Mapping and Analysis of Crime in Aurangabad City using GIS

Shahebaz M. Ansari¹, Dr. K. V. Kale²
¹, ²(Department of Computer Science and IT; Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, M.S., India)

Abstract: The study of Earth and its features with the help of Remote Sensing has lead to a fast growing and effective technology called Geographic Information System (GIS) where digital Maps and spatial data can be used for visualizing, analyzing and taking decision based on the analysis. Crime Analysis has become an integral part of GIS where the Analyst uses the spatial and temporal aspect of Crime information for analysis and forecasting. Using Hotspot analysis technique the Hotspot means the area where the concentration of Crime is more can be found. There are different classes of Hotspot detection such as Spatial Analysis, Interpolation and Spatial Autocorrelation for finding out the Crime Hotspot. In this paper Kernel Density Estimation, Inverse Distance Weighted and Getis-Ord Gi* methods from each classes respectively are discussed. These methods are applied on Aurangabad city of Maharashtra state, India for finding out the hotspots for crime incidence like Murder, Day House Break and Night House Break.

Keywords: Crime Mapping, Getis-Ord Gi*, GIS, Hotspot, KDE

I. Introduction

A computer framework which is used for capturing, preserving, querying, exploring and displaying geospatial data is known as GIS. The geospatial data also known as geographically referenced data describes both the location and the characteristics of the feature called spatial feature which may consist of roads, land extents, and vegetation on the surface of the Earth [1] [2]. The thing that differentiates GIS from other information system is its capability of handling and performing operations on geospatial data. The spatial data may be the location while the attribute data is the characteristics possessed by that location [1] [3].

In India, Police is a critical component of civil administration. There has been an enormous increase in crime in the recent past. The cops in their role to catch criminals are required to remain convincingly ahead in the eternal race between law breakers and law enforcers. One of the key concerns of the law enforcers is how to enhance investigative effectiveness of the police. There is an urgent need to analyze the increasing number of crimes as approximately 17 lakhs Indian Penal Code (IPC) crime, and 38 lakhs local and Special Law crimes per year. The police should use the current technologies to give themselves the much-needed edge [4].

Crime analysis mapping is the process of utilizing the geographic information system possessing together crime analysis techniques to concentrate on the spatial context of crime and other law enforcement procedures [2]. Maps offer graphical representation of the crime and its related issues. To improve the tact of fighting crime one must understand why and where the crime took place. Patrolling can be improved by providing the maps displaying the crime location or the area where the concentration of crime is high. A map consisting of trends of criminal activity, high density areas, and temporal information can be very much useful for the policy makers of the police departments [5].

1.1 Hotspot

The term hot spot has become an integral part of the study called crime analysis and has is popular with most of the analyst. A hot spot as the name suggests is a state of indicating some arrangements of clusters in a spatial distribution. Nevertheless, not all clusters are hot spots since the environments that provide existence crime; the residence places are—also apt to be clusters [6]. The cluster of geographical areas consisting of habitually high number of crime events are known as Hotspots [7]. It will help to easily determine the site where the action can be taken.

GIS based hot spot identification in the criminal cases is to precisely state the characteristics of criminal cases, the use of spatial analysis tools and techniques of criminal case which had took place to determine the spatial model, hot form of calculation action [8].

II. Study Area and problem definition

We have selected Aurangabad City for our research work as it is an International tourist place because Ajanta and Ellora Caves, and many more. Aurangabad district comprises of 9 Talukas and 1344 villages, covers an area of 10,100 sq. km. with the population of 36.95 Lakhs. Aurangabad city is at Decimal Degree19.880 N and 75.320 E. Aurangabad Municipal Corporation is the Prime planning organization consisting of six zones
with a population of approximately 11,89,376 as per census 2011. Its total area is nearly 138.44 sq. km. Aurangabad city consist of three main Police Zones and having total twenty Police Stations [9].

In today day to day life the crime around us has significantly increased. There are different types of crimes such as burglary, theft, murder, etc. whose First Information Report (FIR) is filed. The officer has to look into the lists of records for analysis. This is time consuming and will result in the delay of action to be taken during Emergency situations [10].

As the Aurangabad City is one of the fastest growing cities in Maharashtra, the crime rate is also increased in the past few years. As it is a Tourist spot, fraud cases and theft of personal belongings of the tourist also comes into the picture. The concept of Hot Spot has become an integral part of the study called crime analysis. A Hot Spot is a state of suggesting some arrangements of clusters in a spatial distribution [11]. The clusters consisting of extremely high number of crime events within a geographical area are known as Hot Spots [7]. Getting the knowledge of concentration of crime within an area, law enforcement agencies can enact timely and effective judgement of assigning police resources. In addition, Crime Hot Spots mapped by using GIS technology can admirably communicate crime patterns and crime prevention policies to decision makers and the public [8].

