creative commons
COMMONS DEED

저작자표시-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학석사 학위논문

상업적 테마거리를 탐지하기 위한 핫 스트리트 군집화 기법 연구
: LBSNS 데이터를 활용하여

A Hot Street Clustering Method for Detection of Commercial Themed Street
: Using LBSNS Data

2014 년 2 월

서울대학교 대학원

건설환경공학부

지 병 석

상업적 테마거리를 탐지하기 위한 핫 스트리트 군집화 기법 연구
: LBSNS 데이터를 활용하여

A Hot Street Clustering Method for Detection of Commercial Themed Street
: Using LBSNS Data

지도교수 유 기 윤

이 논문을 공학석사 학위논문으로 제출함

2013 년 12 월

서울대학교 대학원

건설환경공학부

지 병 석

지병석의 공학석사 학위논문을 인준함

2013 년 12 월

위 원 장 _____

부위원장 _____

위 원 _____

국 문 초 록

최근 지역경제를 활성화시키고, 그 지역의 정체성과 장소성(placeness)을 확립시키기 위해 다양한 테마거리(themed street)를 조성하는 사례가 많아지고 있다. 하지만 사람들의 인식 속에 자연스럽게 형성된 그 지역의 특성을 충분히 고려하지 않고 경제적 측면만을 생각하여 무분별하게 진행한 지방자치단체의 테마거리 조성사업들은 그 효과가 미미한 경우도 발생하고 있다. 지방자치단체에서 조성한 테마거리 이외에 사람들의 자연스러운 활동으로 인해서 테마거리가 자생적으로 형성될 수 있고, 급변하는 도시환경에서 아직 알려지지 않은 테마거리가 있을 수 있다. 그리고 번화가 내부에서도 지역적 분화에 따라 업종별 테마거리가 형성되어 있을 수 있다. 이러한 테마거리를 객관적으로 탐지한다면 사람들에게 더욱 풍부한 사회·문화적인 공간에 관한 정보를 제공하고 지역경제의 활성화를 기대할 수 있을 것이다.

따라서 본 연구는 테마거리 유형 중에서도 높은 비중을 차지하고 있는 상업적 테마거리를 탐지하기 위한 핫 스트리트 군집화 기법(hot street clustering method)을 제안하였다. 특히 본 연구는 도로 중심적인 도시환경을 반영하고 테마거리의 영역을 직관적으로 표현하기 위해 도로를 기본 단위로 하여 분석 및 결과를 도출하였다. 그리고 활성화된 테마거리를 탐지하기 위해 체크인 기능 등을 통해 장소에 대한 사람들의 인식과 인기도를 알 수 있는 LBSNS(Location Based Social Network Service) 데이터를 활용하였다.

테마거리를 탐지하기 위해서 본 연구는 공간조인 연산(spatial join)과 ISOVIST 등을 활용하여 LBSNS 데이터를 통해 알 수 있는 업소의 인기도 정보를 업소가 위치한 건물, 그리고 건물이 영향을 미치는 도로에 순차적으로 연결시켰다. 각 도로마다 연결된 인기도 정보와 도로의 길이, 주변 건물 수 등을 종합하여 핫 스트리트(hot street)로 판단될 고유값(핫 밸류, hot value)을 계산하고, 공간 클러스터 기법인 Getis-Ord의 G_i^*를 활용하여 통계적으로 유의한 핫 스트리트 군집(hot street cluster)을 탐지하였다. 이렇게 대

상지역에서 탐지된 특정 테마의 핫 스트리트 군집은 그 지역의 테마거리로 선정한다.

본 연구에서는 신림 사거리, 강남역 번화가 두 지역을 대상으로 핫 스트리트 군집화 기법의 실험을 수행하였다. 실험 결과 본 기법을 통해 음식점 거리로 변질된 신림 패션문화의 거리와 최근에 형성된 강남 언덕길 카페거리 등을 탐지할 수 있었으며, 기존에 조사된 각 지역의 상권영역과 탐지된 상업적 테마거리의 영역을 비교·분석하였다.

주요용어 : 테마거리 탐지, 핫 스트리트(hot street), LBSNS(Location Based Social Network Service), Getis-Ord의 G_i^*, ISOVIST

학번 : 2012-20923

< 목 차 >

초 록 ····· i
목 차 ····· iii
그림 차례 ····· v
표 차례 ····· vii

1. 서론 ····· 1
1.1 연구 배경 및 목적 ····· 1
1.2 연구 방법 및 범위 ····· 4
1.3 관련 연구 ····· 6
1.4 용어 정리 ····· 15

2. 핫 스트리트 군집화 기법 ····· 17
2.1 활용 데이터 수집 ····· 18
2.1.1 LBSNS 데이터(포스퀘어 베뉴 데이터) 수집 ····· 18
2.1.2 LBSNS 데이터의 테마 별 카테고리 분류 ····· 20
2.1.3 도로명 주소 기본도 데이터 활용 ····· 21
2.2 LBSNS 데이터와 건물 데이터 간의 공간조인 ····· 22
2.2.1 LBSNS 데이터 위치보정 ····· 22
2.2.2 건물-베뉴 데이터의 공간조인 연산 ····· 29
2.3 건물과 도로 세그멘트간의 매칭 ····· 31
2.3.1 도로 세그멘트 생성 ····· 32
2.3.2 건물-도로 세그멘트의 매칭 조건 및 기법 ····· 34
2.3.3 ISOVIST를 이용한 건물-도로 세그멘트 매칭 과정 ····· 41
2.4 핫 스트리트 군집 탐지 ····· 47
2.4.1 도로 세그멘트의 핫 밸류(hot value) 계산 ····· 47
2.4.2 Getis-Ord의 G_i^*를 이용한 핫 스트리트 군집 탐지 ····· 52

3. 실험 결과 및 분석 ····· 59
3.1 실험 결과(1) - 신림 사거리 ····· 59
3.1.1 음식점 테마 핫 스트리트 군집 ····· 62
3.1.2 카페 테마 핫 스트리트 군집 ····· 64
3.1.3 패션 테마 핫 스트리트 군집 ····· 65
3.1.4 유흥주점 테마 핫 스트리트 군집 ····· 66
3.1.5 상권보고서와 비교 분석 ····· 66
3.2 실험 결과(2) - 강남역 번화가 ····· 67
3.2.1 음식점 테마 핫 스트리트 군집 ····· 71
3.2.2 카페 테마 핫 스트리트 군집 ····· 72
3.2.3 패션 테마 핫 스트리트 군집 ····· 73
3.2.4 유흥주점 테마 핫 스트리트 군집 ····· 75
3.2.5 상권보고서와 비교 분석 ····· 76

4 결론 ····· 78

참고문헌 ····· 81
부록 ····· 85
Abstract ····· 92

< 그림 차례 >

그림 1-1. 도로 중심으로 표현한 테마거리의 지도 예시 ····· 3
그림 1-2. LBSNS 데이터를 이용한 도로기반의 핫 스트리트 군집화 기법 프로세스 ····· 5
그림 1-3. 서초구 상권 추출 결과 ····· 8
그림 1-4. 미국 뉴욕 지역의 Livehoods 결과 화면 ····· 10
그림 1-5. 중국 우한 지역의 체크인 핫 스팟 지도 ····· 11
그림 1-6. 범죄 발생 점 데이터에 STAC과 NT-STAC을 적용하여 핫 스팟을 탐지한 결과 ····· 12
그림 1-7. 미국 버팔로 지역의 차량 도난이 집중적으로 발생한 핫 스트리트 탐지 결과 ····· 14
그림 2-1. LBSNS 데이터를 이용한 도로기반의 핫 스트리트 군집화 기법 흐름도 ··· 17
그림 2-2. LBSNS 앱 선호도 조사 ····· 18
그림 2-3. 포스퀘어 서비스의 연령별 사용자 분포 ····· 19
그림 2-4. 신림역 사거리 지역의 포스퀘어 메뉴 데이터 분포 현황 ····· 23
그림 2-5. 포스퀘어 데이터의 유클리드 거리 오차 히스토그램 ····· 24
그림 2-6. 포스퀘어 데이터 오차 벡터의 방위각 히스토그램 ····· 25
그림 2-7. 상세한 수준의 주소 정보가 등록되어 있는 포스퀘어의 메뉴 예시 ····· 27
그림 2-8. 신림 사거리 지역의 음식점 메뉴 데이터를 위치보정 적용한 결과 ····· 29
그림 2-9. 건물 내부에 위치하지 않은 메뉴 점 좌표 예시 1 ····· 31
그림 2-10. 건물 내부에 위치하지 않은 메뉴 점 좌표 예시 2 ····· 31
그림 2-11. 신림 사거리 지역의 도로구간 예시 ····· 33
그림 2-12. 교차로를 기준으로 도로구간을 8개의 도로 세그멘트로 분절시킨 예시 ··· 33
그림 2-13. 건물과 도로가 맞닿아 있는 상태를 설명하는 예시 ····· 35
그림 2-14. 도로명 주소 기본도의 출입구와 연결선 레이어 적용 예시 ····· 36
그림 2-15. 출입구 데이터가 없는 다른 건물면에 존재하는 실제 상점들과 출입구 ··· 37
그림 2-16. 도로 중심의 버퍼 연산시 발생되는 매칭 오류 ····· 38

그림 2-17. 건물 중심의 버퍼 연산 수행 시 발생되는 매칭 오류 ····· 39
그림 2-18. ISOVIST 공간 예시 ····· 40
그림 2-19. ISOVIST를 이용한 건물-도로 세그멘트 매칭 예시 ····· 41
그림 2-20. ISOVIST를 이용한 건물-도로 세그멘트 매칭 프로세스 ····· 43
그림 2-21 *view_dist* 설명 ····· 44
그림 2-22. 건물 중심의 ISOVIST 공간 예시 ····· 46
그림 2-23. 베뉴 건물들의 구성비가 같지만 밀집도가 다른 두 도로 세그멘트 ···· 50
그림 2-24. 교차로에서 연결된 9개의 도로 세그멘트 예시 ····· 54
그림 2-25. 신림 사거리 지역 152개 도로 세그멘트의 음식점 핫 밸류 분포 ····· 56
그림 2-26. [그림 2-24]을 로그 변환한 분포 ····· 57
그림 2-27. [그림 2-25]의 $\ln(H_i)$ 분포를 정규분포와 Q-Q plot을 수행한 결과 ·· 57
그림 3-1. 신림 사거리 실험 지역 범위 ····· 60
그림 3-2. 신림 사거리 지역의 핫 스트리트 군집 탐지 결과 ····· 61
그림 3-3. 신림동 순대타운 위치 ····· 63
그림 3-4. 관악구에서 조성한 신림동 “패션문화의 거리” ····· 64
그림 3-5. B군집 아래에 위치한 카페 밀집 지역 ····· 65
그림 3-6. 상권보고서의 상권 영역과 핫 스트리트 군집 영역과의 비교 ····· 66
그림 3-7. 강남역 번화가 실험 지역 ····· 68
그림 3-8. 강남역 번화가의 핫 스트리트 군집 탐지 결과 ····· 70
그림 3-9. 인기도가 상당히 높은 소수의 업소에 의해 핫 스트리트가 탐지된 그림 ··· 72
그림 3-10. 강남역 번화가에 위치하고 있는 명소 “언덕길” ····· 73
그림 3-11. 강남대로변을 따라 위치한 패션 매장들의 모습 ····· 74
그림 3-12. 실제 패션 매장의 출입구 방향을 반영하지 못하고 핫 스트리트 군집이 탐지된 그림 ····· 75
그림 3-13. 상권보고서의 상권 영역과 핫 스트리트 군집 영역과의 비교 ····· 76
그림 부록 F. 신림 사거리 지역의 각 테마 별 $\ln(H_i)$의 정규 Q-Q plot 결과 ····· 90
그림 부록 G. 강남역 번화가 지역의 각 테마 별 $\ln(H_i)$의 정규 Q-Q plot 결과 ··· 91

< 표 차례 >

표 1-1. 특화거리에 관한 최근 연구동향 ····· 7
표 2-1. 포스퀘어 API 활용 정보 목록 ····· 19
표 2-2. 테마 & 지역별 정확한 주소 정보를 획득한 베뉴의 비율 ····· 28
표 2-3. [그림 2-22]의 공간 가중 행렬 생성 예시 ····· 55
표 3-1. 신림 사거리 실험 지역의 테마 별 위치 보정된 베뉴 수, 베뉴 건물 수, 베뉴 건물과 매칭된 도로 세그멘트 수 비교 ····· 60
표 3-2. 강남역 번화가 실험 지역의 테마 별 위치 보정된 베뉴 수, 베뉴 건물 수, 베뉴 건물과 매칭된 도로 세그멘트 수 비교 ····· 69
표 부록 A. 테마 별 포스퀘어 카테고리 목록 ····· 85
표 부록 B. 도로명 주소 기본도 건물 레이어 주요 속성 ····· 86
표 부록 C. 도로명 주소 기본도 도로구간 레이어 중 속성 ····· 88
표 부록 D. 도로명 주소 기본도 출입구 레이어 주요 속성 ····· 89
표 부록 E. 도로명 주소 기본도 연결선 레이어 주요 속성 ····· 89

1. 서론

1.1 연구 배경 및 목적

최근 정체되어 있는 지역경제를 활성화시키고, 그 지역의 정체성과 장소성을 확립시키고자 다양한 테마거리(themed street)를 조성하는 사례가 많아지고 있다. 특히 상업적 성격이 강한 테마거리는 동일 혹은 유사한 테마의 업종들이 집적의 이익을 창출하기 위해 일정한 장소에 군집을 이루어 형성된 특화거리와 두 개 이상의 단일테마가 복합되어 형성된 번화가 지역으로 구분된다(김경수, 2011). 예를 들어, 서울시에는 강남역 번화가, 건대 먹자골목, 서래마을 까페거리 등 상업적 성격의 테마거리들이 많이 알려져 있고 해당지역들은 인터넷 검색이나 웹 지도 서비스 등을 통해 쉽게 확인해 볼 수 있다. 한범수(2007)의 서울시 테마거리 자원 기초 조사 결과에 따르면 서울시의 141개의 거리 표본 중 상업 활동 유형의 성격을 지닌 거리가 3분의 1정도(57개)의 높은 비중을 차지하며 테마거리의 핵심적인 테마 유형임을 설명한 바 있다. 이러한 상업적 성격의 테마거리는 단순한 기능적 물품구매의 장소를 넘어 커뮤니티 공유의 장과 일상적 생활을 즐기는 장소로서의 환경과 기능을 제공하고 있기 때문에 일정기간 계속적으로 증가될 것으로 전망하고 있다(박천보, 2009).

하지만 지방자치단체가 타 지역 테마거리의 경제적 측면만을 보고 무분별한 테마거리 조성사업을 진행함으로써 투자대비 효과가 매우 미미하고 도심 활성화에 기여하지 못하는 경우도 발생하고 있다(박천보, 2009). 이는 그 지역의 정체성과 차별성, 그리고 사람들의 인식 속에 자연스럽게 형성된 그 지역의 특성을 충분히 고려하지 않았기 때문이다. 이러한 문제가 더 이상 발생하지 않기 위해서는 단지 특정 테마의 업종들이 밀집한 지역뿐만 아니라 사람들이 실제로 많이 방문하고, 유명하다고 생각되는 지역을 대상으로 테마거리를 조성해야 할 것이다.

지방자치단체의 주도로 테마거리가 조성되기도 하지만, 이와 달리 '가로수길'은 최근에 자생적으로 형성된 패션 테마거리로 알려진 대표적인 곳이다(우종국, 2008). 이처럼 사람들의 자발적인 활동으로 인해 형성된 자생적 테

마거리가 있을 수 있고, 급변하는 도시환경에서 아직 알려지지 않은 테마거리가 형성 되어 있을 수 있다. 이렇게 자생적으로 형성되거나 아직 알려지지 않은 테마거리를 객관적이고 효율적으로 탐지할 수 있다면 사람들에게 더욱 풍부한 사회·문화적인 공간에 관한 정보를 제공하고 지역경제의 활성화를 기대할 수 있을 것이다. 그리고 두 개 이상의 단일테마가 복합되어 형성된 번화가 지역 내부에서도 지역적 분화에 따라 여러 테마거리들이 형성되어 있을 수 있다. 이러한 테마거리를 탐지함으로써 번화가 내에서도 유명 음식점, 패션매장들이 밀집한 테마거리의 고급정보를 제공하여 사람들의 효율적인 공간적 의사결정을 돕고, 해당 지역의 도시 계획 및 사업을 추진할 때 기초적인 자료로써 사용될 수 있을 것이다.

테마거리가 실재하는 것은 많은 사람들에게 인식되어 있지만, 해당 테마거리의 명확한 공간적 범위를 표현하는 데에는 어려움이 있다. 이러한 어려움은 비단 테마거리 뿐만 아니라 공간적 범위를 지닌 특정 장소들도 해당된다. 명확한 공간적 범위를 표현하기 위한 연구들 중 중소기업청에서는 상권이라는 공간적 범위를 지정하기 위해 현장 실사를 통해 보고서를 작성하였으며, 다른 상권 관련 연구들은 업종별 건물분포 또는 건물 블록단위로 상권의 경계를 구분하였다. 그 외의 공간적 범위를 결정하는 많은 연구들은 공간상에 발생한 현상들을 점 데이터로 수집하고 거리 기반 혹은 밀도 기반의 클러스터 기법(density based clustering)을 활용하여 폴리곤(polygon) 영역 기반의 핫 스팟(hot spot) 지도로 표현하는 것이 일반적이다.

하지만 이러한 방법을 통해 얻어진 결과물들은 현대의 도시환경을 충분히 고려하지 못한 결과이다. 실제 세계에서 사람들의 이동은 도로(street)를 통해 이루어지며, 그 활동 반경은 도로를 중심으로 확장되고 제한되는 것이 일반적이다. Jacobs(1961)는 도시에서 최초로 떠오르는 것은 가로(도로)이며, 가로에서 경험하는 흥미성은 결국 도시의 흥미성과 밀접한 관계를 지닌다고 주장하면서 가로에서 얻는 이미지가 도시이미지를 형성하는 중요한 역할을 한다고 주장하였다. 김경수(2011)는 도로는 아무것도 없는 장소에서 존재하지 못하고, 주위환경과 분리할 수 없는 것이라고 주장하며, 도로가 단순히 물리적인 시설물에서 그치는 것이 아니라 사람들에게 특별한 공간, 장소로서의 근본적인 단위가 될 수 있음을 시사하였다. 특히 테마거리라는 용어와 더불어

로데오 거리, 까페거리, 먹자골목 등 명소의 이름들 대부분이 도로를 지칭하는 현상은 위의 주장을 강하게 뒷받침 하는 근거이다. 이러한 관점에서 봤을 때 도시환경에서는 특정 범위의 장소를 표현할 때 도로를 중심으로 표현하는 방법이 사람들에게 좀 더 직관적으로 이해될 수 있다고 보여 진다. 신사동 가로수길과 강남역 언덕길 테마거리 지도인 [그림 1-1]은 도로를 중심으로 장소를 표현한 대표적인 예시이다.

그림 1-1. 도로 중심으로 표현한 테마거리의 지도 예시
출처 : 박근희(2008)(좌), 데님스토어 제작·배포(2013)(우)

따라서 위와 같은 배경으로 본 연구에서는 활성화된 상업적 성격의 테마거리를 탐지하는 것을 목적으로 하여, 도로 단위의 분석과 시각화가 이루어 질 수 있는 '핫 스트리트(hot street) 군집화 기법'을 제안하였다. 핫 스트리트는 공간적 집중 현상을 표현한 '핫 스팟(hot spot)'과 대응되는 개념으로써, 본 연구에서는 사람들이 많이 방문하고 유명한 업소들이 밀집하여 테마거리로서 특별한 장소성(placeness)을 가질 수 있는 도로를 의미한다. 특히 제안된 핫 스트리트 군집화 기법을 LBSNS 데이터에 적용하여 상업적 성격의 테마거리를 자동으로 탐지해 낼 수 있음을 보이고자 한다.

1.2 연구 방법 및 범위

사람들이 많이 방문하고 활성화된 테마거리를 찾기 위해 본 연구에서는 참여형 지리정보(VGI, Volunteered Geographic Information)[1]의 일종인 LBSNS(Location Based Social Network Service) 데이터를 활용하였다. LBSNS는 특정 장소에 체크인(check-in) 등의 기능을 이용하여 자신의 흔적을 남기고, SNS기술과 연결하여 사용자 간의 관계를 형성하며 정보를 공유, 생산하고 이를 통해 소통하게 하는 지역적 관계서비스를 말한다. LBSNS 데이터를 활용하는 가장 큰 이유는 기존의 공간정보와 달리 풍부하고 실시간성의 데이터라는 특징과 LBSNS 데이터 그 자체가 사람들의 장소에 대한 인식과 반응 정보를 직접적으로 나타낸 것이기 때문이다.

이러한 LBSNS 데이터의 특성을 통해 사람들의 관심과 경험기록이 많은 장소일수록 사람들에게 대중적이고 인기 있는 장소임을 유추할 수 있다. 장소에 대한 관심과 경험기록은 현장조사 또는 단순히 위치 정보만 담고 있는 기존의 POI(Point Of Interest) 데이터에서는 볼 수 없는 LBSNS 데이터만의 핵심적인 특징이다. 따라서 LBSNS 데이터는 활성화된 테마거리를 찾기 위한 가장 적합한 데이터이기 때문에 이를 본 연구에서 활용하였다.

본 연구에서 제안한 핫 스트리트 군집화 기법은 아래 [그림 1-2]와 같이 ①활용 데이터 수집, ②LBSNS 데이터와 건물 데이터 간의 공간조인(spatial join), ③건물과 도로 세그멘트 간의 매칭(matching), ④핫 스트리트 군집(hot street cluster) 탐지 등 크게 4단계로 진행된다.

---

1) VGI(Volunteered Geographic Information) : 크라우드 소싱 기법의 일환으로, 일반인 누구나 지리·공간 정보의 생성에 참여하고 수집하며 전파하는 활동을 통해 구축된 데이터들을 의미한다(Goodchild, 2007).

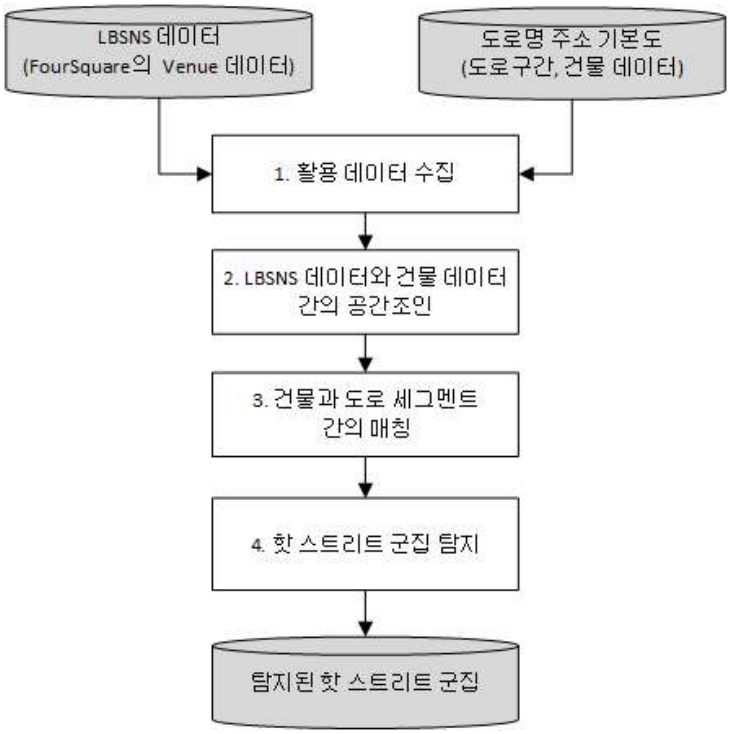


그림 1-2. LBSNS 데이터를 이용한 도로기반의
핫 스트리트 군집화 기법 프로세스

첫 번째 단계에서는 위에서 설명한 LBSNS 데이터를 수집하고 테마 별 분석을 위해 LBSNS 데이터를 테마 별로 분류한다. 도로 기반의 결과를 도출하기 위해 도로명주소 기본도의 건물 레이어, 도로구간 레이어 또한 활용한다.

두 번째 단계에서는 부정확한 LBSNS 데이터의 위치를 보정시키고, 건물 레이어와의 공간조인 연산을 통해 LBSNS 데이터를 통해 알 수 있는 업소의 인기도 정보를 업소가 위치한 건물에 연결시킨다.

세 번째 단계에서는 연구의 기본 단위인 도로 세그멘트를 도로구간 레이어의 교차로를 기준으로 분절하여 생성한다. 그리고 ISOVIST를 활용하여 각 도로 세그멘트와 맞닿아 있는 건물들을 매칭 시키고 공간조인 연산을 통해 건물에 연결된 업소의 인기도 정보를 도로 세그멘트에도 연결시킨다.

마지막 단계에서는 도로 세그멘트에 연결된 업소의 인기도 정보를 종합하

여 도로 세그멘트가 핫 스트리트로서 판단될 고유의 값인 핫 밸류(hot value)를 계산한다. 계산된 도로 세그멘트의 핫 밸류를 대상으로 공간 클러스터 기법(spatial clustering)인 Getis-Ord의 G_i^*[2]를 활용하여 통계적으로 유의한 핫 스트리트 군집을 탐지하였다. 이렇게 대상지역에서 탐지된 특정 테마의 핫 스트리트 군집은 그 지역의 테마거리로 선정한다.

본 연구에서 제안한 기법을 통해 탐지된 핫 스트리트 군집이 테마거리로서 적절한지 판단하기 위해 신림 사거리, 강남역 번화가 두 지역을 대상으로 실험을 실시하였다. 그리고 상권 보고서, 관련 논문, 보도된 자료 등을 참고하여 실험 결과를 비교·분석하였다.

1.3 관련 연구

테마거리에 관련된 연구는 주로 하나의 테마로 구성된 특화거리에 대한 연구로 많이 진행되어 왔다. [표 1-1]은 특화거리에 관한 최근 연구동향을 정리한 표이다. [표 1-1]과 같이 기존의 연구들은 특화거리의 조성과정과 변천과정에 관한 연구, 물리적 환경개선방안과 활성화에 관한 연구가 주를 이루며, 최근에는 가로 이미지의 평가방법에 관한 연구가 진행되어 왔다(김경수, 2011).

