

Heuristic Function Influence to the Global Optimum Value in Shortest Path Problem

Andysah Putera Utama Siahaan

Faculty of Computer Science, Universitas Pembangunan Panca Budi

Jl. Jend. Gatot Subroto Km. 4,5 Sei Sikambing, 20122, Medan, Sumatera Utara, Indonesia

Abstract: *Determination of the optimum route is often encountered in daily life. The purpose of the optimum route itself is to find the best trajectory of the two pairs of vertices contained in a map or graph. The search algorithm applied is A*. This algorithm has the evaluation function to assist the search. The function is called heuristic. Two methods which have been introduced as a step to obtain the value of heuristic function are by using Euclidean and Manhattan distance. Both of these methods create the optimum distance in shortest path problem, but these functions gain the different results. This research performs the development of the heuristic function using Euclidean, Manhattan, Euclidean Square and the author method (Andysah).*

Keywords: *Shortest Path, Artificial Intelligence, A**

I. Introduction

The optimum route is always required to travel in achieving a goal [7], but it needs a way how it is reached. To determine the optimum route, there are advantages and disadvantages associated with the method used where each method produces different output. The gap will affect the outcome of a process of finding the optimum route. In the A* algorithm, heuristic function is a value that affects the optimum value; the route length depends on the heuristic functions given [1][2].

In the previous research, the search for the optimum path can be done by two methods: conventional methods and heuristic methods. The conventional method is a method that uses a simple calculation used to resolve a few vertex alone. As for vertex more, more varied methods of heuristics for heuristic method directional approach and conduct searches. Some heuristic methods that have been used to get the value of the heuristic function are Euclidean and Manhattan distance. Euclidean will provide the results to the heuristic function by measuring the direct distance between two vertices. It is done by drawing a straight line without a hitch and then determine the distance between two vertices, while Manhattan, value heuristic function is obtained by mapping the vertex to a coordinate plane two dimensions and then summing the difference in the second coordinate of the vertex.

From some heuristic methods to calculate the optimum distance of a path, then there is a comparison and calculations to determine which method is the best that works for the determination of the optimum route. The author will discuss the methods used in the heuristic function to get the global optimum value. The author will try to create a method to expect this method to have a better result than the method that has ever existed. The method used is to determine the initial vertex to vertex distance and the purpose of the vertex after the initial vertex to vertex of interest. This method will get the new distance measurement models that can be applied to the heuristic function.

II. Related Work

In a heuristic search method, used a method used to evaluate the circumstances of individual problems and determine how far they could be used to obtain the desired solution [4]. The heuristic function is different for each destination, so this function is often used as an important parameter in resolving problems. The heuristic function is implemented to find the distance between two vertices, which typically uses Euclidean and Manhattan distance. In determining the value of the heuristic function will involve the coordinates x and y , where x and y are the meeting point of the two-dimensional plane. Determining the value of x and y are assumed to be taken of latitude and longitude which has been converted into two-dimensional coordinates.

Optimum Search Using Genetic Algorithm

This study describes the application of genetic algorithms to find the most optimum from the point of origin to the point of destination. Genetic algorithm system that has been designed using chromosome representation in the form of a bit string. Therefore this kind of mutation that is used is a bit mutation. The system also uses several methods of selection, namely roulette wheel, elitism and a combined method of the roulette wheel and elitism. There are two types of crossover used, one cut point and two cut point crossover [6].

Optimum Route Finders On Map to Increase Efficiency Travel Time Road Users By Method A* and BFS

This study develops an application for the determination of the optimum route using a digital map where in this application there is congestion factor, road condition, and mileage. To search routes used method A* and Best First Search, which uses a heuristic function to 'direct' the search on the map which is represented in the concept graph [5].

Shortest Variations Determination Function Using Heuristic Algorithm A* In Mobile Devices

This study also aims to build an application on mobile devices, especially cell phones that can provide the shortest route from a starting location to a destination location. It also compares the execution time of the algorithm and the number of nodes are examined for each heuristic function in the search for the shortest route by using algorithms A*.

III. Proposed Work

A. Graph Design

The graph that will be used in this study was obtained by taking a cut from the fields of the digital map obtained with the help of Google Maps rather take the city of Medan Baru subdistrict of Medan. The streets on the digital map will be formed into a graph by hand by way of drawing a straight line on its lines (remapping).