III. Literature Survey

The GIS and Remote Sensing can be used for visualizing the data, analyzing the facts, and to take firm decision based on the analysis. This can be used to map the Police stations and to identify the Crime Zone as Hot Spot and to statistically analyze the reported Crime which will help to take effective measures to control the crime [12]. There are different Hotspot detection techniques such as KDE, Kriging, Anselin Local Moran's I, Getis-Ord Gi* which is shown in Fig. 3.1

![Figure 3.1 Shows different Methods for Hot Spot Analysis](image)

3.1 Spatial Analysis Methods

Spatial Analysis is the process of checking the locations, attributes, and connection of features in spatial data among overlay and other analytical techniques; which is used for acquiring knowledge that can be used in different aspect. Spatial analysis creates or extracts different new information from spatial data [13]. Different techniques under this category are Kernel Density Estimation (KDE) [14] [15], Point Density [15], Line Density.
3.2 Interpolation Methods

Interpolation is the process of using points with better-known values to propose values at alternate unknown points. It is often used to foretell unknown values at any geographic point data, which can be used in the field of rainfall, noise levels, chemical concentrations, elevation, or other spatially-based phenomena. It is the approximate judgment of surface values at the points which are un-sampled based on the surface values of surrounding points which are known. Interpolation is usually used as a raster operation, but using a TIN (Triangulated Irregular Networks) surface model it can be used as vector operation [13] [16]. There are several well-known interpolation techniques such as Inverse Distance Weighted (IDW), Kriging, Spline, and Natural Neighbor.

3.3 Spatial Autocorrelation

Spatial Autocorrelation is an amount of the degree to which a set of spatial features and the data values associated with it. It can be clustered together in Space (positive spatial autocorrelation) or Scatter Widely (negative spatial autocorrelation) [13]. Different methods under Spatial Autocorrelation are Anselin Local Moran's I [14] [17] and Getis-Ord Gi* [1].

IV. Methodology

To implement the Hotspot techniques the methodology shown in the Fig 4.1 is followed. It consist of six steps before reaching to the Result and Analysis part. Data Collection is the basic step for implementation, after that Georeferencing of the raster dataset takes place. In the third step the Spatial features from the dataset are digitized and then the police stations and their boundaries are mapped. In the fifth step crime data base is created and are mapped after which the Hotspot methods are applied on to the dataset and then the Result and Analysis is done.

4.1 Data Collection

The Toposheet of Aurangabad (M.S., India) has been used for implementation; the detail for the same is given below:
- Creator: Survey Of India
- Year: 2005-06
- Scale: 1:50,000
- Sheet: E43DM (47 M/5), Aurangabad, Maharashtra
- Projection: UTM, 43N
- Datum: WGS 84
- Other: Magnetic Variation from True North about $\frac{3}{4}$ West in 2005 (Decreasing by about 1’ annually)
Crime data including the Spatial information of the Crime is required form the FIR filed from the Commissioner of Police office Aurangabad City has been collected. For the crime mapping hypothetical data which is based on the coordinates of the Toposheet and assumption is used for experimental purpose. The Crime of Murder, Day House Break and Night House Break has considered. Data about the name and location of Police Stations under Aurangabad City and their zone boundary has gathered. [13] [14].

### Figure 4.2 Toposheet of Aurangabad City of the year 2005-06

![Toposheet](image)

4.2 Georeferencing of the Toposheet

In the process of Georeferencing registration of the digitized Toposheet into the Geographic Coordinate System (GCS) takes place. The GCS for Aurangabad is WGS_84. Aurangabad comes under the Universal Transverse Mercator (UTM) zone 43 N [18].

4.2.1 Adding the dataset into the ArcMap

While adding a raster dataset into an ArcMap that does not have pyramids, you will be prompted to build pyramids. Pyramids are necessary because they improve the speed of drawing for raster datasets which are displayed at less than their full resolution. [1] [19].

4.2.2 Fundamentals for georeferencing a raster dataset

Raster data is generally obtained by scanning the maps or collecting aerial photographs and satellite images. Scanned map datasets lack spatial reference information. Thus, the dataset is to be aligned or georeferenced to a map coordinate system for using it with other datasets. A map coordinate system is defined using a map projection; a method by which the Earth’s curved surface is portrayed on a flat surface [19] [20]. Following are the general steps for georeferencing a raster dataset:

a) In ArcMap, add the raster dataset that need to be projected

b) Add control points that link the known positions of raster dataset with the known positions in map coordinates.

c) Register the dataset means save the Georeferencing information all the control points are properly placed.
d) Transform the raster dataset permanently using Rectify (this is optional).