2) 일반적으로 공간 클러스터를 탐지하기 위해서, LISA(local indicators of spatial association, 국지적 공간적 자기상관 통계량)(Anselin, 1995)를 활용하는 방법론을 사용한다. Getis-Ord의 G_i^*는 이 국지적 공간적 자기상관 통계량 중 하나이다. Getis-Ord의 G_i^*에 대해서는 이후 본문에서 자세히 설명되어 있다.

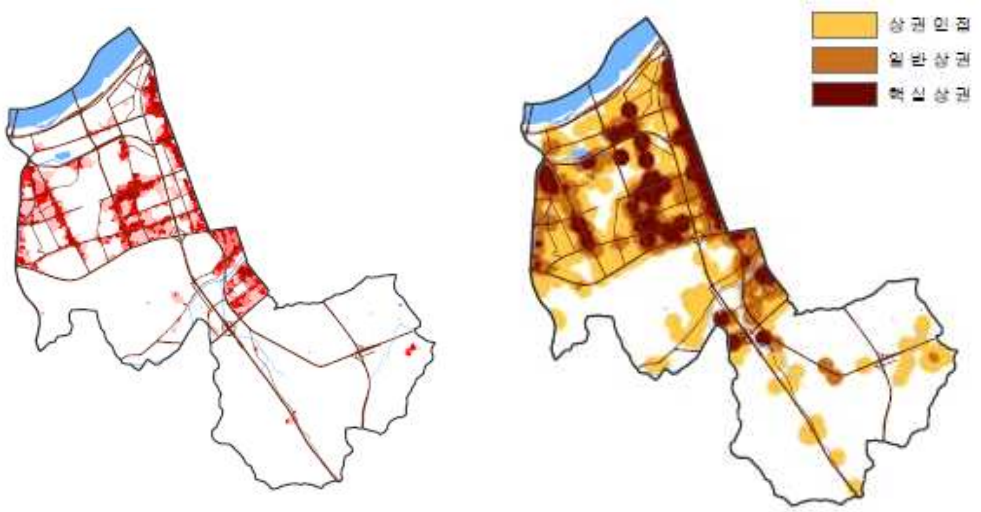
표 1-1. 특화거리에 관한 최근 연구동향 (김경수, 2011)

분류		연구자	내 용
조성 과정에 관한 연구		손재룡(2003)	특화거리 조성의 변천과정에서 나타난 Network의 중요성에 관한 연구
물리 환경 개선	장소성 부여	이수옥(2000)	가로 공간 개선방안에 관한 연구
	개선방안	박현규(2003)	도심재생을 위한 특화거리의 조성과 개선방안
	활성화 방안	박천보(2009)	물리환경적 도심재생 관점의 특화거리 활성화 방안 연구
	활성화 요인연구	김홍순(2010)	신사동 가로수 길과 삼청동 길의 활성화 요인 연구
평가를 위한 이미지 도출		김지혜(2009)	특화거리 조성을 위한 환경디자인 구성 요소의 평가에 관한 연구
장소적 특성		최지영(2007)	신도시 상업가로의 장소적 특성에 관한 연구

이 외에 김경수(2011)는 특화거리의 객관적 평가를 위한 물리적 환경 분석의 체크리스트를 구축하였다. 이러한 특화거리에 관한 기존 연구들은 현재 형성되어 있는 특화거리에 한정되어 종합 분석한 연구들이며 특화거리, 즉 테마거리 탐지 기법에 관한 연구는 선행되지 않은 것으로 판단된다. 테마거리 탐지 기법에 관련된 연구를 한다면 한정된 지역에만 적용되지 않고 지역마다 다른 물리적인 환경과 시간에 따른 지역적 변화를 대처할 수 있도록 어느 지역에서나 객관적이고 재현 가능한 테마거리 탐지 기법을 제시하여야 한다.

상업적 성격의 테마거리는 그 지역의 핵심 상업 중심지로서 지역 경제 발전에 중요한 역할을 한다. 상업 중심지의 속성, 형성과 발전과정 등에 관한 연구들은 기존에 많이 이루어졌으며, 개별 또는 집단의 상업시설이 흡수할 수 있는 소비자의 범위를 뜻하는 상권을 단위로 분석이 이루어져 왔다. 일반적으로 상권은 정의에서 조차 명확하지 않으며 구체적인 수치로 나타내는 것도 용이하지 않아, 상권의 범위는 500m, 1km와 같이 거리로 표시하거나 전문가의 경험에 의해 결정되었다(이용직, 2005). 이병길(2003)은 보다 정확한 상권의 속성, 형성, 발전과정 등의 지속적인 연구가 지속되기 위해서는 지역

이나 시간에 따른 차이가 없는 객관성 있고 재현성 있는 방법이 필요하다고 주장하며, 공간상에서 일어나는 경제 현상을 객관적으로 설명할 수 있는 상권[3]추출 기법을 제안하였다. 이병길(2003)의 연구에서는 카드사 가맹점 데이터와 회원 정보를 가공하여 얻어진 점 형태의 동적 데이터를 활용하여, 공간 클러스터 방법(k-medoid 기법, SMTIN 기법 및 DENCLUE 기법)을 적용시키고 상권 추출을 위한 최적의 클러스터 기법으로 DENCLUE 기법을 추천하였다. [그림 1-3]은 서초구를 대상으로 SMTIN 기법과 DENCLUE 기법을 수행한 결과이다. [그림 1-3]에서 제시한 모호한 형태의 상권은 사람들에게 명확한 상권 영역을 제시하지 못하고 있으며, 그림에서도 살펴볼 수 있듯이 도로를 따라 형성된 핵심 상권의 특성을 전혀 반영하지 못하는데 한계가 있다.



(a) SMTIN 기법에 의한 가맹점 상권추출 결과

(b) DENCLUE 기법에 의한 가맹점 상권추출 결과 (단순 가맹점 매출 밀도)

그림 1-3. 서초구 상권 추출 결과 (이병길, 2003)

3) 이병길(2003) 연구에서의 상권은 일반적인 상권 의미와는 달리 해당 지역에 상업 시장이 형성되어 있는지 판단하는 것으로, 현재 소비가 일어나고 있거나 잠재적으로 소비가 일어날 수 있는 지역을 상권으로 정의하였다. 따라서 본 연구에서 탐지하고자 하는 테마거리의 영역은 이병길(2003)의 연구 대상인 상권의 부분집합이다.

본 연구에서 활용하는 LBSNS의 베뉴 데이터, 특히 체크인 정보는 그 지역에 위치한 장소들의 분포와 상업적인 활동 등을 간접적으로 확인할 수 있는 동적 데이터이다. 이러한 특성 때문에 최근에는 체크인 데이터를 활용한 관련 연구가 활발히 진행되고 있다(Wang, 2012). Cranshaw et al.(2012)는 'Livehoods'라는 프로젝트를 통해 사람들의 공간에 따른 생활 패턴과 특성을 살펴보기 위한 공간 영역 분할 기법을 제시하였다. Livehoods 프로젝트에서 제시한 기법은 포스퀘어(Foursquare) 서비스의 체크인 데이터를 통해 체크인 속성과 위치 속성을 이용한 베뉴 데이터 간의 유사도 값을 계산하고, 스펙트럴 군집 기법(spectral clustering)[4]을 통해 지역적으로 근접해 있고 유사하다고 판단되는 데이터들을 군집화 시켜 Livehoods라고 불리는 고유의 영역으로 분할시켰다. 뉴욕, 몬트리올, 시애틀 등 주요도시를 대상으로 실시한 Livehoods 결과물은 웹사이트를 통해 제공되고 있으며, [그림 1-4]는 웹사이트에서 제공하는 뉴욕의 Livehoods 결과 그림이다.[5]

4) 스펙트럴 군집기법(spectral clustering)은 라플라시안(Laplacian) 행렬을 이용한 그래프 분할(graph partitioning) 기반의 군집분석 기법이다. 이 방법은 k-means와 같은 전통적인 군집분석 기법과는 대조적으로 볼록한 모양뿐 아니라 다양한 군집의 형태에 대해 좋은 성능을 보이는 것이 특징이다.

5) http://livehoods.org/maps/nyc, Livehoods 결과 화면에서 베뉴를 선택하면 군집된 영역이 표시되고, 군집에 따라 점 형태의 베뉴 데이터가 색깔별로 분류되어 있다. 왼쪽 상자란에는 군집을 형성한 베뉴들의 카테고리 비율, 요일별 체크인 수, 시간당 체크인 수 등 체크인 데이터 통계 결과도 함께 나타내고 있다.

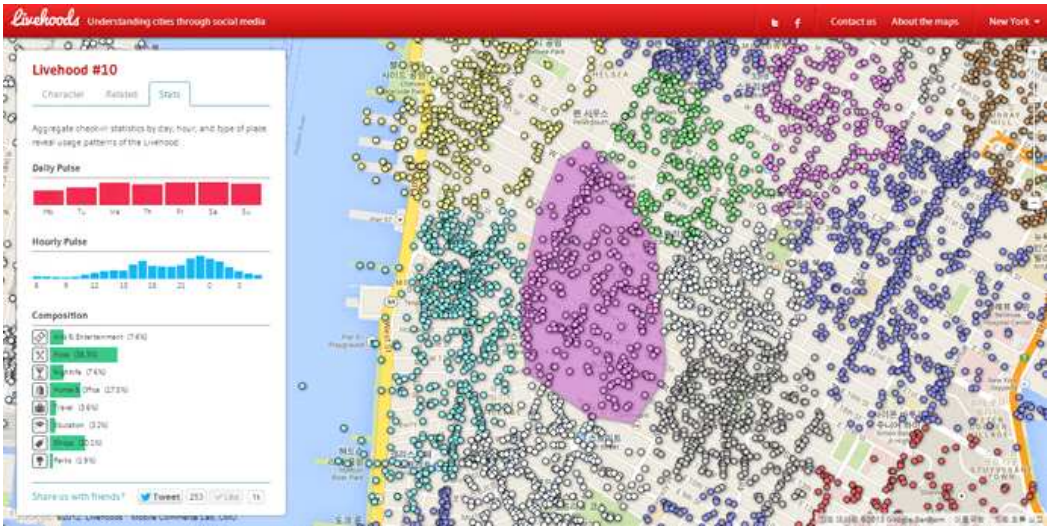


그림 1-4. 미국 뉴욕 지역의 Livehoods 결과 예시

Livehoods가 베뉴들의 유사도 값을 통해 군집된 영역이지만, 각 군집이 의미하는 뚜렷한 특성을 알 수는 없다. 모든 카테고리의 베뉴들을 연구에 사용함으로써 분할된 군집 영역은 공통적으로 음식점과 상점 카테고리의 비율이 높게 나타나고 중요치 않은 지역 또한 군집으로 적용되기 때문이다. 따라서 Livehoods는 대상 도시에서 확인할 수 있는 특정 현상을 탐지하는 목적보다 사람들의 활동을 반영한 지방 경계를 제안하는데 의의가 있다.

Wang(2012)은 중국의 지에팡(Jiepang) 서비스의 체크인 데이터를 이용하여 대상 지역의 핫 스팟 분석을 수행하였다. 대상 지역을 균일한 격자로 분할한 뒤 각 격자에 포함된 체크인 수를 계산하고, 핫 스팟 분석을 위해 Getis-Ord의 G_i^*를 이용하였다. [그림 1-5]는 Wang(2012)의 연구 결과로서 유의수준에 따라 붉은색으로 음영 처리한 핫 스팟 그림이다. 그리고 Wang(2012)은 6개 실험 지역의 인구 데이터와 체크인 데이터간의 상관관계가 있음을 보이며, 체크인 데이터가 간접적으로 그 지역의 경제와 인구를 반영한다고 주장하였다.

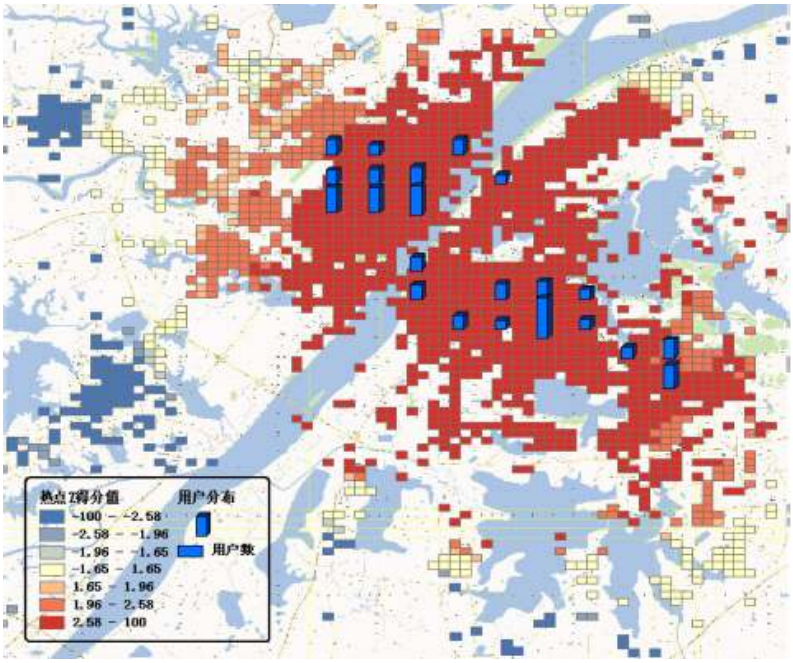


그림 1-5. 중국 우한 지역의 체크인 핫 스팟 지도 (Wang, 2012)

[그림 1-5]의 핫 스팟 지도는 사람들의 사회·경제적 활동의 집중지역을 대략적으로 판단할 수 있는 좋은 자료이지만, 격자 단위로 분석이 이루어져 격자 크기에 따라 핫 스팟 지도 결과가 다르게 표현되는 단점이 있다. 그리고 체크인한 장소에 대한 정보를 전혀 고려하지 않음으로써 테마 별 분석이 이루어지지 않고, 단순히 체크인 집중 지역을 표현한 것에 지나지 않는데 한계가 있다.

위의 연구들은 공통적으로 모호한 경계의 영역으로 결과를 표현하는데 한계가 있다. 도시환경에서는 특정한 장소를 표현할 때 사람들에게 좀 더 직관적으로 이해될 수 있도록 도로 기반의 분석과 시각화가 이루어져야 한다. 이러한 문제점을 인지하고 도로 및 네트워크 기반의 군집 기법, 핫 스팟 기법에 관한 선행 연구가 많이 이루어졌다.

Mohaymany(2013)는 차량사고가 자주 발생하는 도로를 탐지하기 위해 NKDE(Network Kernel Density Estimation)를 활용하였다. 도로 위에서 발생한 차량사고 데이터를 기존의 KDE(Kernel Density Estimation)방법으로 수행하게 되면, 도로가 아닌 지역에 대해서도 밀도 추정이 이루어지는 등 의

도치 않은 결과를 초래하게 된다. 이 때문에 도로 네트워크상에서 KDE를 분석할 수 있는 Okabe(2009)의 NKDE을 활용하여 차량사고가 자주 발생하는 지점을 탐지하였다.

Shiode(2011)는 도로 상에서 발생하는 범죄(강도, 차량 도난, 마약 밀매 등)들이 자주 발생하는 거리를 탐지하기 위한 방법으로 NT-STAC(Network Spatial and Temporal Analysis of Crime)과 NT-SaTScan(Network Spatial Scan Statistic)을 제안하였다. STAC과 SaTScan은 탐색 윈도우(search window)를 이용하여 핫 스팟을 탐지하는 기법들로서 각각 포아송 분포와 최대우도법을 이용하여 통계적으로 유의미한 결과를 도출해낸다. STAC과 SaTScan을 통해 탐지된 핫 스팟 결과는 일반적으로 2차원 평면상에 원형으로 표시되지만, 도로 네트워크상에 발생하는 범죄들을 탐지 대상으로 하였을 때는 선형 공간, 도로 네트워크 상에서 분석이 이루어져야 한다고 주장하였고 NT-STAC와 NT-SaTScan을 통해 주장의 타당성을 검증하였다[그림 1-6].

(a) STAC 적용 (b) NT-STAC 적용

그림 1-6. 범죄 발생 점 데이터에 STAC과 NT-STAC을 적용하여 핫 스팟을 탐지한 결과 (Shiode, 2011)

위의 도로 기반의 핫 스팟 탐지 연구들은 도로의 선형 공간 상에서 핫 스팟을 탐지하는데 의의가 있지만, 주로 도로 위에 있는 점 데이터에만 초점을 맞춘 군집 기법들이고 도로 객체 자체에 대한 의미를 고려하지 않는다. 따라

서 위 연구들의 핫 스팟 영역 결과는 핫 스팟인 거리의 의미보다 점 데이터들을 단순히 선형으로 연결한 영역에 가깝다. 본 연구에서는 개별 도로가 장소로서 인지되는 사람들의 공간 인식을 반영하여, 도로 객체를 분석단위로 고려하고 테마거리인 핫 스팟 지역을 탐지하는 측면에서 위의 연구들과 차이점이 있다.

개별 도로를 분석단위로 수행한 연구는 본 연구 이외에 Lu(2005)의 선행연구에서도 찾아볼 수 있다. Lu(2005)는 본 연구와 마찬가지로 도시 환경에서 사람들의 사회 경제적 활동은 도로에 의해 공간적 제약을 받는 현상을 설명하며, 2차원 유클리드 평면상에 군집 결과를 표시하던 점 데이터의 군집 기법들을 도로 네트워크 환경에서 적용할 수 있도록 확장시켰다. Lu(2005)는 점 데이터의 군집 기법을 크게 거리 기반(distance criterion)과 영역 기반(zoning technique)의 두 그룹으로 분류하였고, 특히 차량 도난 데이터를 이용한 영역 기반의 군집 기법 실험에서 도로의 핫 스팟을 뜻하는 '핫 스트리트(Hot Street)'란 용어를 사용하였다. 그리고 교차로를 기준으로 도로를 분절하여 군집 분석의 단위가 되는 '도로 세그멘트(street segment)'를 정의하고, 각 도로 세그멘트는 그 위에 있는 점 데이터의 개수를 속성 정보로 사용하였다. 최종적으로 도로 세그멘트의 포아송 분포를 이용하여 통계적으로 유의한 핫 스트리트를 탐지하였다[그림 1-7]. Lu(2005)의 연구에서 정의한 도로 세그멘트와 핫 스트리트의 개념은 본 연구에서도 똑같이 이용된다.

그림 1-7. 미국 버팔로 지역의 차량 도난이 집중적으로
발생한 핫 스트리트 탐지 결과 (Lu, 2005)

Lu(2005)의 영역 기반의 군집 기반에서 사용하는 데이터는 도로 위에서 차량 도난이 발생한 위치만을 나타낸 점 데이터이다. 그리고 Lu(2005)의 연구에서 핫 스트리트를 탐지하기 위해 포아송 분포(Poisson distribution)를 사용하는데, 이는 도로 세그멘트 위의 점 데이터의 개수 등 이산적인 값을 지닌 데이터만을 분석할 수 있는 한계점을 보여준다. 또한 각 도로 세그멘트의 길이는 제각각이기 때문에 길이에 따른 차량 도난 사건 수의 평균화가 이루어졌어야 한다. 이와 달리 테마거리를 탐지하기 위해서는 베뉴의 위치 데이터뿐만 아니라 베뉴의 카테고리, 체크인 수 그리고 도로 세그멘트의 길이 등 다양한 정보를 종합적으로 활용하여야 할 것이다.

이상 선행 되었던 관련 연구들을 종합해 보았을 때, 다음과 같이 3가지의 한계점을 확인할 수 있었다.

■ 객관성 있고 재현성 있는 테마거리 탐지 기법의 부재
■ 공간적 단위로서 하나의 장소로 인지되는 개별 도로를 기반으로 한 핫 스트리트 분석과 시각화의 필요성
■ 다양한 속성 정보를 종합적으로 활용 가능한 핫 스트리트 탐지 기법의 필요성

따라서 본 연구는 선행 연구에서 확인한 위의 한계점들을 해결할 수 있는 테마거리를 탐지하기 위한 효율적인 기법을 제시하고자 한다.

1.4 용어 정리

핫 스트리트 군집화 기법에 대한 설명에 앞서 정확한 의미전달을 위해 본 연구에서 사용하는 용어(terminology)에 대해 정리하였다. 용어에 대한 설명은 아래와 같다.

■ 테마거리(themed street)
: 특정 주제(음식, 카페, 패션, 유흥주점 등)를 정해 독특하게 꾸며 사람들의 관심을 이끄는 길거리. 본 연구에서는 상업적 성격의 테마거리만을 의미함.

■ VGI(Volunteered Geographic Information)
: 크라우드 소싱 기법의 일환으로, 일반인 누구나 지리·공간 정보의 생성에 참여하고 수집하며 전파하는 활동을 통해 구축된 데이터.

■ LBSNS(Location Based Social Network Service)
: 특정 장소에 체크인(check-in) 등의 기능을 이용하여 자신의 흔적을 남기고, SNS기술과 연결하여 사용자 간의 관계를 형성하며 정보를 공유, 생산하고 이를 통해 소통하게 하는 지역적 관계서비스.

■ 베뉴(venue)

: 사용자들이 포스퀘어(Foursquare) 서비스의 데이터베이스 상 기존에 없던 장소를 등록하고 싶은 경우 새로이 장소(e.g. 레스토랑, 지하철 역, 영화관, 음식점, 개인 주거지 등)를 생성하여 등록하거나, 해당 장소와 관련된 정보들을 입력 및 수정, 공유, 체크인 할 수 있는 특정 지점.

■ 베뉴 건물

: 베뉴가 위치하고 있는 건물.

■ 도로구간

: 2014년부터 전면 시행되는 도로명 주소 체계에서 정의된 도로의 시작지점부터 끝지점을 연결한, 가상의 연속된 구간. 각 도로구간마다 고유의 도로명을 가지고 있음.

■ 도로 세그멘트(street segment)

: 교차로를 기준으로 연속된 도로를 분절시킨 군집 분석의 단위

■ ISOVIST

: 공간 내에서 특정 지점(조망점)으로부터 장애물에 의해 가려지지 않고 보이는 모든 지점들을 연결한 선들의 집합으로 이루어진 가시 공간.

■ 핫 밸류(hot value)

: 특정 도로 세그멘트가 핫 스트리트인지 판단하기 위해 계산된 고유의 값. 핫 밸류는 테마거리로서 필요한 요소들인 베뉴들의 평균 인기도, 베뉴 건물들의 구성비, 베뉴 건물들의 밀집도를 고려하여 계산됨.

■ 핫 스트리트(hot street)

: 공간적 집중 현상을 표현한 '핫 스팟(hot spot)'과 대응 되는 개념으로써, 사람들이 많이 방문하고 유명한 업소들이 밀집하여 테마거리로서 특별한 공간성을 가질 수 있는 도로.

■ 핫 스트리트 군집(hot street cluster)

: 인접해 있는 핫 스트리트들로 구성된 공간적 응집체. 연구 대상 지역에 형성된 것으로 예상되는 테마거리를 의미함.

2. 핫 스트리트 군집화 기법

본 장에서는 LBSNS 데이터를 활용한 핫 스트리트 군집화 기법에 대해서 자세히 설명한다. [그림 2-1]은 전체적인 연구의 이해를 돕기 위한 핫 스트리트 군집화 기법의 흐름도를 나타낸 것이다.

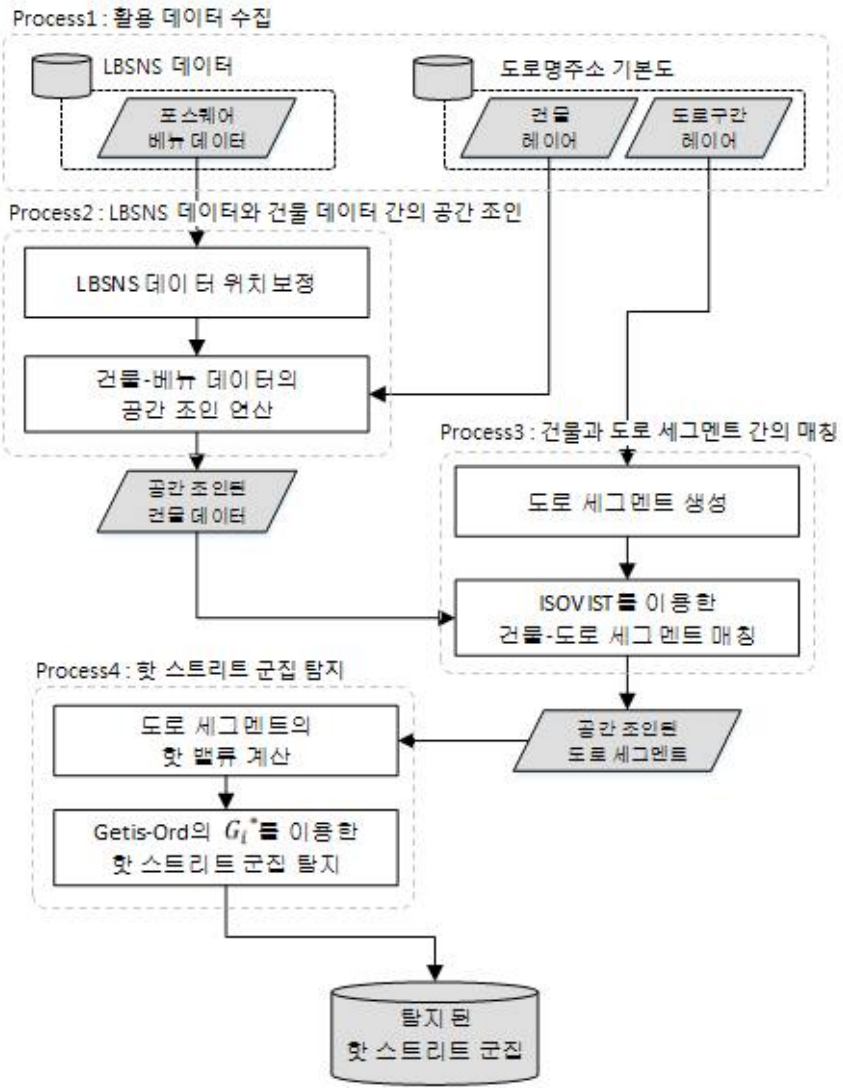


그림 2-1. LBSNS 데이터를 이용한 도로기반의
핫 스트리트 군집화 기법 흐름도

2.1 활용 데이터 수집

2.1.1 LBSNS 데이터(포스퀘어 베뉴 데이터) 수집

본 연구에서는 체크인 기능 등을 통해 사람들의 장소 이용 패턴과 인기도 및 장소 분포 특성들을 쉽게 분석할 수 있는 LBSNS 데이터를 이용하였다.

