Using Random Forest to Analyze Spatial Location of Motorcycle Involved Collision

Tien-Pen Hsu [a], Ku-Lin Wen [b]

[a,b] *Department of Civil Engineering, National Taiwan University, Taipei, 106, Taiwan*
[a] *E-mail: hsutp@ntu.edu.tw*
[b] *E-mail: wengulin@ntu.edu.tw*

Abstract

Conventional studies created the safety improvement countermeasures for intersection mostly based on the analysis of the relation between the accident collision types and the accident risk factors directly. They used to ignore the real accident occurred location within an intersection. However, different collision locations of same collision types should correspond different accident causes and different countermeasures accordingly, especially, if we deal with the motorcycle (scooter) involved accidents due to the widespread occurrence location at an intersection. The study discussed the accident risk by taking the real occurrence location into account by applying random forest analysis and taking a typical collision type of left turn opposing other angle collision as example. Through such involvement, the research results show that the highest related three major causes of such kind of accident collision type consist of the intersection geometry intersection angle, the yellow time and the width of curb slow traffic lane, which are difficult to be identified without taking the accident occurrence location into consideration. It depicted more clearly to identify the causes and then to generate the improvement design principles for corresponding issues of accident to improve the intersection safety.

Keywords: Motorcycle, Accident, Random forest, Location

1. INTRODUCTION

The motorcycle ownership of Taiwan is the highest in the world with 608 motorcycle per 1000 people. It causes the high accident occurrence between motorcycle and other motor vehicle under mixture of high motorcycle traffic composition. The registered vehicle number of car reached about 7.7 million, and motorcycle is about 13.6 million now. In recent years (2002 to 2014), the annual growth rate of car was 2.0%, and motorcycle was 1.6% [**1**]. However, the number of casualties of accident is growing rapidly. The accident annual growth rate of car is 4.3%, and motorcycle is 12.7% [**2**]. The accident occurrence growth rate is larger than the registered number growing rate of car and motorcycle. Wherein, the number of casualties caused by motorcycle accident is much higher than car accident. For analyzing the causes of increasing accident occurrence in order to create the improvement strategies, in Taiwan, every year , the ministry has conducted a nationwide accident prone intersection safety improvement program since a long time[**3**], and in general the local county and city government authorities also conducted accident improvement works regular[4]. The safety improvement consequent is limited. Therefore, the study aimed at using new method to find out the causes more accurately, especially, by taking the location into account and using random forest method.

2. TRAFFIC ACCIDNET FEATURES

For managing the mixed traffic with high motorcycle composition, in Taiwan, there is a regulation rule called “segregated flow” is conducted since decades in order to reduce the conflict between car and motorcycle [**7**]. The segregated flow strategies consist of following measures:

1. Motorcycle cannot use express lane:
   Motorcycle only can use mixed traffic lane, slow traffic lane, motorcycle only lane, and permissive motorcycle lane.
2. Motorcycle should make two stage left-turn:

In principle, motorcycle should left turn by two stage let-turn method at more than three-lane intersection.

3. Motorcycle head start holding zone:
   There is a motorcycle waiting zone behind the stop line to separate the motorcycle and other vehicle.

Therefore, a typical layout of intersection on urban street will consist of motorcycle-only lane, motorcycle two-stage left-turn waiting zone, motorcycle head start holding zone, and motorcycle forbidden inner lane, as illustrated in Fig. 1.

