

# 시공간 분석 기반 연쇄 범죄 거점 위치 예측 알고리즘<sup>†</sup>

## Base Location Prediction Algorithm of Serial Crimes based on the Spatio-Temporal Analysis

홍동숙\* / Dong-Suk Hong, 김정준\*\* / Joung-Joon Kim, 강홍구\*\*\* / Hong-Koo Kang  
이기영\*\*\*\* / Ki-Young Lee, 서종수\*\*\*\*\* / Jong-Soo Seo, 한기준\*\*\*\*\* / Ki-Joon Han

### 요약

고급 GIS 및 복잡한 공간 분석 기술이 발전함에 따라 다양한 의사 결정 지원 시스템에서 지리적 혹은 공간적 문제 해결을 위한 고급 지식을 지원하기 위해 더욱 강력한 기술이 필요하게 되었다. 또한, 법집행 기관 및 수사 기관 등을 중심으로 효율적인 수사 및 향후 범죄 예방을 위해 과학 수사, 법 과학에 관한 연구의 필요성이 증대되고 있다. 특히, 연쇄 범죄의 공간적 패턴을 분석함으로써 범죄자의 거점 위치를 예측하기 위한 지리적 프로파일링(Geographic Profiling)에 대한 연구가 활발하다. 그러나, 기존의 지리적 프로파일링 연구에서는 공간적 패턴 분석을 위해 단순히 통계적 방법만을 사용하고 있고, 연쇄 범죄에 대한 다양한 공간적, 시간적 분석 기술을 지원하지 않으므로 거점 예측시 낮은 정확도를 보인다.

그러므로, 본 논문에서는 범행 위치의 공간적 분포와 범죄 발생의 시간적 분포 특성에 따라 연쇄 범죄의 시공간 패턴을 유형화하고, 이를 기반으로 연쇄 범죄의 거점 위치를 보다 정확하게 예측하는 알고리즘으로 STA-BLP(Spatio-Temporal Analysis based Base Location Prediction)을 제안한다. STA-BLP는 하나의 거점으로부터 특정 방향을 선호하여 이동하며 발생하는 연쇄 범죄의 비등방성 패턴을 고려하고, 동일한 경로에 대한 반복 이동에 대한 범죄자의 학습 효과를 고려함으로써 예측 정확도를 개선시킨다. 또한, 다수의 군집화된 범행 위치들로부터 각 군집에 소속된 범행 위치들에 대한 지역적 거점 위치 예측과 모든 범행 위치에 대한 전역적 거점 위치 예측을 통해 거점이 다수 존재하는 연쇄 범죄의 경우에도 보다 정확한 예측을 수행한다. 마지막으로 다양한 실험을 통해 기존에 제시된 알고리즘과 STA-BLP의 예측 정확도를 비교하여 제안 알고리즘의 우수성을 입증하였다.

† 본 연구는 건설교통부 첨단도시기술개발사업-지능형국토정보기술혁신 사업과제의 연구비지원(07국토정보C05)에 의해 수행되었음.

- 논문접수 : 2008.3.10      ■ 심사완료 : 2008.5.7
- \* 연세대학교 TMS 정보기술사업단 박사후연구원(dongsukh@yonsei.ac.kr)
- \*\* 교신저자 건국대학교 대학원 컴퓨터공학과 박사과정(jkim9@db.konkuk.ac.kr)
- \*\*\* 건국대학교 대학원 컴퓨터공학과 박사과정(hkkang@db.konkuk.ac.kr)
- \*\*\*\* 을지대학교 의료산업학부 교수(kylee@eulji.ac.kr)
- \*\*\*\*\* 연세대학교 전기전자공학과 교수(jsseo@yonsei.ac.kr)
- \*\*\*\*\* 건국대학교 컴퓨터공학부 교수(kjhan@db.konkuk.ac.kr)

## Abstract

With the recent development of advanced GIS and complex spatial analysis technologies, the more sophisticated technologies are being required to support the advanced knowledge for solving geographical or spatial problems in various decision support systems. In addition, necessity for research on scientific crime investigation and forensic science is increasing particularly at law enforcement agencies and investigation institutions for efficient investigation and the prevention of crimes. There are active researches on geographic profiling to predict the base location such as criminals' residence by analyzing the spatial patterns of serial crimes. However, as previous researches on geographic profiling use simply statistical methods for spatial pattern analysis and do not apply a variety of spatial and temporal analysis technologies on serial crimes, they have the low prediction accuracy.

Therefore, this paper identifies the typology the spatio-temporal patterns of serial crimes according to spatial distribution of crime sites and temporal distribution on occurrence of crimes and proposes STA-BLP(Spatio-Temporal Analysis based Base Location Prediction) algorithm which predicts the base location of serial crimes more accurately based on the patterns. STA-BLP improves the prediction accuracy by considering of the anisotropic pattern of serial crimes committed by criminals who prefer specific directions on a crime trip and the learning effect of criminals through repeated movement along the same route. In addition, it can predict base location more accurately in the serial crimes from multiple bases with the local prediction for some crime sites included in a cluster and the global prediction for all crime sites. Through a variety of experiments, we proved the superiority of the STA-BLP by comparing it with previous algorithms in terms of prediction accuracy.

**주요어 :** 시공간 분석, 범죄 패턴, 연쇄 범죄, 거점 위치 예측

**Keyword :** Spatio-Temporal Analysis, Crime Patterns, Serial Crimes, Base Location Prediction

## 1. 서론

최근에 고급 GIS 및 복잡한 공간 분석 기술이 발전함에 따라 다양한 의사 결정 지원 시스템에서 보다 효율적인 지식 검색과 고급 지식의 활용을 위해 지리적 혹은 공간적 문제 해결을 위한 강력하고 획기적인 기술이 필요하게 되었다. 특히, 공간 의사 결정 지원 시스템에서 비주얼하고 인터랙티브한 인터페이스 지원에 관한 연구와 지리적 공간의 복잡한 특성과 암시적인 제약의 효율적인 처리를 위한 연구가 많이 진행되었다[1,2,3].

또한, 검찰청, 경찰청 등의 법집행 기관 및 수사 기관 등을 중심으로 효율적인 수사 및 향후 범죄

예방을 위해 과학 수사, 법 과학(Forensic Science)에 관한 연구의 필요성이 증대되고 있다 [4,5,6]. 실제로 범죄 발생 장소, 용의자의 거주지 등 공간적 특성을 포함하는 범죄 데이터를 분석함으로써 수사 대상 지역과 용의자를 우선순위화하기 위해 다수의 범죄 정보 관리 및 분석 시스템이 개발되어 활용되고 있다[7,8,9].

범죄 발생 사후의 효과적 수사 대응과 사전 예방을 위한 의사 결정 지원 시스템으로써 범죄 정보 관리 및 분석 시스템은 범죄 발생 장소, 용의자의 거주지 등 공간적 특성을 포함하는 복잡한 범죄 데이터를 분석함으로써 수사 대상 지역과 용의자를 우선순위화한다[10,11,12]. 예를 들어, 다음의 질

의를 생각해 볼 수 있다.

질의 예) 미해결 연쇄 범죄에 대한 범죄자를 검거하기 위해 어느 지역에서 순찰을 시작해야 하는지?

이러한 질의에서 보는 바와 같이, 범죄자의 위치를 파악하는 것은 범죄 수사를 위해 가장 기본적이고 핵심적인 사항이다. 특히, 동일인에 의해 반복적으로 수행되는 연쇄 범죄의 경우, 관련된 다수 범행 위치들로부터 거주지나 회사와 같은 범죄자의 거점(Base)을 예측하는 일은 효과적으로 범죄자를 검거하기 위해 매우 중요하다. 다수 연구에서 연쇄 범죄의 공간적 패턴을 분석하여 범죄자의 거점 위치를 예측하기 위한 지리적 프로파일링(Geographic Profiling) 알고리즘[11,13,14,15]을 제시하였다.

그러나 기존의 지리적 프로파일링 연구에서는 공간적 패턴 분석을 위해 대부분 단순한 통계적 방법만이 사용되고 있고, 실제 환경에서 연쇄 범죄 분포에 대한 다양한 공간적, 시간적 분석이 어려운 형편이다. 또한 범행 위치의 공간적 특성에 대한 비현실적인 가정으로 인해 다양한 연쇄 범죄에 대한 적용이 어려우며, 범죄자의 범행 위치 선택 비용을 위해 단순히 범행 위치와의 거리만을 활용함으로써 결과적으로 거점 예측 결과 또한 낮은 정확도를 보이고 있다.