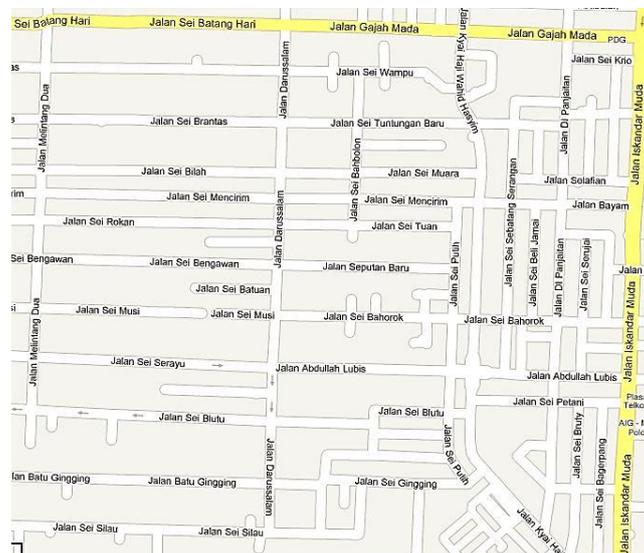


Fig. 1 Pieces of the field map obtained from Google Maps

Once a map is obtained as shown in Figure 1, then it can be that each line has a junction and will be placed vertex. From the point of intersection of these lines will be established weighted graph. The goal is that defined distances between vertices. This distance will then be processed as a determinant of the direction of the road users who want to move from the starting point to the end point. Branching on the map is so complex that there are many routes to be passed from the destination point to the end point. After the mapping phase is completed, then the vertex will be formed at each branch and at the end of a dead end road. Then the vertex indexed for easy programming.

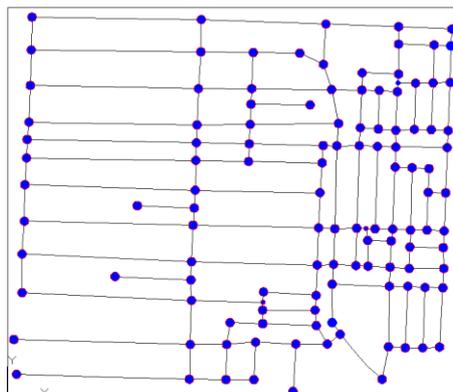


Fig. 2 A Graph that has been formed

Figure 2 is a map that has been converted into a graph, but the graph has yet to do the best distance calculation of specific points. The next step is to give weight to contain the shortest distance (straight line) from the two points and the names on the vertex. The number of vertices in the graph reaches 120 pieces so it is a fairly complex graph for examination. It must be emphasized, rather than relying on the truth of little or number of items involved, but depending on the programming algorithms.

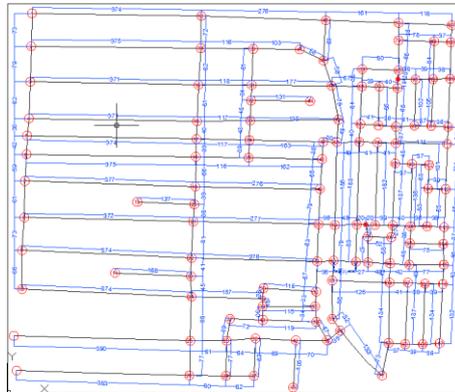


Fig. 3Final result of the mapping graph

Figure 3 is already a graph, and more precisely as a map as well as have a real comparison on each of the vertices. However, to do a search by using the A* algorithm takes a heuristic function that supports the movement of the election.

B. Heuristic Function

The heuristic function will determine the cost to get to the goal by using the Euclidean distance, Manhattan, Euclidean Square and a new method. In the discussion, will the tests on methods of heuristic functions on some models of cases, which method is superior to other methods, and how it functions heuristic influence the outcome of the global optimum.

Euclidean Distance and Euclidean Squares

Heuristic value for h (n) is obtained by the calculation using the Euclidean distance is the distance that determines the direct relationship freeway from two points while Euclidean Squared takes the value of the sum of squares difference of x and y coordinates.