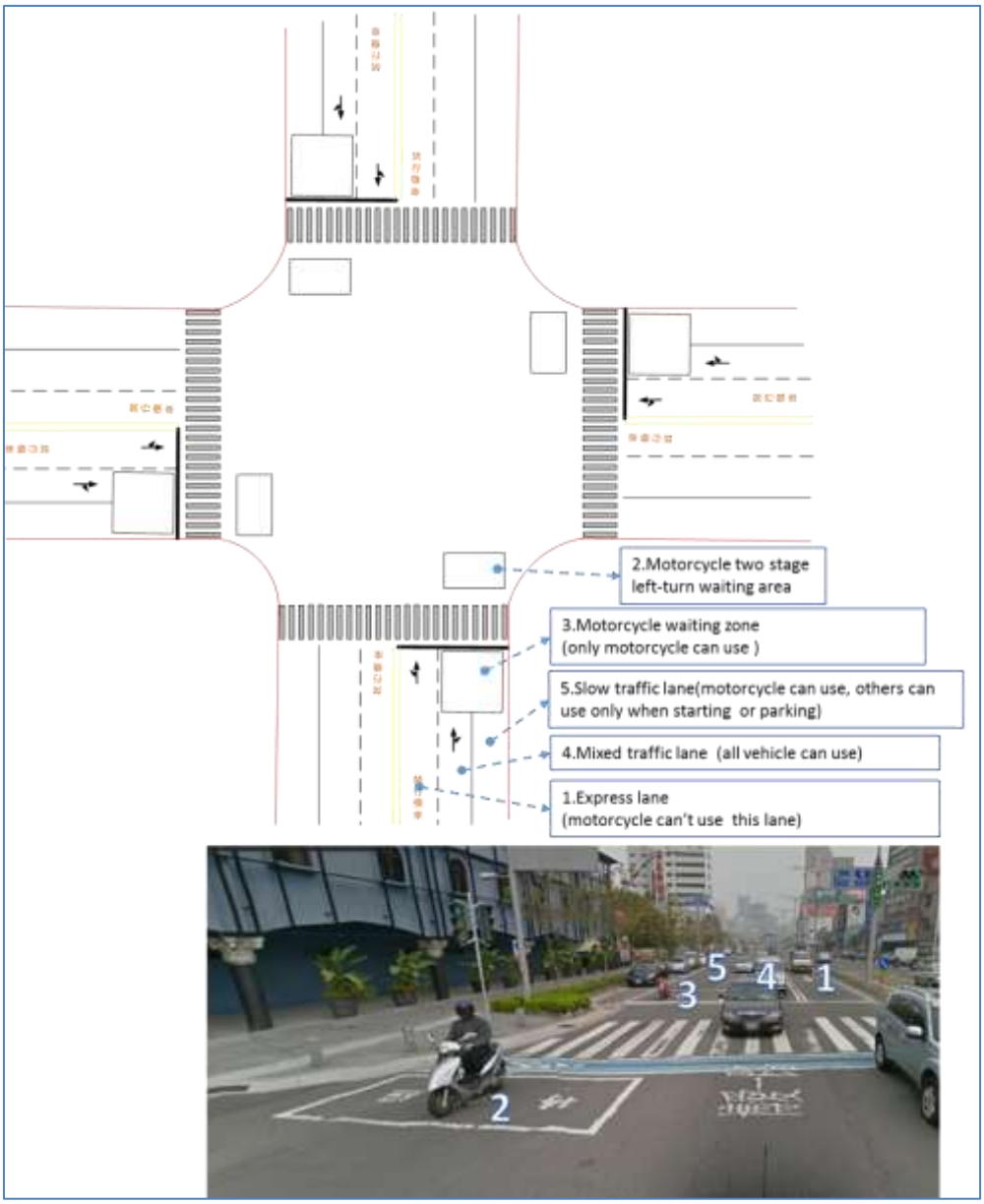

Figure 1 Basic layout of intersection

Under such geometry layout, most of the motorcycles would ride on curb side lane. The typical collision types of motorcycle involved accidents will have following 5 types as illustrated in Fig. 2.

Type 1: Collision between straight motorcycle and opposing left turn car.
Type 2: Collision between straight motorcycle and near approach's motorcycle.
Type 3: Collision between straight motorcycle and right turn car.
Type 4: Collision between straight motorcycle and direct motorcycle.
Type 5: Collision between straight motorcycle and direct vehicle.