그러므로 본 논문에서는 보다 높은 정확도를 갖는 연쇄 범죄 위치 예측을 위해 범죄 발생 장소에 대한 공간적 분포와 범죄 발생 시각에 대한 시간적 분포 특성에 따라 연쇄 범죄의 시공간 패턴을 유형화하였다. 또한 본 논문에서는 보다 정확한 예측을 위해 특정 방향을 선호하여 이동하면서 발생하는 비등방성 패턴의 연쇄 범죄를 고려하였고, 다수 거점이 존재하는 패턴의 연쇄 범죄도 고려하였다. 특히, 비등방성 패턴의 연쇄 범죄의 경우 동일한 경로에 대한 반복 이동에 대한 범죄자의 학습 효과를 적용하였고, 다수 거점에 의한 연쇄 범죄의 경우 다수의 근접화된 범행 위치들로부터 각 근접에 소속된 범행 위치들에 대한 지역적 거점 위치 예측과

모든 범행 위치에 대한 전역적 거점 위치 예측을 수행함으로써 연쇄 범죄 거점 예측의 정확도를 개선하였다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 관련 연구로서 기존의 지리적 프로파일링 알고리즘에 대하여 분석한다. 제 3 장에서는 본 논문에서 제시한 연쇄 범죄 시공간 패턴의 유형을 기술한다. 제 4 장에서는 시공간 분석 기반 연쇄 범죄 거점 위치 예측 알고리즘(STA-BLP: Spatio-Temporal Analysis based Base Location Prediction)을 제시하고, 제 5 장에서는 다양한 실험 결과로서 기존에 제시된 예측 방법과 본 논문에서 제시한 STA-BLP의 예측 정확도를 비교 설명한다. 마지막으로, 제 6 장에서 본 논문의 결론을 언급한다.

## 2. 관련 연구

기존의 지리적 프로파일링 알고리즘은 크게 두가지 접근 방법으로 나뉜다. 첫 번째는 범죄 발생 위치들로부터 지리적인 중심 위치를 찾는 공간 분포(Spatial Distribution) 접근 방법이고, 두 번째는 영역안의 모든 위치에서 범죄 발생 위치와의 거리에 따른 확률을 계산하여 거점이 존재할 확률값을 구하는 확률 거리(Probability Distance) 접근 방법이다[16].

공간 분포 접근 방법은 Centographic 방법이라고도 하며, 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 x 좌표값에 대한 산술적 평균을 x 좌표로 하고, 범행 위치들의 y 좌표값에 대한 산술적 평균을 y 좌표로 하는 위치이다. 중간점은 범행 위치들의 x 좌표값 중 중간 크기인 값을 x 좌표로 하고, 범행 위치들의 y 좌표값 중 중간 크기인 값을 y 좌표로 하는 위치이다. CMD란 범행의 거점 위치를 예측하기 위해 각 위치에서 모든 범행 위치들까지의 유클리디언

거리의 총합을 구한 후 이 총합이 가장 작은 위치이다[14].

확률 거리 접근 방법은 범죄자의 거점으로부터 거리가 먼 지역일수록 범죄 발생 수가 줄어든다는 거리 감퇴 모델(Distance Decay Model)을 기반으로 한다. 거점으로부터의 거리와 범죄 발생 수에 관한 정량적 관계에 대한 다양한 수식을 이용하여 검색 영역안의 모든 위치에 대해 거점이 될 확률값을 구하므로 수학적 방법이라고도 한다. 확률 거리 접근 방법으로는 Criminal Geographic Targeting, Canter Mode, Journey-to-Crime 등이 있다.

ECRI(Environmental Criminology Research Inc.)에서 개발되어 RIGEL Analyst에 적용된 Criminal Geographic Targeting(CGT)[15]은 임의의 셀과 범죄 위치간의 맨하탄 거리(Manhattan Distance)를 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.

David Canter에 의해 개발되어 Dragnet에 적용된 Canter Model(CM)[11]은 임의의 셀과 범죄 위치간의 유클리디언 거리(Euclidean Distance)를 기본으로 거리 감퇴 함수를 정의하고 있다. 정의된 함수에 따라 셀과 범죄 위치간 거리에 대한 음의 지수 함수(Negative Exponential Function)에 의하여 범죄 위치에서 멀리 있는 셀일수록 작은 확률값을 가진다.

Ned Levine & Associates에서 개발되어 Crime-Stat에 적용된 Journey-to-Crime (JTC)[14]은 Criminal Geographic Targeting과 Canter Model의 알고리즘을 수정하여 사용자에게 의해 선택될 수 있는 몇 가지 거리 감퇴 함수를 정의하고 있다. 대표적으로 절단 음의 지수 함수(Truncated Negative Exponential Function)는 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 확률값은 해당 셀과 범죄 위치간의 유클리디언 거리에 의해 선형으로 커지거나 해당 셀과 범죄 위치간의 유클리디언 거리를 지수로 하는 음의 지수 함수에 의해 작아진다.

위에서 언급한 이러한 기존의 방법들은 다음의 가정으로 인해 문제점을 가진다.

- 가정1) 범인은 연쇄 범죄를 수행하는 동안 하나의 거점에서 이동하지 않는다.
- 가정2) 거점으로부터 모든 방향이 범행에 대한 동일한 기회를 갖는다.

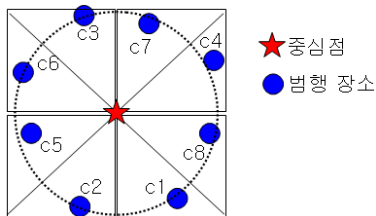
가정1은 연쇄 범죄로 연관된 모든 범행들이 하나의 거점으로부터 발생되었다는 것을 의미한다. 따라서, 집, 회사, 부모님의 집, 친구의 집 등 범죄자의 일상 생활에서 실제 근거지가 다수일 때 거점 예측의 한계를 보인다. 가정2는 범죄자가 범행 위치를 선택할 때 거점을 중심으로 모든 방향이 같은 확률을 가진다는 것을 의미한다. 따라서, 지리적·지형적 특성 등으로 인해 특정한 방향성을 갖는 연쇄 범죄에 대해 거점 예측의 한계를 보인다. 또한, 임의의 위치와 범죄 발생 위치들 간의 직선 거리를 기반으로 하고, 범죄 발생들 간의 시간적 관계를 고려하지 않기 때문에 효과적으로 연쇄 범죄의 거점을 구할 수 없다.

### 3. 연쇄 범죄 데이터 시공간 패턴

본 장에서는 시간 및 공간 분포에 따른 연쇄 범죄 데이터의 시공간 패턴에 대해 정의한다. 본 논문에서 정의하는 연쇄 범죄 데이터의 시공간 패턴은 다수 범행 위치들의 분포가 원형에 가까운지 혹은 선형에 가까운지에 따라 크게 두 가지, 즉 등방성 패턴과 비등방성 패턴으로 분류된다.

### 3.1 등방성 패턴

등방성 패턴이란 특정한 한 점 혹은 다수 점들을 중심으로 모든 방향에 유사한 선호를 보이는, 즉 원형 분포를 보이는 연쇄 범죄의 시공간 패턴을 뜻한다. 그림 1은 등방성 패턴 연쇄 범죄의 한 예로써 한 점을 중심으로 8방위로 동일한 분포를 보이는 8개 범죄(c1~c8)에 대한 발생 장소와 그 범행 위치들의 중심점을 보여준다.



<그림 1> 등방성 패턴의 예

이와 같은 등방성 패턴의 연쇄 범죄에서는 원형으로 분포된 범죄 발생 위치들로부터 중심점의 식별이 가능하다. 중심점이란 등방성 패턴인 연쇄 범죄의 경우 범행 위치까지의 이동을 위한 공간적 중심인 한 점(Point)을 의미한다. 본 논문에서 제안하는 거점 위치 예측 알고리즘에서는 등방성 패턴의 연쇄 범죄 범행 위치들로부터 이러한 중심점을 찾는다. 즉, 등방성 패턴의 연쇄 범죄에서는 중심점이 곧 거점이 된다.

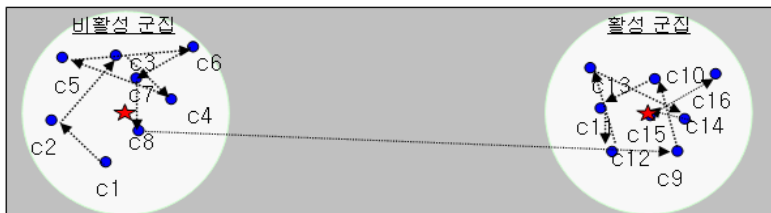
다수 범행 위치들을 포함하는 하나의 그룹을 군집(Cluster)이라 한다. 전체 범행 위치들이 등방성 패턴을 보일 때 모든 범행 위치들이 하나의 군집으로 표현되지만 단지 일부 범행 위치들이 등방성 패

턴을 보일 때는 해당하는 일부 범행 위치들도 하나의 군집으로 표현될 수 있다. 그러므로 모든 범행 위치가 하나의 군집으로 간주되는 경우와 다수 군집으로 간주되는 경우에 따라 등방성 패턴은 다시 세부 패턴으로 구분된다. 등방성 패턴의 세부 패턴을 소개하기 위해 먼저 활성 군집에 대하여 정의한다. 범행 발생 순서를 의미하는 번호를 범행번호라 하면, 활성 군집의 정의는 다음과 같다.