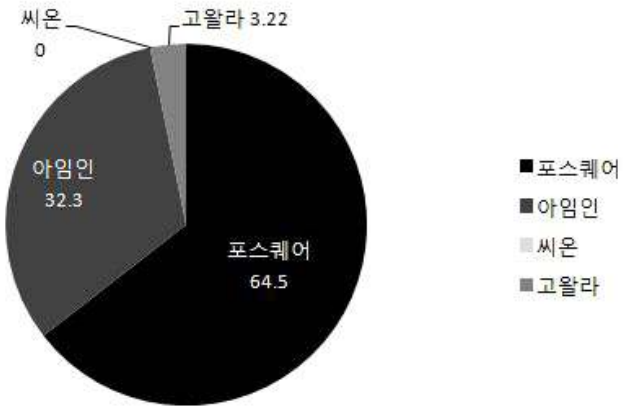


그림 2-2 LBSNS 앱 선호도 조사 (신인지, 2012)

현재 국내의 LBSNS 시장 현황은 포스퀘어, 아임인(2013년 5월 20일 서비스 종료), 고왈라, 씨온 등 유사한 서비스들이 경쟁을 벌이고 있지만, 이 중에서도 본 연구에서는 2009년 Dennis Crowley와 Naveen Selvadurai에 의해 개발된 포스퀘어(Foursquare) 서비스의 베뉴(venue) 데이터를 이용하였다.

한국 사용자들을 대상으로 앱 선호도 조사를 2011년 10월 10일부터 10월 12일까지 설문을 통해 실시한 결과, [그림 2-2]에서 살펴볼 수 있듯이 포스퀘어 64.5%, 아임IN 32.3%, 씨온은 무응답의 결과를 알 수 있었다(신인지, 2012). 이처럼 상당한 비율의 LBSNS 사용자들이 포스퀘어를 사용한다고 생각할 수 있으며, 충분한 데이터의 활용과 다수의 의견을 대변하고 신뢰할 수 있는 핫 스트리트 군집을 탐지하기 위해서는 선호도가 가장 높은 포스퀘어 서비스의 데이터를 활용하는 것이 적절하다고 판단되었다.

하지만 포스퀘어 서비스는 2010년 Ignite Social Media의 포스퀘어 사용자 통계에 따르면 [그림 2-3]과 같이 25~34세, 35~44세 두 연령대에서 집중적으로 사용되어 지고 있다. 포스퀘어 서비스는 모든 연령대에서 고르게 사용되지 못하고 있기 때문에, 포스퀘어 데이터를 활용한 본 연구의 결과는 주로 25~44세 연령대의 사람들을 만족시키는 결과가 될 것이다.

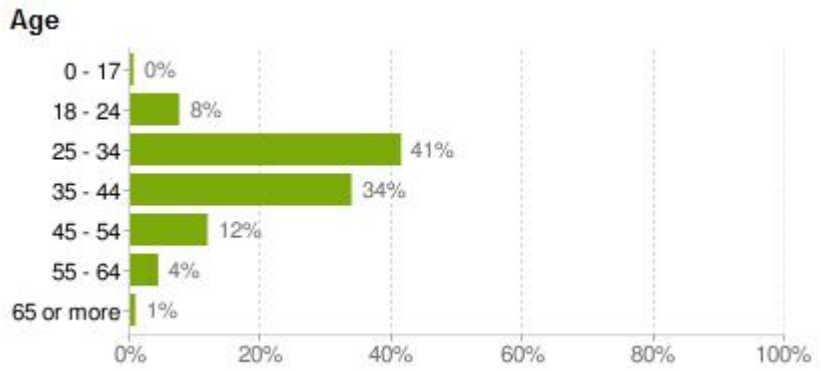


그림 2-3. 포스퀘어 서비스의 연령별 사용자 분포
출처 : Ignite Social Media의 2010 Social Network Analysis Report

이러한 포스퀘어 데이터를 수집하기 위해서 핫 스트리트를 탐지할 대상 지역 범위를 설정(남서쪽 좌표와 북동쪽 좌표를 이용하여 사각형 범위 설정)하고, 포스퀘어 측에서 제공하는 API를 활용하여 그 지역 범위 안에 생성된 포스퀘어 데이터를 수집하였다. 포스퀘어 API[6]는 개발자 및 사용자가 원하는 기능과 목적에 따라 다양한 API를 제공하고 있으며, 본 연구는 Venues API의 search와 likes 기능을 활용하여 원하는 대상 지역의 베뉴 정보를 수집하였다. '베뉴' 란 사전적 의미로는 (콘서트·스포츠 경기·회담 등의) 장소[7]를 의미한다. 포스퀘어 서비스에서는 사용자들이 포스퀘어 데이터베이스 상 기존에 없던 장소를 등록하고 싶은 경우 새로이 장소(e.g. 레스토랑, 지하철 역, 영화관, 음식점, 개인 주거지 등)를 등록할 수 있고, 이 장소와 관련된 정보들을 입력 및 수정, 공유, 체크인 할 수 있는 특정 지점을 베뉴라고 일컫는다.

6) https://developer.foursquare.com/에서 제공
7) 네이버 영어사전 참고

포스퀘어 API 요청 시 수집할 수 있는 베뉴 관련 정보는 포스퀘어 서비스와 관련된 자세한 수준의 정보까지 제공해 주고 있지만, 본 연구에서는 그 중 연구과정에서 필요하다고 생각되는 핵심적인 정보만 활용하였으며 해당 정보 속성은 [표 2-1]과 같다. 특히 usersCount(특정 베뉴에 체크인 한 유저 수), tipCount(베뉴에 팁이 작성된 수), count(베뉴를 라이크로 등록한 수) 속성은 사람들이 베뉴에 대한 관심을 남긴 기록들이며 베뉴의 인기도를 직관적으로 알려주는 수치이다.

표 2-1. 포스퀘어 API 활용 정보 목록

기능	취득 정보 속성	정보 내용 설명	데이터 형식
search	venue_id	베뉴의 고유 ID	string
	name	베뉴의 명칭	string
	categories_id	베뉴가 속한 카테고리 ID	string
	categories_name	베뉴가 속한 카테고리의 명칭	string
	lat	위도 좌표	float
	lng	경도 좌표	float
	address	베뉴가 위치한 주소	string
	crossStreet	베뉴의 위치와 교차하는 도로	string
	checkinsCount	총 체크인 수	int
	usersCount	체크인 한 유저 수	int
	tipCount	팁(tip)이 작성된 수	int
likes	count	해당 베뉴를 라이크(like)로 등록한 수	int

2.1.2 LBSNS 데이터의 테마 별 카테고리 분류

현재 국내 포스퀘어 서비스는 사용자들이 새로운 베뉴를 생성할 때 그 베뉴의 카테고리(category)를 의무적으로 설정하도록 한다(카테고리 설정 기능은 2010년 3월 10일 이후에 시행되었기 때문에, 그 이전에 생성된 베뉴들은 카테고리가 분류되어 있지 않을 수 있다). 포스퀘어의 카테고리는 총 3단계의 계층적 구조로 구성되어 있다. 가장 상위 수준의 카테고리는 예술 및 오락, 대학 및 대학교, 음식, 유흥 명소, 야외 활동 및 오락, 전문점 및 기타 장소, 거주지, 상점 및 서비스, 여행 및 교통 등 총 9개로 분류 되어있다. 각 카

테고리는 음식 카테고리를 88개의 하위 카테고리로 분류 할 정도로 세부 카테고리를 상세히 분류하고 있다.

서비스 이용 시 베뉴의 카테고리를 의무적으로 설정해 놓았기 때문에 모든 베뉴 데이터는 카테고리를 이용한 분석에 빠짐없이 사용될 수 있고, 카테고리의 조합을 어떻게 하느냐에 따라서 다양한 테마 별 분석을 가능케 한다. 이를 활용하여 본 연구는 포스퀘어 데이터를 활용하여 상업적 테마거리의 주요 테마로 생각되는 ①음식점, ②패션, ③카페, ④유흥주점, 4 가지의 테마 별 분석을 수행하였다.

포스퀘어 자체에서 테마를 의미하는 카테고리를 최상위 단계의 9가지 카테고리로 분류하고 있지만, 음식 카테고리에 '고급 술집'의 세부 카테고리를 포함하고 있는 등 원하는 테마에 어울리지 않는 카테고리들이 혼재 하고 있다. 따라서 본 연구 대상 테마에 적절한 데이터를 사용하기 위해 [부록 A]와 같이 4가지의 테마 별로 포스퀘어 카테고리 그룹을 자체적으로 구성하였다. 즉, 특정 테마의 핫 스트리트 군집을 탐지할 때 해당 테마로 분류된 카테고리 그룹의 베뉴 데이터를 이용하였다.

단. 포스퀘어 베뉴 데이터 중 usersCount 수가 1이하인 베뉴들은 본 연구에서 사용하지 않았다. usersCount 수가 1인 베뉴들은 베뉴를 생성한 사람 이외에 다른 사람들에게 베뉴의 존재를 검증받지 못한 상태이기 때문이다. 그리고 이러한 usersCount 수가 1이하인 베뉴들은 개인적인 용도로 생성하거나 베뉴 이름에 오타가 심한 경우가 많다.

2.1.3 도로명 주소 기본도 활용

본 연구에서는 도로 기반의 군집을 탐지하기 위해, 도로명 주소 기본도 데이터 또한 활용하였다.

도로명 주소는 모든 도로에 도로명을 부여하고 건물에는 도로를 따라 일정한 규칙으로 건물번호를 부여하는 새로운 주소 체계이며, 도로명주소 기본도는 이러한 주소체계를 관리하기 위해 구축한 벡터 형태의 지도 자료이다. 도로명 주소 기본도는 도로명 주소안내를 위해 건물, 건물이력, 건물군, 실폭도로, 도로구간, 도로구간 이력, 도로명판, 기초구간, 출입구, 연결선, 교차로, 지

하철선로, 지하철역사, 지하철출입구, 철도역사, 철도선로, 터널, 교량, 하천호수, 공원, 고가도로, 자동차전용도로, 지하차도, 기타시설물, 기타표지판 이상 25개의 레이어를 포함한다. 본 연구에서는 도로명 주소 기본도의 '건물'과 '도로구간' 두 개의 레이어를 활용하였으며, 각 레이어의 속성 스키마는 <부록 B>, <부록 C> 와 같다.

2.2 LBSNS 데이터와 건물 데이터 간의 공간조인

베뉴 데이터는 점 형태의 데이터로써 해당 업소가 위치한 지점을 대략적으로 알려준다. 하지만 실제로 해당 업소는 특정한 물리적인 영역을 가지고 있고, 특히 건물 내에 위치하고 있다. 그리고 점 데이터와는 달리 폴리곤 형태의 건물 데이터는 각 건물 객체의 크기나 주변 건물, 도로와의 위상 관계 등을 고려할 수 있어 좀 더 현실적인 분석을 가능케 한다.

따라서 본 연구는 업소의 인기도를 알 수 있는 베뉴 정보를 공간 조인 연산을 통해 건물 데이터에 연결시켰다.

2.2.1 LBSNS 데이터 위치보정

건물과 포스퀘어 베뉴 데이터와의 공간조인을 실행하기에 앞서, VGI의 일종인 LBSNS 데이터가 지닌 정보의 품질에 대한 불확실성 문제를 확인해야 한다. Goodchild는 VGI가 실시간으로, 그리고 꽤 자세한 공간·지리정보를 매우 적은 비용에 구축할 수 있는 효과적인 방법인 반면에, 그 정보의 품질 및 정확도 결함에 대해서 지적하고 그에 대한 해결 방안들을 크라우드 소싱, 사회적 측면, 지리학적 측면으로 접근하여 제시한 바 있다(Goodchild et al., 2012).

[그림 2-4]은 신림역 사거리 지역의 도로명 주소 기본도(건물, 실폭 도로 레이어)와 포스퀘어 베뉴 데이터를 함께 중첩시켜 나타낸 그림이다. 그림에서 베뉴 데이터들은 도로 한가운데에 분포하고 있거나 빈 공간에도 베뉴 데이터가 위치하는 등 현실적이지 못한 분포를 보여주고 있다. 이를 통해 실제로 현재 구축된 포스퀘어 베뉴 데이터들은 공간·지리 정보의 핵심인 위치 정

확도(location accuracy)가 현저히 떨어지는 것을 알 수 있다.

그림 2-4. 신림역 사거리 지역의 포스퀘어 베뉴 데이터 분포 현황

포스퀘어 데이터의 위치가 정량적으로 어느 정도의 오차가 있는지 신림 사거리 지역(SW 좌표 : 126.928, 37.482, NE 좌표 : 126.931, 37.486 내의 사각 범위)을 대상으로 측정하였다. 대상 범위 내에서 취득한 412개 베뉴 데이터 중 랜덤 샘플링한 128개의 정확한 위치를 알기 위해, 네이버와 다음 두 개의 주요 포털 사이트에서 베뉴 명칭을 검색하여 해당 베뉴의 지번 주소를 조사하였다. 조사한 지번 주소 정보들을 웹 기반의 GIS 어플리케이션인 'X-Ray Map for Web'[8]의 지오코딩(geocoding) 툴을 활용하여 베뉴 샘플들의 주소를 해당 경위도 좌표로 변환하였다. 변환된 정확한 경위도 좌표와 포스퀘어 데이터의 경위도 좌표간의 유클리드 거리오차를 측정하고, 위치오차가 일정한 방향성을 지니고 있는지 오차 벡터의 방위각 또한 측정하였다.

8) BIZ-GIS.com에서 제공하는 웹 기반 GIS 어플리케이션

$$\overline{P_1P_2}\ (m) = \sqrt{(x_{P_1} - x_{P_2})^2 + (y_{P_1} - y_{P_2})^2} \qquad \text{식 (2-1)}$$

$$\angle P_1P_2\ (\mathrm{degree}) = \arctan\left(\frac{y_{p1} - y_{p2}}{x_{p1} - x_{p2}}\right) \times \frac{180}{\pi} \qquad \text{식 (2-2)}$$

본 연구에서 지오코딩 된 정확한 좌표와 포스퀘어 데이터의 좌표는 모두 경위도 좌표이므로 식을 적용하기에 앞서 미터(m) 단위인 도로명 주소 기본도와 같은 좌표체계(UTM-K 좌표계, Bessel 1841 타원체, TM 투영)로 투영(project)시켰다. P_1은 포스퀘어 데이터의 좌표를 의미하며, P_2는 지오코딩을 통해 조사한 정확한 위치 좌표를 의미한다. 식 (2-1)은 P_1과 P_2 두 점 좌표를 이용하여 유클리드 거리(Euclidean distance)를 측정하는 식이며, 식 (2-2)는 P_1과 P_2가 이루는 방위각을 측정하는 식이다.

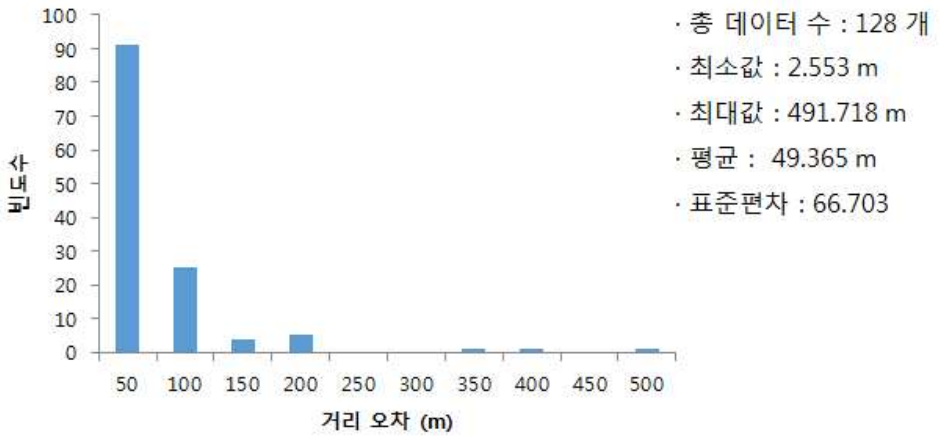


그림 2-5. 포스퀘어 데이터의 유클리드 거리 오차 히스토그램

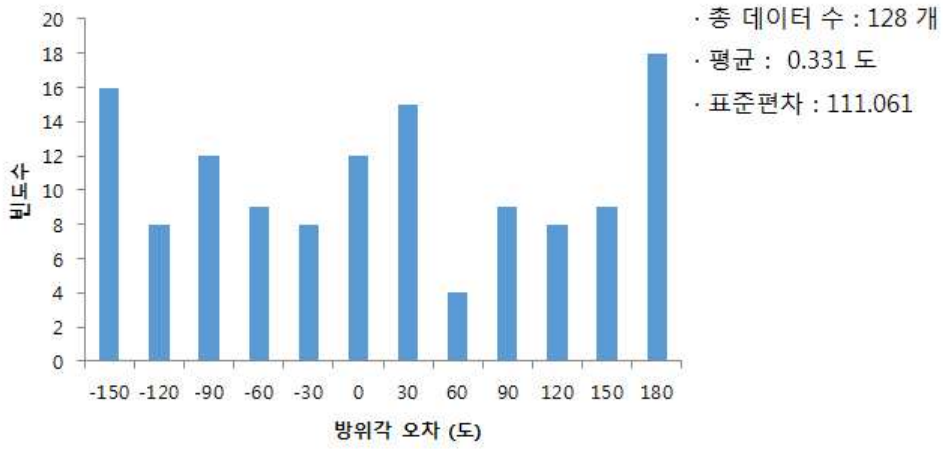


그림 2-6. 포스퀘어 데이터 오차 벡터의 방위각 히스토그램

128개의 포스퀘어 데이터 샘플에 위의 식을 적용한 결과, 유클리드 거리 오차의 평균은 약 50m 로 측정되었다[그림 2-5]. 오차 벡터의 방위각은 의미 있는 특정 분포를 나타내지 않으며, 모든 방향(-180도~180도)에서 랜덤한 방향성을 나타내고 있음을 알 수 있다[그림 2-6].

유클리드 거리 오차가 약 50m 정도로 발생하는 이유는 무선 인터넷 환경과 사용자의 포스퀘어 서비스 사용 패턴에 의한 것으로 예상된다. 보통 사용자들은 포스퀘어 서비스에 베뉴를 등록하거나 체크인을 하는 경우, 그 베뉴가 위치한 건물의 실내에서 실행하는 경우가 많다. 포스퀘어는 베뉴를 초기 생성할 때 지도 상에 모바일 기기의 GPS 장치를 통해 측정된 위치를 핀 형태로 우선적으로 보여주며, 사용자는 핀을 드래그 하여 원하는 위치로 수정할 수 있고 해당 위치에 베뉴를 생성할 수 있다. 지도학적인 전문 지식이 적은 일반인들은 해당 건물 실내에 위치하고 있기 때문에, 베뉴 초기 생성 시 GPS 장치를 통해 측정된 자신의 위치가 건물 위치라고 간과하기 쉬우며 그 위치 그대로 베뉴를 생성하는 경우가 많다. 따라서 GPS 오차가 심한 경우에는 베뉴의 위치가 부정확하게 생성될 수밖에 없다. 모바일 기기의 GPS 위치 정확도는 건물 및 지형·지물들로 인해 위성으로부터 송신된 GPS 신호가 굴절, 반사되고, 기타 다른 요인 등 주변 환경에 따라 낮아지게 된다. 통상적으

로 좋은 환경 조건에서 GPS와 WiFi 기반의 측위 시스템은 각각 10m 와 40m 정도의 위치 오차를 지닌 것으로 알려져 있다(Paek, J. et al., 2010). 그리고 사람들이 해당 장소를 다른 베뉴로 착각하여 매우 부정확한 위치에 베뉴를 생성하는 등 여러 요인들을 고려하였을 때, 약 50m의 유클리드 거리 오차가 포스퀘어 베뉴 데이터에서 충분히 발생할 수 있음을 유추할 수 있었다.

이와 같은 위치 오차를 지니고 있는 포스퀘어 데이터를 그대로 이용한다면 최종 연구 결과에서 심각한 오류를 발생시킬 우려가 있다. 따라서 본 연구에서는 포스퀘어 데이터의 위치 오차를 보정해주기 위한 전처리 과정으로써 ① 포스퀘어 데이터에 등록되어 있는 주소활용, ②네이버 지역 검색 API 활용, ③웹 포털 지도서비스를 활용한 주소 검색, ④수집된 주소 정보들을 지도 좌표로 지오코딩 등 이상 4가지 방법을 단계적으로 실행하였다.

Yanhua Li et al에 의하면 베뉴의 프로필 정보(주소, 전화번호, 카테고리 등)의 완결성이 높을수록, 해당 베뉴의 체크인 수나 팁 수가 많은 유명한 장소임을 실험을 통해 설명하였다. 유명한 장소들은 그 만큼 많은 사람들이 참여하여 해당 장소와 관련된 정보들을 등록, 수정하며 정확하고 최신의 정보가 유지되는 것이다. 유명한 베뉴들은 프로필 정보 중 주소 역시 자세하고 신뢰할만한 수준으로 정보가 등록되어 있다[그림 2-7]. 따라서 우선적으로 포스퀘어 데이터에 등록되어 있는 상세한 수준의 주소들을 지오코딩 하여 정확한 경위도 좌표를 수집하였다.

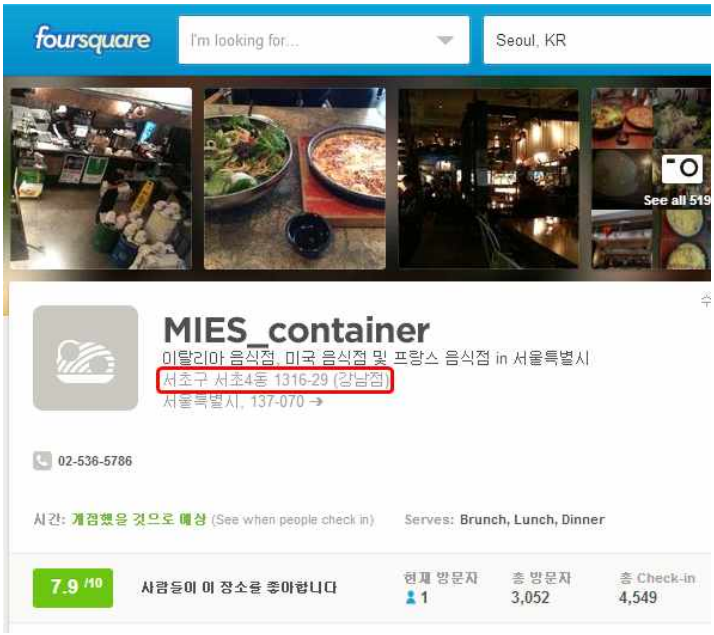


그림 2-7. 상세한 수준의 주소 정보가
등록되어 있는 포스퀘어의 베뉴 예시

다음 단계는 베뉴 데이터의 프로필 정보 자체만으로 해당 주소를 알지 못한 베뉴들을 네이버 지역 검색 API[9]를 활용하여 주소를 수집하였다. 네이버 지역 검색 API는 원하는 키워드 검색을 URL(Uniform Resource Locator) 형태로 요청하여 네이버 지역 검색 서비스에 등록한 업체들의 정보를(검색 결과 업체명, 업체 설명, 전화번호, 주소, 카텍 x좌표, 카텍 y좌표) XML(Extensible Markup Language) 형태로 제공해준다. 본 연구는 베뉴 이름을 키워드로 검색하여 주소 결과 값을 수집하였다.

베뉴 데이터에 입력된 주소 정보와 네이버 지역 검색 API를 활용하여도 주소 정보를 얻지 못한 베뉴들은 주요 웹 포털 서비스인 네이버와 다음 지도 서비스의 검색을 활용하여 주소 정보를 수집하였다.

이렇게 수집된 베뉴의 주소 정보들은 웹 기반의 GIS 어플리케이션인 'X-Ray Map for Web'의 지오코딩 툴을 통해 일괄적으로 경위도 좌표로 변환되었다. 하지만 위의 과정을 거쳐도 모든 베뉴들의 주소 정보를 수집할 수

9) http://dev.naver.com/openapi/apis/search/local 에서 제공

는 없었다. 그 주요 원인으로는 주소 정보 수집 과정 상 웹 포털 정보에 많이 의존하고 있기에 잘 알려지지 않은 곳은 검색이 되지 않는 경우가 많고, 이미 폐점한 베뉴들이 포스퀘어 상에 상당히 존재하기 때문이다. 이러한 문제로 정확한 주소를 얻지 못한 베뉴들의 비율이 상당부분 차지한다면 연구 데이터로서의 가치가 현저히 떨어질 것이다. 하지만 [표 2-2]에서 확인할 수 있듯이 위의 전처리 과정을 거쳐 베뉴의 정확한 주소 정보를 알 수 있는 데이터의 비율이 약 80~90%로 조사되었으며, 연구 결과를 도출할 만한 충분한 양·질의 데이터임을 확인할 수 있었다.

표 2-2. 테마 & 지역별 정확한 주소 정보를 획득한 베뉴의 비율

테 마 / 대상지역	음식점(%)	패션(%)	카페(%)	유흥주점(%)
신림 사거리	312/357 (87)	24/31 (77)	65/71 (92)	108/131 (82)
강남 번화가	1570/1815 (87)	187/207 (90)	400/447 (89)	534/624 (86)

본 연구에서는 위의 과정을 거쳐 위치 오차 보정을 마친 베뉴 데이터들만을 활용하였으며, 최종적으로 베뉴들의 경위도 좌표를 도로명 주소 기본도의 좌표체계(UTM-K 좌표계, Bessel 1841 타원체, TM 투영)로 투영시켜 사용하였다. [그림 2-8]은 앞의 전처리 과정을 거쳐 신림 사거리 지역의 음식점 베뉴 데이터들이 정확한 좌표로 위치보정 된 결과를 보여준다. 위치보정 후 시각적으로 베뉴 데이터 수가 적어진 것처럼 보이는 이유는 같은 주소 및 건물에 위치한 여러 베뉴들이 겹쳐서 한 지점에 존재하고 있기 때문이다.

(a) 위치보정 적용 전 베뉴 분포 (b) 위치보정 적용 후 베뉴 분포

그림 2-8. 신림 사거리 지역의 음식점 베뉴 데이터를 위치보정 적용한 결과

2.2.2 건물-베뉴 데이터의 공간 조인 연산

도로명 주소 기본도의 데이터는 시·도를 경계로 하여 생성되었으며, 건물 데이터 또한 전국이 아닌 시 단위로 관리되고 있다. 시도 내의 모든 건물 데이터를 이용하여도 상관없지만 불필요한 데이터 이용을 방지하기 위해, 핫스트리트 군집을 탐지할 대상 지역 범위에 속한 건물들만 선택하여 별개의 레이어로 생성하고 이를 활용하였다.