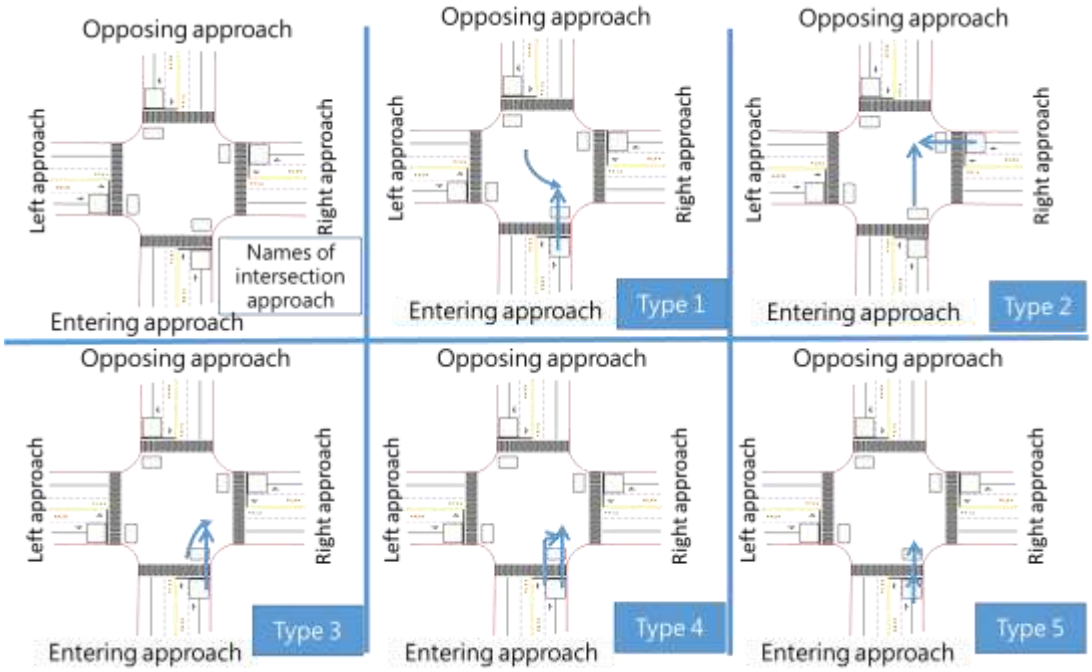

Figure 2 Main motorcycle involved collision type

Counting from 85 accident prone intersections selected from Kaohsiung city in Taiwan the percentage of collision types are illustrated in Table 1. Most accident occurred due to the conflict between motorcycle and car, which has 42.91%. Left turn other angle against opposing collision and right turn other angle against straight Collison are also more than 10%, s. Table 1.

Table 1 Statistical analysis of collision type at intersection

Collision Diagram (Unit:%)	Right angle collision	Through right angle collision	Left turn opposing through collision	Through left angle collision	Rear end collision	Sideswipe collision	Head-on collision	Merge collision	Other	Total
Car-Car	0.85%	0.91%	1.53%	0.82%	12.82%	7.68%	0.12%	1.32%	1.53%	27.58%
Car-Motorcycle	5.24%	12.21%	9.32%	1.76%	3.41%	5.21%	0.29%	1.76%	3.71%	42.91%
Motorcycle-Motorcycle	5.94%	1.09%	1.09%	1.12%	5.18%	4.44%	0.32%	0.53%	0.68%	20.39%
Others	1.18%	0.82%	0.71%	0.12%	2.00%	2.88%	0.12%	0.47%	0.82%	9.12%

Because the causes of the left-turn other angle against opposing collision are difficult to be identified, therefore, in this study, it is chosen as example. First, this type of collision has the high proportion, and this collision type often causes more serious casualties. Second, in spite of the clear path of straight motorcycle, left-turn car has more wide space of path to make left-turn. Because the complex of path, the collision position area is distributed wider, and it's harder to find the accident causes. Third, it often needs to change the geometry layout and road intersection angle to remedy this kind of accidents effectively. The safety enhancement work is often hard to be implemented. Finally, motorcycle frequently ride in the blind area by the left-

turn car. In the past, most researches of accident analysis at intersections discussed the accident factors by modeling predict accident risk or discuss improvement measures accordingly. The model considered usually geometry, environment characteristics, and other current situation at intersection into accident prediction model [**8**]. Most frequently applied model are Poisson regression and negative binomial regression etc. These methods could also be used to compare the accident risk level among different intersections [**9**]. In addition, some other researches also discussed the driver, vehicle, and road environment factors at accident hotspots by fuzzy pattern [**10**]. In recent years, Markov chain Monte Carlo method was invented. It makes that using Bayesian statistical method to quickly seek numerical solutions possible. The application of Bayesian methods in accident analysis is increasing. Currently, Bayesian method is widely used in traffic safety analysis [**11**]. Moreover, using data mining to investigate the accident factor is also becoming popular in recent years [**12**]. Some studies using random forest tree or other decision tree methods to explore important accident factors [**13**, **14**]. Most analysis methods aimed at finding out main accident factors, and then establish the accident risk model. However these researches are insufficient to apply to the collision of high mixed traffic combined with car and motorcycle and taking the real occurred location into account. Therefore, the study applies random forest method to explore important accident factors, and combines with the coordinate analysis at collision moment in order to investigate the correlation between the location of the collision and accident factors. Finally, the study find the potential collision problems and propose the improvement design methods.

3. METHODOLOGY

3.1 Location Spatial Analysis and Risk Analysis

First, for analyzing accident location, it has to establish a two dimensional coordinate system at intersection. The coordinate system could be the foundation to analyze accident causes by applying random forest tree. It gets the coordinate of a rectangle's four points at collision image by the intersections' AutoCAD file which could locate the collision position on the video images. Then using the spatial analysis software to calculate the transformation principle between the coordination of video image and the coordination of real intersection status. The analysis coordination's origin point sets the left side of express lane be $X = 0$ and near approach's curb extending line rearward 50 meters line be $Y = 0$, as shown in Fig. 3.

Figure 3 Coordination and image position analysis

As sample, 27 real left-turn through collision accident videos are collected at 6 approaches, and used to get collision locations. These videos can show the whole picture of accident occurrence process. It is record by the CCTV at intersections. According to the peak hour traffic volume at different approach, the research defines the risk (R) of each collision point.

$$R = \frac{No.\ of\ \ collision}{Direct\ motorcycle\ volume \times Opposing\ left\ turn\ car\ volume} \quad (1)$$

All samples are ordered sequentially according to the risk level and then divided into three groups, as follows:

a. High risk (R3): R3 ≥ 0.31969

b. Medium risk (R2): 0.1 1431≤ R2 < 0.31969

c. Low risk (R1): R1< 0.11431

Besides, the X and Y coordinate of samples are grouped into different sections, s. Table 2.