[정의 1] 활성 군집

- 1) 모든 범행 위치들을 포함하는 1개 군집 Ca는 활성 군집이다.
- 2) Ca에 소속된 범행 위치들 중 최소인 범행번호가 다른 군집 Cb에 소속된 범행 위치들 중 최대인 범행번호보다 큰 경우 군집 Ca는 활성 군집이다.
- 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 경우 Ca와 Cb는 모두 활성 군집이다.

등방성 패턴은 군집이 1개일 때 단일 중심점 패턴, 군집이 2개 이상이면 다중 중심점 패턴이라 한다. 다중 중심점 패턴은 다시 군집이 2개 이상이면서 활성 군집이 1개일 때 단일 활성 다중 중심점 패턴, 군집이 2개 이상이면서 활성 군집도 2개 이상일 때 다중 활성 다중 중심점 패턴으로 구분된다. 그림 2는 단일 활성 다중 중심점 패턴인 연쇄 범죄



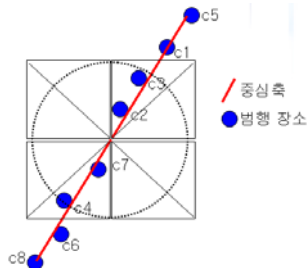
<그림 2> 단일 활성 다중 중심점 패턴의 예

의 간단한 예를 보여준다.

그림 2에서 보는 바와 같이, 총 16개의 범행 위치 중 8개의 범행 위치(1번째 범행부터 8번째 범행 위치)는 한 점을 중심으로 원형 분포를 보이고, 나머지 8개의 범행 위치는 또 다른 한 점을 중심으로 원형 분포를 보인다. 따라서, 전체 범행 위치는 8개의 범행 위치를 포함하는 군집과 나머지 8개의 범행 위치를 포함하는 또 다른 군집으로 분류될 수 있다. 범행의 발생 순서에 따르면 8번째 범행과 9번째 범행을 기준으로 두 군집간 이동이 존재하지 않는다. 따라서, [정의 1]의 2)에 의해 9번째 이후 범행들을 포함하는 군집이 활성 군집이다.

### 3.2 비등방성 패턴

비등방성 패턴이란 특정한 방향 혹은 방향들에 대한 선호를 보이는, 즉 선형 분포를 보이는 연쇄 범죄의 시공간 패턴을 뜻한다. 그림 3은 비등방성 패턴 연쇄 범죄의 한 예로써 북북동-남남서 방향으로 분포된 8개 범죄(c1~c8)에 대한 발생 장소와 그 범행 위치들의 중심축을 보여준다.

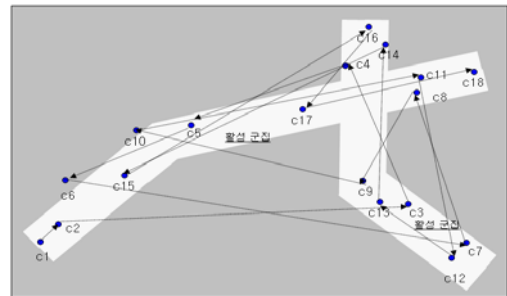


<그림 3> 비등방성 패턴의 예

이와 같은 비등방성 패턴의 연쇄 범죄에서는 선형으로 분포된 범죄 발생 위치들로부터 중심축의 식별이 가능하다. 중심축이란 비등방성 패턴인 연쇄 범죄의 경우 범행 위치까지의 이동을 위한 공간적 중심인 선(Line)을 의미한다. 본 논문에서 제안하는 거점 위치 예측 알고리즘에서는 비등방성 패턴의 연쇄 범죄 범행 위치들로부터 중심축을 찾고 그 중심축상의 반복적인 이동을 고려하여 최소 비용

지점을 거점으로 찾는다. 즉, 비등방성 패턴의 연쇄 범죄에서 중심축은 거점을 찾기 위해 활용된다.

비등방성 패턴에서도 범행 위치를 다수의 군집으로 구분하는 군집화가 가능하다. 즉, 모든 범행 위치가 하나의 군집으로 간주되는 경우와 다수 군집으로 간주되는 경우에 따라 비등방성 패턴은 다시 세부 패턴으로 구분된다. 비등방성 패턴은 군집이 1개일 때 단일 중심축 패턴, 군집이 2개 이상이면 다중 중심축 패턴이라 한다. 다중 중심축 패턴은 다시 군집이 2개 이상이면서 활성 군집이 1개일 때 단일 활성 다중 중심축 패턴, 군집이 2개 이상이면서 활성 군집도 2개 이상일 때 다중 활성 다중 중심축 패턴으로 구분된다. 그림 4는 다중 활성 다중 중심축 패턴인 연쇄 범죄의 간단한 예를 보여준다.



<그림 4> 다중 활성 다중 중심축 패턴의 예

그림 4에서 보는 바와 같이, 총 18개의 범행 위치 중 10개의 범행 위치(1, 2, 5, 6, 8, 10, 11, 15, 17, 18번째 범행 위치)는 일정한 방향으로 선형 분포를 보이고, 나머지 8개의 범행 위치는 또 다른 선형 분포를 보인다. 따라서, 전체 범행 위치는 10개의 범행 위치를 포함하는 군집과 나머지 8개의 범행 위치를 포함하는 또 다른 군집으로 분류될 수 있다. 이때, 연속적인 범행에 있어서 반복적인 군집간 이동을 보이고 있다. 따라서, 두 군집은 [정의 1]의 3)에 의해 모두 활성 군집이다.

## 4. STA-BLP 알고리즘

본 장에서는 연쇄 범죄의 시공간 패턴에 따라 보

다 정확하게 거점 위치를 예측할 수 있는 시공간 분석 기반 연쇄 범죄 거점 위치 예측 알고리즘인 STA-BLP(Spatio-Temporal Analysis based Base Location Prediction)을 제시한다. 거점(Base)은 범죄 활동의 근거지를 의미하며 범죄자의 거주지, 회사, 친구 집, 친척 집 등이 예가 될 수 있다. 본 논문에서는 거점이 다음과 같은 특성을 가질 수 있다고 가정한다.

- 1) 두 개 이상의 거점으로부터 연쇄 범죄의 범행들이 수행될 수 있다.
- 2) 범죄자는 거점으로부터 특정 방향을 선호하여 범행 위치를 선택할 수 있다.

STA-BLP 알고리즘은 거점 위치의 예측 정확도를 높이기 위해 연쇄 범죄의 시공간 패턴에 따라 차별화된 예측 기법을 제시하며, 예측 시 시공간 데이터로 표현되는 범죄 데이터에 대하여 다양한 시공간 분석을 적용한다.

‘범죄자는 다른 모든 조건이 동일하다면, 가장 적은 노력을 들이면서 범행을 저지른다. 즉, 가장 적은 비용이 드는(예를 들어, 가장 가까운) 장소를 선택한다.’는 최소 노력 원칙(Least-effort Principle) [17]을 근거로 등방성 패턴의 연쇄 범죄의 경우에

는 거점 위치를 예측하기 위해 중심점으로써 CMD(Center Of Minimum Distance) [14]를 찾는다. 반면, 비등방성 패턴의 연쇄 범죄의 경우에는 중심축을 기반으로 최소 비용 지점인 CMC(Center of Minimum Cost)를 찾는 방법을 제시한다. 특히, 다수의 군집화된 범행 위치들로부터 군집 단위의 거점 위치 예측 기법을 제시함으로써 두 개 이상의 거점으로부터 발생된 연쇄 범죄의 거점 위치를 보다 정확하게 예측할 수 있다.

본 논문에서는 우선 STA-BLP 알고리즘에서 사용되는 자료 구조에 대하여 살펴보고, STA-BLP 알고리즘을 크게 네 부분(단일 중심점 패턴, 다중 중심점 패턴, 단일 중심축 패턴, 다중 중심축 패턴)으로 구분하여 각 패턴의 연쇄 범죄에 대한 거점 예측 기법을 설명한다.

#### 4.1 자료 구조

본 절에서는 STA-BLP 알고리즘에서 사용되는 자료 구조에 대하여 설명한다. 표 1은 연쇄 범죄의 공간 및 시공간 표현을 위해 사용되는 자료 구조를 보여주는데, 연쇄 범죄란 동일인에 의해 반복적으로 수행되는 일련의 범죄를 의미한다.