포스퀘어 베뉴 데이터는 각종 업소 및 상점들을 점 데이터로 표현하였지만, 이들은 실세계에서는 건물 내부에 존재하거나 배치되어 있다. 따라서 각 건물들은 포스퀘어 베뉴 데이터가 가진 정보를 대표하는 객체로서 생각될 수 있으며, 이는 이전 과정을 거쳐 정확한 위치로 위치 보정된 점 형태의 베뉴 데이터와 폴리곤 형태인 건물 데이터간의 공간 조인 연산을 통해 처리할 수 있다. 공간 조인 연산은 어떤 객체의 속성 정보에 그 객체와 특정 공간 조건(포함, 교차, 겹침 등)을 만족하는 다른 객체의 속성 정보를 연결시키는 연산이다. 이를 통해 생성된 공간 조인 결과물은 건물 레이어 정보에 각 건물에 위치한 포스퀘어 베뉴의 [표 2-1] 속성 정보들을 추가로 지니게 된다.

공간 조인 연산을 제대로 수행하기 위해서는 모든 베뉴 데이터가 건물 폴리곤 내부에 위치해야 한다. 하지만 건물 내부에 위치하지 않은 베뉴 데이터들이 몇몇 존재하였다. 이러한 현상이 발생하는 이유는 지오코딩을 하기 위한 주소 정보가 지번 주소이고 도로명 주소 기본도가 부정확하기 때문이다.

지번이란 부동산공시법에 의하면 필지(하나의 지번에 붙는 토지의 등록단위로 토지소유자의 권리를 구별하기 위한 표시단위)에 부여하여 지적공부에 등록한 번호를 말한다. 따라서 건물단위로 주소를 결정한 도로명 주소와 달리 지번주소로 지오코딩을 실행하면 필지 중심의 점 좌표로 변환이 되고, 필지의 일부분만 사용하는 건물일 경우에 지오코딩 결과가 건물 외부에 존재하게 되는 것이다[그림 2-9]. 이러한 베뉴 데이터들은 수동으로 해당 건물 객체 내부로 위치를 수정하여, 공간 조인 연산에 오류가 생기지 않도록 하였다.

[그림 2-10]는 도로명 주소 기본도의 건물 데이터가 실제 건물의 모양과 다르게 생성된 것을 나타내고 있다. [그림 2-10]에서 사용한 도로명 주소 데이터는 2013년 2월 데이터로서 본 연구 시점을 기준으로 최신의 데이터이지만, 실제 존재하는 건물과 도로명 주소 기본도의 건물이 상이하다. 이러한 지도 데이터 상의 오류로 인해 베뉴 점 데이터가 건물 내부에 위치하지 않는 경우가 발생하는 것이다. 도로명 주소 기본도는 정부에서 지속적으로 갱신을 하여 지도의 정확도를 향상시키고 있어 장기적으로 보았을 때 이런 문제는 해결될 것으로 보인다. 하지만 연구 당시에는 실세계와 다른 건물 데이터를 이용함으로써 핫 스트리트 군집 탐지 시에 정확도를 떨어뜨릴 수 있는 한계점을 시사한다.

그림 2-9. 건물 내부에 위치하지 않은 베뉴 점 좌표 예시 1

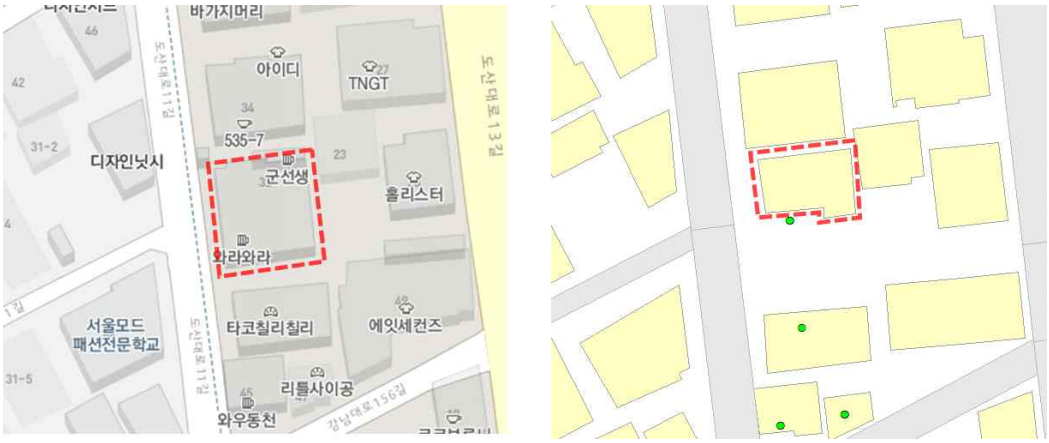


그림 2-10. 건물 내부에 위치하지 않은 베뉴 점 좌표 예시 2

2.3 건물과 도로 세그멘트간의 매칭

연구의 최종 목표는 핫 스트리트 군집을 탐지하는 것으로써, 군집 분석을 하기 위해서는 도로 구간 레이어를 연구 기본 단위인 도로 세그멘트로 재구성해야 한다. 그리고 김경수(2011)가 언급하였듯이 도로 주변의 건물들은 명백히 도로의 특성과 공간성을 형성하는 중요한 요소로 작용한다. 따라서 본

연구에서는 도로 세그멘트와 그 도로 세그멘트에 영향을 미치는 주변 건물들을 결정하고 매칭 시킨다. 궁극적으로는 공간 조인된 건물 데이터의 베뉴 정보를 매칭 쌍을 통해 도로 세그멘트에 연결시킨다.

2.3.1 도로 세그멘트 생성

도로명 주소 기본도의 도로구간 레이어는 폴리 라인 형태로 도로구간 정보를 구축한 벡터 지도이며, 건물 레이어와 마찬가지로 시·도 단위로 데이터가 구축되어 있다. 따라서 효율적인 데이터 처리를 위해 분석하고자 하는 지역 내의 도로들만 선택하여 별개의 레이어를 생성하고 이를 활용하였다.

도로구간은 도로의 시작지점부터 끝지점을 연결한 가상의 연속된 구간을 말하며, 각 도로구간마다 고유의 도로명을 가지고 있다. 도로구간 레이어를 그대로 활용하기에는 도로구간 한 객체의 길이가 연구 대상지역에 비해 과도한 크기를 가지고 있는 문제점이 있다. [그림 2-11]에서는 대로 하나의 객체가 연구대상범위를 한 번에 가로지르거나, 길 하나의 객체가 4~5개의 블록을 가로지르고 있는 것을 살펴볼 수 있다. 이러한 특성을 지닌 도로구간 레이어의 데이터를 그대로 이용하기에는 연구 분석 단위로서 연구 대상 범위에 비해 도로의 길이가 너무 크다. 또한 다양하고 자세한 수준의 핫 스트리트 군집을 생성하기보다, 도로구간이 길수록 주변 건물과 베뉴 데이터에 영향을 더 많이 받기 때문에 긴 도로구간이 핫 스트리트로 탐지되는 무의미한 결과가 도출될 수 있다. 따라서 본 연구에서는 연구 분석 단위를 좀 더 세분화 시킬 필요가 있으며, 교차로를 분절 기준점으로 하여 각 도로구간을 도로 세그멘트 단위로 세분화 시켰다[그림 2-12]. 교차로를 기준으로 도로구간을 분절한 이유는 사람들이 도로를 따라 통행할 때 일반적으로 교차로에서 통행의 방해를 받고, 이는 일관된 장소의 인식을 단절시키기 쉽기 때문이다(Lu, 2005).

단, 도로 세그멘트의 길이가 10m 이하인 객체들을 제거하고, 이웃한 도로 세그멘트와의 위상을 유지하도록 수정하였다. 길이가 10m 이하인 도로 세그멘트를 제거한 이유는 10m 이하의 도로 세그멘트들은 길이가 너무 짧아 하나의 독립적인 분석단위로써 부적절하고, 도로명 주소 기본도 제작 시 발생

한 오버슈트(overshoot) 등의 디지타이징(digitizing) 오류에 의해 생성된 미세한 도로 세그멘트 객체들을 제거하기 위해서이다.

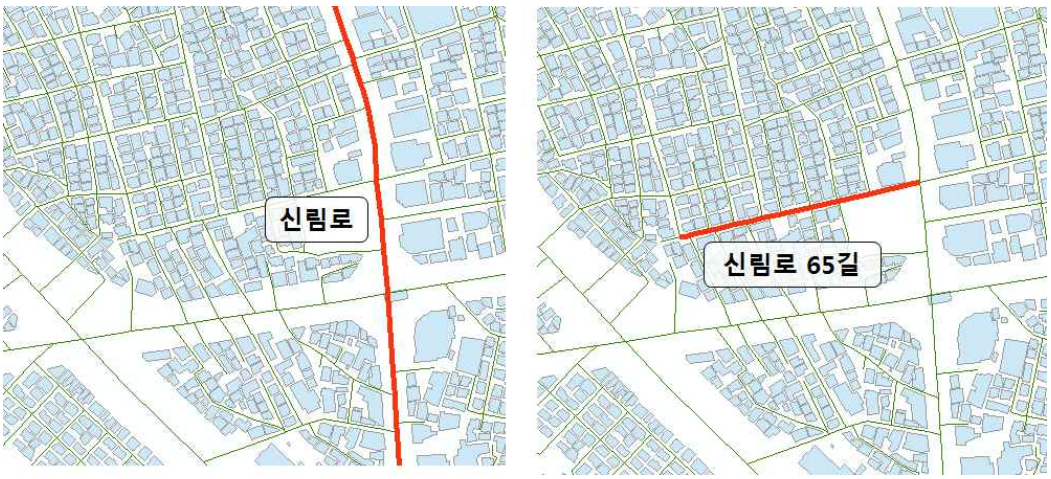


그림 2-11. 신림 사거리 지역의 도로구간 예시

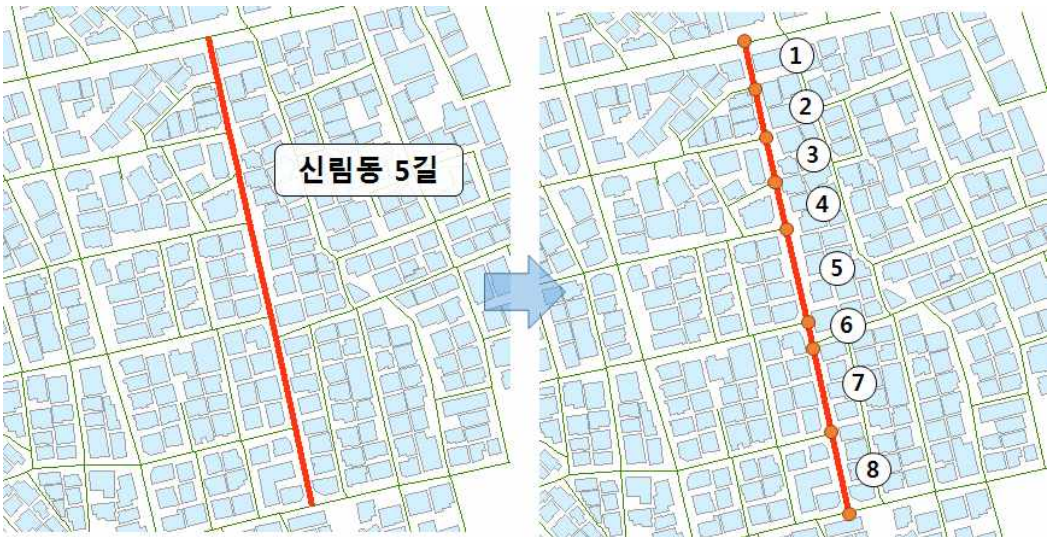


그림 2-12. 교차로를 기준으로 도로구간을
8개의 도로 세그멘트로 분절시킨 예시

2.3.2 건물-도로 세그멘트의 매칭 조건 및 기법

본 연구에서는 도로 세그멘트와 그 도로 세그멘트에 영향을 끼친다고 생각되는 주변 건물을 매칭 시키는 조건을 '건물이 도로에서 정면으로 보여야 하며, 건물과 도로는 바로 맞닿아 있어야 한다' 로 정의하였다.

본 연구에서 건물과 도로가 맞닿는다고 하는 의미는 물리적, 기하학적으로 접촉한 상태를 의미하지는 않는다. 건물과 도로 사이에는 보도와 건물전면공간[10]에 의해서 실제로 맞닿기에는 여유 공간이 있기 때문이다. 따라서 본 연구에서 건물과 도로가 맞닿는다는 의미는 거리상 가까이 있으면서 건물의 중심을 기준으로 다른 도로나 건물에 의해 막힘없이 연결될 수 있는 상태를 의미한다. [그림 2-19]는 건물과 도로 세그멘트가 맞닿은 상태를 설명하는 그림이다. 그림 가운데에 있는 건물은 ①번 도로 세그멘트에만 맞닿아 있고 나머지 도로 세그멘트에는 주변 건물에 의해 시야나 진입로가 제한되어 있는 것을 알 수 있다.

10) 건물전면공간 : 건물전면공간은 미관지구나 지구단위계획 등의 규제에 의해 건축선이 대지 경계선에서 3~5m 정도 후퇴 지정되면서, 보도와 건물 사이에 형성되는 공지를 말한다.

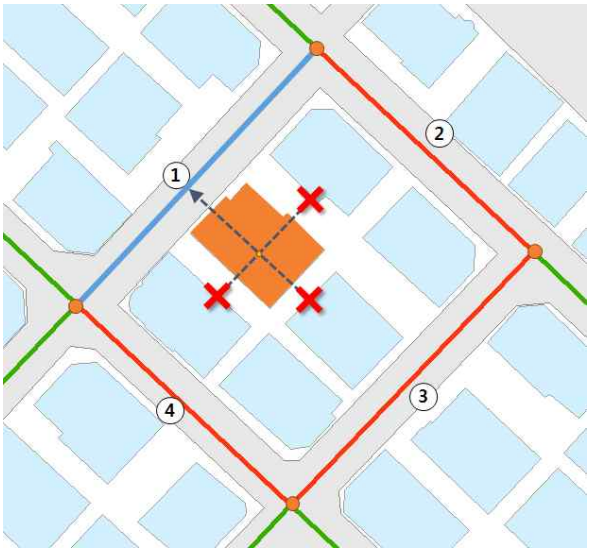


그림 2-13. 건물과 도로가 맞닿아 있는 상태를 설명하는 예시

본 연구는 위의 매칭 조건을 만족시키는 건물과 도로 세그멘트를 매칭 시키기 위해 ①도로명 주소 기본도의 출입구 정보 이용, ②버퍼(buffer) 연산, ③ISOVIST 등 이상 3가지 방법들을 비교·분석하여 가장 적절한 방법을 선택하였다.

1) 도로명 주소 기본도의 출입구 정보 이용

도로명 주소 기본도에는 건물의 출입구 정보를 출입구 레이어와 연결선 레이어를 통해서 제공하고 있다. 출입구 레이어는 각 건물의 출입구를 점 데이터로 표시하고 있으며, 연결선 레이어는 이 출입구 점 데이터와 도로구간을 연결하는 선형 데이터이다[그림 2-14].

출입구와 연결선 레이어는 도로와 건물을 매칭 시키기 위한 좋은 데이터임에는 틀림없지만, 이 데이터를 이용하기에는 한계점이 있다. 각 건물마다 대부분 하나의 출입구 정보만을 제공하고 있고, 이는 각 건물마다 도로명 주소를 부여할 때 그 주소에 부여된 도로의 방향으로만 출입구(주출입구)를 구축해 놓은 것으로 판단된다. <부록 D>를 참고하여 살펴보면 출입구 레이어의

주요 속성에서 주출입구와 보조출입구 등을 구분할 수 있는 'ENT_SE'(구분코드) 필드가 존재하고 있음을 알 수 있지만, 서울시를 기준으로 약 60만개의 출입구 중 300개 정도만이 보조출입구로 분류되어 있는 실정이다.

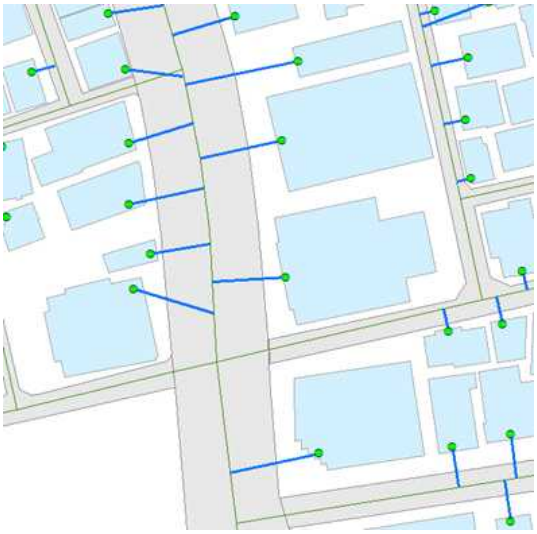


그림 2-14. 도로명 주소 기본도의 출입구와
연결선 레이어 적용 예시

하지만 [그림 2-15]와 같이 실제로는 그 건물로 출입할 수 있는 출입구는 건물의 다른 면에 존재할 수도 있고, 하나 이상인 경우가 대부분이다. 따라서 도로명 주소 기본도의 출입구 정보만 이용한다면 굉장히 부족하고 부정확한 정보를 이용하는 것이며, 도로명 주소 데이터에 너무 의존적인 핫 스트리트 군집 결과가 나올 것이다.

그림 2-15. 출입구 데이터가 없는 다른 건물면에
존재하는 실제 상점들과 출입구

2) 버퍼 연산 이용

버퍼 연산이란 특정 객체를 중심으로 일정 거리 혹은 시간만큼 확장시킨 영역을 말한다. 건물과 도로 세그멘트를 매칭시키기 위해서 도로 세그멘트 중심 혹은 건물 중심으로 버퍼 연산을 할 수 있다. 도로 세그멘트 중심으로 버퍼연산을 한다면, 도로 세그멘트의 실제 도로 폭 보다 더 여유있게 버퍼 연산을 수행하여 그 버퍼 영역과 교차(intersect)하는 건물들이 도로와 맞닿아 있는 건물들로 판단, 매칭 시켜주는 것이다. 그리고 반대로 건물 중심의 버퍼연산을 실행한다면, 건물 크기보다 더 크게 버퍼 연산을 하여 그 버퍼 영역과 교차하는 도로들과 매칭 시켜준다.

하지만 버퍼 연산을 통해 매칭시키는 방법은 잘못된 매칭과 과도한 매칭 결과를 도출하기 쉽다. [그림 2-16]는 도로 중심의 버퍼 연산 시 발생되는 매칭 오류를 보여준다. 그림은 중심에 위치한 도로 세그멘트를 중심으로 실폭 + 10m의 버퍼 연산을 적용시켰으며, 녹색으로 색칠된 건물들은 버퍼 영역과 교차하고 있으며 도로와 맞닿는다고 올바르게 판단된 건물들이다. 하지만 빗금면 건물들은 도로 세그멘트와 맞닿는다고 판단되기에는 어려운 건물들이지만 버퍼 영역과 교차한 건물들이다. 특히 도로 세그멘트 양 끝단(교차로 지점)에서 이러한 오류가 발생하기 쉬우며, 그림 중간에 위치한 빗금면 건물은

버퍼영역과 미세하게 교차하여 매칭 건물로 선택된 결과를 보여준다.

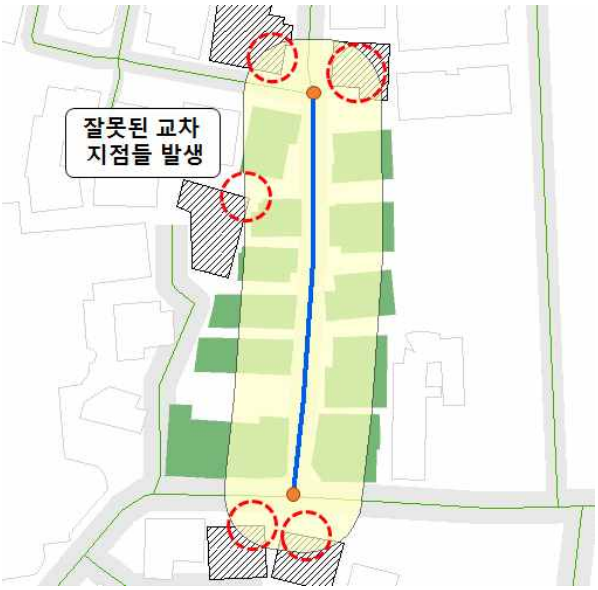


그림 2-16. 도로 중심의 버퍼 연산시 발생되는
매칭 오류

[그림 2-17]은 건물 중심의 버퍼 연산 수행 시 발생하는 매칭 오류를 보여준다. 그림은 색칠된 건물 객체 주변으로 10m의 버퍼 연산을 적용시켰다. 굵은 녹색 선으로 표시된 도로 세그멘트는 버퍼 영역과 교차하고 있으며 건물과 맞닿는다고 올바르게 판단된 도로이다. 하지만 붉은색 점선으로 표시된 도로 세그멘트는 건물과 맞닿는다고 판단되기에는 어려운 도로지만 버퍼 영역과 교차함으로써 매칭 대상으로 선택된 도로이다. 이렇게 교차로 지점에서는 건물 중심의 버퍼 영역에 따라 교차로의 모든 도로 세그멘트가 매칭 대상으로 선택되는 오류가 발생한다.

그림 2-17. 건물 중심의 버퍼 연산 수행 시
발생되는 매칭 오류

이러한 오류들은 버퍼 적용 거리를 조절해가며 해결할 수 있겠지만, 연구 대상 지역에 따라 건물 크기, 외형 또는 건물전면공간이 천차만별이기 때문에 일정 거리의 버퍼를 적용하는 것은 근본적인 해결책이 될 수 없다. 그리고 이 결과를 그대로 적용하게 되면, 교차로 부근에 있는 건물들이 잘못된 도로 세그멘트와 매칭 되었기 때문에 정확한 핫 스트리트 군집을 탐지할 수 없을 것이다.

3) ISOVIST 이용

ISOVIST는 Benedikt(1979)에 의해 제안되었으며, 공간 내에서 특정 지점(조망점)으로부터 장애물에 의해 가려지지 않고 보이는 모든 지점들을 연결한 선들의 집합으로 이루어진 가시 공간을 말한다. [그림 2-18]은 건물들이 주변에 위치하고 있을 때 임의의 조망점에서 바라본 ISOVIST 공간을 나타낸 그림이다.

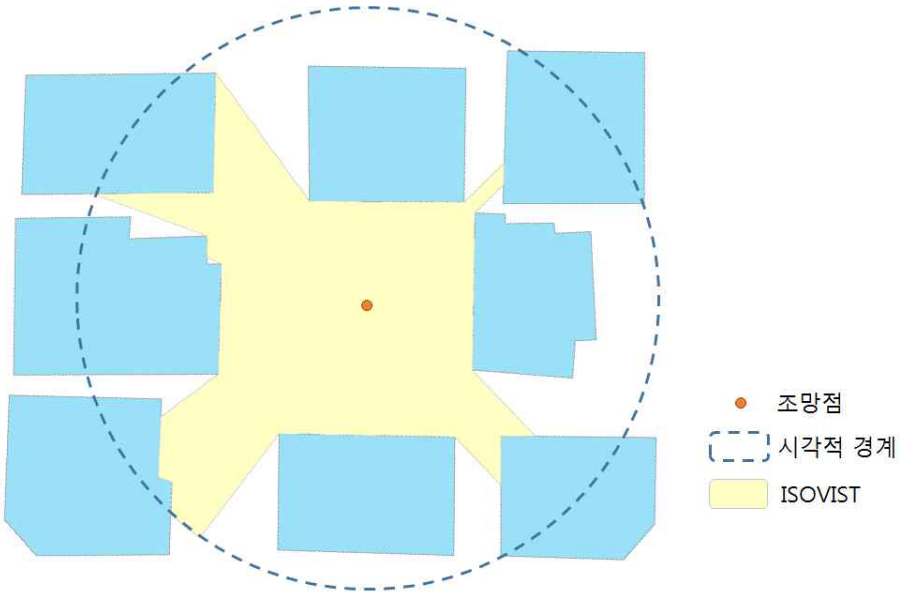


그림 2-18. ISOVIST 공간 예시

ISOVIST 공간을 이용하여 건물과 도로 세그멘트를 매칭하는 과정을 간단히 설명하면 다음과 같다. 우선 대상 지역의 건물들과 실폭 도로를 병합하여 ISOVIST 공간에 장애가 될 레이어를 생성한다. 그리고 실폭 도로 객체를 제외하고 건물들의 중심에서 각 건물의 크기를 고려한 시야거리만큼 ISOVIST 공간을 생성한다. 그 결과 생성된 ISOVIST 공간과 교차하거나 만나는 실폭 크기의 도로 세그멘트는 ISOVIST 공간 안에 있는 건물과 매칭 쌍이 된다.

건물과 도로 세그멘트의 매칭을 위해 ISOVIST를 이용함으로써 도로 위에 있는 사람들이 시야 범위 내에 보이는 업소의 간판이나 건물을 인지하여 원하는 건물로 향하는 과정 또한 직관적으로 설명된다. 그리고 건물 중심에서 생성된 ISOVIST 공간은 다른 건물들에 의해 시야가 제한이 되며 버퍼 연산을 이용하였을 때 문제시 되던 오류인 도로와 맞닿아 있지 않은 건물과 도로 세그멘트의 잘못 된 매칭을 좀 더 효과적으로 줄일 수 있다[그림 2-19(a)]. 또한 ISOVIST를 이용하는 과정에서 실폭 도로를 가시권에 장애가 되는 객체로 간주하였기 때문에 교차로 지점의 모든 주변 도로가 매칭 대상으로 선

정되는 오류를 해결할 수 있다[그림 2-19(b)].