Table 2 The coordinate of section of X and Y

Y section	Y1	Y2	Y3	Y4	Y5	Y6
Y coordinate range	47≦Y<50	50≦Y<53	53≦Y<56	56≦Y<59	59≦Y<62	Y≥62
No.of sample	1	2	4	4	0	16
X section	**X1**		**X2**		**X3**	
X coordinate range	3≦X<6		6≦X<9		X≥9	
No. of sample	12		5		11	

3.2 Random Forest and Data

In order to find out the causes and influence factors of accident, random forest is adopted to build model. Random forest is a kind of decision tree method which builds the classification rules by learning from training samples. And the main features of random forest contain:

a. Not easy to be overfitting.
b. Excellent anti-noise capability.
c. Discrete and continuous data could be processed.
d. The important ranking of factors would be provided.

The screening factors would be measured by their error of out of bag (OOB). On average, there are 37 % of all samples wouldn't be used at each training. These data isn't selected to build model. Therefore, it can be tested for evaluation of model. The samples of accident are selected from accident hotspots with the same road type. The total accident number of the 85 intersections is 3805 in a year. However, not every intersection exists CCTV. Besides, the video from CCTV system would overwrite after a period of time. Therefore, only some video data could be captured. The layout of intersections selected consist of slow traffic lane, mixed traffic lane, and express lane and have all of motorcycle control facilities, such as motorcycle waiting zone and two-stage left-turn zone. These accident prone intersections have not diferent geometric design. By analyzing these data, this research selects not only the coordinate of collision, but also chooses other factors including all red time, yellow time, slow traffic lane width, intersection angle, parking on roadside, stop line position, road type of near approach, road type of opposing approach.[5-7]. The basic characteristics of intersections are shown in Table3. The variables are listed in Table4 and Table5. Because track of left turn car is related to intersection angle, the research defines three accident factors (X1R, X2R, X3R) to discuss the geometry of intersection, as illustrated in Fig. 4.

Table 3 Characteristics of Intersection

Intersection name	Intersection type	Approach direction	width of slow traffic lane(m)
Bo-ai 1st Rd. & Long De Rd.	4 legged	North	5
Da Shun 2nd Rd. & Jiangong Rd.	4 legged	East	4.3
		North	2.9
		South	5.5
		West	3.3
Guanghua. 2nd/3rd Rd. & Yisin 1st Rd.	4 legged	North	2.8