<표 1> 연쇄 범죄의 공간 및 시공간 표현을 위한 자료 구조

공간 표현	시공간 표현
S_POINT 범죄 발생 위치 표현 예측된 거점 위치 표현	ST_POINT 범죄 발생 위치 및 시각 표현
S_MULTIPPOINT 관련된 범죄 발생 위치들 표현 예측된 거점들의 위치들 표현	ST_MULTIPPOINT 관련된 범죄들의 발생 횟수, 위치, 시각 표현
S_CIRCLE 원형인 예측 거점 영역 표현	ST_LINESTRING 관련된 범죄들의 중심축 표현 (중심축을 구성하는 점들의 시간적 순서 및 위치 표현)
S_POLYGON 다각형인 예측 거점 영역 표현	ST_MULTILINESTRING 관련된 범죄들의 중심축들 표현
S_GEOMETRY S_POINT, S_MULTIPPOINT, S_CIRCLE, S_POLYGON에 대한 추상화된 자료 구조	

표 1에서 S\_POINT는 하나의 범죄 발생 위치 혹은 관련된 범죄 발생 위치들로부터 예측된 거점의 위치(즉, 좌표)를 표현하기 위해 사용되는 자료 구조이다. S\_MULTIPPOINT는 연쇄 범죄로서 관련된 다수 범죄의 발생 위치들과 예측된 거점들의 위치들을 표현하기 위해 사용되며, 2개 이상의 S\_POINT를 포함한다. S\_CIRCLE은 원형인 예측 거점 영역을 표현하기 위해 사용되는 자료 구조이고, S\_POLYGON은 다각형인 예측 거점 영역을 표현하기 위해 사용되는 자료 구조이다. S\_GEOMETRY는 공간 표현을 위한 추상화된 자료 구조이다.

범죄 데이터는 발생한 범죄의 현상 및 특성을 표현함으로써 근본적으로 범죄 발생 시간 정보도 포함한다. 따라서, 하나의 범행 발생 위치와 시각을 표현하기 위해 표 1에서 정의한 ST\_POINT가 사용되고, 관련된 범죄들의 발생 횟수, 발생 위치, 발생 시각을 표현하기 위해 ST\_MULTIPPOINT가 사용된다. ST\_LINestring은 관련된 범죄들의 중심축을 표현하기 위한 자료 구조로서 발생 순서에 따라 순서화된 범죄 발생 위치를 표현하지만 발생 시각은 포함하지 않는다. ST\_MULTILINestring은 관련된 범죄들의 다수 중심축들을 표현하기 위한 자료 구조로서 2개 이상의 ST\_LINestring을 포함한다.

#### 4.2 단일 중심점 패턴 연쇄 범죄의 거점 위치 예측

최소 노력의 원칙에 따라 모든 범행을 수행하기 위해 최소의 비용이 요구되는 지점인 CMD가 거점

으로 고려될 수 있으므로 본 논문에서는 단일 중심점 패턴의 연쇄 범죄에 대한 거점 위치 예측을 위해 CMD 방법을 활용하여 중심점을 찾는다. 즉, 모든 범행 위치들로부터 공간상 가장 가까운 한 점을 찾고 그 점과 다른 모든 범행 위치들과의 평균 거리만큼의 반지름을 갖는 원 영역을 생성한다. 이러한 원 영역을 예측된 거점 영역이라 한다.

CMD를 찾기 위해 먼저 전체 영역은 같은 크기를 갖는 셀들로 정규화된다. 전체 영역의 모든 정규화된 셀들 중에서 CMD를 찾게 되므로 이러한 전체 영역은 거점을 예측하기 위한 검색 영역이라 한다. 즉, 검색 영역내의 모든 셀들은 CMD인지를 판단하기 위한 후보 CMD가 된다. 모든 범행 위치들로부터 공간상 가장 가까운 하나의 CMD를 찾기 위하여 각 후보 CMD에 대하여 모든 범행 위치들에 대한 거리합을 구한다.

그림 5는 단일 중심점 패턴 연쇄 범죄의 거점 예측을 위한 알고리즘을 보여준다. 알고리즘의 입력으로는 예측 대상인 범죄 데이터(pts)를 가지고, 출력으로는 예측된 거점 영역(baseArea)을 반환한다.

이 알고리즘에서는 우선 CMD를 찾고, 이 CMD와 모든 범행 위치들과의 평균 거리를 구한다. 그리고 CMD를 중심점으로 평균 거리를 반지름으로 하는 원 영역을 거점 영역으로 생성한다.

#### 4.3 다중 중심점 패턴 연쇄 범죄의 거점 위치 예측

본 논문에서는 다중 중심점 패턴의 연쇄 범죄에 대한 거점 위치 예측을 위해서 군집 단위의 지역적 거점 위치 예측 기법과 모든 범행 위치에 대하여

```

ALGORITHM : STA_BLP_oneCenterPoint( ST_MULTIPPOINT pts )
Search and return the base location from crime site points on their center point.
BEGIN
  S_POINT CMDPt ← ∅
  S_GEOMETRY baseArea ← ∅
  DOUBLE avgDistanceToCMD ← 0
1:  CMDPt ← Search_CMD(pts) //CMD인 지점을 검색
2:  avgDistanceToCMD ← GetAvgDistanceCMD(CMDPt, pts) //CMD와 범행 장소들간의 평균 거리 계산
3:  baseArea ← Create_baseArea(CMDPt, avgDistanceToCMD) //CMD를 중심점으로 하여 거점 영역 생성
4:  RETURN baseArea
END
    
```

<그림 5> 단일 중심점 패턴 연쇄 범죄의 거점 위치 예측 알고리즘



군집화 특성에 따라 결정되는 전역적 거점 위치 예측 기법을 제시한다. 이를 통해 각 거점이 각 활성 군집의 중심 역할을 하는 경우와 한 거점이 모든 다수의 활성 군집의 중심 역할을 하는 경우를 모두 고려할 수 있다.

이때 지역적 거점 위치 예측은 앞 절에서 언급한 단일 중심점 패턴 연쇄 범 죄의 거점 위치 예측 알고리즘을 각 활성 군집에 소속된 범 죄 위치들에 대하여 적용함으로써 수행된다. 또한, 전역적 거점 위치 예측은 군집 내 범 죄들의 발생 빈도, 범 죄 발생 시간의 최근 여부를 고려함으로써 수행된다. 예를 들어, 범 죄자가 상대적으로 빈번히 범 죄를 수행했던 군집 혹은 상대적으로 최근 범 죄를 수행했던 군집은 다른 군집에 비하여 거점으로부터의 이동이 용이할 수 있으므로 전역적 거점 위치 예측을 위해 높은 거점 가중치가 부여된다.

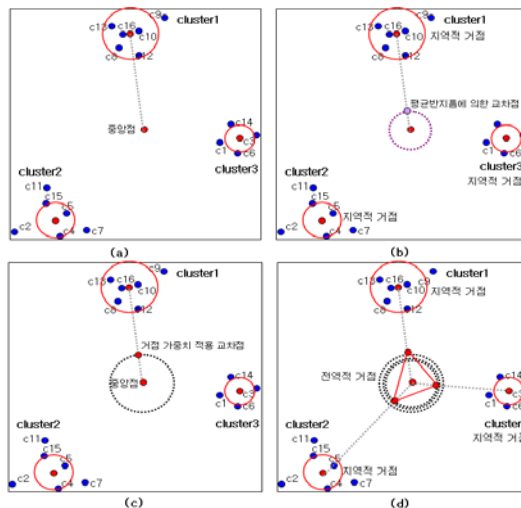
전역적 거점 위치 예측을 위해 먼저 모든 활성 군집에 포함된 범 죄 위치들의 중앙점을 찾는다. 중앙점의 x 좌표는 각 활성 군집의 CMD들의 x 좌표 값에 대한 평균값이고, 중앙점의 y 좌표는 각 활성 군집의 CMD들의 y 좌표값에 대한 평균값이다. 그리고, 중앙점과 각 군집의 CMD를 연결하는 선을 생성하고, 각 군집의 거점 가중치를 적용하여 얻은

원과 교차하는 점을 찾는다. 마지막으로 활성 군집의 개수만큼 얻어진 교차점들을 연결하여 생성된 다각형이 예측된 거점 영역이 된다. 거점 가중치는 [수식 1]을 통해 구해진다.

$$[수식 1] \text{ 거점 가중치(baseWeight)} = (\text{occurrenceRatio} * (1 - \text{recentWeight})) + (\text{recentnessRatio} * \text{recentWeight})$$

여기서 occurrenceRatio는 전체 범 죄들에 대하여 해당 군집에서 발생된 범 죄의 발생 횟수에 대한 비율을 의미하고, recentnessRatio는 최근 범 죄들 중에서 해당 군집에서 발생된 최근 범 죄의 횟수에 대한 비율을 뜻한다. occurrenceRatio와 recentnessRatio는 군집별 범 죄 위치 데이터를 분석하면 자동으로 알 수 있는 값이며, recentWeight는 occurrenceRatio와 recentnessRatio에 대한 상대적인 중요도를 결정하기 위해 부여되는 파라미터 값이다. occurrenceRatio와 recentnessRatio가 동일한 중요도를 가질 때 recentWeight는 기본값으로 0.5를 가진다.