(a) 다른 건물들에 의해 제한된 가시권 형성으로 굵은 녹색선의 도로 세그멘트에만 매칭된 올바른 결과

(b) 실폭 도로에 의해 제한된 가시권 형성으로 굵은 녹색선의 도로 세그멘트에만 매칭된 올바른 결과

그림 2-19. ISOVIST를 이용한 건물-도로 세그멘트 매칭 예시

하지만 ISOVIST를 이용하는 방법도 건물과 도로 세그멘트의 완벽한 매칭 결과를 달성하지 못한다. ISOVIST 또한 버퍼 연산과 마찬가지로 시야 거리가 과도하게 크면 맞닿지 않는 도로 세그멘트에 매칭이 될 수 있고, 시야거리가 작으면 도로와 아예 매칭이 안 될 수 있다. 그리고 건물 중심을 기준으로 원형의 가시권을 형성하기 때문에 건물 외형의 기하학적인 특성을 반영하지 못하는 한계점을 지니고 있다.

2.3.3 ISOVIST를 이용한 건물-도로 세그멘트 매칭 과정

본 연구에서는 위에서 설명한 매칭 기법들을 비교·분석한 결과, 도로명 주소 기본도의 출입구 정보를 이용하면 실제의 정보와 달리 부족한 출입구 정보를 제공함으로써 충분한 매칭 결과를 생성하지 못하고, 버퍼 연산을 사용하면 주변 건물과 도로의 존재를 무시하는 가상의 버퍼 영역과 그 안에 교차하는 모든 도로 세그멘트와 건물 객체를 매칭대상으로 선택하게 되어, 부정

확한 매칭 결과를 초래하는 것을 살펴 볼 수 있었다. 그리고 ISOVIST를 이용하면 시야에 바로 보이는 건물과 도로 세그멘트를 매칭시킨다는 직관적인 개념을 적용시킬 수도 있고 부정확한 매칭 결과를 다른 방법에 비해 줄일 수 있을 것으로 판단되었다. 따라서 본 연구에서는 건물과 도로 세그멘트를 매칭시키는 방법 중 가장 최적의 방법이라고 생각되는 ISOVIST를 이용하여 건물과 도로 세그멘트 간 매칭을 실행하였다.

ISOVIST를 이용하여 건물과 도로 세그멘트를 매칭 시키는 프로세스는 [그림 2-20]와 같다.

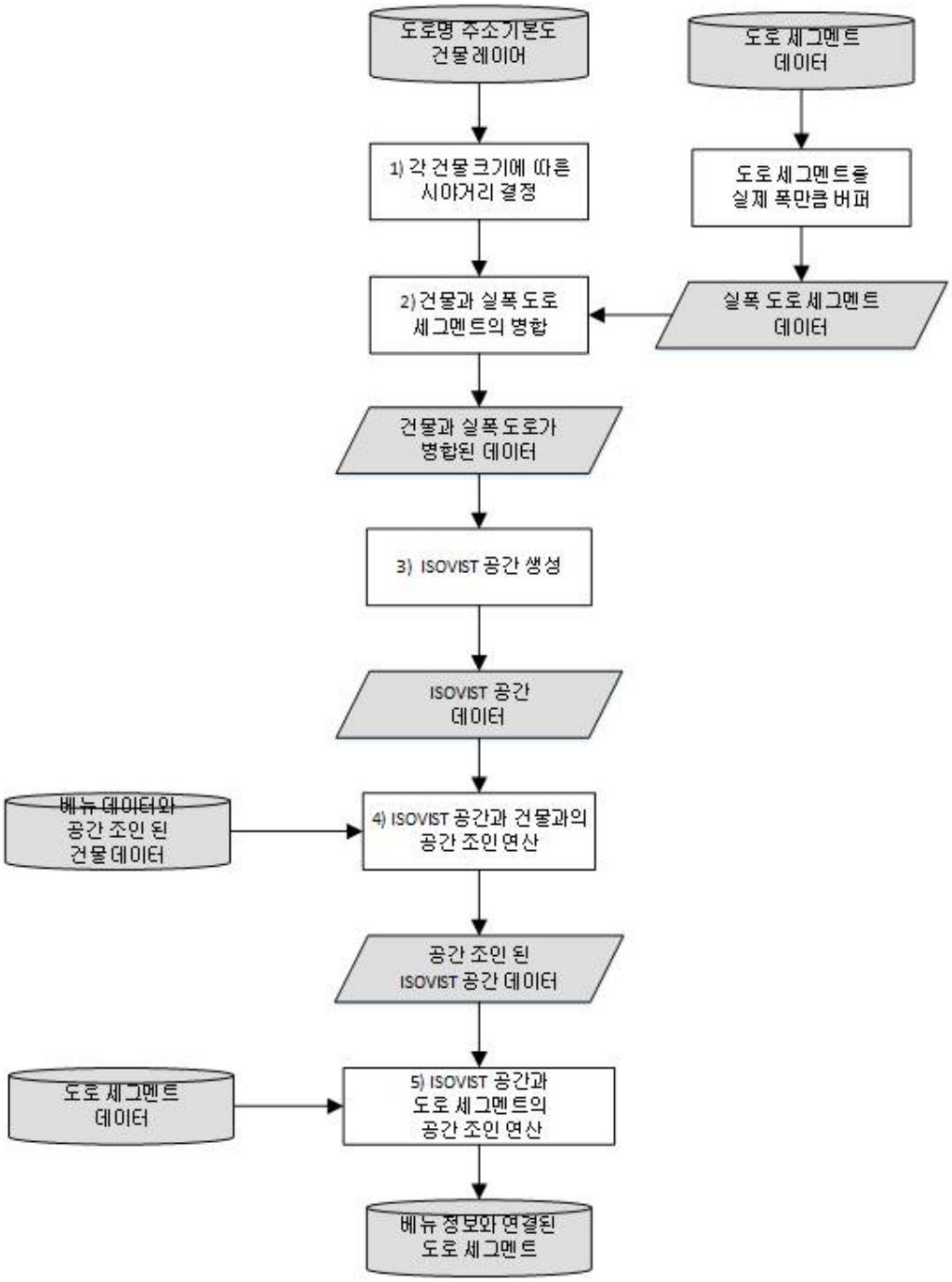


그림 2-20. ISOVIST를 이용한 건물-도로 세그멘트 매칭 프로세스

1) 각 건물 크기에 따른 시야거리 결정

ISOVIST를 이용하기 위해서는 조망점에서 보이는 시야거리를 설정해야 하고, 본 연구에서는 각 건물의 중심을 ISOVIST의 조망점으로 설정한다. 이 때 일정한 크기의 시야거리를 일괄적으로 설정하면, 건물의 외형과 크기가 천차만별이기 때문에 건물 내부에 ISOVIST 공간이 생기거나 아니면 건물 크기보다 과도한 크기의 ISOVIST 공간을 형성하게 된다. 이러한 문제를 방지하고 건물과 도로 세그멘트의 알맞은 매칭을 달성하기 위해 각 건물 크기에 따라 조절된 시야거리를 결정하였다.

현대식 건물 대부분은 사각형 형태의 부지를 차지하고 있고, 각 건물을 정사각형으로 가정한다면 정사각형 한단면의 길이, l은 건물 면적의 제곱근 길이이다. 건물 외곽선 내부에 가시권이 생기는 것을 방지하기 위해서는 시야거리가 $l(\text{정사각형한단면의길이}) \times \frac{\sqrt{2}}{2}$ 보다는 커야 한다. 그리고 건물과 도로 사이에는 건물전면공간과 보도가 존재할 수 있다. 건물전면공간은 보도와 건물 사이에 약 3~5m 정도의 간격을 차지하고, 보도는 “도로의 구조·시설 기준에 관한 규칙”에 따라 보도 유효폭 최소 2m 이상 차지하게 된다. 이 같은 도로와 건물 환경들을 반영하여 각 건물의 시야거리, *view_dist*를 식 (2-3)와 같이 계산하였다[그림 2-21].

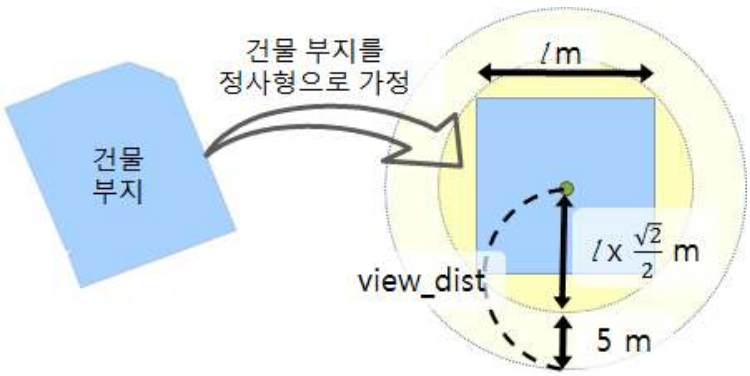


그림 2-21. *view_dist* 설명

$$view_dist\,(m) = l \times \frac{\sqrt{2}}{2} + 5 \qquad \text{식 (2-3)}$$

2) 건물과 실폭 도로 세그멘트의 병합

버퍼 연산과 같이 교차로 지점에 연결된 모든 도로 세그멘트가 건물의 ISOVIST 공간 안에 교차되어 잘못된 매칭 결과가 발생하는 것을 방지하기 위해([그림 2-17] 참고), 주변 건물 이외에도 실폭 도로 세그멘트 또한 장애물로 인식하여 ISOVIST 공간을 생성해야 한다. 따라서 도로명 주소 기본도의 건물 데이터와 도로 세그멘트를 실제 폭만큼 버퍼 시킨 실폭 도로 세그멘트 데이터를 병합(merge)시키고 이를 활용하였다.

3) ISOVIST 공간 생성

앞의 병합 과정을 통해 생성된 데이터를 이용하여 ISOVIST 공간을 생성하였다. 단, 실폭 도로 세그멘트는 가시권 생성을 제한하는 역할을 할 뿐이므로 실폭 도로 세그멘트를 제외한 건물 객체들만을 대상으로 ISOVIST 공간을 생성한다. [그림 2-22]은 ISOVIST를 실행하여 ISOVIST 공간을 생성한 결과이다. 그림과 같이 ISOVIST 공간은 해당 건물을 충분히 포함하고 주변 건물과 도로에 의해 시야가 제외된 형태로 생성된다.

그림 2-22. 건물 중심의 ISOVIST 공간 예시

4) ISOVIST 공간과 건물과의 공간 조인 연산

각 건물은 고유의 ISOVIST 공간을 가지게 된다. 이렇게 생성된 ISOVIST 공간은 공간 조인 된 건물 데이터가 지닌 베뉴 정보를 도로 세그멘트에 연결해주는 매개체의 역할을 해야 한다.

앞의 과정을 통해 생성된 ISOVIST 공간은 별다른 속성 정보를 지니고 있지 않는다. 따라서 ISOVIST 공간과 공간 조인 된 건물 데이터간의 공간 조인 연산을 통해 건물이 가지고 있는 베뉴 정보를 ISOVIST 공간 데이터에 부여하였다.

5) ISOVIST 공간과 도로 세그멘트의 공간 조인 연산

건물과 도로 세그멘트와의 실질적인 매칭 과정이다. ISOVIST 공간과 접하는 실폭 도로 세그멘트는 그 ISOVIST 공간의 중심 건물과 맞닿아 있는 상태이고, 매칭 조건을 충족시킨다. 그리고 건물의 베뉴 정보들은 앞의 과정에서 각각의 ISOVIST 공간에 부여되었다. 공간 조인 연산을 통해 ISOVIST

공간이 가지고 있는 건물의 베뉴 정보를 실폭 도로 세그멘트에, 그리고 도로 세그멘트로 연결시켰다.

즉, ISOVIST를 이용하여 건물과 도로 세그멘트를 매칭 시키고, 최종적으로 도로 세그멘트의 속성정보에 공간 조인된 건물 데이터가 가지고 있던 베뉴 정보[표 2-1]를 연결시켰다. 이렇게 편집된 도로 세그멘트 데이터를 활용하여 다음 핫 스트리트 군집 탐지 과정을 수행한다.

2.4 핫 스트리트 군집 탐지

2.4.1 도로 세그멘트의 핫 밸류(hot value) 계산

앞의 ISOVIST를 이용한 건물-도로 세그멘트 매칭 과정을 통해 생성된 결과는 도로 세그멘트 속성 정보에 [표 2-1]의 베뉴 정보가 단순히 추가된 것이다. 하지만 핫 스트리트 군집을 탐지하기 위해서는 각 도로 세그멘트가 핫 스트리트로 판단될 고유의 값을 가지고 있어야 한다. 본 연구에서는 이 고유값을 '핫 밸류(hot value)' 로 명명하였다.

사람들이 많이 방문하고 활동하여 활성화된 테마거리를 찾기 위해 각 도로 세그멘트의 핫 밸류는 다음과 같은 요소들을 고려하여 계산하였다.

① 도로 주변의 베뉴들을 사람들이 얼마나 많이 방문하는가?
(베뉴들의 평균 인기도)
② 도로와 맞닿은 주변 건물들 중 베뉴가 위치한 건물들이 얼마나 많이 차지하고 있는가? (베뉴 건물들의 구성비)
③ 도로 세그멘트의 단위 길이 당 베뉴가 위치한 건물 수는 얼마나 많은가?
(베뉴 건물들의 밀집도)

1) 베뉴들의 평균 인기도(P_i)

포스퀘어 서비스 이용자들은 특정 베뉴에 대한 관심과 흥미를 체크인과 팁을 남김으로써 표현한다. 특히 체크인 수가 높은 베뉴는 그만큼 사람들이 그

베뉴를 많이 방문했다는 것을 의미한다. 그리고 팁이 많이 남겨진 베뉴는 사람들이 그 베뉴에 특별한 관심을 가지고 베뉴를 방문했을 때의 경험과 정보를 지인들과 공유하고 싶어하는 것을 의미한다(Yanhua Li et al., 2013). 또한 최근에 추가된 like or dislike 기능을 통해 특정 베뉴의 평판을 살펴볼 수 있다.

위와 같이 특정 베뉴에 등록된 체크인, 팁, like의 수치는 사람들이 해당 베뉴를 얼마나 선호하고 관심있어 하는지를 직관적으로 알려주는 척도이다. 본 연구에서는 이 수치들을 이용하여 도로 주변에 위치한 베뉴의 인기도를 반영한다. 체크인, 팁, like 각각의 세부적인 의미는 다르지만 공통적으로 사람들이 해당 베뉴에 관심을 갖고 긍정적인 생각을 표현한 행동이기 때문에 세 수치의 합을 통해 베뉴의 인기도($N_{i,j}$)를 나타내었다.

$$N_{i,j} = check_{i,j} + tip_{i,j} + like_{i,j} \qquad \text{식 (2-4)}$$

$N_{i,j}$는 분석 대상 지역에 n개의 도로 세그멘트가 존재할 때 i번째의 도로 세그멘트 주변에 위치한 베뉴들 중 j번째 베뉴의 인기도를 의미한다. $check_{i,j}$와 $tip_{i,j}$, $like_{i,j}$는 각각 j번째 베뉴의 등록된 체크인, 팁, like 수를 의미한다.

도로 세그멘트와 매칭된 건물들에는 여러 베뉴들이 동시에 위치하고 있다. 해당 도로 세그멘트 주변에 위치한 베뉴들의 종합적인 인기도를 고려하기 위해서 본 연구에서는 베뉴 인기도($N_{i,j}$)의 평균을 사용하였다. 즉, 베뉴들의 평균 인기도가 높은 도로 세그멘트 일수록 주변에 유명한 업소들이 많이 위치하고 사람들이 그 도로를 상대적으로 특수하게 생각할 것이다. 베뉴들의 평균 인기도는 다음과 같이 계산된다.

$$P_i = \frac{\sum_{j=1}^{k} N_{i,j}}{k} \qquad \text{식 (2-5)}$$

P_i는 분석 대상 지역에 n개의 도로 세그멘트가 존재할 때 i번째 도로 세그

멘트의 평균 인기도를 의미한다. k는 i번째 도로 세그멘트 주변에 위치한 베뉴들의 전체 수를 나타낸다.

2) 베뉴 건물들의 구성비(R_i)

도로 주변의 건물들 중 동일 테마의 베뉴들이 위치한 건물들이 많을수록 그 도로는 해당 테마의 특성을 대표할만한 도로로서 사람들에게 인식될 것이다. 이러한 도로들이 모여 단순히 도로가 아닌 특수한 장소성을 획득하고 더 나아가 테마거리, 특화거리를 형성하게 될 것이다.

이러한 도로의 특성은 다음의 식으로 계산 될 수 있다.

$$R_i = \frac{BV_i}{BN_i} \qquad \text{식 (2-6)}$$

R_i는 i번째 도로 세그멘트의 베뉴 건물들의 구성비를 의미한다. BN_i는 i번째 도로 세그멘트와 맞닿은 주변 건물의 수를 나타내고, BV_i는 i번째 도로 세그멘트와 매칭된 건물 중 베뉴가 위치한 건물의 수이다. R_i가 높을수록 해당 도로 세그멘트 주변에는 분석 테마의 베뉴 건물들이 많이 밀집해 있다고 판단할 수 있고, 핫 스트리트로 선별될 가능성이 크다.

3) 베뉴 건물들의 밀집도(D_i)

베뉴들의 평균 인기도와 건물들의 구성비 두 가지만을 고려하여 핫 밸류를 계산한다면, 다음과 같은 잘못된 결과를 얻을 수 있다.

(a) 큰 건물이 위치한 도로 (b) 평균 크기의 건물들이 밀집한 도로

그림 2-23. 베뉴 건물들의 구성비가 같지만 밀집도가 다른 두 도로 세그멘트

[그림 2-23]은 비슷한 길이의 도로 세그멘트가 위치한 지역의 스트리트 뷰 사진이다. 왼쪽 하단의 지도에서 실선 테두리와 점이 위치한 건물은 베뉴가 위치한 건물을 의미하고, 점선 테두리의 건물은 베뉴가 위치하지 않은 건물이다. [그림 2-23(a)]는 도로 주변에 총 3개의 건물과 2개의 베뉴 건물을 나타내고 있고, [그림 2-23(b)]는 도로 주변에 총 6개의 건물과 4개의 베뉴 건물을 나타내고 있다. 각 도로 세그멘트의 베뉴 건물들의 구성비(R_i)는 $\frac{2}{3}$로 같다. 만약 각 도로 세그멘트의 평균 인기도 또한 우연히 같다면, 두 도로 세그멘트의 핫 밸류는 같은 값을 지니게 될 것이다.

하지만 [그림 2-23(b)] 지역은 [그림 2-23(a)] 지역보다 더 많은 건물들이 밀집해 있고, [그림 2-23(b)]지역의 사람들은 각 건물마다 설치한 업소들의 간판 또는 출입구를 시각적으로 더 많이 확인할 수 있다. 그리고 비슷한 크기의 건물들이 밀집 할수록 도로 구간 전체에 걸친 건물들의 조화와 질서, 통일감이 더욱 형성될 수 있고, 가로공간을 형성하는데 주요한 영향을 끼친다. 따라서 밀집된 건물의 경관은 사람들로 하여금 특별한 장소성을 도로에 좀 더 쉽게 형성할 수 있도록 할 것이다.

이러한 이유로 도로 주변에 건물이 밀집되어 있는 정도 또한 핫 밸류를 계산하는데 고려해 주어야 한다. 따라서 본 연구에서는 밀집도를 도로 세그멘

트 단위 길이당 베뉴 건물의 수로 정의하고 적용하였다. (베뉴의 인기도, 베뉴 건물의 구성비 등 베뉴의 정보를 이용한 연구의 방향에 맞도록 주변 건물의 수 대신 베뉴 건물 수를 활용하여 밀집도를 계산하였다.) 베뉴 건물들의 밀집도(D_i)의 계산은 다음과 같다.

$$D_i = \frac{BV_i}{length_i} \qquad \text{식 (2-7)}$$

BV_i는 i번째 도로 세그멘트와 매칭된 건물 중 베뉴가 위치한 건물의 수이고, $length_i$ 는 I번째 도로 세그멘트의 길이이다.

베뉴 건물들의 밀집도(D_i)는 도로 세그멘트 주변에 얼마나 많은 베뉴 건물들이 위치하는지 다른 도로 세그멘트와 비교할 수 있는 직접적인 척도이다. 또한 베뉴들의 평균 인기도와 구성비가 같은 도로 세그멘트들을 차별할 수 있도록 해주는 역할을 한다.

4) 핫 밸류(H_i) 계산

도로 세그멘트의 핫 밸류는 앞서 설명한 베뉴들의 평균 인기도(P_i), 베뉴 건물들의 구성비(R_i), 베뉴 건물들의 밀집도(D_i)를 모두 반영한 값이어야 하며, 다음과 같이 세 지수의 곱으로 계산하였다.

$$H_i = P_i \times R_i \times D_i = \frac{BV_i^2}{BN_i} \frac{P_i}{length_i} \qquad \text{식 (2-8)}$$

H_i 는 i번째 도로 세그멘트의 핫 밸류 이다. 핫 밸류 값이 높을수록 분석 테마의 상점 및 업소들이 도로 주변에 밀집하고 있으며, 사람들에게 인기 있는 도로를 뜻하고 핫 스트리트로서 탐지될 가능성이 커진다.

본 연구에서는 앞의 LBSNS 데이터-건물 데이터 공간조인, 건물-도로 세그멘트 매칭 과정을 거쳐 도로 세그멘트에 연결된 베뉴 정보를 이용하여 각

도로 세그멘트의 핫 밸류를 계산하였다. 간단한 코딩 작업을 통해 각 도로 세그멘트의 속성정보에 베뉴들의 평균 인기도(P_i), 베뉴 건물들의 구성비(R_i), 베뉴 건물들의 밀집도(D_i) 필드를 새로 생성하고 측정하였다. 그리고 최종적으로 식(2-8)을 이용하여 핫 밸류(H_i)를 계산하였다.

2.4.2 Getis-Ord의 G_i^*를 이용한 핫 스트리트 군집 탐지

1) Getis-Ord의 G_i^*

핫 밸류가 높은 도로 세그멘트들은 위의 조건들(베뉴들의 평균 인기도, 베뉴 건물들의 구성비, 베뉴 건물들의 밀집도)을 높은 수준에서 만족시키는 도로 세그멘트들이다. 개개의 도로 세그멘트 자체만으로도 유명한 거리로 이해할 수 있지만, 본 연구는 핫 밸류가 높은 핫 스트리트의 군집(hot street cluster)을 탐지한다. 그 이유는 특정 테마를 특성으로 한 상업적 성격의 테마거리는 군집을 이루려는 경향이 크기 때문이다. 군집을 이룬 테마거리는 거리극복비용 및 상품의 교환을 최소화 하여 생산과 분배의 효율성 및 상호작용을 극대화시키며, 같은 기능끼리 집단을 이룸으로써 접촉의 기회를 높여 상점상호간의 정보교환과 편의성을 증진시키는 이점이 있다(김지혜, 2009).

군집을 탐지하기 위한 군집화 기법에 대해 다양한 연구가 되어 왔지만, 공간 데이터의 군집을 분석할 때에는 공간 클러스터 기법을 이용해야 한다. 일반 통계 및 수치 데이터와는 달리 공간 데이터는 '모든 것은 관련되어 있지만 공간적으로 가까운 것일수록 더 밀접하게 관련되어 있다.'는 Tobler(1970)의 '지리학 제1법칙'을 적용해야하기 때문이다.

본 연구에서는 공간 클러스터를 탐지하기 위한 방법론 중 특히, LISA (Local Indicators of Spatial Association, 국지적 공간 연관성 지표)(Anselin, 1995)를 활용한 방법론을 사용한다. LISA는 국지적 공간적 자기상관 통계량을 의미하는 것으로 국지적 Moran의 I_i 와 국지적 Geary의 c_i, 그리고 Getis-Ord의 G 와 G_i^*를 의미한다(이상일 등, 2010).

LISA를 이용해 공간 클러스터를 탐지하는 방법은 이미 표준화 되어 있다. 특정 LISA를 산출한 후 그 값에 통계적 유의성 검정을 적용시키면 통계적으로 유의한 공간적 응집체를 추출할 수 있고 그것을 공간 클러스터(핫 스팟 혹은 콜드 스팟)라 부를 수 있다(이상일 등, 2010).

본 연구에서는 LISA 중 Getis-Ord의 G_i^*를 이용한다. G_i^*가 가지는 가장 큰 장점은 통계치로부터 직관적으로 핫 스팟인지 콜드 스팟인지 확인할 수 있다는 점이다. 즉, 통계치가 양수면 핫 스팟과 관련되고, 음수면 콜드 스팟과 관련된다. 이러한 특성은 다른 통계량에서도 찾아볼 수 없는 G_i^*만의 장점이다(이상일 등, 2010). 따라서 Getis-Ord의 G_i^*를 이용한 공간 클러스터 방법은 높은 핫 밸류를 지닌 핫 스트리트 군집, 일종의 핫 스팟을 찾는 본 연구 목적에 잘 부합한다.

Getis-Ord의 G_i^*의 계산식은 다음과 같다.

$$G_i^* = \frac{\sum_{j=1} w_{ij}x_j - \bar{x}\sum_{j=1} w_{ij}}{s\sqrt{\frac{n\sum_{j=1} w_{ij}^2 - \left(\sum_{j=1} w_{ij}\right)^2}{n-1}}} \qquad \text{식 (2-9)}$$

이 식에서 s는 표준편차를, w_{ij}는 공간 가중 행렬의 요소 값을, n은 전체 데이터 수를 의미한다. i와 j는 개별 공간단위(혹은 위치)를 의미하는 것으로 두 공간단위가 이웃으로 정의되면 $w_{ij}=1$이, 그렇지 않으면 $w_{ij}=0$이 되며, 자신 또한 이웃으로 간주하기 때문에 $w_{ii}=1$로 주어진다. 이 통계량의 기대값은 0이고, 분산은 거의 1이다(Aldstadt and Getis, 2006). 따라서 이 통계량의 유의성 검정은 정규분포를 상정한 표준화 점수에 대한 것과 거의 동일하게 이루어진다(이상일 등, 2010).

2) 도로 세그멘트의 공간 가중 행렬 생성

G_i^*를 계산하기 위해서는 공간 가중 행렬이 필요하다. 본 연구의 공간분석

단위는 이전 과정에서 생성한 도로 세그멘트이다. 따라서 두 공간분석 단위가 이웃한다는 의미는 도로 세그멘트 2개 이상이 교차로에서 서로 연결되어 있다는 것을 말한다. 교차로에서 연결된 도로 세그멘트들은 공간 가중 행렬 요소 값(w_{ij})이 1로 주어지고, 나머지 도로들과의 공간 가중 행렬 요소 값은 0으로 주어진다. [그림 2-24]는 교차로에서 연결되어 있는 9개의 도로 세그멘트 예시 그림이며, 이 대상 지역의 공간 가중 행렬(w)은 [표 2-3]와 같다.