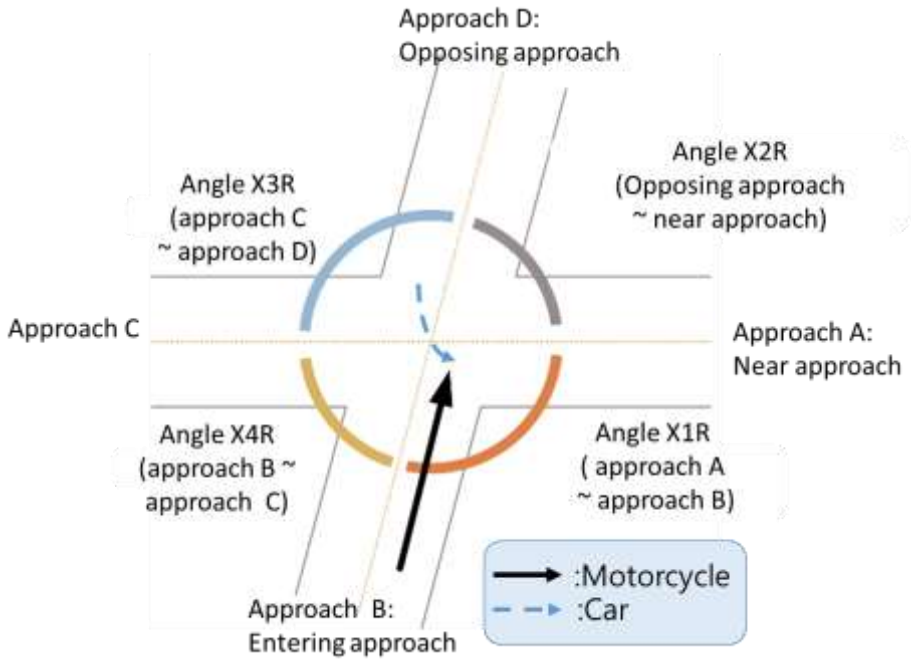

Figure 4 Intersection angles and geometry locations

Table 4 Summary statistics for selected continuous variables

Variable	Label	Mean	Standard deviation	Minimum	Maximum
XlocO	Collision location's X coordinate(m)	7.51	2.63	3.28	11.95
YlocO	Collision location's Y coordinate(m)	61.16	5.82	49.50	70.60
TheroyYSecondDirect	Theory yellow time for the approach of direct motorcycle(sec)	2.44	0.51	2.00	3.00
TheoryRSecondDirect	Theory red time for the approach of direct motorcycle(sec)	5.63	1.04	4.00	7.00
Ydiff	Lack of yellow time for the approach of direct motorcycle(sec)	1.04	0.71	0.00	2.00
Rdiff	Lack of red time for the approach of direct motorcycle(sec)	2.81	0.88	1.00	4.00
OutSideLaneWidth	Slow lane width(m)	3.79	0.98	2.78	5.90
IntergreenDiff	Lack of intergreen time for the approach of direct motorcycle(sec)	1.56	0.80	0.00	2.00

Table 5 Summary for selected dummy variables

Variable	Label	Number of observations for each value(%)						
		0	1	2	3	4	5	6
RiskLevel	Accident risk level	-	9(33.3)	9(33.3)	9(33.3)	-	-	-
Xloc	Location segment of X	-	11(40.7)	5(18.5)	11(40.7)	-	-	-
Yloc	Location segment of Y	-	1(3.7)	2(7.4)	5(18..5)	4(14.8)	-	15(55.6)
ApproachNo	Number of approach at intersection	-	-	-	-	24(88.9)	3(11.1)	-
RightAngle	Right angle intersection(0 if no,1 if yes)	17(63.0)	10(37.0)	-	-	-	-	-
X1R	The angle range between the approach of direct motorcycle and its counterclockwise approach(1:<80,2:80~100,3:>80)	-	7(25.9)	12(44.4)	8(29.6)	-	-	-
X2R	The angle range between the approach of left-turn vehicle and its clockwise approach(1:<80,2:80~100,3:>80)	-	7(25.9)	-	-	-	-	-
X4R	The angle range between the approach of direct motorcycle and its clockwise approach(1:<80,2:80~100,3:>80)	-	8(29.6)	9(33.3)	10(37.0)	-	-	-
NearHurSlowIsland	Express/slow traffic divider on the counterclockwise approch of direct motorcycyle approach(0 if no,1 if yes)	23(85.2)	4(14.8)	-	-	-	-	-
NearCentralIsland	Medium island on the counterclockwise approach of direct motorcycle approach (0 if no, 1if yes)	9(33.3)	18(66.7)	-	-	-	-	-
OppCentralIsland	Left-turn vehicle approach with central island(0 if no, 1 if yes)	18(66.7)	9(33.3)	-	-	-	-	-
NearApproachLaneNo	No.of total lanes of the counterclockwise of the direct motorcycyle approach	-	2(7.4)	2(7.4)	20(74.1)	-	3(11.1)	-
OppApproachLaneNo	No. of total lanes of the left turn car approach	-	-	2(7.4)	25(92.6)	-	-	-
OppLeftTurnGuideLine	Left-turn vehicle approach with left-turn guide line (0 if no,1 if yes)	17(63.0)	10(37.0)	-	-	-	-	-
Parking	Direct motorcycle approach with parking(0 if no, 1 if yes)	15(55.6)	12(44.4)	-	-	-	-	-
IntergreenLacking	The intergreen time of direct motorcycle approach (0 if enough, 1 not enough)	5(18.5)	22(81.5)	-	-	-	-	-
Stoplineover	The stopline of direct motorcycle approach is over 9m from the curb extending line of near approach(0 if no,1 if yes)	3(11.1)	24(88.9)	-	-	-	-	-

4. ANALYSIS RESULTS AND DISCUSSION

4.1 X Section Analysis

The X coordinate of collision mainly depends on the X coordinate of the arriving motorcycle at intersection. This study explores the relation between the X section of collision and other factors by random forest analysis. It provides an overall result that the error rate is 22.22%. From the analysis result, the connection of X section and intersection angle could be found, as illustrated in Fig. 5.