그림 6은 거점 가중치를 적용하여 생성한 전역적 거점 영역의 예를 보여준다. 그림 6(a)는 중앙점과



<그림 6> 거점 가중치를 적용한 전역적 거점 영역 생성 예

cluster1 군집에 소속된 범행 위치들의 CMD를 연결하여 생성한 선을 보여준다. 그림 6(b)는 중앙점을 중심으로 하여 평균 반지름에 의해 생성된 원과 그림 6(a)에서 생성된 선이 교차하는 점을 보여준다. 이때, 평균 반지름은 세 군집 cluster1, cluster2, cluster3에 소속된 범행 위치들의 지역적 거점 영역인 원들의 반지름에 대한 평균이다.

다음으로는 그림 6(c)에서 보여주는 바와 같이 거점 가중치를 적용하여 교차점을 구한다. 거점 가중치를 적용한 교차점이란 그림 6(b)에서의 평균 반지름에 [수식 1]의 거점 가중치를 곱한 값만큼 확장된 반지름을 갖는 새로운 원을 생성하고 그것이 그림 6(a)에서의 선과 교차하는 점이다. 전역적 거점 영역을 예측하기 위해서는 그림 6(a), (b), (c)에서 보여준 과정이 활성 군집의 수만큼 반복적으로 수행되어야 한다. 그림 6(d)는 cluster1뿐만 아니라 cluster2, cluster3 군집에 대하여 거점 가중치 적용 교차점을 구하고 그 교차점들을 연결하여 생성된 전역적 거점 영역을 보여준다.

그림 7은 다중 중심점 패턴 연쇄 범죄의 거점 예측을 위한 알고리즘을 보여준다. 알고리즘의 입력

으로는 예측 대상인 범죄 데이터(pts), 최근 범행 발생 비율에 대한 가중치(recentWeight)를 가진다. 예측 대상 범죄 데이터에서는 동일한 pts 배열에 저장된 범죄들은 동일한 군집으로 간주된다. 그리고 출력으로는 예측된 거점 영역(baseArea)을 반환한다.

이 알고리즘에서는 각 활성 군집에 대하여 지역적 거점 영역을 생성한 후 모든 활성 군집에 대한 평균 반지름을 계산하고 중앙점을 구한다. 그리고, 각 활성 군집에 대해서 거점 가중치를 적용하여 전역적 거점 영역을 생성한다.

#### 4.4 단일 중심점 패턴 연쇄 범죄의 거점 위치 예측

본 논문에서는 단일 중심점 패턴의 연쇄 범죄에 대한 거점 위치 예측을 위해서 범행의 중심점으로써 연속적인 범행 위치들을 연결하는 초기 이동선(MovingLine)을 생성한다. 이동선을 기반으로 범죄자가 동일한 방향 및 동일한 경로로 이동한 경우 범행을 위한 비용을 이동선 진입 비용, 이동선상 이

```

ALGORITHM : STA.BLP_multiCenterPoint( ST.MULTIPOINT[] pts, DOUBLE recentWeight )
    Search and return the base location from clustered crime sites on their center points.
BEGIN
    S.MULTIPOINT subPts ← ∅
    S.POINT[] candidateCMDPts ← ∅
    S.POINT[] candidateCMDPt ← ∅
    S.GEOMETRY baseArea ← ∅
    INT[] activeClusters ← ∅
    INT clusterNums ← 1
    DOUBLE avgDistanceToCMD ← 0
    DOUBLE avgRadius ← 0
    clusterNums ← pts.count
1: activeClusters ← Analyze_connectivity(clusterNums, pts) //활성군집 분석
2: FOR EACH cluster clust of activeClusters //각 활성군집에 대한 지역적 거점 위치 예측
3:     subPts ← Get_points(clust, pts)
4:     candidateCMDPt ← Search_CMD(subPts)
5:     avgDistanceToCMD ← Get_avgDistanceCMD(candidateCMDPt, subPts)
6:     baseArea ← Create_union(baseArea, Create_baseArea(candidateCMDPt, avgDistanceToCMD))
7:     avgRadius ← Get_avgRadius(avgRadius + avgDistanceToCMD)
8: END FOR
9: FOR EACH cluster clust_a in activeClusters //모든 활성군집에 대한 전역적 거점 위치 예측
10:     FOR EACH cluster clust_b of the next clusters of cluster_a in activeClusters
11:         candidateCMDPts ← candidateCMDPts U Get_centroid(cluster_a, cluster_b, pts)
12:     END FOR
13: END FOR
14: IF ( candidateCMDPt != NULL )
15:     candidateCMDPt ← Get_centroid(candidateCMDPts)
16:     baseArea ← Create_union(baseArea, Create_baseArea(candidateCMDPt, avgRadius, recentWeight))
17: END IF
18: RETURN baseArea
END
    
```

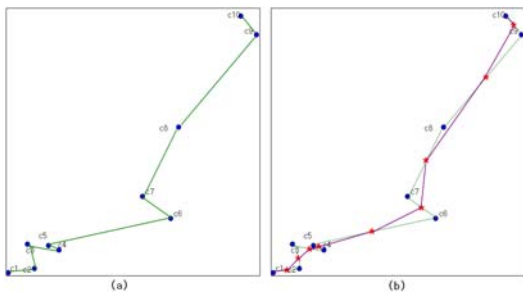
<그림 7> 다중 중심점 패턴 연쇄 범죄의 거점 위치 예측 알고리즘

동 비용, 이동선에서 범행 위치까지의 이동 비용의 총합으로 정의하고 이러한 비용이 최소가 되는 지점을 CMC(Center of Minimum Cost)라 한다. CMC는 다음과 같이 정의된다.

[정의 2] CMC(Center of Minimum Cost)  
 CMC란 모든 범행에 대하여 이동선 진입 비용, 이동선상 이동 비용, 이동선에서 범행 위치까지의 이동 비용의 총합이 최소가 되는 지점이다.

전체 검색 영역에 대해서 CMC를 찾기 위해 우선 후보 CMC들이 선정된다. 검색 영역은 모든 범행 위치들에 대한 최소 x, 최소 y, 최대 x, 최대 y로 표현되는 사각형 영역이 되며 이러한 사각형 영역은 모든 범행 위치들에 대한 최소 경계를 의미한다. 최소 경계 사각형으로 표현되는 검색 영역은 같은 크기를 갖는 셀들로 분할되고, 모든 셀들이 일단 후보 CMC들이 된다.

기존 이동선(초기 이동선 포함)을 구성하는 선분들의 중앙점들을 연결함으로써 새로운 이동선을 계속적으로 생성할 수 있다. 그림 8(a)는 초기 이동선을 보여주고, 그림 8(b)는 초기 이동선에 대해서 새로운 이동선을 생성하는 예를 보여준다.



<그림 8> 이동선 생성의 예

연속된 범죄 위치를 서로 연결한 그림 8(a)의 초기 이동선은 그림 8(b)에서와 같이 초기 이동선을 구성하는 선분들의 중앙점들을 연결한 이동선으로 갱신된다. 이러한 이동선의 갱신은 주어지는

movingLnWeight 파라미터 값에 의해 반복 횟수가 결정된다. 예를 들어 movingLnWeight가 0.8 이라면 범죄 위치 개수 10개 중 8개의 정점을 포함하는 이동선을 생성할 때까지 반복된다.

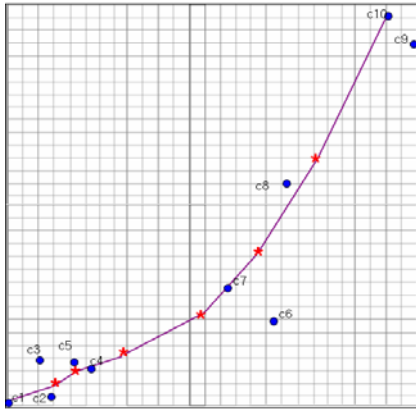
후보 CMC에서 범행 위치까지의 이동 비용은 [수식 2]를 통해 구해진다.

[수식 2] 후보 CMC에서 범행 위치까지의 비용 = 후보 CMC에서 이동선 진입 비용 + 이동선상 이동 비용 + 이동선 진출 후 범행 위치 도달 비용

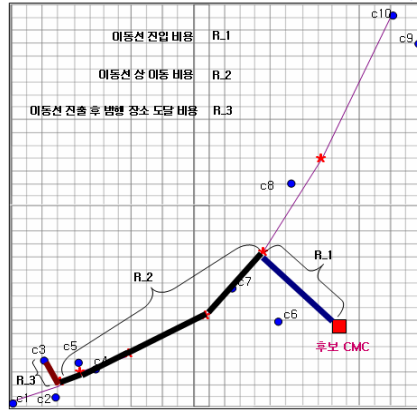
그림 9(a)는 그림 8(a)의 예에 대하여 movingLnWeight가 0.8일 때 생성된 이동선을 보여주고, 그림 9(b)는 그림 9(a)의 예에 대하여 임의의 후보 CMC에서 범행 위치 c3까지의 이동 경로를 보여준다. 이러한 이동 경로의 길이가 곧 후보 CMC에서의 범행 위치 c3에 대한 비용이다. 그림 9(b)에서 보는 바와 같이, 후보 CMC로 부터 범행 위치 c3까지의 이동 경로는 세 부분으로 구분된다. 즉, 이동선 진입을 위한 이동 경로(R\_1), 이동선 상 이동 경로(R\_2), 이동선 진출 후 범행 위치까지의 이동 경로(R\_3)로 나뉜다.