Known:

- Vertex coordinate 12: 287, 424
- Vertex coordinate 44: 669, 648

Asked:

- Euclidean Distance
- Euclidean Squares

Answer:

- x1 : 287
- y1 : 424
- x2 : 669
- y2 : 648

- ED2 : $(x_2 - x_1)^2 + (y_2 - y_1)^2$
- : $(669 - 287)^2 + (648 - 424)^2$
- : 3822 + 2242
- : 145.924 + 50.176
- : **196.100**

- ED : $\sqrt{196.100}$
- : 442,8318
- : **443**

Manhattan Distance

The value of $h(n)$ will be obtained from the calculation of the difference of the coordinates of the vertices start, and end vertices are represented in the form of 2D coordinates.

Known:

- Vertex coordinate 12: 287, 424
- Vertex coordinate 44: 669, 648

Asked:

- Manhattan Distance

Answer:

- x_1 : 287
- y_1 : 424
- x_2 : 669
- y_2 : 648

- MD : $(x_2 - x_1) + (y_2 - y_1)$
- : $(669 - 287) + (648 - 424)$
- : $382 + 224$
- : **606**

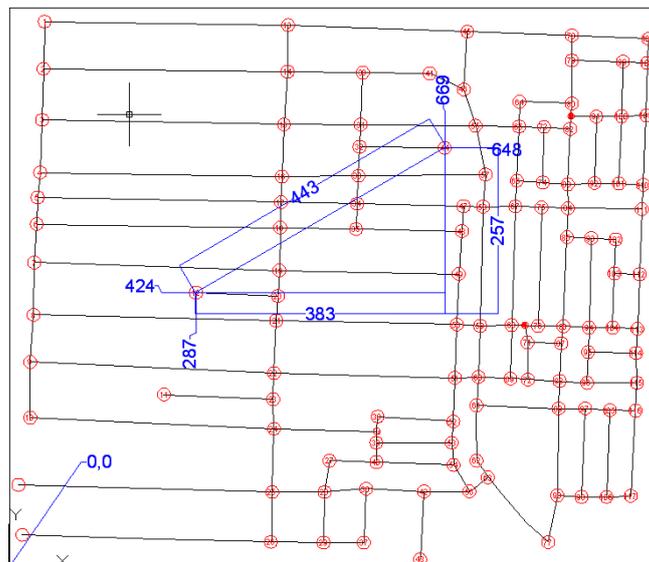


Fig. 4 Euclidean and Manhattan calculation

The result of the calculation of Euclidean and Manhattan earlier showed the same results as in figure 3.5. The position of the origin point is at the lower left of the image. The values of x_1 , y_1 , x_2 and y_2 taken by the shift of the position of the origin point. To determine the point coordinates, use the help of AutoCAD software that was used to draw a map or maps. Moreover, having done the math, Euclidean calculation results have a smaller value compared with the results obtained from the calculation of Manhattan for calculating Euclidean diagonal from the difference between x and y coordinates.

C. Andysah Method

The method used in this study is Euclidean, Manhattan, and Euclidean Squares. The third method of generating optimum values vary. As a complement to the study, the authors created a new method that will be used as a comparison to the three methods that have been done, and it is hoped this method has better results than previous methods. The new method is called Andysah where the calculation result of this formula will be given to the heuristic function. The purpose of this method is to direct the movement towards the destination point by selecting the vertex is better under the terms are given to this method. Selection of the best vertex raised by way of calculating three-point coordinates, i.e., the coordinates of origin, and the next destination.

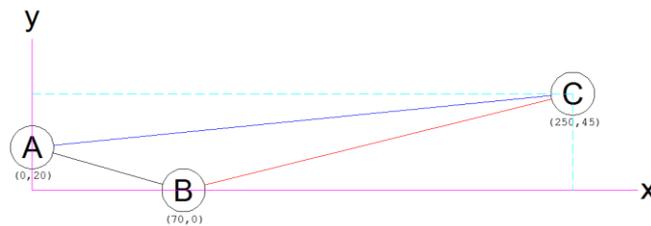


Fig. 5 Andysah method

Figure 5 shows how to perform calculations on Andysah method. To determine the functionality of the heuristic A to C obtained through an intermediary B is $H = BC + AC$, while AB is not used. To find a long-BC and AC can use Euclidean Distance (equation 2.3). Flowchart can be found in Appendix E. It is formed with the formulation as follows:

$$H_{a-b-c} = BC + AC$$

$$H_{a-b-c} = \sqrt{(xc - xb)^2 + (yc - yb)^2} + \sqrt{(xc - xa)^2 + (yc - ya)^2}$$