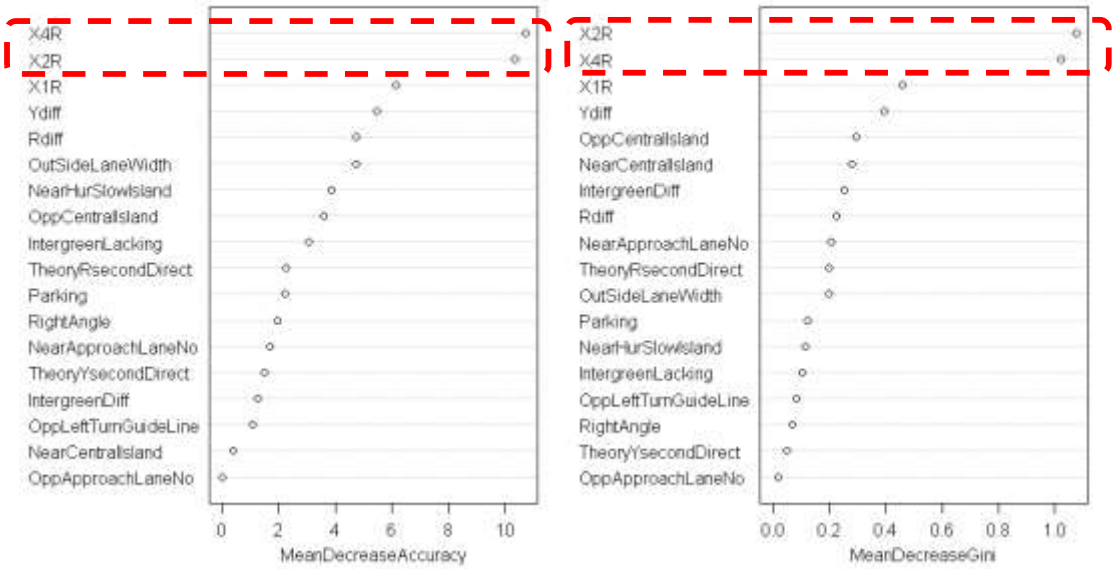

FIGURE 5 Relative importance of factors about X section

Mean decrease Gini is more stable than Mean decrease accuracy, and most of samples' X2R is equal to X4R. Therefore, the research uses the most important factor X2R for further analysis. According to different angle classification of X2R, the collision position for X coordinate could be analyzed, as shown in Table 6.

Table 6 Statistical analysis of intersection and X coordinate

	No.	Range	Min.	Max.	Average	Standard error	skewness		kurtosis	
X2R=1	8	5.36	3.73	9.09	6.51	2.15	-0.18	0.75	-1.73	1.48
X2R=2	9	8.67	3.28	11.95	7.64	3.24	-0.06	0.72	-1.65	1.40
X2R=3	10	6.60	5.35	11.95	8.20	2.38	0.01	0.69	-1.52	1.33

About the result of statistics, the research further discusses as below:

1. Range

Statistics show that the range of "X2R=1" and "X2R=3" are relatively small than "X2R=2". The reason should be related to oblique intersection. At skew intersection, sight distance of left-turn car would be blocked, some straight through motorcycle cannot be seen by left-turn car. Therefore, the range is different between skew intersection and orthogonal intersection.

2. Minimum

The accident location on mixed traffic lane is defined as the minimum X coordinate minus 3 meter which is the width of express lane. Statistics show that the minimum of X coordinate is 5.35m when the intersections angle are "X2R=3". In other words, the section of 0～2.25m left side of mixed traffic lane of is relatively safe. The width is wider than the intersections angle are "X2R=1" or "X2R=2". The reason should be that the curvature of left-turn car is smaller at angle "X2R=3". Therefore, it's easier for left-turn car to observe the opposing straight motorcycle on mixed traffic lane.

3. Maximum

The statistics show that the maximum X coordinate of angle "X2R=1" is smaller than "X2R=2" and "X2R=3". The reason should be that the curvature of left-turn car is larger at angle "X2R=1". Therefore, straight through motorcycle on slow traffic lane, i.e., curb lane, is easier to be observed by the left-turn car than the motorcycle on mixed traffic lane.

4. Average

The statistics show that the larger of angle X2R has also the larger of the average collision X coordinate. The reason is that the curvature of left-turn car decreases while the angle X2R increases. Besides, the overall of minimum and maximum X coordinate will increase. Therefore, the overall of X coordinate is larger when the angle X2R increases.

5. Skewness

The statistics show that X coordinate distribution of "X2R=1" is left skewness. The reason should be that the curvature of left-turn car is larger at angle "X2R=1". For motorcycle on the mixed traffic lane, the response time of angle "X2R=1" is shorter than angle of "X2R=2" or "X2R=3" if conflict happened.

4.2 Y Section Analysis

The Y coordinate of collision mainly depends on the path of left-turn car. And the path of left-turn car is related to geometry, lane choice of left-turn car…etc. This study explores the relation between the Y section of collision and other factors by random forest analysis. It provides an overall result that the error rate is 22.22%. From the analysis result, it is found that the yellow time, and width of slow lane are most relevant to the Y section of collision position, as shown in Fig. 6.