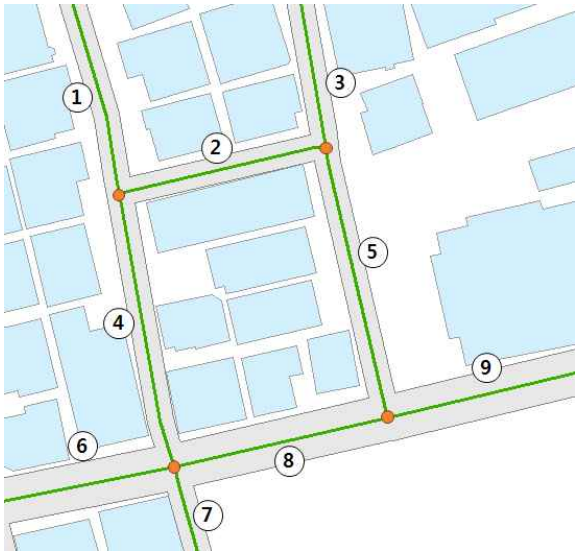


그림 2-24. 교차로에서 연결된 9개의 도로 세그멘트 예시

표 2-3. 그림 2-24의 공간 가중 행렬 생성 예시

Segment	1	2	3	4	5	6	7	8	9
1	0	1	0	1	0	0	0	0	0
2	1	0	1	1	1	0	0	0	0
3	0	1	0	0	1	0	0	0	0
4	1	1	0	0	0	1	1	1	0
5	0	1	1	0	0	0	0	1	1
6	0	0	0	1	0	0	1	1	0
7	0	0	0	1	0	1	0	1	0
8	0	0	0	1	1	1	1	0	1
9	0	0	0	0	1	0	0	1	0

단, 연구 대상 지역의 모든 도로 세그멘트를 대상으로 공간 가중 행렬(w)을 생성하지 않는다. 분석 테마의 베뉴가 위치한 건물과 매칭된 도로 세그멘트들 즉, 이전 과정에서 핫 밸류를 가지고 있는 도로 세그멘트들만을 대상으로 공간 가중 행렬(w)을 생성한다.

그 이유는 분석 테마의 유명한 업소와 장소가 밀집되어 있는 도로 군집을 찾기 위해서는 핫 밸류(H_i)를 지닌 도로들 간의 비교가 이루어져야 하고 그 통계량을 측정하는 것이 알맞다고 판단하였다. 그리고 핫 밸류가 0인 도로 세그멘트들로 인해서 영과잉 분포를 이루고 전체 평균을 많이 낮추는 문제를 피하기 위함이다.

3) 핫 밸류의 로그 정규 분포 변환

Getis-Ord의 G_i^* 통계량이 기댓값 0, 분산 1인 정규성을 갖기 위해서는 공간 데이터 또한 정규분포를 따라야 한다. 단, 공간 데이터 수가 매우 많고 공간단위의 이웃 수가 적어도 30개 이상인 경우에는 공간 데이터가 왜도를 갖는 분포를 따르더라도 G_i^* 통계량의 정규성을 가정할 수 있다. 반대로 공간 데이터 수가 적고 공간 단위의 이웃 수가 8개 정도로 적은 경우에는 분포의

왜도가 심하지 않는 이상 G_i^* 통계량의 정규성을 가정할 수 있다(Getis and Ord, 1996).

본 연구의 공간 단위인 도로 세그멘트는 교차로에서 분절시켰기 때문에 최소 4개 이상의 적은 이웃 수를 가진다. 또한 공간 데이터 수는 연구 대상 지역 범위에 따라 유동적이며, 적은 수의 공간 데이터 또한 다룰 수 있어야 한다. 따라서 위의 조건에 따라 도로 세그멘트의 핫 밸류(H_i) 분포는 왜도가 심하지 않아야 한다.

하지만 도로 세그멘트의 핫 밸류(H_i) 분포는 [그림 2-25]과 같이 보통 왼쪽으로 기울어진 왜도가 심한 분포를 나타낸다. 이러한 분포를 정규 분포화 시키기 위해서 핫 밸류에 자연로그를 취해 값을 로그 변환하였다. [그림 2-25]의 핫 밸류(H_i)를 로그 변환한 분포는 [그림 2-26]와 같이 정규성을 나타내었다. 추가로 SPSS 통계 프로그램의 Q-Q plot을 통해 살펴본 결과, 로그 변환된 핫 밸류 분포가 정규분포를 따른다고 가정할 수 있었다[그림 2-27].

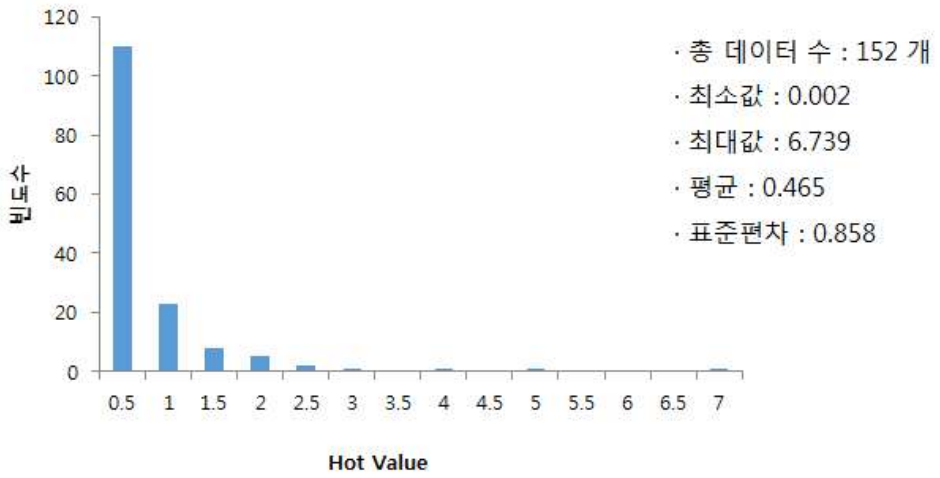


그림 2-25. 신림 사거리 지역 152개 도로 세그멘트의 음식점 핫 밸류 분포

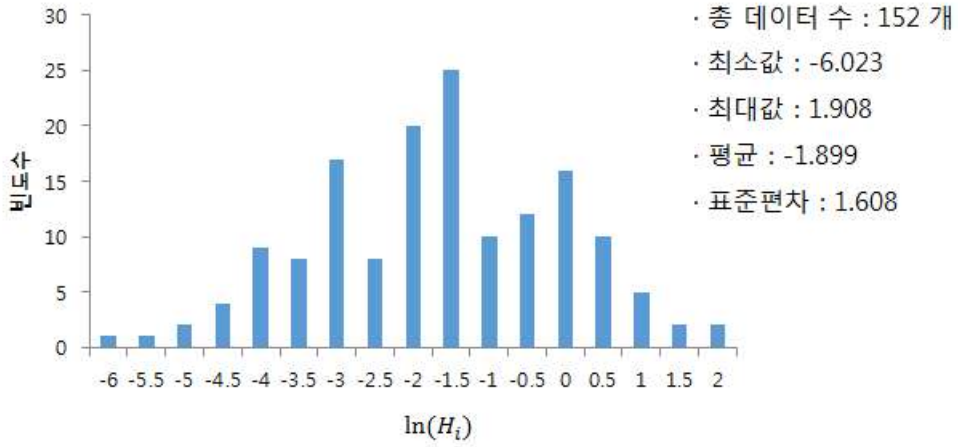


그림 2-26. [그림 2-24]을 로그 변환한 분포

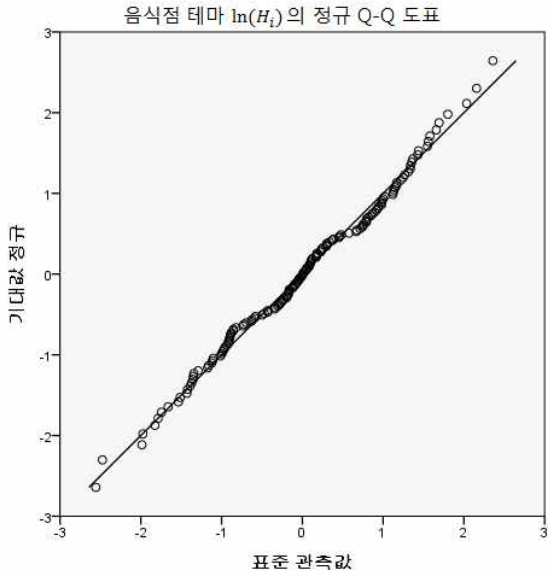


그림 2-27. [그림 2-25]의 $\ln(H_i)$ 분포를
정규분포와 Q-Q plot을 수행한 결과

4) Getis-Ord의 G_i^*를 이용한 핫 스트리트 군집 탐지

마지막으로 핫 밸류를 로그 변환한 값, $\ln(H_i)$을 이용하여 각 도로 세그멘트의 G_i^* 통계량을 계산한다. 앞에서 설명하였듯이 G_i^* 통계량 값에 통계적 유의성 검정을 적용시키면 통계적으로 유의한 공간 클러스터(군집)를 탐지할 수 있기 때문에 이를 적용한다.

본 연구에서는 공간 클러스터 중 G_i^*가 양의 값을 갖는 핫 스팟만을 탐지 대상으로 한다. 그리고 G_i^* 통계량은 정규분포를 따른다고 가정하기 때문에, 0.05의 통계적 유의수준을 적용하여 G_i^*의 z 점수가 1.96 이상인 도로 세그멘트들을 핫 스팟 클러스터라고 판단한다. 즉, 최종적으로 탐지된 이 핫 스팟 클러스터가 본 연구에서 탐지된 핫 스트리트 군집이다.

3. 실험 결과 및 분석

본 연구에서 제안한 핫 스트리트 군집화 기법을 적용하고 평가하기 위해 신림 사거리, 강남역 번화가 지역을 대상으로 실험을 수행하였다. 그리고 2.1.2장에서 설명하였듯이 음식점, 패션, 카페, 유흥주점 등 4가지 테마에 해당하는 핫 스트리트 군집을 탐지하였다.

실험을 수행하는 과정에서 ESRI사의 ArcGIS 프로그램 및 Visual Studio 2013의 F#, Python, Java 등의 프로그래밍 언어를 활용하여 수행하였다. ArcGIS는 베뉴 데이터와 도로명 주소 기본도 DB의 편집과 공간조인 연산, 버퍼 연산, Getis-Ord의 G_i^* 계산 등의 기능을 위해 활용되었으며, 산출된 결과물의 시각적인 표현을 수행하는 데에도 이용되었다. 그리고 포스퀘어 베뉴 데이터 수집, ISOVIST 공간 생성 및 부가적인 연산들은 F#, Python, Java 언어로 구현하여 실험에 적용하고 최종 결과물을 산출하였다.

본 장에서는 신림 사거리, 강남역 번화가 두 지역을 대상으로 실험한 결과를 살펴보고, 그 지역의 상업분포와 특색을 알 수 있는 상권 보고서, 관련 논문, 보도된 자료 등을 참고하여 실험 결과를 분석하였다.

3.1 실험 결과(1) - 신림 사거리

신림역 상권 보고서에 따르면 신림 사거리 지역은 '순대촌' 정도로만 알려져 있다가 1984년 지하철 2호선이 개통되면서 급속도로 상권이 발전하였다. 순대타운을 중심으로 고가의 음식점보다 저가의 음식점이 주를 이루며 입점해 있고, 현재는 지하철 1, 2번 출구에 위치한 포도몰이 오픈함에 따라 젊은이들이 많이 찾는 지역 중에 하나로 손꼽힌다.

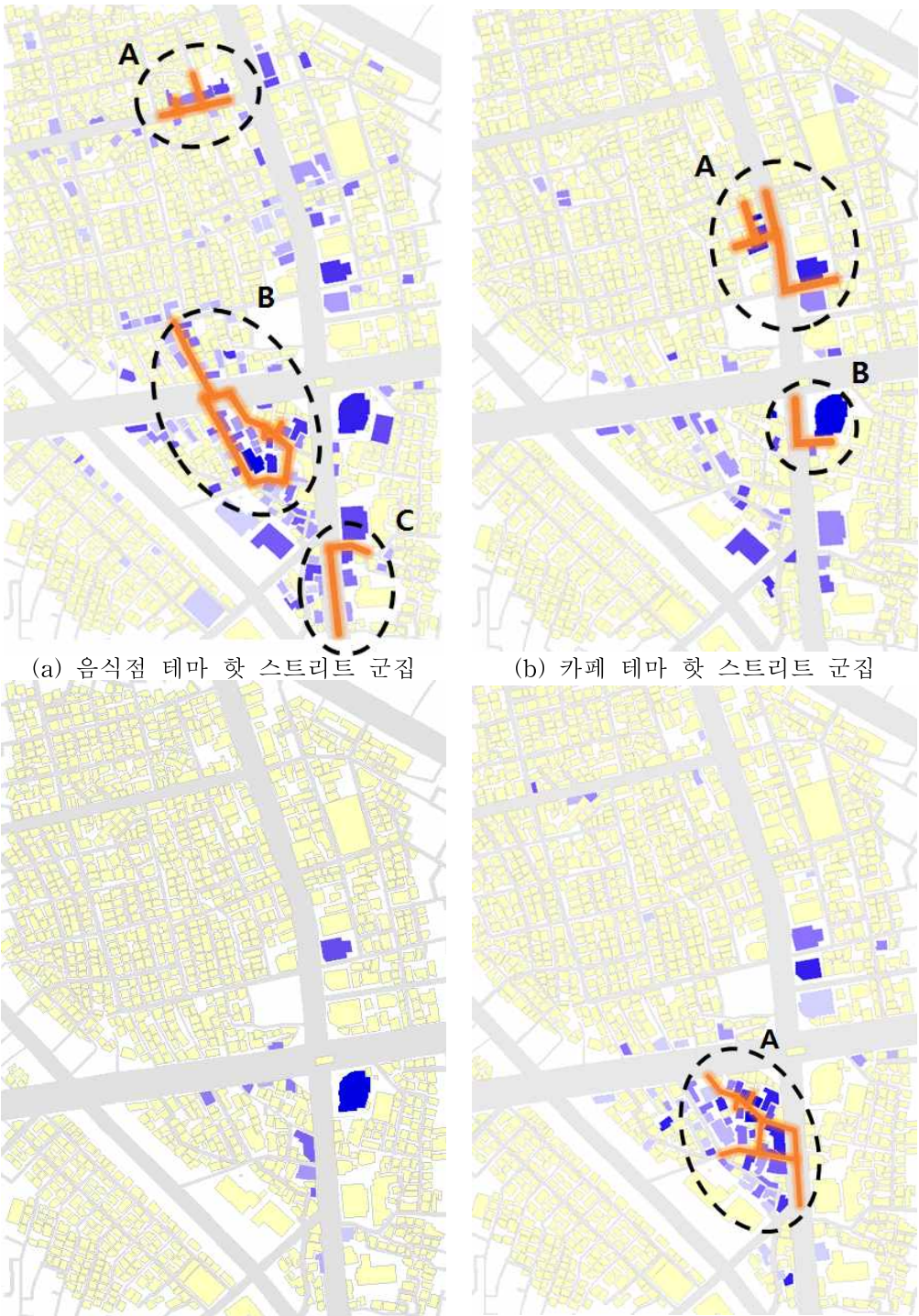
본 연구는 이러한 신림 사거리 주변을 대상으로 음식점, 카페, 패션, 유흥주점 테마에 따라 핫 스트리트 군집을 탐지하였다. 실험지역의 구체적인 범위는 경도 126.9256~126.931265, 위도 37.481567~37.488276 이고[그림 3-1], 2013년 7월 3일에 포스퀘어 API를 이용하여 해당 지역의 베뉴 데이터를 수집하였다.

그림 3-1. 신림 사거리 실험 지역 범위

위의 신림 사거리 실험 지역에서 수집된 포스퀘어 베뉴 데이터는 총 1330개 이다. 이 데이터를 테마 별로 구분하고 활용하여, 2.1~2.3 과정을 거쳐 생성된 위치 보정된 베뉴 수, 베뉴 건물 수, 베뉴 건물과 매칭된 도로 세그멘트 수는 아래의 [표 3-1]과 같다.

표 3-1. 신림 사거리 실험 지역의 테마 별 위치 보정된 베뉴 수, 베뉴 건물 수, 베뉴 건물과 매칭된 도로 세그멘트 수 비교

테 마 / 구 분	음식점(개)	카페(개)	패션(개)	유흥주점(개)
위치 보정된 베뉴 수	312	65	24	108
베뉴 건물 수	154	40	12	65
베뉴 건물과 매칭된 도로 세그멘트 수	152	54	27	65



(a) 음식점 테마 핫 스트리트 군집 (b) 카페 테마 핫 스트리트 군집

(c) 패션 테마 핫 스트리트 군집 (d) 유흥 주점 핫 스트리트 군집

그림 3-2. 신림 사거리 지역의 핫 스트리트 군집 탐지 결과

최종적으로 2.4 과정을 거쳐 신림 사거리 실험 지역의 각 테마 별 핫 스트리트 군집을 탐지한 결과는 [그림 3-2]와 같다. 도로 중심에 표시한 붉은 선들은 유의 수준 0.05를 만족시키는 G_i^*값을 지니며, 각 테마의 핫 스트리트라고 판단된 도로 세그멘트들이다. 파란색으로 음영 처리된 건물들은 그 건물에 해당 테마의 베뉴들이 위치하고 있음을 나타내고, 베뉴들의 인기도(체크인, likes, tip 수를 합한 수치)에 따라 진하게 표시하였다. 각 테마 별 $\ln(H_i)$의 분포가 정규 분포를 가정할 수 있는지 살펴보기 위해, SPSS 통계 프로그램을 이용하여 정규 Q-Q plot를 수행하였다[부록 F]. 각 테마 별 정규 Q-Q plot을 수행한 결과, $\ln(H_i)$의 분포가 모두 정규분포를 따른다고 가정하고 G_i^*를 적용할 수 있었다.

3.1.1 음식점 테마 핫 스트리트 군집

[그림 3-2(a)]에서 살펴볼 수 있듯이 신림 사거리 주변에는 크게 3개의 음식점 테마 핫 스트리트 군집이 존재한다. 특히 B 군집에는 순대타운이 위치하고 있고[그림 3-3], 신림역 상권보고서에서 언급하였듯이 순대타운 주변으로 많은 음식점이 분포하고 있음을 알 수 있다. B 군집의 핫 스트리트 주변에는 총 92개의 음식점 베뉴가 위치하고 있다. 그 중 순대와 관련된 음식점은 44개를 차지하였다. B 군집의 음식점 중 절반가량이 순대 음식점을 차지하는 비정상적인 분포는 이 지역이 순대로 유명한 곳임을 다시 한 번 확인할 수 있다.

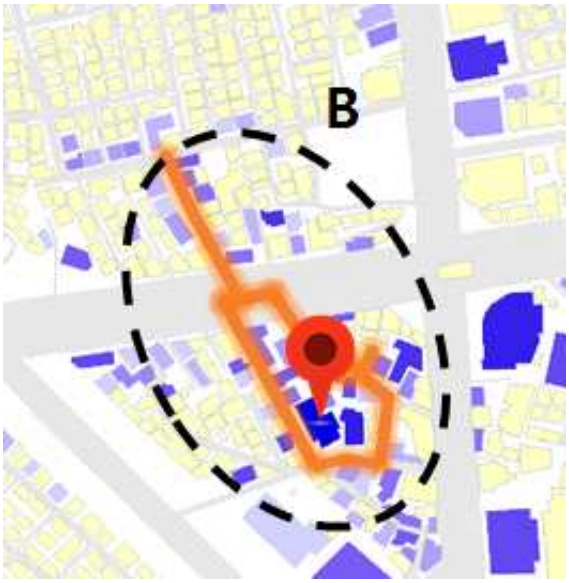


그림 3-3. 신림동 순대타운 위치

C 군집에는 피자헛, 한스델리, 죠스 떡볶이, 장터 쭈꾸미 등 여러 음식점들이 신림로를 따라 집중적으로 모여 있다. B 군집의 아래에 위치한 C 군집은 B 군집과 함께 음식점이 분포한 지역에 위치하고 있지만, 특이하게도 B 군집과 함께 하나의 핫 스트리트 군집을 이루고 있지는 않다. 실험 결과를 보아 B와 C 군집 사이에 분포하고 있는 음식점들은 주변 음식점들에 비해 사람들에게 많은 인기와 관심을 충분히 받지 못하고 있고, 이는 B와 C 군집이 하나의 음식점 거리로서 사람들이 인식하는데 방해 요소로 작용하고 있는 것으로 보여진다.

[그림 3-2(a)]의 A 군집은 B, C 군집과 상당히 멀리 떨어진 곳에 위치하고 있고 소규모로 형성되었다. A 군집일대는 본래 관악구에서 조성한 '패션문화의 거리'이다. 신림동 패션문화의 거리는 신림5동 1439번지 일대 500m 구간의 신림동길을 대상으로 조성되었다[그림 3-4]. 하지만 거리 조성 계획과는 달리 패션관련 상점들은 대부분 사라지고 음식점만 즐비하고 있어 패션문화의 거리 조성 사업의 실패에 대한 불만 섞인 여론을 조사할 수 있었다(정한모, 2006). 이와 같이 현재 저 지역은 패션문화의 거리에서 먹자 골목의 형태로 성격이 많이 변하였으며 A 군집은 현 실태를 그대로 보여주는 결과를 나

타내었다. A 군집의 음식점 메뉴는 8개 정도로 적은 수가 위치하고 있지만, 이 메뉴들은 다른 지역 메뉴들에 비해 평균 인기도가 높아 사람들이 많이 방문하는 맛집 거리임을 알 수 있다.

그림 3-4. 관악구에서 조성한 신림동 "패션문화의 거리"

3.1.2 카페 테마 핫 스트리트 군집

실험을 수행한 결과 신림 사거리 지역에는 카페 핫 스트리트 군집이 두 군데 존재하고 있다[그림 3-2(b)]. A 군집 주변에는 총 6개의 적은 수의 카페가 위치하고 있지만, 스타벅스, 투썸 플레이스, 할리스 커피 등 유명 브랜드 매장이 밀집되어 있어 다른 카페가 위치한 지역보다 사람들이 많이 방문하는 것으로 여겨진다. B 군집 또한 메뉴들이 4개의 건물에 위치하고 있고, 스타벅스, 엔젤리너스, 라떼떼커피 등 유명 브랜드 매장이 밀집되어 있어, 높은 인기도를 통해 탐지된 것으로 보인다.

하지만, 카페가 위치한 건물의 분포 양상을 살펴보았을 때 B 군집 아래에 위치한 지역에 도로를 따라 카페가 늘어선 것을 확인 할 수 있다[그림 3-5]. 핫 밸류를 계산할 때 메뉴가 위치한 건물의 밀집도 등을 고려함으로써 이 도로들 또한 핫 스트리트 군집 후보군으로 생각될 수 있다. 하지만 이 도로들이 핫 스트리트로 탐지되지 않은 것은 본 연구의 목적이 단순히 카페가 밀집되어 있는 도로만을 탐지하는 것이 아닌 사람들이 많이 방문하고 인기 있는

매장들이 존재하는 활발한 테마거리 영역을 제시 해 주는 것이기 때문이다. 따라서 소수 이지만 인기도가 압도적으로 높은 매장들이 위치한 A와 B 군집 지역의 도로들이 핫 스트리트로 탐지된 것이다. 그리고 역 주변과 대로변에 위치한 카페들이 소수임에도 불구하고 높은 인기도를 나타내는 것으로 보아 현재 신림 사거리 지역에 카페 매장이 들어서기 위해서는 대로변과 역과의 거리가 중요한 입지 조건으로 작용하는 것을 유추할 수 있다.

그림 3-5. B군집 아래에 위치한 카페 밀집 지역

3.1.3 패션 테마 핫 스트리트 군집

[그림 3-2(c)]의 그림과 같이 신림 사거리 지역에는 패션 테마로 돋보이는 핫 스트리트 군집이 탐지되지 않았다. 유의 수준 0.05에서 만족시키는 G_i^* 값을 가진 도로가 없을뿐더러, 이 지역에는 패션 베뉴가 위치한 건물 수 자체가 상당히 적다. 이로 인해 핫 스트리트 군집 탐지에 사용된 도로 세그멘트는 총 27개에 불과했다.

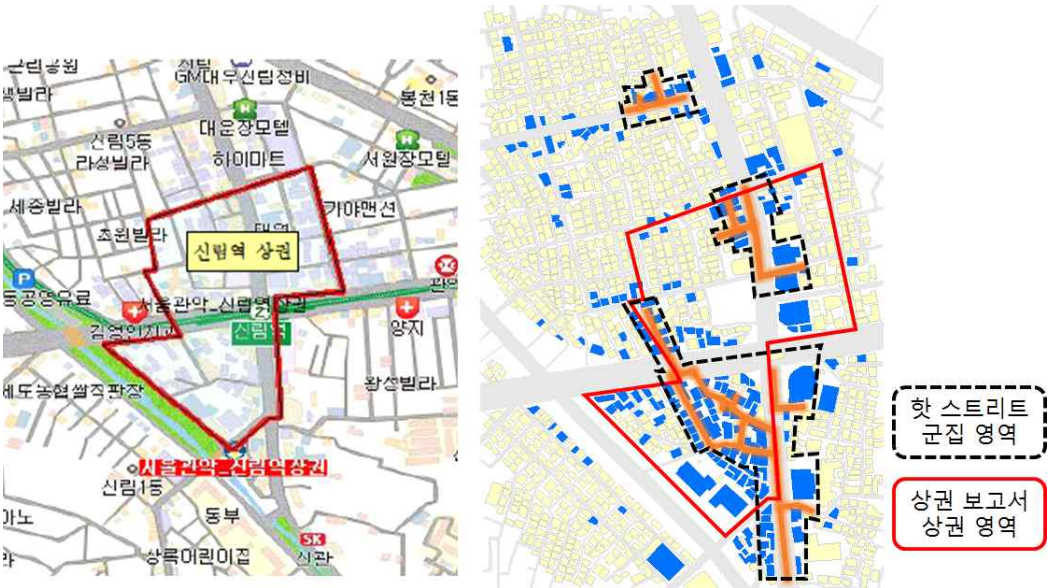
이 지역에 핫 스트리트 군집이 탐지되지는 않았지만, 역 주변에 위치한 포

도몰의 대형 건물이 가장 높은 인기도를 나타내는 것으로 보아, 신림사거리 지역에서 사람들이 인식하는 패션으로 유명한 장소는 특정한 거리나 영역 보다 대형 몰에 집중되어 있을 것으로 판단된다.