Keterangan:

- H_{a-b-c} : total cost a to c via b
- x_a : coordinate x on a
- x_b : coordinate x on b
- y_a : coordinate y on a
- y_b : coordinate y on b

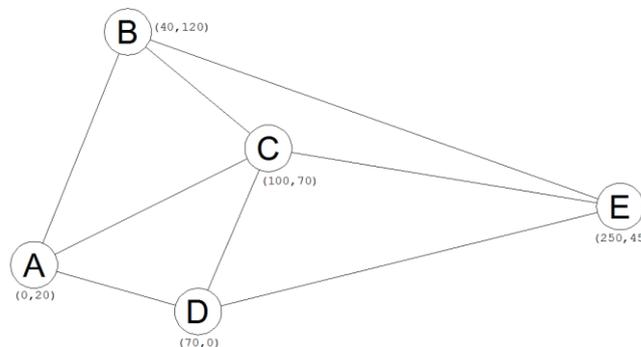


Fig. 6 Sample of Andysah method

In Figure 6, it is known that the coordinates of the point A (0, 20), B (40, 120), C (100, 70), D (70, 0) and E (250, 45). The purpose of this method to get a path from vertex A to vertex E. The position of the current search is in vertex A, while towards the vertex E 3 pieces (B, C, D) track selection. The method given in this heuristic function will calculate the smallest cost in the third vertex selection. The value of the smallest cost will be elected as the track be passed after the vertex A.

The calculation:

Coordinate:

A(0, 20), B(40, 120), C(100, 70), D(70, 0) and E(250, 45)

Vertex A => B

$$BE = \sqrt{(250-40)^2 - (45-120)^2}$$

$$= 222,991$$

$$AE = \sqrt{(250-0)^2 - (45-20)^2}$$

$$= 251,2469$$

$$T = BE + AE$$

$$= 222,991 + 251,2469$$

$$= \mathbf{474,2379}$$

Vertex A => C

$$\begin{aligned} CE &= \sqrt{(250-100)^2 - (45-70)^2} \\ &= 154,0292 \\ AE &= \sqrt{(250-0)^2 - (45-20)^2} \\ &= 251,2469 \\ T &= CE + AE \\ &= 154,0292 + 251,2469 \\ &= \mathbf{405,2761} \end{aligned}$$

Vertex A => D

$$\begin{aligned} DE &= \sqrt{(250-70)^2 - (45-0)^2} \\ &= 185,5398 \\ AE &= \sqrt{(250-0)^2 - (45-20)^2} \\ &= 251,2469 \\ T &= DE + AE \\ &= 185,5398 + 251,2469 \\ &= \mathbf{436,7867} \end{aligned}$$

If compared to the value obtained between AB, AC, and AD, the smallest value that is owned by AC is 405.2761 after going through vertex A, the next track will pass through the vertex C.

IV. Evaluation

A. Observation

The following test will try taking a sample calculation of the graph. Here will be displayed 25 pieces scattered vertex at particular coordinates. By using the A* algorithm, the process looks for the best path from vertex "G" to "S". There is some way from the vertex G towards S. The heuristic function takes a role in choosing the best route. The following calculation tests the heuristic functions by using the same graph model. Here are the coordinates for the vertex position information that is spread in the form of 2D coordinates and the number of branching vertices.

```
Graph = { 'A': [100, 100, ['B', 'F']],
          'B': [100, 250, ['A', 'C', 'G']],
          'C': [100, 400, ['B', 'H', 'D']],
          'D': [100, 550, ['T', 'C', 'E']],
          'E': [100, 700, ['D', 'J']],
          'F': [300, 100, ['A', 'G', 'K']],
          'G': [300, 250, ['B', 'H', 'F', 'L']],
          'H': [300, 400, ['C', 'G', 'I', 'M']],
          'I': [300, 550, ['D', 'H', 'N', 'J']],
          'J': [300, 700, ['E', 'I', 'O']],
          'K': [500, 100, ['F', 'L', 'P']],
          'L': [500, 250, ['G', 'K', 'Q', 'M']],
          'M': [500, 400, ['H', 'L', 'R', 'N']],
          'N': [500, 550, ['I', 'M', 'S', 'O']],
          'O': [500, 700, ['J', 'N', 'T']],
          'P': [700, 100, ['K', 'Q', 'U']],
          'Q': [700, 250, ['L', 'P', 'V', 'R']],
          'R': [700, 400, ['M', 'Q', 'W', 'S']],
          'S': [700, 550, ['N', 'R', 'X', 'T']],
          'T': [700, 700, ['O', 'S', 'Y']],
          'U': [900, 100, ['P', 'V']],
          'V': [900, 250, ['U', 'Q', 'W']],
          'W': [900, 400, ['V', 'R', 'X']],
          'X': [900, 550, ['W', 'S', 'Y']],
          'Y': [900, 700, ['X', 'T']] }
```