During this process a series of ground control points i.e. known \(x, y\) coordinates that is used to link locations on the raster dataset with locations in the spatially referenced target data [20]. Mathematically, the affine transformation is expressed as a pair of first-order polynomial equations [1]:

\[
X = Ax + By + C \\
Y = Dx + Ey + F
\]

where \(X\) and \(Y\) are the input coordinates that are given, \(X'\) and \(Y'\) are the output coordinates to be determined and \(A, B, C, D, E\) and \(F\) are the transformation coefficients. The affine transformation is also called the six-parameter transformation because it involves six estimated coefficients [1] [18] [20].

4.2.3 Interpreting the Root Mean Square error

After the application of transformation on the control points the residual error is returned. The error is the difference between the ‘from point’ to the actual location the ‘to point’ position. Then total error is computed by applying the root mean square (RMS) sum of all the residuals to compute the RMS error. The obtained value describes how consistent the transformation is between the different control points (links). If the error is particularly large the ground control points can be adjusted, removed or added to adjust the error [21]. Mathematically, the input or output error of control point is computed by [1]:

\[
\sqrt{(x_{act} - x_{est})^2 + (y_{act} - y_{est})^2}
\]

where \(x_{act}\) and \(y_{act}\) are the \(x\) and \(y\) values of the actual location and \(x_{est}\) and \(y_{est}\) are the \(x\) and \(y\) values of the estimated location.

The average RMS error can be computed by averaging errors from all control points [22]:

\[
\sqrt{\frac{\sum_{i=1}^{n}(x_{act,i} - x_{est,i})^2 + \sum_{i=1}^{n}(y_{act,i} - y_{est,i})^2}{n}}
\]

where \(n\) is the number of control points \(x_{act,i}\) and \(y_{act,i}\) are the \(x\) and \(y\) values of the actual location and \(x_{est,i}\) and \(y_{est,i}\) are the \(x\) and \(y\) values of the estimated location of control point \(i\) [22].

If more control points of equal quality are used, the polynomial can convert the input data to output coordinates more accurately. Typically, adjust and spline transformations give an RMS of nearly zero or zero; however, this does not mean that the image will be perfectly georeferenced [23].

4.2.4 Resampling the raster dataset

After the satisfied RMS error has reached the next step is to rectify or transform a raster dataset; convert it from one projection to another; or change the cell size, these are geometric transformation. Geometric transformation is the process of changing the geometry of a raster dataset from one coordinate space to another [23].

The three most common resampling techniques are Nearest Neighbor Assignment, Bilinear Interpolation, and Cubic convolution. These techniques assign a value to each empty cell by examining the cells in the uncoherent raster dataset [24].

Nearest neighbor assignment is the fastest resampling technique and is appropriate for categorical or thematic data, since it does not alter the value of the input cells. Once the location of the cell's center on the output raster dataset is located on the input raster, nearest neighbor assignment determines the location of the closest cell center on the input raster and assigns the value of that cell to the cell on the output raster [1] [19]. Bilinear interpolation uses the value of the four nearest input cell centers to determine the value of the output raster. This interpolation method results in a smoother-looking surface than can be obtained using nearest neighbor [1]. Cubic convolution is similar to bilinear interpolation, except the weighted average is calculated from the 16 nearest input cell centers and their values. Cubic convolution tends to sharpen the data more than bilinear interpolation, since more cells are involved in the calculation of the output value. [1].
4.3 Digitization

Digitization of Spatial feature means to extract the Road network using the Editor toolbar. We can select from the available features from the set like Point, Line and Polygon. As for Road Network the Line Feature is required. Network Index of Road network is calculated for finding out the optimal path to reach the Crime scene and the Police station.

4.4 Mapping Police Stations and their Boundary

In this phase the Police stations are placed on the map according to their coordinates. For coordinates Field Survey is performed. The Police station is the Point Feature. The zone boundary information gathered in the first phase is used for mapping the boundary according to the police stations. The polygon feature is used for representing the boundary.
4.5 Mapping different types of Crimes
After mapping the Police stations and their zone boundary now the Crime data is mapped on to the map. For Crime mapping Field Survey is required in order to collect the coordinates of the Crime scene. The Crime scene is mapped as a Point feature.

![Figure 4.7 Murder Crime, Day House Break Crime and Night House Break Crime in Aurangabad](image)

**V. Result and Analysis**
In the last phase different Statistical analysis is performed to analyze the crimes took place at different places. The performance of different methods can be judged from the output they provide. In this phase from the different Hotspot detection techniques discussed in Phase 6 some are implemented which shows the satisfied result for identifying the Hotspot. The methods that are implemented are KDE, IDW and Getis-Ord Gi* for the Murder Crime.