3.1.4 유흥주점 테마 핫 스트리트 군집

[그림 3-2(d)]는 신림 사거리 지역의 유흥주점 핫 스트리트 군집을 나타낸 그림이다. 유흥주점은 일반적으로 음식점들과 인접하고 있는 경향이 있다. 이는 유흥주점들이 음식점과 가까운 곳에 위치함으로써 상업 활동에 시너지 효과를 발생시킬 수 있기 때문이다. 신림 사거리 지역도 마찬가지로 유흥주점 핫 스트리트 군집이 음식점과 같이 남서쪽 일대에 형성되면서 상업 활동에 서로 도움을 줄 것으로 예상된다.

3.1.5 상권보고서와 비교 분석



(a) 2008년 조사된 상권 영역 (b) 핫 스트리트 군집 영역

그림 3-6. 상권보고서의 상권 영역과 핫 스트리트 군집 영역과의 비교

본 연구에서는 음식점, 카페, 패션, 유흥주점 4가지 테마를 대상으로 핫 스

트리트 군집을 탐지하였으며, 이 4가지 테마는 특화거리의 주요 테마로도 많이 사용되지만 상권을 형성하는데 중요한 상업 분야이다. 따라서 4가지 테마의 핫 스트리트 군집들을 종합한 영역은 그 지역의 핵심 상권을 제시할 수 있고, 소상공인지원센터에서 2008년 6월에 조사·작성한 상권 보고서에서 제시한 상권 영역과 비교하였다.

[그림 3-6]은 신림사거리 지역 상권보고서의 상권영역(빨간색 테두리)과, 핫 스트리트 군집과 이에 영향을 끼친 건물들을 포함하는 영역(검은색 점선 테두리)을 나타내었다. 파란색으로 표시된 건물들은 베뉴들이 위치한 건물들이다.

실험 결과를 통해 얻어진 핫 스트리트 군집 영역이 상권보고서의 상권 영역 내부에 대부분 포함되어 있는 것으로 보아 상권과 거리가 먼 지역을 탐지하지 않는 등 실험 결과의 신뢰성을 확인할 수 있었다. 그리고 상권보고서의 상권영역은 블록단위로 광범위한 지역을 설정한데 반해, 핫 스트리트 군집 영역은 사람들의 인기도가 높은 핵심적인 지역만을 구체적으로 설정하였다.

또한 상권보고서는 2008년 시점의 상권을 조사하였기 때문에 현재와 조금 다른 양상을 띠고 있음을 확인할 수 있다. 신림역 남동쪽에 위치한 포도몰은 상권 조사 이후에 완공(2008년 11월)되었기 때문에 상권보고서에서는 포도몰을 상권에 포함시키지 못한 것으로 예상된다. 그리고 핫 스트리트 군집 영역은 이전보다 신림역 남쪽 방향으로 상권이 확장되었음을 제시한다. 패션문화의 거리 지역은 상권보고서에서는 조사를 실시하지 않아 비교분석을 시행하지 못하였다.

3.2 실험 결과(2) - 강남역 번화가

강남역 상권 보고서에 따르면 강남역 지역은 현재 대한민국 최대의 상권으로 손꼽히고 있다. 강남역 일대에는 강남대로를 중심으로 의류, 잡화, 음식점, 유흥, 식품 등 다양한 업종이 활성화 되어 있으며, 대형 업무시설과 학원시설 및 주변 극장 등으로 인해 인구유발효과가 뛰어난 지역이다.

그리고 분당, 용인, 수원 등에 신도시가 생겨나면서 강남지역과 이들 지역간의 직접적인 연결이 확대되어 수도권 신도시와 서울간의 연결을 위한 관문

역할을 강남역이 하게 되면서 더욱 거대한 상권으로 발전하고 있다.

본 연구는 이러한 강남역 번화가 지역을 대상으로 음식점, 카페, 패션, 유흥주점 테마에 따라 핫 스트리트 군집을 탐지하였다. 실험지역은 경도 127.02482~127.033412, 위도 37.492566~37.499138의 범위와 경도 127.02361~127.031215, 위도 37.499138~37.506049 범위를 합한 지역이고[그림 3-7], 2013년 7월 2일, 7월 10일에 걸쳐 포스퀘어 API를 이용하여 해당 범위의 베뉴 데이터를 각각 수집하였다.

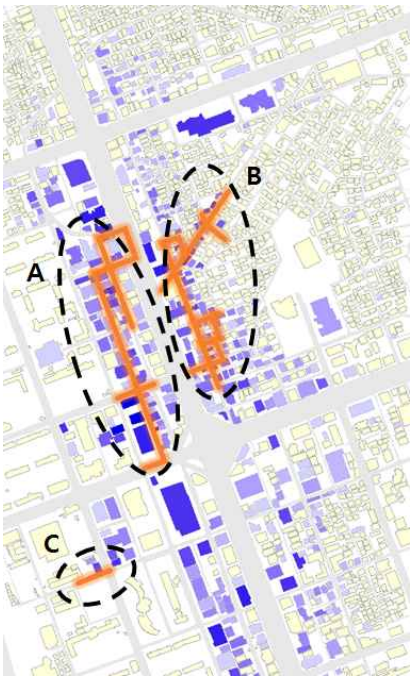
그림 3-7. 강남역 번화가 실험 지역

위의 강남역 번화가 실험 지역에서 수집된 포스퀘어의 베뉴 데이터는 총 8556개 이다. 이 데이터를 테마 별로 구분하고 활용하여, 2.1~2.3 과정을 거쳐 생성된 위치 보정된 베뉴 수, 베뉴 건물 수, 베뉴 건물과 매칭된 도로 세그멘트 수는 아래의 [표 3-2]와 같다.

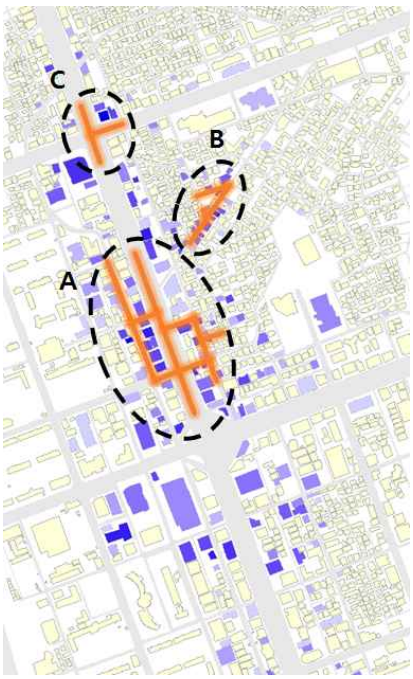
표 3-2. 강남역 번화가 실험 지역의 테마 별 위치 보정된 베뉴 수, 베뉴 건물 수, 베뉴 건물과 매칭된 도로 세그멘트 수 비교

구 분 \ 테 마	음식점(개)	카페(개)	패션(개)	유흥주점(개)
위치 보정된 베뉴 수	1570	400	187	534
베뉴 건물 수	425	192	83	204
베뉴 건물과 매칭된 도로 세그멘트 수	242	175	93	162

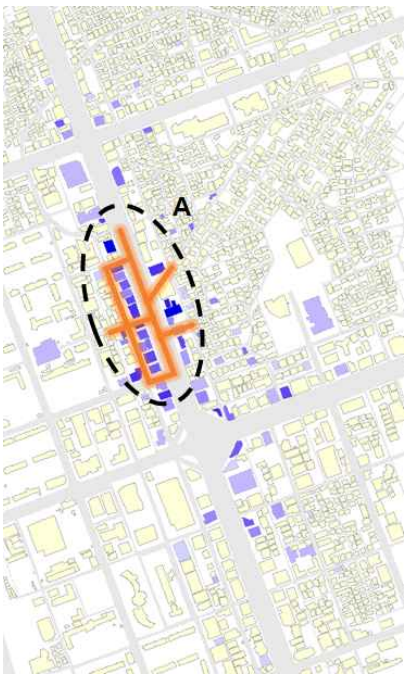
최종적으로 2.4 과정을 거쳐 강남역 번화가 실험 지역의 핫 스트리트 군집을 탐지한 결과는 [그림 3-8]과 같다. 그리고 강남역 번화가 지역의 각 테마 별 $\ln(H_i)$의 분포가 정규 분포를 가정할 수 있는지 살펴보기 위해, SPSS 통계 프로그램을 이용하여 정규 Q-Q plot를 수행하였다[부록 G]. 각 테마 별 정규 Q-Q plot을 수행한 결과, $\ln(H_i)$의 분포가 모두 정규분포를 따른다고 가정하고 G_i^*를 적용할 수 있었다.



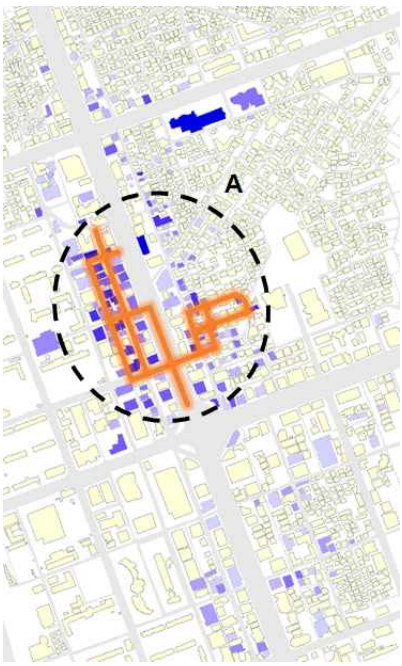
(a) 음식점 테마 핫 스트리트 군집



(b) 카페 테마 핫 스트리트 군집



(c) 패션 테마 핫 스트리트 군집



(d) 유흥 주점 핫 스트리트 군집

그림 3-8. 강남역 번화가의 핫 스트리트 군집 탐지 결과

3.2.1 음식점 테마 핫 스트리트 군집

실험을 수행한 결과, 강남역 번화가 지역에는 총 3군데의 음식점 핫 스트리트 군집이 탐지되었다[그림 3-8(a)]. 눈여겨 볼 특징으로는 A, B군집이 강남대로를 기준으로 양 쪽 이면도로 지역에 형성되었다는 점이다. 곽갑신(2010)의 강남역 상업지의 업종분포 연구에 따르면 강남대로보다 이면도로를 중심으로 음식점이나 유흥시설이 밀집되어 있어 이 지역은 젊은이들의 여가문화 욕구를 충족시켜주는 도심 상업지로 분석하고 있다. 이를 참고하였을 때 실험 결과가 객관적인 정확도를 지니고 있음을 확인할 수 있다.

강남역의 남쪽 지역에 위치한 음식점들 역시 대로 주변에 밀집한 분포를 보여주고 있지만, A와 B 군집의 인기도가 상대적으로 높아 핫 스트리트로 탐지되지 않았다. 이는 사람들이 보편적으로 강남역 번화가의 음식점을 방문할 때 강남역 북쪽 지역의 음식점들을 더욱 선호하고 자주 방문하며, A와 B 군집지역이 강남역 번화가 내부에 위치한 음식점 테마거리로 인식되기에 충분하다고 판단된다.

만족할 만한 결과를 보여준 A, B 군집과 달리 C 군집은 실험 방법의 한계를 보여주는 하나의 예이다. C 군집은 정확히 말하면 군집이 아니라 하나의 핫 스트리트가 탐지된 것이다. 이 도로 세그멘트의 G_i^*의 z 값은 약 1.98로 0.05의 유의수준을 간신히 만족시켰다. 그리고 해당 도로 세그멘트와 매칭된 음식점 베뉴 건물은 하나만 존재하고 있다[그림 3-9]. '흑돈가'라는 음식점이 그 건물에 위치하고 있고, 1045의 인기도를 기록하고 있다. 강남역 번화가 지역에서 측정된 음식점 베뉴의 평균 인기도가 약 80정도인 것과 비교하였을 때 '흑돈가' 음식점은 사람들에게 상당히 높은 인기를 얻고 있는 음식점인 것을 알 수 있다. 이렇게 베뉴의 높은 인기도는 도로 세그멘트의 핫 밸류를 높이고 핫 스트리트로서 탐지될 가능성을 높여준다.

하지만 본 연구는 하나의 업소 또는 대형 몰과 같은 랜드마크를 탐지하기보다, 동종 업소가 밀집하여 특정한 거리공간을 형성하고 있고 다수의 사람들을 이끄는 활성화된 거리를 찾는 것이다. 따라서 2.4.1장에서 서술하였듯이 인기도뿐만 아니라 베뉴 건물들의 구성비, 베뉴 건물들의 밀집도 등 거리공간 형성을 위한 요소들도 고려하였다. 그럼에도 불구하고 비정상적으로 매우

높은 인기도는 C 군집과 같이 핫 스트리트 군집 탐지에 지배적인 영향을 끼칠 수 있다는 실험 방법의 한계점을 나타낸다.

그림 3-9. 인기도가 상당히 높은 소수의
업소에 의해 핫 스트리트가 탐지된 그림

3.2.2 카페 테마 핫 스트리트 군집

[그림 3-8(b)]와 같이 강남역 번화가 지역에는 3군데의 카페 핫 스트리트 군집이 탐지되었다. A군집은 강남역에서 신논현역 방향의 강남대로를 중심으로 형성되었다. 강남역 번화가 일대에서 강남대로가 유동인구가 가장 많기 때문에 강남대로 주변에 스타벅스, 투썸 플레이스 등 유명 브랜드의 카페들이 많이 입점하고, 강남역으로부터 접근이 용이한 이 카페들을 사람들이 많이 방문하고 인기가 있는 것으로 생각된다.

B 군집은 음식점 테마의 B군집과 함께 상당히 의미 있는 결과를 나타낸다. 이 지역은 일명 “강남역 언덕길”이라고 불리며 최근 강남역 CGV 뒤편으로 신사동 가로수길처럼 트렌디한 카페와 맛집이 운치있게 들어서서 새로운 명소로 부상하고 있는 곳이다(류근원, 2012). 이 지역은 2~3년 전만 해도 오래된 상점들과 사주 카페가 주를 이루었던 곳이었지만, 2008년에 “에스프레소

퍼블릭” 카페가 처음으로 이곳에서 문을 열었고 색다른 분위기의 카페들이 연달아 입점하면서 강남역 언덕길 카페거리로 새롭게 태어났다(정아람, 2013). [그림 3-10]의 왼쪽 그림과 같이 강남역 언덕길만의 지도가 존재할 만큼 명소로서 많이 홍보되고 있으며, 통상적으로 알려진 언덕길의 범위는 [그림 3-10]의 오른쪽 그림과 같다.

그림 3-10. 강남역 번화가에 위치하고 있는 명소 “언덕길”

C 군집은 신논현역을 중심으로 유명 브랜드 카페들이 밀집되어 있다. 이 지역에는 테이크 어반, 할리스 커피 등 총 7개의 카페 매장이 위치하고 있고, 이 카페들의 평균 인기도는 약 726을 기록하였다. 카페 베뉴의 전체 평균 인기도인 약 137과 비교했을 때 C 군집의 카페들은 소수이지만 상당히 인기 있는 명소임을 알 수 있다.

3.2.3 패션 테마 핫 스트리트 군집

강남역 번화가 지역에서 패션으로 유명한 거리는 강남대로변을 따라 형성되어 있는 것으로 판단된다. [그림 3-8(c)]에서 살펴볼 수 있듯이 강남역과 신논현역 사이의 강남대로를 중심으로 패션 매장이 밀집되어 있고 이를 따라

핫 스트리트 군집으로 탐지되었다. 실제로 다음 지도 포털 서비스의 로드뷰[11]를 통해 강남대로변을 따라 위치한 대형 패션 매장들을 [그림 3-11]과 같이 쉽게 확인할 수 있다.

그림 3-11. 강남대로변을 따라 위치한 패션 매장들의 모습

[그림 3-8(c)]의 A 군집을 자세히 살펴보면, 강남대로 뿐만 아니라 왼쪽에 있는 서초이면도로도 핫 스트리트로 같이 탐지되었다. 강남역 대로변에 위치한 패션매장들의 출입구와 간판 대부분은 모두 대로변을 향하고 있고, 같은 건물이지만 다른 업종의 매장간판과 출입구 방향은 서초이면도로를 향하고 있다. 이런 상황을 고려하였을 때 서초이면도로가 패션 테마 핫 스트리트로

11) 다음 지도 서비스에서 제공하는 2013년 4월에 측정된 로드뷰(2013년 9월 30일)

탐지된 것은 부정확하다고 판단된다. 이렇게 서초이면도로가 핫 스트리트로 탐지된 것은 베뉴가 위치한 건물과 도로 세그멘트를 매칭 시키는 방법의 한계를 보여주는 예이다. 모든 베뉴의 출입구 정보가 100% 구축되어 있지 않기 때문에 본 연구는 ISOVIST를 활용하여 건물과 맞닿아 있는 도로 세그멘트를 탐색하고 서로 매칭 시켰다. 이 방법은 건물에 위치한 베뉴들의 출입구 방향이 건물의 모든 방향에서 가능하다는 것을 가정하기 때문에, A 군집과 같이 업종에 따라 출입구 방향이 특정한 방향성을 가지고 있는 경우 부정확한 결과를 나타낼 수 있다[그림 3-12]. 하지만 공간적 자기 상관성을 고려한 군집 기반의 탐지가 이루어진 본 연구의 결과는 핵심 도로와 전혀 무관한 도로가 아닌 인근 도로가 추가로 탐지되는 것이기 때문에 결과의 신빙성을 크게 해치지 않을 것으로 판단된다.

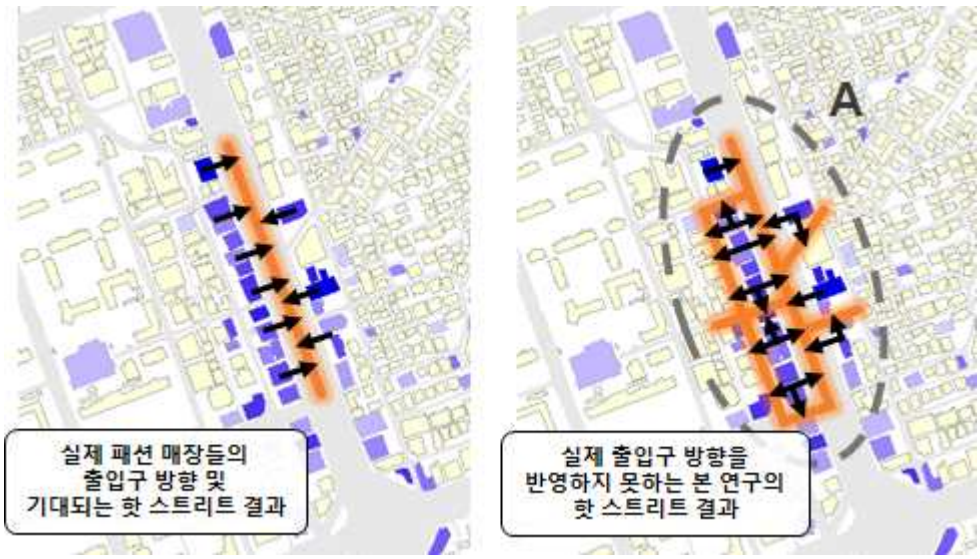


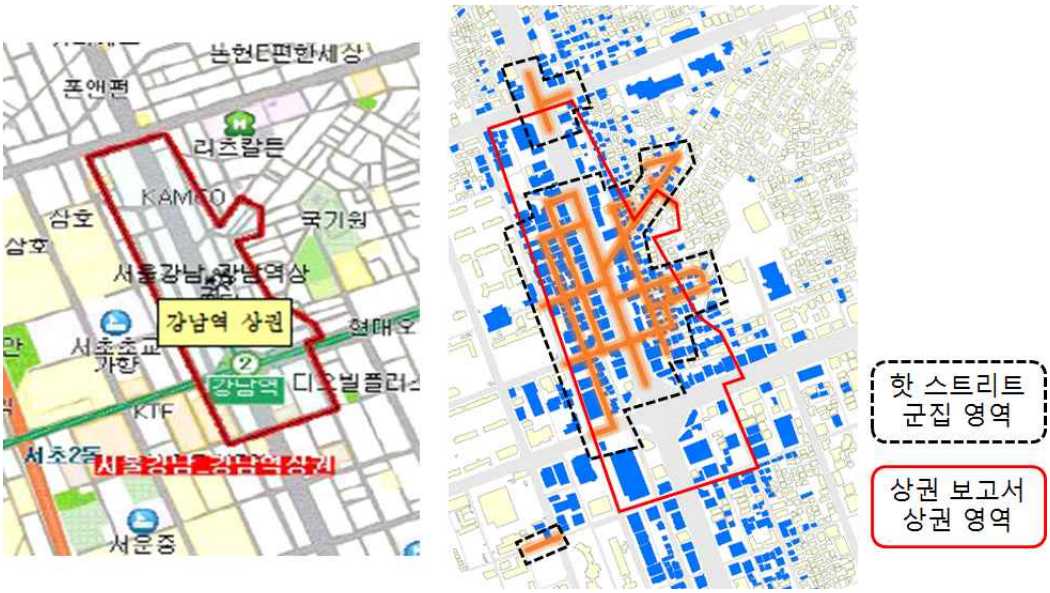
그림 3-12. 실제 패션 매장의 출입구 방향을 반영하지 못하고
핫 스트리트 군집이 탐지된 그림

3.2.4 유흥주점 테마 핫 스트리트 군집

[그림 3-8(d)]는 강남역 번화가 지역의 유흥주점 핫 스트리트 군집을 나타내고 있다. A 군집은 강남대로와 서초이면도로, 역삼이면도로의 부분적인 지

역을 포함하고 있고, 특히 서초이면도로에 음식점들과 같이 유흥주점이 밀집해 있어 이 지역은 많은 사람들에게 인기 있는 명소로 판단된다.

3.2.5 상권보고서와 비교 분석



(a) 2008년 조사된 상권 영역 (b) 핫 스트리트 군집 영역

그림 3-13. 상권보고서의 상권 영역과 핫 스트리트 군집 영역과의 비교

신림 사거리 지역과 마찬가지로 강남역 번화가 지역 4가지 테마의 핫 스트리트 군집들을 종합한 영역과 2008년 5월에 조사·작성된 상권 보고서의 상권 영역과 비교·분석하였다. [그림 3-38]은 강남역 상권보고서의 상권영역(빨간색 테두리)과, 핫 스트리트 군집과 이에 영향을 끼친 건물들을 포함하는 영역(검은색 점선 테두리)을 나타내었다. 파란색으로 표시된 건물들은 베뉴들이 위치한 건물들이다.

핫 스트리트 군집 영역은 실험 방법의 한계에 의해 탐지된 “흑돈가” 음식점 지역의 개별 핫 스트리트를 제외하고, 상권보고서의 상권 영역 내부에 대부분 포함되어 있는 것을 보아 실험결과의 신뢰성을 살펴볼 수 있다. 베뉴들이 위치한 건물들의 분포는 강남역을 중심으로 넓게 분포하고 있고, 상권보

고서의 상권 또한 강남역 남쪽 지역까지 상권을 정의하고 있다. 하지만 핫 스트리트 군집 영역을 살펴보았을 때 강남역 전체 상권 중에서 사람들이 자주 방문하고 번화가로서의 핵심적인 지역은 강남역과 신논현역 사이임을 알 수 있다. 그리고 2008년에 조사된 상권은 핫 스트리트 군집 영역과 달리 최근에 명소로 유명해진 강남역 “언덕길” 지역을 반영하지 못하고 있음을 알 수 있다.

4. 결론

본 연구에서는 지역 경제를 활성화 시키고 사람들에게 사회·문화적인 공간을 제공하는 상업적 테마거리를 탐지하기 위한 핫 스트리트 군집화 기법을 제안하였다. 단순히 동일한 업종의 상점이 밀집되어 있는 거리를 찾는 것이 아니라, LBSNS 데이터의 장소에 대한 인기도 정보를 활용하여 실제로 사람들이 그 거리를 많이 방문하고 테마거리로서 특별한 장소성이 형성된 거리를 탐지하였다. 먹자골목, 카페거리 등 특별한 장소로서 인지되는 도로의 특성 또한 고려하여 도로 기반의 공간 클러스터 기법을 사용하였다. 본 기법은 활용 데이터 수집, LBSNS 데이터와 건물 데이터 간의 공간조인, 건물과 도로 세그멘트간의 매칭, 핫 스트리트 군집 탐지 등 크게 4단계의 프로세스로 구성된다.

핫 스트리트 군집화 기법의 효용성을 알아보기 위해 본 연구는 신림 사거리, 강남역 번화가 지역을 대상으로 실험하여 음식점, 카페, 패션, 유흥주점 등 4개의 테마에 대해서 테마거리를 탐지하였고, 그 지역의 산업분포와 특색을 알 수 있는 상권 보고서, 관련 논문, 보도된 자료 등을 참고하여 실험결과와 비교·분석하였다.

특히 현지 조사를 통해 결정된 상권 보고서의 상권 영역과 4가지 테마거리를 종합한 영역을 비교한 결과, 상권 영역 내부에 테마거리 영역이 대부분 포함되어 있어 실험 결과의 신뢰성을 확인할 수 있었다. 그리고 광범위한 상업지역의 상권에 비해 테마거리 영역은 사람들의 인기도가 높은 핵심적인 상업 지역을 제시하는데 의의가 있다.

본 연구는 전문가의 경험과 현지 조사 결과에 의존하여 일관된 결과를 얻을 수 없는 기존의 테마거리 결정 방법의 한계점을 극복하여 객관성 있고 재현성 있는 테마거리 탐지 기법을 제시한 데 가장 큰 의의가 있다. 본 기법은 LBSNS 데이터와 도로명주소 기본도 등 수치적인 데이터만을 이용하고 정형화된 방법을 수행하기 때문에 어느 지역에서 수행하여도 객관적인 결과를 기대할 수 있고 재현 가능하다.

그리고 본 연구는 공간적 단위로서 하나의 장소로 인지되는 개별 도로를 기반으로 핫 스트리트 군집을 탐지하였다. 이를 통해 테마거리를 기존 핫 스

팟 지도나 밀도 추정 결과와 같은 모호한 영역 대신 사람들에게 좀 더 직관적이고 분간할 수 있는 도로 기반의 영역을 제시하였다는 점에서도 의의가 있다. 특히 공간 통계기법인 Getis-Ord의 G_i^*는 일반적으로 면형데이터(areal data)를 대상으로 수행하지만, 본 연구에서는 테마거리의 군집 특성과 유기적으로 연결되어 있는 도로의 특성을 고려하여 선형데이터(line data)를 대상으로 G_i^*를 수행하였다.