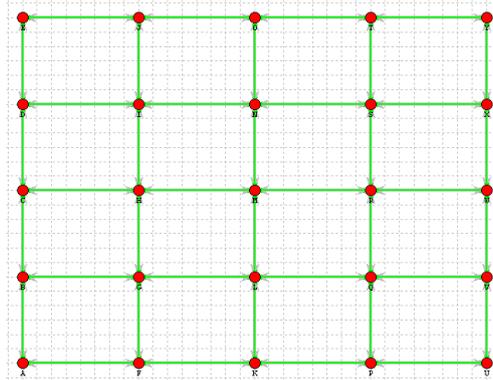


Fig. 7 Graph model

Figure 7 describes the results forming the vertex position by using data on the graph above information. The formation of this graph deliberately made irregular by the number five columns and five rows to see the path that will be skipped program using four methods described previously. Can be seen the problems that occur is how to get the optimum out of the vertex G to the vertex S. In eyesight, the possibility of the path there are five, namely G-L-Q-R-S, G-L-M-R-S, G-H-M-R-S, G-H-M-N-S, and G-H-I-N-S. The fifth possibility is to have the same total cost is 700. The laying of point coordinates with regularly aims to obtain a search pattern (searching pattern) of the four methods used.

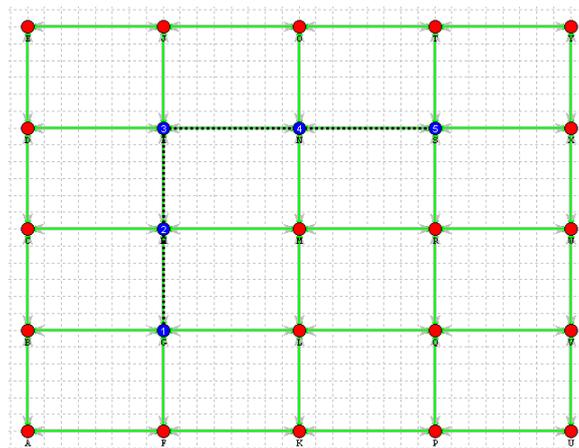


Fig. 8 Euclidean result

The result shown in Figure 8 states that by using the Euclidean distance in the method of use of heuristic functions, the A* algorithm will approach the first vertically and horizontally eventually approach after the vertex vertically movement has traversed one by one. The distance G to H closer than the distance G to L and while the cost $(GH + HS) < (GL + LS)$, thus making G rather than L H to move to the next vertex. Moreover, when the process is in the first position, will move horizontally in the direction N due to the costs involved $(IN + NS) < (IJ + JS)$. At the moment the process is in the S search will stop because the vertex of interest has been found.

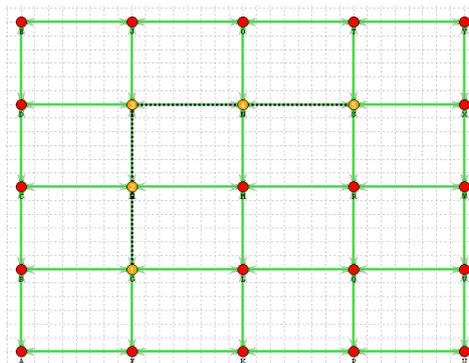


Fig. 9 Manhattan result

The result shown in Figure 9 stated that by using the method of use of Manhattan distance heuristic function, the A * algorithm have the same result by using the Euclidean distance journeys will be made vertically advance and eventually approached horizontally as vertically vertex movement has traversed one by one. Results from Manhattan distance calculation produces a value not very different from the Euclidean distance. Can be seen $(GH + HS) < (GL + LS)$, so that G prefers H instead of L to move to the next vertex. Moreover, when the process is in the first position, will move horizontally in the direction N due to the costs involved $(IN + NS) < (IJ + JS)$. At the moment the process is in the S search will stop because the vertex of interest has been found.