단일 중심축 패턴의 연쇄 범죄에 대한 거점 영역을 구하기 위해서는 우선 각 후보 CMC에 대해서 모든 범행 위치들까지의 이동 비용을 합한 총비용을 계산한 후, 이것이 최소인 후보 CMC 혹은 후보 CMC들을 선택한다. 따라서, 이 예에서 CMC를 구하기 위해서는 [수식 2]를 기반으로 후보 CMC에서 모든 범행 위치들(c1~c10)까지의 총비용을 계산해야 한다.

합리적 선택 이론(Rational Choice Approach) [17]에 따르면, 범죄자는 자신의 경험이나 학습한 지식을 기초 자료로 활용하여 범죄 장소를 합리적으로 선택한다. 즉, 범죄자는 범행시 이동했던 경로에 대하여 학습이 가능하며 학습된 경로를 보다 쉽게 다시 선택할 수 있음을 알 수 있다. 예를 들어, 그림 9(b)에서 c1 이후 c3에서의 범행시에는 이미 학습된 경로를 따라 이동하므로 동일한 경로라고



(a) 단일 중심축 패턴 연쇄 범죄의 이동선 예



(b) 후보 CMC로 부터 범행 위치 c3까지의 이동 경로 예

<그림 9> 이동선 기반의 범행 경로 예

하더라도 c1에서의 비용만큼 들지 않는다. 이러한 학습 효과는 범죄자의 지능이나 해당 경로에 대한 반복 횟수에 따라 달라질 수 있다. 따라서, 이동선상의 동일 경로에 대해 반복 이동이 있는 경우 진입점과 진출점간 유클리디언 거리에 대하여 해당 경로의 누적 방문 횟수와 범죄자 학습률을 고려한 가중치를 적용할 수 있다. 이동선상 이동 비용은 [수식 3]을 통해 구해진다.

[수식 3] 이동선상 이동 비용 =  
 진입점과 진출점의 거리 - {진입점과 진출점의 거리 \* (해당 진출점에 대한 누적 방문 횟수/전체 진출점 개수) \* 학습률}

그림 10은 단일 중심축 패턴 연쇄 범죄의 거점 예측을 위한 알고리즘을 보여준다. 알고리즘의 입력으로는 예측 대상인 범죄 데이터(pts), 이동선 생성을 위한 가중치(movingLnWeight), CMC를 검색하기 위한 범죄자 학습률 가중치(learning-Weight), 거점 영역 생성을 위한 결과 영역 비율(topRate)을 가진다. 그리고 출력으로는 예측된 거점 영역(baseArea)을 반환한다.

이 알고리즘에서는 모든 범행 위치들로부터 이동선을 생성하고 후보 CMC들을 추출한다. 그리고, 전체 영역의 topRate 비율에 해당하는 최소 비용의 후보 CMC들을 찾아서 거점 영역을 반환한다.

```

ALGORITHM : STABLP_oneCenterMovingLine( ST_MULTIPPOINT pts, DOUBLE movingLnWeight,
                                           DOUBLE learningWeight, DOUBLE topRate )
    Search and return the base location from crime sites on their center linestring.
BEGIN
    ST_MULTILINESTRING movingLn ← ∅
    ST_POINT[] candidateCMCs ← ∅
    S_GEOMETRY baseArea ← ∅
1: movingLn ← Get_movingLine(pts, movingLnWeight) //이동선 생성
2: candidateCMCs ← Get_candidateCMCs(movingLn, pts) //후보 CMC 결정
3: baseArea ← Create_baseArea(candidateCMCs, movingLn, learningWeight, topRate) //CMC인 지점을 검색하여 거점 영역 생성
4: RETURN baseArea
END
    
```

<그림 10> 단일 중심축 패턴 연쇄 범죄의 거점 위치 예측 알고리즘

#### 4.5 다중 중심축 패턴 연쇄 범죄의 거점 위치 예측

본 논문에서는 다중 중심축 패턴의 연쇄 범죄에 대한 범죄자 거점 위치 예측을 위해서 군집 단위의 지역적 거점 위치 예측 기법과 모든 범행 위치에 대하여 군집화 특성에 따라 결정하는 전역적 거점 위치 예측 기법을 제시한다. 따라서, 다중 중심축 패턴 연쇄 범죄의 거점 예측은 각 거점이 각 군집의 중심 역할을 하는 경우와 한 거점이 모든 다수의 활성 군집의 중심 역할을 하는 경우 모두 적용이 가능하다.

지역적 거점 위치 예측의 경우에는 각 활성 군집에 소속된 범행 위치들에 대하여 단일 중심축 패턴 연쇄 범죄의 거점 위치 예측 알고리즘을 적용한다. 즉, 각 활성 군집에 소속된 범행 위치들로부터 생성된 군집별 이동선을 기반으로 각 후보 CMC에서 그 군집내 범행 위치들까지의 총비용을 계산한다. 전역적 거점 위치 예측의 경우에는 모든 활성 군집에 소속된 범행 위치들에 대하여 단일 중심축 패턴 연쇄 범죄의 거점 위치 예측 알고리즘을 적용한다.

즉, 모든 활성 군집들에 소속된 모든 범행 위치들로부터 생성된 이동선을 기반으로 각 후보 CMC에서 모든 활성 군집내 범행 위치들까지의 총비용을 계산한다.

그림 11은 다중 중심축 패턴 연쇄 범죄의 거점 예측을 위한 알고리즘을 보여준다. 알고리즘의 입력으로는 예측 대상인 범죄 데이터(pts), 이동선 생성을 위한 가중치(movingLnWeight), CMC를 검색하기 위한 범죄자 학습률 가중치(learningWeight), 거점 영역 생성을 위한 결과 영역 비율(topRate)을 가진다. 그리고 출력으로는 예측된 거점 영역(baseArea)을 반환한다.

이 알고리즘에서는 각 활성 군집에 대하여 지역적 거점 영역을 생성한 후, 모든 활성 군집에 대한 전역적 거점 영역을 생성한다. 지역적 거점 영역 생성을 위해서는 각 활성 군집마다 이동선을 생성하고, 이를 기반으로 후보 CMC들을 구하여 각각의 거점 영역을 생성한다. 전역적 거점 영역 생성을 위해서는 모든 활성 군집을 위한 이동선을 생성하고, 생성된 이동선을 기반으로 후보 CMC들을 구하여 거점 영역을 생성한다.

```

ALGORITHM : STA.BLP_multiCenterMovingLine(ST.MULTIPOINT pts, DOUBLE movingLnWeight,
                                             DOUBLE learningWeight, DOUBLE topRate )
Search and return the base location from clustered crime sites on their center linestrings.
BEGIN
  ST.MULTILINESTRING movingLn ← ∅
  ST.MULTIPOINT activePts ← ∅
  ST.POINT[] candidateCMCs ← ∅
  S_GEOOMETRY baseArea ← ∅
  INT clusterNums ← 1
  INT[] activeClusters ← ∅
1: clusterNums ← pts.count
2: activeClusters ← Analyze_connectivity(clusterNums, pts) //활성군집 분석
3: clusterNums ← activeClusters.count
4: FOR EACH cluster clust of activeClusters //각 활성군집에 대한 지역적 거점 위치 예측
5:   movingLn ← Get_movingLine(clust, pts, movingLnWeight)
6:   candidateCMCs ← Get_candidateCMCs(movingLn, pts)
7:   baseArea ← Create_union(baseArea, Create_baseArea(candidateCMCs, movingLn, learningWeight, topRate, clusterNums))
  END FOR
8: IF ( clusterNums >= 2 ) //모든 활성군집에 대한 전역적 거점 위치 예측
9:   FOR EACH cluster clust of activeClusters
10:    activePts ← activePts U Get_points(clust, pts)
  END FOR
11: movingLn ← Get_movingLine(activePts, movingLnWeight)
12: candidateCMCs ← Get_candidateCMCs(movingLn, pts)
13: baseArea ← Create_union(baseArea, Create_baseArea(candidateCMCs, movingLn, learningWeight, topRate))
  END IF
14: RETURN baseArea
END
    
```

<그림 11> 다중 중심축 패턴 연쇄 범죄의 거점 위치 예측 알고리즘

### 5. 실험 및 평가

본 장에서는 실험에서 사용된 연쇄 범죄 데이터와 평가 지표에 대하여 언급하고, 다양한 실험을 통하여 기존에 제시된 알고리즘들과 본 논문에서 제시한 알고리즘의 정확도를 비교 설명한다. 먼저, 단일 중심축 패턴 연쇄 범죄와 다중 중심축 패턴 연쇄 범죄의 경우 예측 결과에 큰 영향을 끼치게 되는 이동선의 중앙점 조정 비율과 범죄자의 동일 경로 학습률에 따라 거점 위치 예측 오차를 분석함으로써 STA-BLP 알고리즘의 예측 정확도를 확인한다. 또한, 다양한 시공간 패턴을 보이는 연쇄 범죄 데이터에 대한 실험을 통하여 기존에 제시된 알고리즘들과 STA-BLP 알고리즘에 따른 거점 위치 예측 오차를 분석함으로써 알고리즘 간의 예측 정확도를 비교한다.