5.1 KDE Method
Density Estimation measures cell densities in a raster by using a sample of known points. Kernel Density Estimation associates each known point with a Kernel function. This can be expressed as a bivariate probability density function, a kernel function looks like a “bump”, centring at a known point and tapering off to 0 over a defined bandwidth or window area [14] [15].

The Hotspot detected using the KDE is shown in the dark Red color. It can be seen that the area with different intensity of crime can be easily detected. KDE shows the smoothing effect and can be applied only on the Points and Polylines in Fig. 5.1. For Murder Crime it is seen that Hotspot areas are the area with Dark Red color concentration and the Coldspot areas are the area having Light Yellow color. Fig. 5.1 (a) and (b) shows the same result despite of images from different source. Fig 5.1 (a) is a Toposheet Map and (b) is a map of Aurangabad City on Google Map.

![Figure 5.1 The Hotspot identified by the KDE method (a) shows result on Toposheet and (b) shows result on the Google Map’s image](image)
5.2 IDW Method

Inverse Distance Weighted (IDW) Interpolation is an exact method that enforces the condition that the estimated value of a point is influenced more by nearby known points than by those farther away [1]. IDW creates layered shaded area of the crime incidence. Similar to the KDE the Dark Red color area shows the Hotspot and the Light Yellow color shows the Coldspot. Fig. 5.2 (a) and (b) shows the same result despite of images from different source i.e. a Toposheet Map and a map of Aurangabad City on Google Map respectively. IDW works only on the Point feature.

![Figure 5.2 The Hotspot identified by the IDW method on Toposheet (a) and on Google Map’s image (b)](image)

5.3 Getis-Ord Gi* Method

G-statistics, which can separate clusters of high values from clusters of low values [25]. The result of Getis-Ord Gi* method is shown in Fig. 5.3. The circles with the Dark Red color shows the Hotspot and that with Dark Green color shows the Coldspot area. The Gi* results in point wise Hotspot detection. It does not create the smoothing or the layer effect.

![Figure 5.3 The Hotspot identified by the Getis-Ord Gi* method on Point](image)

The Fig. 5.4 shows the Police station boundary with high crime incidence in the Dark Red Color. The Boundary with Dark Green Color is the Coldspot. Gi* can be applied to Point and the Polygon features. For Murder Crime it is seen that Hotspot area is of Dark Red color and Coldspot area is of Dark Green color. Similarly for Day House Break and Night House Break the Hotspots and Coldspots are identified.
The graph in the Fig. 5.6 shows different Police station Boundaries with their crime incidence showing the rate of Murder, Day House Break and Night House Break crimes under each boundary. The maximum and minimum crime rate areas can be identified which is based on the assumed hypothetical data based on the coordinates of the Toposheet.

In the Fig. 5.7 shortest path from Police station up to the Hotspot is seen, this can help the police personal to take quick and timely action based on the analysis. Dijkstra’s algorithm is used for calculating the shortest path from the Police station to the Hotspot.
VI. Conclusion

Mapping of Crime such as Murder, Day House Break and Night House Break is done and different Crime Analysis techniques are used to find out the Crime Hotspot in Aurangabad City. It will help the Police personal as well as the Tourist visiting these places by providing information regarding which area is safe and where the concentration of particular crime is more. For Point feature it is identified that KDE and IDW provide satisfactory result both on the Toposheet map and the Google Map and provide same result despite of change in the image acquiring process of same area. Similarly Getis-Ord Gi* provide satisfactory results for Polygon feature.

During the implementation it is found that KDE, IDW and Getis-Ord Gi* methods from Spatial Analysis, Interpolation and Spatial Autocorrelation respectively are more precise than the others. KDE and IDW can be applied on Point and Polyline features and produces smoothing effect layered effect respectively. Getis-Ord Gi* can be used with Point and Polygon features and it produces simple point like result on Point feature.

VII. Future Scope

To improve the quality of the result and its preciseness real time data of crime incidence can be used. Using temporal data of crime occurrence can help predicting the future crime occurrence and reducing crime incidence. The real time Global Positioning System (GPS) data can help the patrolling team to reach the crime scene from the shortest path which can reduce the impact of crime.

Acknowledgements

Authors would like to acknowledge and extend heartfelt gratitude to UGC (University Grant Commission) for development of Geospatial Technology Laboratory to complete this work under UGC SAP (II) DRS Phase-1 F. No.-3-42/2009 and One Time Research Grant F. No. 4-10/2010 (BSR) to Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, M.S., India.

References