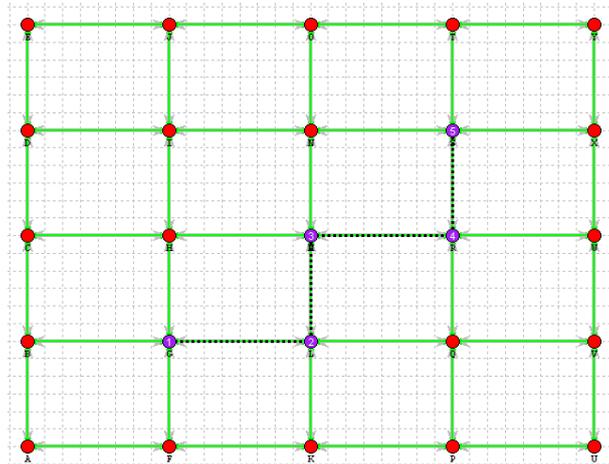


Fig. 10 Euclidean Square result

The results shown in Figure 10 has a clear difference compared with the results obtained in figure 4.4 and 4.5. The use of the Euclidean distance squared has a different approach than the previous method. The trip will be made in a zig-zag movement starting horizontal and then vertical movement. The results of the calculation of the Euclidean distance squared calculation produces much different from the Euclidean distance and Manhattan. Can be seen $(GH + HS) > (GL + LS)$, so that G prefers L than H to move to the next vertex. When the process is at stake L, the thing done is the opposite, namely $(MN + NS) > (MR + RS)$, so I would prefer R than N. And finally, when the process is in the S search will stop because the vertex of interest have been found.

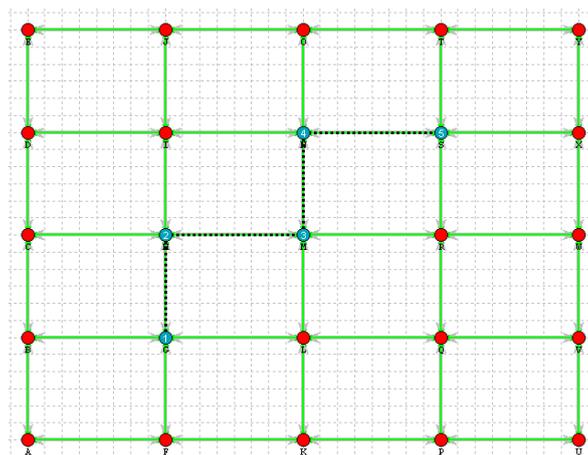


Fig. 11Andysah result

The results shown in Figure 11 is inversely proportional to the results obtained in Figure 4.6. The use of this method is not much different from the Euclidean distance squared. The trip will be made in a zig-zag starting the vertical movement and then the horizontal movement. The results of calculations using this method produce a different calculation so as to have different results. Can be seen $(GH + HS) < (GL + LS)$, so that G prefers H instead of L to move to the next vertex. When the process is at stake H, the thing done is the opposite, namely $(HM + MS) < (HI + IS)$, so that H would prefer M than I. And finally, when the process is in the S search will stop because the vertex of interest have been found.

B. Andysah Complexity

To determine the efficiency of the Andysah method on A* algorithm, it must perform the calculation and analysis of algorithms. The complexity of the algorithm A* depends on the method applied to the heuristic function. In the worst case, the number of nodes that can be developed exponentially in the solution, but its complexity can be polynomial if the heuristic function meets the following conditions:

$$|h(x) - h^*(x)| = O(\log h^*(x))$$

Where h^* is heuristic an optimal or the cost of which is certain to get to the destination point. With the use of heuristic function $H(n)$, the A* algorithm can focus the search on the vertex closer to the purpose of the search. Then the search will be dismissed if the vertex of interest is encountered. This method can minimize the number of vertices to be examined and the time it takes to get on line directly proportional to the number of vertices is checked, then the search time can be optimized.