#### 5.1 실험 환경

본 논문에서 실험은 6번 이상 범행을 저지른 16건의 연쇄 범죄 데이터를 이용하여 실제 거점 위치에 대한 예측 정확도를 비교하였다. 실험에 사용된 연쇄 범죄 데이터는 살인, 강간, 강도 등 다양한 유형의 범죄 데이터이며 따라서 다양한 공간적 분포 특성을 갖는다. 각 범행의 발생 시간 및 발생 장소를 포함하는 실험 데이터는 상대 좌표로 먼저 변환되었고, 또한 발생 순서에 따라 순서화되었다.

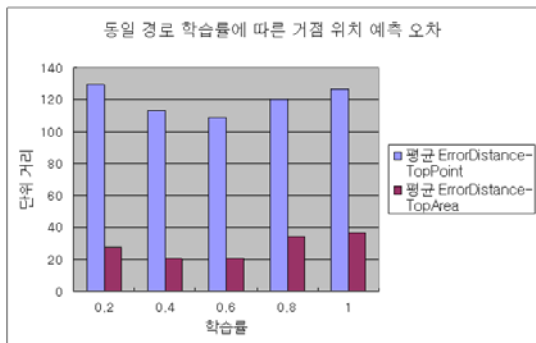
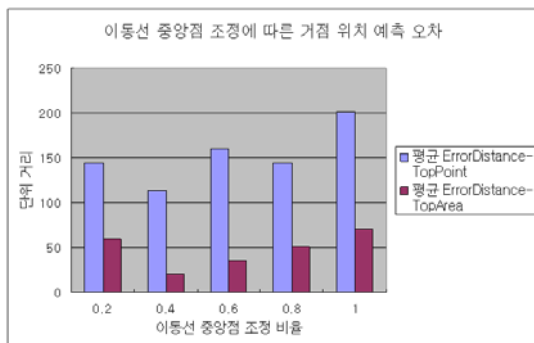
본 논문은 시공간 분석 기반 연쇄 범죄 거점 위

지 예측 알고리즘의 실험 평가를 위해서 두가지 평가 지표를 정의한다. 첫 번째로, ErrorDistance-TopPoint는 한 점 혹은 다수 점으로 표현되는 예측 거점 위치와 실제 거점 위치와의 최소 거리로써 예측 결과에 대한 정량적 오차 분석을 위해 사용된다. 여기서 TopPoint는 거점으로써 가장 확률이 높다고 판단되는 한 점 혹은 다수 점들을 의미한다. 두 번째로, ErrorDistance-TopArea는 영역으로 표현되는 예측 거점과 실제 거점간의 최소 거리를 의미하는 또 다른 지표이다. 여기서 TopArea는 거점으로써 확률이 높다고 판단되는 영역을 의미한다. 본 실험에서 ErrorDistance-TopArea는 모든 CMC 후보 지점들 중에서 비용이 적은 0.01% CMC에 대하여 실제 거점과의 최소 거리를 계산한 것이다.

#### 5.2 예측 정확도 비교

그림 12는 동일 경로 학습률을 0.4로 고정하고 중앙점 조정 비율을 늘려가면서 STA-BLP의 평균 ErrorDistance-TopPoint 및 평균 ErrorDistance-TopArea를 분석한 것이다. 여기서 중앙점 조정 비율은 이동선 생성 가중치 movingLnWeight 값이다.

그림 12의 결과를 통해 평균 ErrorDistance-TopPoint 값과 평균 ErrorDistance-TopArea 값을 기준으로 이동선 조정 비율이 1일 때, 즉 중앙점 조정 없이 범행 위치 점들을 이용하여 거점 위



<그림 12> 이동선 중앙점 조정에 따른 거점 위치 예측 오차 <그림 13> 동일 경로 학습률에 따른 거점 위치 예측 오차

치 예측을 수행하는 경우에 예측 오차가 가장 크며, 0.4에서 예측 오차가 가장 작은 것을 알 수 있다. 이는 중앙점 조정 횟수가 늘수록 이동선상의 동일 경로를 공유하는 횟수도 늘기 때문에 중심축 기반의 범행이 많을수록 이동선상 이동을 기반으로 하는 CMC가 거점과 더욱 가까워지기 때문이다. 또한 중앙점 조정 비율이 0.2인 경우 중앙점 조정이 지나치게 많이 이루어져 오히려 정확도가 낮아지는 것을 알 수 있다.

그림 13은 중앙점 조정 비율을 0.4로 고정하고 동일 경로 반복에 대한 학습률을 늘려가면서 STA-BLP의 평균 ErrorDistance-TopPoint와 평균 ErrorDistance-TopArea를 분석한 것이다.

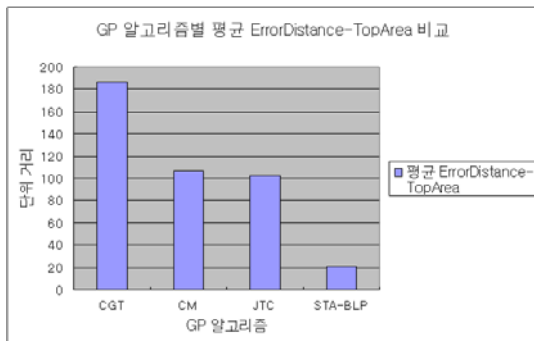
그림 13의 결과를 통해 평균 ErrorDistance-TopPoint값은 학습률이 0.6일 때 가장 작고, 평균 ErrorDistance-TopArea값은 학습률이 0.4 혹은 0.6일 때 가장 작다는 것을 알 수 있다. 즉, 본 실험 데이터에서는 평균적으로 동일 경로에 대한 0.4에서 0.6 사이의 학습률을 가진 범죄자들로 구성되어 있다는 것을 알 수 있다.

그림 14는 기존의 지리적 프로파일링 알고리즘인 Criminal Geographic Targeting(CGT), Canter Model(CM), Journey-to-Crime(JTC), CMD와 STA-BLP 알고리즘의 평균 ErrorDistance-TopPoint 및 평균 ErrorDistance-TopArea를 비교한 것을 보여준다. Criminal Geographic Targeting, Canter Model, Journey-to-Crime

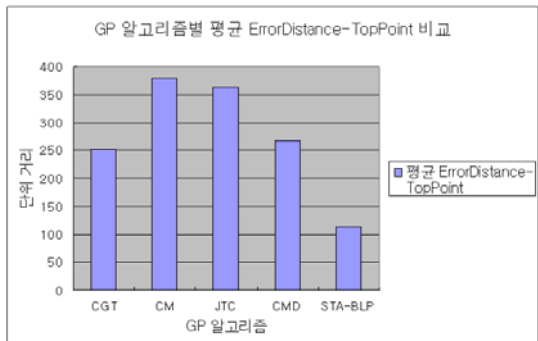
에서는 거리 감쇠 함수의 입력 파라미터들에 대해 기본값을 적용하였고, STA-BLP에서는 중앙점 조정 비율 0.4, 동일 경로 학습률 0.4를 적용하였다. 평균 ErrorDistance-TopArea 값을 기준으로 STA-BLP는 Canter Model 및 Journey-to-Crime에 비해 5배 정도 높은 정확도를 보이며, Criminal Geographic Targeting에 비해 9배 이상 높은 정확도를 보인다. 이는 거점이 범행 발생 위치와 공간상 크게 떨어져 있거나 거점이 범행 발생 위치들의 중앙에 위치하지 않을 경우나 거점이 2개인 경우에도 기존 지리적 프로파일링 방법들은 단순히 거리를 기반으로 모든 범행들의 무게 중심, 즉 분포 상 밀도가 가장 높은 지점을 거점으로써 높은 확률을 주기 때문이다.

