 대한 패턴을 활용하기 때문이다.

6. 결론

본 논문에서는 연쇄 범죄의 시공간 패턴을 기반으로 하는 거점 및 다음 범행 위치 예측 기법을 제안하였다. 예측 정확도를 개선하기 위하여 기존 연구에서 고려하지 않은 비등방성 패턴과 군집화된 연쇄 범죄 패턴을 고려하며 각 패턴에 따라 거점과 범행 장소간 이동 특성, 연속적 범행 장소간 이동 특성 등을 활용한 시공간 분석 기법을 적용한다. 또한, 범행 발생 공간뿐만 아니라 시간도 활용하였다. 마지막으로 실험을 통해 제안 기법이 기존 방법보다 우수함을 입증하였다.

참고문헌

- [1] Ashby, D. I., and Longley, P. A., "Geocomputation, Geodemographics and Resource Allocation for Local Policing," *Transactions in GIS*, 9(1), 2005, p. 53-72.
- [2] Laukkanen, M., and Santtila, P., "Predicting the Residential Location of a Serial Commercial Robber," *Forensic Science International*, Vol.157, 2005, pp. 71-82.
- [3] Rossmo, K., *Geographic Profiling*, CRC Press, Boca Raton, 2000.
- [4] Canter, D., "Predicting Serial Killers Home Base using a Decision Support System," *Journal of Quantitative Criminology*, 16(4), 2000, pp.457-478.
- [5] Levine, N., "CrimeStat II," *Crime Mapping News*, 5(2), 2003, pp.2-4.
- [6] Levine, N., *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations*, <http://webapp.icpsr.umich.edu/cgi-bin/bob/newark?study=2824&path=ICPSR>, 2005.
- [7] Catalano, P., Hill, B., and Long, B., "Geographical Analysis and Serial Crime Investigation: A Case Study of Armed Robbery in Phoenix, Arizona," *Security Journal*, 14(3), 2001, pp. 27-41.
- [8] Beauregard, E., Proulx, J., and Rossmo, K., "Spatial Patterns of Sex Offenders: Theoretical, Empirical, and Practical Issues," *Aggression and Violent Behavior*, 10, 2005, pp.579-603.