또한 본 연구는 단순히 업소의 위치 정보만 이용하여 핫 스팟 탐지를 수행하지 않고, 다양한 속성 정보를 종합적으로 활용하였다. LBSNS 데이터의 인기도 정보를 통해 인기 있는 업소가 밀집한 도로를, 그리고 도로의 길이, 주변 건물 수 등을 통해 동종 업소가 밀집한 도로를 동시에 고려하여 테마거리로서 적합한 거리들을 탐색할 수 있었다.

이러한 특징을 바탕으로 본 연구의 핫 스트리트 군집화 기법은 크게 3가지의 활용방안을 기대할 수 있다. 첫째, 사람들이 인지하지 못하고 있는 자생적 테마거리를 객관적으로 탐지하여 사람들에게 더욱 다양한 사회·문화적인 공간을 제공해줄 수 있을 것으로 예상된다. 단일 테마가 복합적으로 형성된 번화가 지역에서는 각 테마 별로 상업 활동이 집중적으로 활성화된 거리를 추천, 제시 해줌으로써 해당 지역을 방문한 사람들이 효율적인 공간적 의사결정을 내리는데 도움을 줄 것이다. 둘째, 본 기법을 통해 탐지된 테마거리 결과는 테마거리의 영역 또는 실태 등을 조사하기 위한 현장조사를 어느 정도 대체하여 비용적인 이득을 취할 수 있고, 해당 지역의 도시계획 및 관련 사업을 추진할 때 기초적인 자료로써 참고 될 수 있을 것이다. 셋째, 도로 기반으로 결과물이 생성되기 때문에 차량용 또는 보행자용 내비게이션의 네트워크 데이터에 접목시켜, 길안내 시 인접해 있는 테마거리를 안내하는 등 더욱 풍부한 내비게이션 서비스를 제공해 줄 수 있을 것이다.

본 연구가 가지는 한계로는 첫째, 탐지된 테마거리에 대한 정량적 평가가 부족한 점이다. 테마거리라는 장소가 보편적으로 인식되기는 하지만 개인마다 인식하는 장소의 특징과 범위가 다르기 때문에 실험결과를 평가할 객관적인 자료를 찾을 수 없었다. 둘째, LBSNS 데이터 자체의 결함 문제이다. 본 연구에서는 LBSNS 데이터의 가장 큰 문제점인 위치 오차를 보정하여 실험하였지만, 위치 오차 보정 과정에서 베뉴 데이터를 100% 활용하지 못하였고,

중복된 베뉴 데이터들을 처리하지 못하였다. 또한 [그림 2-3]과 같이 현재 포스퀘어 서비스는 모든 연령대에서 고르게 사용되지 못하고 있으며, 이는 본 연구 결과가 주로 25~44세 연령대의 사람들을 만족시키는 결과라는 한계점을 시사한다. 셋째, 실험 지역 범위 설정에 관한 문제이다. 핫 스트리트 군집을 통계적으로 판단하기 위해 사용한 Getis-Ord의 G_i^*는 사용된 데이터의 수에 따라 전체 평균과 표준편차 등이 달라지기 때문에, 실험 지역 범위에 따라 결과가 달라질 수 있다. 또한 너무 범위를 좁게 설정하면 데이터 수가 굉장히 적어 유의미한 결과를 기대하기 어렵게 된다. 넷째, 각 업소의 실제 출입구 방향을 완벽히 반영하지 못한 문제이다. 본 연구의 기법은 업소들의 출입구 방향이 해당 건물의 모든 방향에서 가능하다는 것을 가정하기 때문에, 출입구 방향이 특정한 방향성을 가지고 있는 경우에도 불필요한 거리까지 테마거리로 탐지하게 된다. 다섯째, 소수의 업소 및 건물에 의해서만 도로 세그멘트가 탐지되는 문제이다. 다른 업소들과 동떨어져 있는 유명 맛집이나 대형 몰 등은 군집을 형성하는데 기여하지는 않지만, 인기도가 비정상적으로 높아 그 주변 도로가 테마거리로 탐지되는 경우가 발생한다.

이와 같은 문제를 극복하기 위해서는 본 기법을 통해 탐지된 테마거리의 적정성을 정량적으로 평가할 수 있는 기법과 LBSNS 데이터와 건물 출입구 정보 등 활용 데이터의 정확도를 향상 시킬 수 있는 방안을 개발하여 이를 활용해야 할 것이다. 그리고 소수의 업소 및 건물에 의해서만 테마거리가 탐지되는 오류가 발생하지 않도록 실제 테마거리의 특성을 반영할 수 있는 핫 밸류에 대한 연구도 지속되어야 할 것이다.

참고 문헌

국토교통부. (2013). 도로의 구조·시설 기준에 관한 규칙. 국토교통부.

곽갑신. (2010). 도심 상업지 업종분포의 유형에 관한 연구 : 강남, 서초지역의 상업지 업종을 중심으로. 광운대학교 석사 학위논문.

김지혜. (2009). 특화가로 조성을 위한 환경디자인 구성 요소의 평가에 관한 연구. 충남대학교 석사 학위 논문.

김경수. (2011). 특화거리의 물리적 환경 분석을 위한 체크리스트 연구. 한남대학교 석사 학위논문.

류근원. (2012. 09. 22). 강남역 모임장소 아시안 요리 레스토랑은. 스포츠월드[온라인]. <http://www.sportsworldi.com/Articles/LeisureLife/Article.asp?aid=20120921024298>. (접속일 2013. 11. 05).

박근희. (2008. 03. 10). 봄바람이 안내하는 맛과 멋의 거리. 조선닷컴[온라인]. <http://danmee.chosun.com/site/data/html_dir/2008/03/10/2008031000996_2.html>. (접속일 2013. 12. 10).

박천보. (2009). 물리환경적 도심재생 관점의 특화거리 활성화 방안 연구. 대한건축학회, 25(8), 285-292.

신인지. (2012). 위치기반소셜네트워크 앱 서비스의 기능이 장소성 형성과 앱 이용 동기에 미치는 영향에 관한 연구 : 포스퀘어, 아임IN을 중심으로. 홍익대학교 석사 학위논문.

우종국. (2008. 09. 23). 새얼굴 불쑥불쑥, 여기가 어디야. 한국경제매거진[온라인]. <http://magazine.hankyung.com/main.php?module=news&mode=sub_view&mkey=1&vol_no=669&art_no=24&sec_cd=1701>. (접속일 2013. 11. 06).

이병길. (2003). 비지니스 GIS 에서 공간 데이터마이닝 (Spatial Data Mining) 기법을 이용한 상권추출. 한국 GIS 학회지, 11(2), 171-184.

이상일, 조대헌, 손학기, & 채미옥. (2010). 공간 클러스터의 범역 설정을 위한 GIS-기반 방법론 연구-수정 AMOEBA 기법. 대한지리학회지, 45(4), 502-520.

이용직. (2005). 아날로그 기법을 이용한 상권분석에 관한 연구 : 서울시 합정 균형발전촉진지구를 중심으로. 홍익대학교 석사 학위논문.

정아람. (2013. 05. 31). 도심 속 변두리길 저마다 개성이 있어 좋다. 세계일보[온라인]. <http://www.segye.com/content/html/2013/05/30/20130530004370.html>. (접속일 2013. 11. 02).

정한모. (2006. 10. 02). 혈세만 낭비한 '패션문화 거리'. YTN 뉴스[온라인]. <http://www.ytn.co.kr/_ln/0103_200610020814127112>. (접속일 2013. 11. 06).

한범수, 김희영. (2007). 문화관광자원으로서의 테마거리의 발굴에 관한 연구-서울의 테마거리를 중심으로. 여가관광연구, 10(단일호), 77-90.

Aldstadt, J., & Getis, A. (2006). Using AMOEBA to create a spatial weights matrix and identify spatial clusters. Geographical Analysis, 38(4), 327-343.

Anselin, L. (1995). Local indicators of spatial association－LISA. Geographical analysis, 27(2), 93-115.

Benedikt, M. L. (1979). To take hold of space: isovists and isovist fields. Environment and Planning B, 6(1), 47-65.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. M. (2012, June). The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In ICWSM.

Getis, A., & Ord, J. K. (1996). Local spatial statistics: an overview. Spatial analysis: modelling in a GIS environment, 264-265.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211-221.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. Spatial statistics, 1, 110-120.

Jacobs, J. (1961). The death and life of great American cities. Random House Digital, Inc..

Li, Y., Steiner, M., Wang, L., Zhang, Z. L., & Bao, J. (2013). Exploring Venue Popularity in Foursquare. In NetSciCom, 2013 5th International Workshop on Network Science for Communication Networks. IEEE.

Lu, Y. (2005). Approaches for Cluster Analysis of Activity Locations along Streets: from Euclidean Plane to Street Network Space. Cartography and Geographic Information Society.

Mohaymany, A. S., Shahri, M., & Mirbagheri, B. (2013). GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. Geo-spatial Information Science, 1-7.

Okabe, A., Satoh, T., & Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS based tool. International Journal of Geographical Information Science, 23(1), 7-32.

Paek, J., Kim, J., & Govindan, R. (2010, June). Energy-efficient rate-adaptive GPS-based positioning for smartphones. In Proceedings of the 8th international conference on Mobile systems, applications, and services (pp. 299-314). ACM.

Shiode, S. (2011). Street level Spatial Scan Statistic and STAC for Analysing Street Crime Concentrations. Transactions in GIS, 15(3), 365-383.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 46, 234-240.

Wang, M., Qin, L., & Hu, Q. (2012, June). Data mining and visualization research of check-in data. In Geoinformatics (GEOINFORMATICS), 2012 20th International Conference on (pp. 1-4). IEEE.

부 록

<부록 A> 테마 별 포스퀘어 카테고리 목록

테마	포스퀘어 카테고리 목록			
음식점	음식	아프가니스탄 음식점	아프리카 음식점	미국 음식점
	아레파 전문점	아르헨티나 음식점	아시아 음식점	호주 음식점
	바비큐 전문점	브라질 음식점	아침식사 식당	부리토 전문점
	케이준 크리올 전문점	카리브해 음식점	중국 음식점	쿠바 음식점
	딤섬 전문점	아메리칸 다이너	만두 전문점	동유럽 음식점
	에티오피아 음식점	팔라펠 전문점	필리핀 음식점	피시 앤 칩스 음식점
	포장마차	프랑스 음식점	프라이드 치킨 전문점	독일 음식점
	글루텐 무함유 식품 전문점	그리스 음식점	핫도그 전문점	인도 음식점
	인도네시아 음식점	이탈리아 음식점	일본 음식점	한국 음식점
	남미 음식점	치즈 없은 마카로니 전문점	말레이시아 음식점	지중해 음식점
	멕시코 음식점	중동 음식 전문점	분자 요리학 전문점	몽고 음식점
	모로코 음식점	뉴 아메리카 음식점	페루 음식점	피자 전문점
	포르투갈 음식점	라면, 국수 전문점	음식점	샐러드 바
	샌드위치 가게	스칸디나비아 음식점	해산물 전문점	스낵 전문점
	수프 전문점	남미 음식점	남부지방 흑인요리 전문점	스페인 음식점
	빠에야 전문점	스테이크 전문점	스시 전문점	스위스 음식점
	타코 전문점	타파스 전문점	태국 음식점	터키 음식점
	채식주의 음식 전문점	베트남 음식점	날개 요리 전문점	
패션	의류매장	액세서리 가게	고급 여성복	아동복
	여성용 속옷 전문점	남성복 매장	신발 매장	여성복 매장
	화장품 매장	백화점	보석 가게	쇼핑몰
	네일샾	양복점		
카페	카페	커피숍	찻집	
유흥 주점	유흥 명소	바	비어 가든	칵테일 바
	동네 주점	게이바	물담배 바	호텔 바
	가라오케바	라운지	나이트클럽	기타 유흥가

	호프집	일본식 주점	무허가 술집	스포츠 바
	스트립 클럽	위스키 바	와인 바	

<부록 B> 도로명 주소 기본도 건물 레이어 주요 속성

필드명	컬럼한글명	설명(C)
BD_MGT_SN	건물관리번호	■ PNU코드(19) + 일련변호(6) - PNU코드(19) 시도(2)+시군구(3)+읍면동(3)+리(2)+ 대장구분(1)+본번(4)+부번(4)+일련번호(6)
RD_MGT_SN	도로구간 관리번호	■ 시스템구분코드(1) + 시군구코드(5) + 일련번호(6) 예) T + 30200 + 000001 * 시스템구분코드 통합센터 : T / 시도지역센터 : I / 시군구자료센터 : P
EQB_MGT_SN	건 물 군 관리번호	■ 시군구코드(5) + 일련번호(6)
RN_MGT_SN	도 로 명 관리번호	■ 시군구코드(5)+대로/로코드(3)+길코드(4)
UDRT_YN	지하여부	■ 0 : 지상, 1 : 지하
BD_MA_SN	건물번호_본번	.
BD_SB_SN	건물번호_부번	.
BD_NM	건물명_한글	■ 건축물대장상의 건물명을 입력
BD_ENG_NM	건물명_영문	.
DET_BD_NM	상세건물명	■ 건물명 외에 추가로 입력할 상세명칭 예) 105동, 본관, 별관, 신관 등
BD_STS_CM	건물상태	■ 건물상태 : 공사중/공사중아님
BD_KD	건물용도코드	■ 건물용도 코드 정의서 참조
BD_IN_SE	건물종속구분	■ M : 주건물, S : 종속건물, (B:부속건물)

ABV_FL_CNT	지상층수	■ 건물의 지상 총 층수를 입력
UDR_FL_CNT	지하층수	■ 건물의 지하 총 층수를 입력
ZIP_CL	우편번호	
EMD_CD	행정구역 읍면동코드	■ 시도(2) + 시군구(3) + 읍면동(3) + 리(2)
MT_YN	산여부	■ 0 : 토지, 1 : 산(임야)
LNDN_MA_SN	지번_본번	■ 대표지번(본번)을 입력 ※ 관련지번은 별도 테이블에서 관리
LNDN_SB_SN	지번_부번	■ 대표지번(부번)을 입력
BD_MK_YMD	준공일자	■ 건물의 준공 일자를 입력
WRK_MGT_SN	작업관리번호	
UPDATE_YMD	갱신일자	■ 자료가 최종으로 갱신된 일자
ANN_YN	고지대상 여 부	■ 0 : 대상아님, 1 : 고지대상
ANN_YMD	고지대상 선정일자	.
NOTICE_YMD	고시일자	.
ANN_DE	고 지 상세설명	.
ANN_SE	고지방법 구 분	■ 1 : 인편, 2 : 우편, 3 : 기타
RESULT_SE	고지결과 구 분	■ 0:고지진행중, 1:고지완료, 2:고시완료
PRECISE_SE	입력방법	■ 1 : 직접입력(묘화입력) 2 : DXF 입력 3 . 측량좌표에 의한 입력 4 : 지적건축물 입력
UP_YMD	입력일자	■ 객체를 입력한 년월일
TRS_YMD	전송일자	■ 객체를 전송한 년월일

<부록 C> 도로명 주소 기본도 도로구간 레이어 중 속성

필드명	컬럼한글명	설명
RD_MGT_SN	도로구간 관리번호	■ 시스템구분(1)+시군구코드(5)+일련번호(6) 예) P + 44750 + 000001 ※ 시스템구분 시군구:P, 시도통합센터:I, 중앙통합센터:T 시도복사본:C, 중앙복사본:B
UNIT_SN	단위구간번호	■ 도로구간을 나누는 경우 (구제시)
RD_NM	도로명_한글	■ 한글 도로명
RD_ENG_NM	도로명_영문	■ 영문 도로명
RN_MGT_SN	도 로 명 관리번호	■ 시군구코드(5)+대로,로 코드(3)+길코드(4) 예) 11450+013+0001
SCALE_SE	도로위계 법적구분	■ 도로의 규모에 따른 법적분류 01:대로, 02:로, 03:길
GRADE_SE	도로위계 기능구분	■ 도로의 기능에 따른 분류 00:고속도로, 01:주간선, 02:보조간선, 03:집산도로, 04:소로, 05:골목길
RD_IN_SE	도로구간 종속구분	■ 0 : 주도로(메인도로) 1 : 1차종속도로 2 : 2차 종속도로
RD_BT	도로폭	■ 도로의 실제 폭
RD_LT	도로길이	■ 도로의 실제 길이

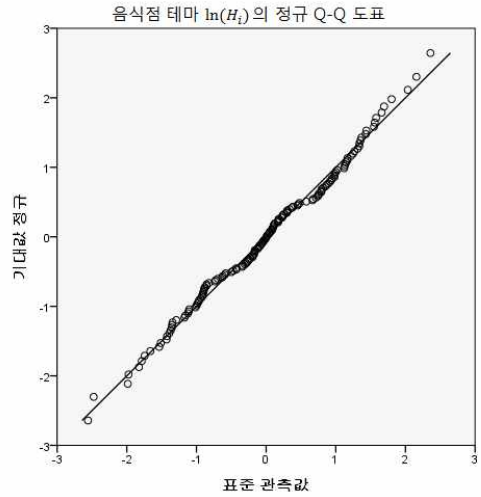
<부록 D> 도로명 주소 기본도 출입구 레이어 주요 속성

필드명	필드 한글명	설명
ENT_MGT_SN	출 입 구 관리번호	■시군구코드(5) + 일련번호(6)
ENT_SE	구분코드	■RM : 주출입구, RS : 보조출입구 VM : 가상주출입구, VS : 가상보조출입구
BD_MGT_SN	건물관리번호	■해당 건물의 관리번호
EQB_MGT_SN	건 물 군 관리번호	■해당 건물군의 관리번호

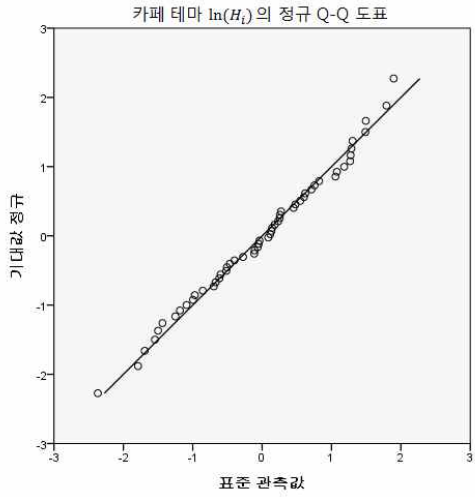
<부록 E> 도로명 주소 기본도 연결선 레이어 주요 속성

필드명	필드 한글명	설명
ENT_MGT_SN	출 입 구 관리번호	■시/구코드(5) + 일련번호(6)
RD_MGT_SN	도로구간 관리번호	■시스템구분(1)+시군구코드(5)+일련번호(6) 예) P + 44750 + 000001 ※ 시스템구분 시군구:P, 시도통합센터:I, 중앙통합센터:T 시도복사본:C, 중앙복사본:B
SP_MGT_SN	기초구간 일련번호	■기초구간 일련번호(6)
DIR_SE	방향성	■L : 왼쪽, R : 오른쪽
DIST_LT	거리	■출입구와 도로 사이의 거리

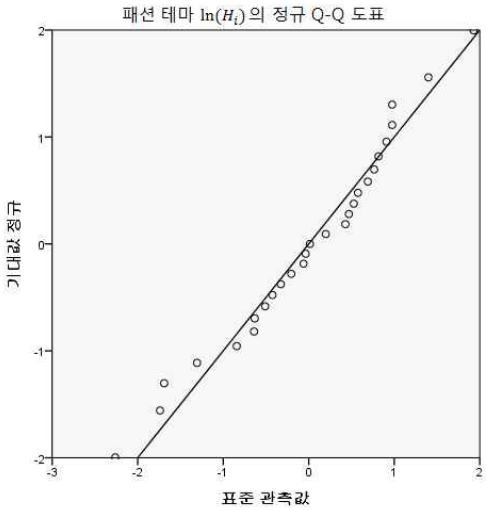
<부록 F> 신림 사거리 지역의 각 테마 별 $\ln(H_i)$의 정규 Q-Q plot 결과



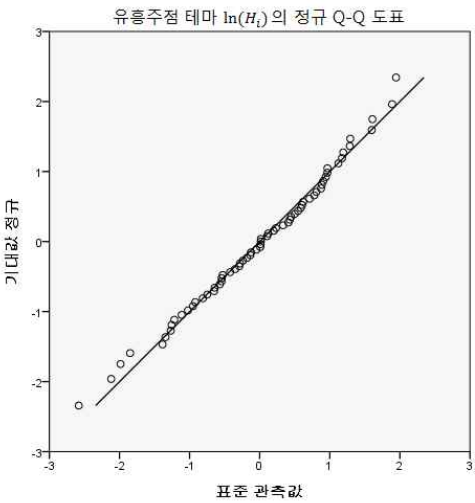
(a) 음식점 테마 정규 Q-Q plot



(b) 카페 테마 정규 Q-Q plot

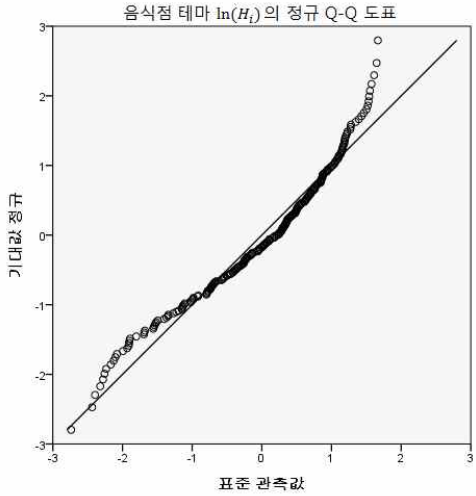


(c) 패션 테마 정규 Q-Q plot

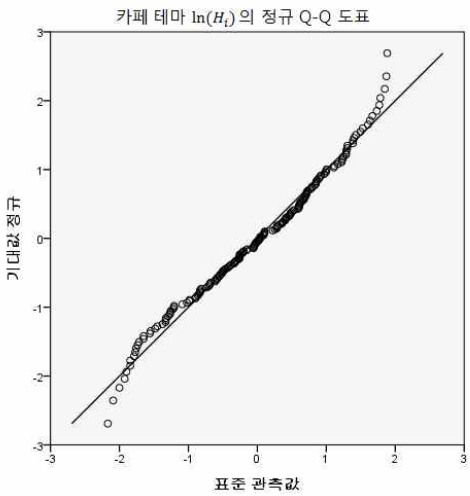


(d) 유흥 주점 테마 정규 Q-Q plot

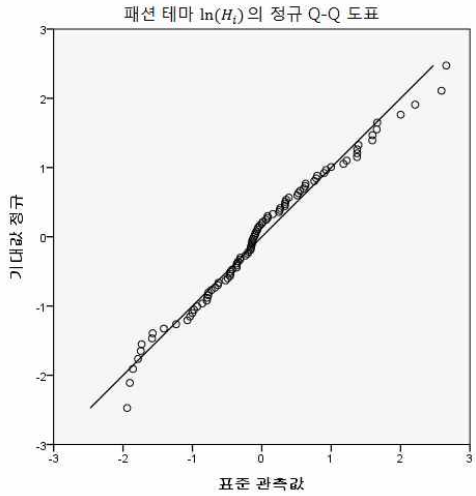
<부록 G> 강남역 번화가 지역의 각 테마 별 $\ln(H_i)$의 정규 Q-Q plot 결과



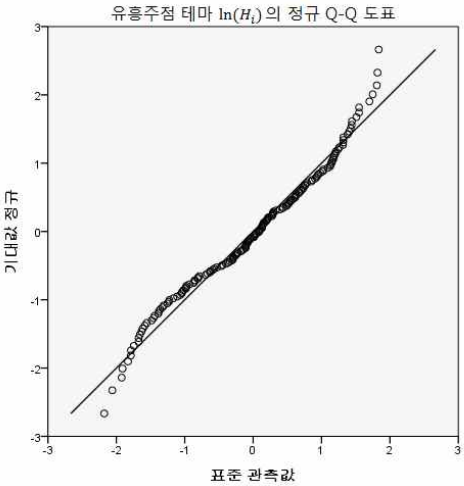
(a) 음식점 테마 정규 Q-Q plot



(b) 카페 테마 정규 Q-Q plot



(c) 패션 테마 정규 Q-Q plot



(d) 유흥 주점 테마 정규 Q-Q plot

ABSTRACT

A Variety of themed streets have recently been developed by local governments in order to activate the local economy and establish the identity of the place. However, some themed street projects, which have been conducted in profit-oriented ways without fully considering people's perceptions of their local area, have consequently resulted in a marginal effect on the local economy. Apart from the themed streets designated by local governments, there are also themed streets spontaneously formed by the locals and still unknown to the public. In addition, various types of themed streets are developed through distinct regional specializations even within the same downtown. If a new method which can detect themed streets in an objective and efficient way provides the public with more information about the street, this will enrich people's sociocultural space and activate the local economy.

Therefore, this study proposes a hot street clustering method to detect the themed street of the specific region, focusing on the commercial themed street which is more prevalent than other types of themed streets. This study especially uses the street as a basic unit for the analysis and the result in order to reflect the street-centered characteristic of the urban system and delineate apparent area of the themed street. In addition, LBSNS data, which informs of people's perception about the place and popularity of the place, is also employed to detect the activated themed street.

A proposed method utilize 'Spatial Join Operation' and 'Isovist Analysis' in order to join the LBSNS data's information, people's perception about the place and popularity of the place, to the buildings and the streets in sequence. This method devises and computes the unique value, named

'hot value', which determine whether the street is a hot street, considering popularity of the place, the street length, number of building around street and etc. This method calculates the Getis-Ord G_i^* statistic(namely z-score) for each street which has the computed hot value, and detect statistically significant spatial clusters. As a result, these clusters are named 'hot street cluster' and are considered themed street of that region.

This study performs an experiment in Sillim intersection and Gangnam station area to evaluate adequacy of the proposed method. The experiment results detected 'Sillim fashion-cultural street' which has been actually turned into the food street and 'Gangnam cafe street' which has recently been developed. By comparing the commercial areas previously surveyed with commercial themed streets detected in this study, the experiment results showed high adequacy of the proposed method.

Keywords : Detection of the Themed Street, Hot Street, LBSNS (Location Based Social Network Service), Getis-Ord G_i^*, ISOVIST

Student Number : 2012-20923