### 6. 결론

최근 강력 범죄 및 지능 범죄에 대한 보다 효과적이고 신속한 수사 대응의 필요성이 증대되고 있다. 특히, 연쇄 살인, 연쇄 강간과 같은 강력 범죄의 경우 범죄 발생 장소들의 공간적 분포에 대한 패턴을 활용하여 연쇄 범죄의 거점을 예측하는 연구가 활발히 진행되었다. 그러나 기존의 연구는 공간적 패턴 분석을 위해 단순히 통계적 방법만을 적용하고 있고, 실제 환경에서의 연쇄 범죄 분포에 대한 다양한 공간적, 시간적 분석을 지원하지 않으며, 또한 다양한 공간 패턴이 존재함에도 불구하고 모든



(a) 평균 ErrorDistance-TopArea 비교



(b) 평균 ErrorDistance-TopPoint 비교

<그림 14> 알고리즘별 예측 오차 비교

경우에 대하여 동일한 방식으로 분석함으로써 거점 예측 결과에 있어 낮은 정확도를 보이고 있다.

이러한 문제점을 해결하기 위해 본 논문에서는 범죄 발생 장소에 대한 공간적 분포와 범죄 발생 시각에 대한 시간적 분포 특성에 따라 연쇄 범죄의 시공간 패턴을 유형화하고, 연쇄 범죄의 거점 위치를 보다 정확하고 효율적으로 예측하는 시공간 분석 기반 연쇄 범죄 위치 예측 알고리즘을 제안하였다. 본 논문에서 제시한 시공간 분석 기반 연쇄 범죄 거점 위치 예측 알고리즘의 특징은 다음과 같다. 첫째, 원형 분포의 연쇄 범죄뿐 아니라 중심축을 따라 나타나는 선형 분포의 연쇄 범죄에서도 정확도 높은 거점 예측이 가능하다. 둘째, 범죄 발생 위치들을 다수개의 서로 다른 군집들로 분류한 후, 각 군집 내 범행에 대한 지역적 거점 예측과 검색 영역의 모든 범행에 대한 전역적 거점 예측이 가능하다. 셋째, 모든 거점 예측 시 범죄들의 발생 공간뿐만 아니라 발생 시간도 활용하여 예측 정확도를 개선하였다.

마지막으로 제안 알고리즘을 구현하여 다양한 실험을 통해 정확도를 평가함으로써 그 우수성을 입증하였는데, 거점 위치 예측에 있어서 기존의 연구들에 비하여 평균 6배 정도 높은 정확도를 보였다. 향후에는 GIS에서 일반적으로 다루는 도로, 강 등의 공간 정보를 고려하고 추가적인 속성들의 연계를 통해 예측 정확도를 향상시키는 연구가 필요하다.

## 참고문헌

1. Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M. J., Maceachren, A., and Wrobel, S., "Geovisual Analytics for Spatial Decision Support: Setting the Research Agenda," *International Journal of Geographical Information Science*, Vol.21 No.8, 2007, pp.839-857.
2. Balram, S., and Dragičević, S., *Collaborative Geographic Information Systems*, IGI Global, Hershey, 2006.
3. Dykes, J. A. and Mountain, D. M., "Seeking Structure in Records of Spatio-Temporal Behaviour: Visualization Issues, Efforts and Applications," *Computational Statistics and Data Analysis*, Vol.43, 2003, pp.581-603.
4. Jennifer J. X., and Hsinchun C., "Fighting Organized Crimes: Using Shortest-Path Algorithms to Identify Associations in Criminal Networks," *Decision Support Systems*, Vol.38, 2004, pp.473-487.
5. Keppens, J., and Zeleznikow, J., "On the Role of Model-based Reasoning in Decision Support in Crime Investigation," *Proc. of the 3rd International Conference on Law and Technology*, 2002, pp.77-83.
6. 전상덕, 홍동숙, 한기준, "디지털 포렌식 기술 동향과 전망," 한국전산원 정보화 정책, 제13권 제4호, 2006, pp.3-19.
7. Environmental Criminology Research Inc., *Rigel: Geographic Profiling Program*, <http://www.ecricanada.com/rigel>, 2006.
8. Laukkanen, M., and Santtila, P., "Predicting the Residential Location of a Serial Commercial Robber," *Forensic Science International*, Vol.157, 2005, pp.71-82.
9. Levine, N., "CrimeStat II," *Crime Mapping News*, Vol.5 No.2, 2003, pp.2-4.
10. Ashby, D. I., and Longley, P. A., "Geocomputation, Geodemographics and Resource Allocation for Local Policing," *Transactions in GIS*, Vol.9 No.1, 2005, pp.53-72.
11. Canter, D., "Predicting Serial Killers' Home Base using a Decision Support System," *Journal of Quantitative Criminology*,



- Vol.16 No.4, 2000, pp.457-478.
12. Getis, A., Drummy, P., Gartin, J., Gorr, W., Harries, K., Rogerson, P., Stoe, D., and Wright, R., "Geographic Information Science and Crime Analysis," URISA Journal, Vol.12 No.2, 2000, pp.7-14.
  13. Canter, D., *Mapping Murder: The Secrets of Geographical Profiling*, Virgin Books, New York, 2003.
  14. 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](http://webapp.icpsr.umich.edu/cgi-bin/bob/newark?study=2824&path=ICPSR), 2005.
  15. Rossmo, K., *Geographic Profiling*, CRCPress, Boca Raton, 2000.
  16. Kocsis, R. N., *Criminal Profiling: International Theory, Research, and Practice*, Humana Press, Totowa, 2007.
  17. Beauregard, E., Proulx, J., and Rossmo, K., "Spatial Patterns of Sex Offenders: Theoretical, Empirical, and Practical Issues," *Aggression and Violent Behavior*, Vol.10, 2005, pp.579-603.

**홍동숙**

1999년 건국대학교 컴퓨터공학과(공학사)  
 2001년 건국대학교 대학원 컴퓨터공학과(공학석사)  
 2000년~2003년 쌍용정보통신 모바일GIS 기술팀  
 2008년 건국대학교 대학원 컴퓨터공학과(공학박사)  
 2008년~현재 연세대학교 TMS 정보기술사업단 박사  
 후연구원  
 관심분야 : 데이터베이스, LBS, 시공간 마이닝, 디지털 포렌식, 차세대 방송  
 e-mail : dongsookh@yonsei.ac.kr

**김정준**

2003년 건국대학교 컴퓨터공학과(공학사)  
 2005년 건국대학교 대학원 컴퓨터공학과(공학석사)  
 2005년~현재 건국대학교 대학원 컴퓨터공학과 박사  
 과정  
 관심분야 : 공간 메인 메모리 데이터베이스, GIS, LBS, 텔레매틱스, USN  
 e-mail : jjkim9@db.konkuk.ac.kr

**강홍구**

2002년 건국대학교 컴퓨터공학과(공학사)  
 2004년 건국대학교 대학원 컴퓨터공학과(공학석사)  
 2004년~현재 건국대학교 대학원 컴퓨터공학과 박사  
 과정  
 관심분야 : 공간 데이터베이스, GIS, LBS, USN, 센서 데이터베이스  
 e-mail : hkkang@db.konkuk.ac.kr

**이기영**

1984년 송실대학교 전자계산학과(공학사)  
 1988년 건국대학교 대학원 컴퓨터공학과(공학석사)  
 2005년 건국대학교 대학원 컴퓨터공학과(공학박사)  
 1984년~1991년 한국해양연구원 연구원  
 1996년~1998년 한국컴퓨터정보학회 이사 및 서울동부지회장  
 1991년~현재 을지대학교 의료산업학부 교수  
 관심분야 : 공간 데이터베이스, GIS, LBS, USN, 텔레매틱스  
 e-mail : kylee@eulji.ac.kr

**서종수**

1975년 연세대학교 전자공학과 졸업(공학사)  
 1983년 Univ. of Ottawa, Canada 전기공학과(공학석사)  
 1998년 Univ. of Ottawa, Canada 전기공학과(공학박사)  
 1975년~1981년 LG정밀 중앙 연구소  
 1987년~1989년 IDC, Canada 책임연구원  
 1990년~1992년 삼성종합기술원 정보 시스템 연구소 수석연구원  
 1992년~1995년 CAL, Canada 책임 연구원  
 1995년~현재 연세대학교 전기전자공학과 교수  
 관심분야 : 디지털 통신/방송 시스템, Resource allocation, CDMA, OFDM  
 e-mail : jsseo@yonsei.ac.kr

**한기준**

1979년 서울대학교 수학교육학과(이학사)  
 1981년 한국과학기술원(KAIST) 전산학과(공학석사)  
 1985년 한국과학기술원(KAIST) 전산학과(공학박사)  
 1985년~현재 건국대학교 컴퓨터공학부 교수  
 1990년 Stanford 대학 전산학과 Visiting Scholar  
 2000년~2002년 한국정보과학회 데이터베이스연구회 운영위원장  
 2004년~2006년 한국공간정보시스템학회 회장  
 2004년~현재 한국정보시스템감리사협회 회장  
 관심분야 : 공간 데이터베이스, GIS, LBS, 텔레매틱스, 정보시스템 감리  
 e-mail : kjhan@db.konkuk.ac.kr