C. Circumstances Worth

Circumstances Worth going if it is found if there are two or more nodes examined has a price $f(n)$ the same. It often happens because $f(n)$ depend on two functions, namely the function $g(n)$ and $h(n)$. It is very likely to occur between nodes located far apart, and that the most likely node located near the goal while the other node is located away from the destination node.

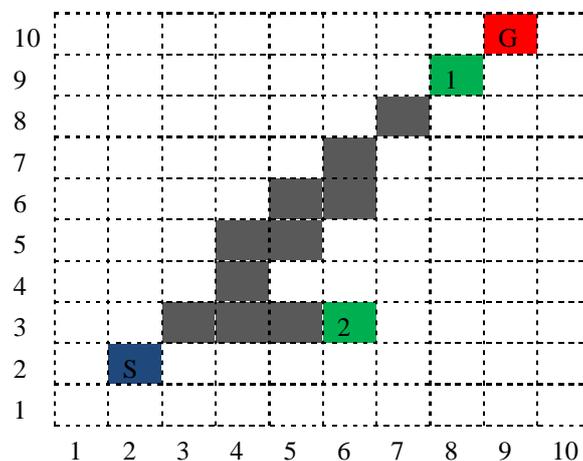


Fig. 12Circumstances Worth

In Figure 12 it can be seen that S is the vertex of the initial (start), G is a vertex end (goal), "1" and "2" is the vertex position that will be compared to the value heuristic functions. From the evaluation function $f(n) = g(n) + h(n)$:

$$\begin{aligned} \text{vertex "1"} : f(n) &= (8 - 2) + (9 - 2) + (9 - 8) + (10 - 9) \\ &= 6 + 7 + 1 + 1 \\ &= 15 \\ \text{vertex "2"} : f(n) &= (6 - 2) + (3 - 2) + (9 - 6) + (10 - 3) \\ &= 4 + 1 + 3 + 7 \\ &= 15 \end{aligned}$$

Because the A* algorithm gives priority based on the price of $f(n)$, then if the state is worth going on, there is more than one node with the same priority. The result is that these nodes will be checked first, which is probably the nodes located far apart the destination node. It causes the performance of the A* algorithm decreased.

V. Conclusion

Based on the discussion and evaluation of the previous chapters and the existing theory, it can be concluded that the function heuristic affects the global optimum results on the search path because different methods were applied to the heuristic function, will produce global optimum results are different, too. The overall method of Euclidean, Manhattan, and Andysah be the best method to use on a heuristic function for optimum yield more lanes than the Euclidean Squares. The Euclidean method generates the shortest route.

While Manhattan is rated second. The results produced by the Manhattan have differences that are not too far away, even in certain cases the result is the same as the Euclidean. Regarding running time, Andysah method has better results than the Euclidean and squared Manhattan. Euclidean generate good execution speed by the way just raised slightly vertices in the search process, but these produced no better than Euclidean, Manhattan, and Andysah.

References

- [1]. L. Biacino and G. Gerla, "Fuzzy Logic, Continuity and Effectiveness," *Mathematical Logic*, vol. 41, p. 643–667, 2002.
- [2]. G. Gerla, "Effectiveness and Multivalued Logics," *The Journal of Symbolic Logic*, vol. 71, pp. 137-162, 2006.
- [3]. H. J. Cho and C. L. Lan, "Hybrid Shortest Path Algorithm For Vehicle Navigation," *Journals Springer Science*, vol. 10, no. 49, pp. 234-247, 2008.
- [4]. F. Lam and A. Newman, "Travelling Salesman Path Problems," *Journal Math Program*, vol. 10, no. 113, pp. 39-59, 2005.
- [5]. R. Adipranata, "Aplikasi Pencari Rute Optimum Pada Peta Guna Meningkatkan Efisiensi Waktu Tempuh Pengguna Jalan Dengan Metode A* dan Best First Search," *Jurnal Teknik Informatika*, vol. 8, no. 2, pp. 100-108, 2007.
- [6]. A. T. Hannawati and Eleazar, "Pencarian Rute Optimum Menggunakan Algoritma Genetika," *Jurnal Teknik Elektro*, vol. 2, no. 2, pp. 78-83, 2002.
- [7]. A. P. U. Siahaan, "Adjustable Knapsack in Travelling Salesman Problem Using Genetic Process," *International Journal of Science & Technoledge*, vol. 4, no. 9, pp. 46-55, 2016.