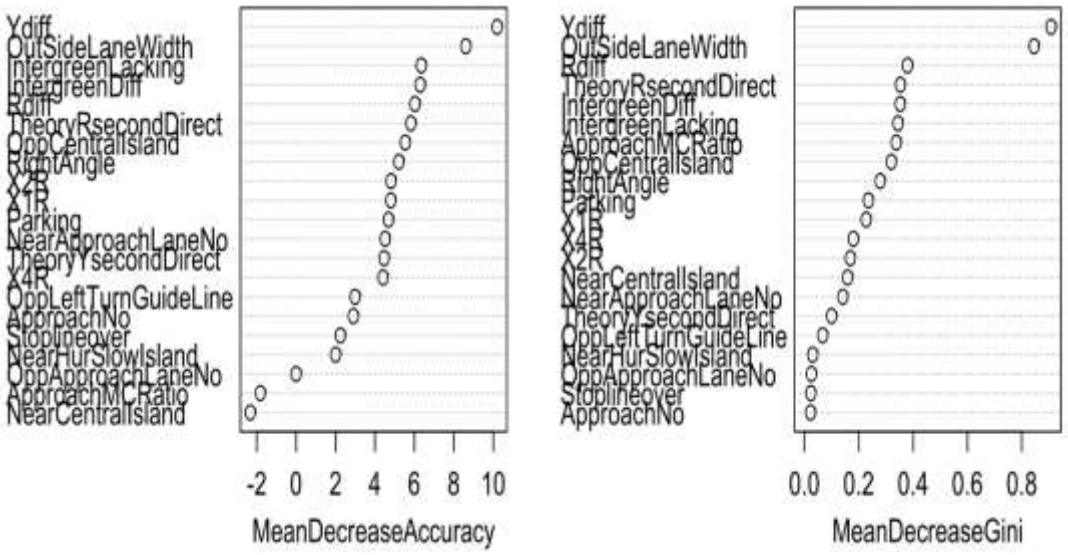

Figure 6 Relative importance of factors about Y section

About the yellow time, we recalculated a new yellow time, defined as theoretical yellow time and to compute the difference between theoretical yellow time and actual yellow time and take it as variable. It means the inadequate yellow time. When motorcycle couldn't stop before stop line during yellow time, motorcycle only can use all red time. However, in order to go through the intersection during inter-green time, opposing left-turn cars which wait at intersection must complete its left-turn as soon as possible during all red time. Therefore, left-turn car and motorcycle are very easy to ignore the conflict between both sides. Besides, the inadequate yellow time would also affect the start-turning time of left-turn car. It would further reflect in the Y section of collision position. Therefore, the research analyzes the correlation of Y section and inter-green difference by Kendall's tau_b analysis, as listed in Table 6. The results show that the correlation between Y section and inter-green time difference is intermediate. And its significance is up to 0.003.

Table 6 Correlation analysis of Y section and Inter-green time difference

			Y section	Intergreen Dif.
Kendall's tau_b	**Y section**	**correlation coefficient**	1.000	0.486**
		significance (one tail)		0.003
		Sample No.	27	27
	IntergreenDif	**correlation coefficient**	0.486**	1.000
		significance (one tail)	0.003	
		Sample No.	27	27

**. Significance level= 0.01 (one tail) ◦

About the width of curb slow traffic lane, supposing the total number of motorcycle on slow traffic lane wouldn't change with the width of slow traffic lane, and motorcycles are homogeneous distribution on the width of slow traffic lane (X_2-X_1). Then it is found that there are about 50% of motorcycle would ride into the blind area of car at $Y=Y_1$ as shown in Fig. 7. However, the ratio of motorcycle at blind area would increase with the increasing width of slow traffic lane. In other words, as X_2 increases, more motorcycles would be at blind area. Finally, motorcycles at different X coordinate use different time to pass through the blind area. Therefore, motorcycles and left-turn cars collide at different Y sections.

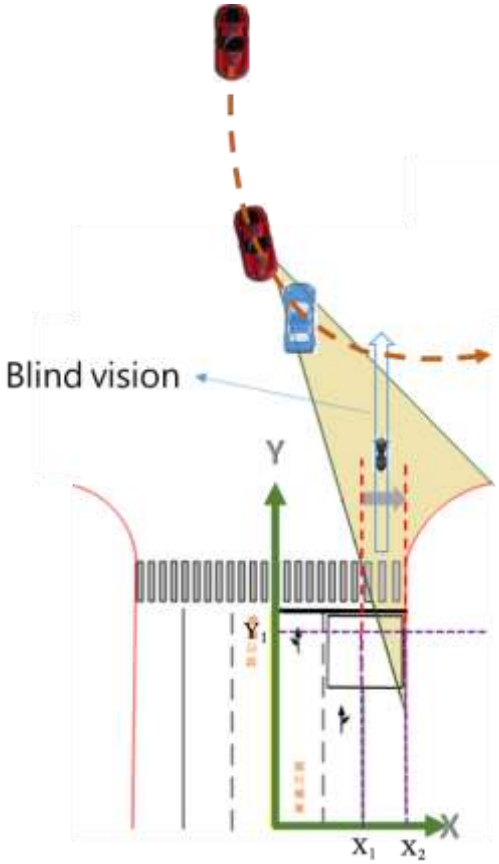

Figure 7 Left-turn car path and blind vision area

5. CONCLUSION AND SUGGESTION

This study proposes the analysis method combined collision location position and random forest. Besides, this study also discusses the accident contribution factors of left-turn through

collision in the view point of accident risk. The X and Y coordinate distribution of collision location are found. After that, random forest method is used to explore the relation of collision positions and accident factors. This study's major findings are as follows:

1. Generally, the range of collision position could be predicted by traffic conflict technique. This study uses accident video image position method to get the coordinate of collision. The method could locate the actual location of collision and use to identify collision position more clearly.
2. Most researches in the past discussed about the influence level of accident factors on accident risk. This study discusses the relation between the X and Y coordinate of collision position and accident factors. Besides, the important accident influence factors are further classified to compare with each other and then to discuss the correlation and its connection of traffic flow features. Not only the importance of accident influence factors are considered, but also the relation between accident factors and collision position is discussed in this analysis process.
3. This study finds that the angle of intersection has the high level of influence on the risk of left-turn through collision. The statistical analysis shows that straight through motorcycle on different X coordinate exists different risk to be involved into left-turn opposing through collision. Therefore, straight through motorcycle should be guided to slow traffic lane when the angle X2R is less than 80 degrees at accident hotspots, On the other hand, straight through motorcycle should be guided to mixed traffic lane when the angle X2R is larger than 100 degrees.
4. Most accident samples from accident hotspots of this study are at signal intersection with simple two phases signal timing. For left-turn through collision, yellow time should be reviewed in order to avoid the inadequate yellow time which would result in motorcycle couldn't stop before stop line and has to pass through the intersection during all red time. Besides, the width of curb slow traffic lane should not be too wide, otherwise the probability of motorcycle at blind area would be too high. The research suggests that the width of slow traffic lane should not be more than 3.5 meters in principle. Then, the systematical safety improvement design countermeasures could be built more effective.

REFERENCES

1. Stastics Inquiry.[cited 2016 June. 20]; Available from: http://stat.motc.gov.tw/mocdb/stmain.jsp?sys=100.
2. National Stastics. [cited 2016 Feb. 20]; Available from: http://statdb.dgbas.gov.tw/pxweb/Dialog/varval.asp?ma=PS0203A1A&ti=%BE%F7%B0%CA%A8%AE%BD%F8%A4%CE%B9D%B8%F4%A5%E6%B3q%A8%C6%ACG-%A6~&path=../PXfile/PublicSafety/&lang=9&strList=L.
3. Lin;, W.-M., et al., The 33th Project for Improving Accident-Prone Locations in the Taiwan Area, N.O.T.M.O.T.A. COMMUNICATIONS, Editor 2016.
4.2012 Kaohsiung City Accident Hotspots Improvement Research, 2012, Transportation Bureau of Kaohsiung City Government.
5. Hsu, T.-P., 2013 Kaohsiung City Accident Hotspots Improvement Research, 2013, Transportation Bureau of Kaohsiung City Government.
6. Hsu, T.-P., 2014 Kaohsiung City Accident Hotspots Improvement Research, 2014, Transportation Bureau of Kaohsiung City Government.
7. Hsu, T.-P., 2015 Kaohsiung City Accident Hotspots Improvement Research, 2015, Transportation Bureau of Kaohsiung City Government.
8. Hsu, T.-P., Segregated motorcycle traffic flow countermeasures at intersections, in 12th REAAA Conference2006: Manila, Philippines.
9. Sadeghi, A., E. Ayati, and M. Pirayesh Neghab, Identification and prioritization of

hazardous road locations by segmentation and data envelopment analysis approach. PROMET-Traffic&Transportation, 2013. 25(2): p. 127-136.
10. Vistisen, D., P. Thyregod, and J.G. Laursen, Models and methods for hot spot safety work, 2002, Technical University of DenmarkDanmarks Tekniske Universitet, Department of PlanningInstitut for Planlægning.
11. Mayora, J.M.P. and R.L. Rubio. Relevant Variables for Crash Rate Prediction in Spains Two Lane Rural Roads. in Transportation Research Board 82nd Annual Meeting. 2003.
12. Sayed, T., W. Abdelwahab, and F. Navin, Identifying accident-prone locations using fuzzy pattern recognition. Journal of Transportation Engineering, 1995. 121(4): p. 352-358.
13. Jiang, X., et al., Investigating macro-level hotzone identification and variable importance using big data: A random forest models approach. Neurocomputing, 2016. 181: p. 53-63.
14. Lin, L., Q. Wang, and A.W. Sadek, A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. Transportation Research Part C: Emerging Technologies, 2015. 55: p. 444-459.
15. Harb, R., et al., Exploring precrash maneuvers using classification trees and random forests. Accident Analysis & Prevention, 2009. 41(1): p. 98-107.