

SKRIPSI

**CLUSTERING DATA LINTASAN BERDASARKAN
KARAKTERISTIKNYA**



Wilson Nathanael

NPM: 6181901053

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2023**

UNDERGRADUATE THESIS

**CLUSTERING FOR TRAJECTORY BASED ON ITS
CHARACTERISTIC**



Wilson Nathanael

NPM: 6181901053

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2023**

LEMBAR PENGESAHAN

CLUSTERING DATA LINTASAN BERDASARKAN KARAKTERISTIKNYA

Wilson Nathanael

NPM: 6181901053

Bandung, 11 Juli 2023

Menyetujui,

Pembimbing

Digitally signed
by Lionov

Lionov, Ph.D.

Ketua Tim Penguji

Digitally signed
by Veronica Sri
Moertini

Dr. Veronica Sri Moertini

Anggota Tim Penguji

Digitally signed
by Keenan
Adiwijaya Leman

Keenan Adiwijaya Leman, M.T.

Mengetahui,

Ketua Program Studi

Digitally signed
by Mariskha Tri
Adithia

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

CLUSTERING DATA LINTASAN BERDASARKAN KARAKTERISTIKNYA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 11 Juli 2023



Wilson Nathanael
NPM: 6181901053

ABSTRAK

Lintasan adalah jalur yang dibuat oleh benda/entitas bergerak. Perkembangan teknologi pelacakan posisi yang pesat menyebabkan kuantitas (jumlah) dan kualitas (ketepatan) data lintasan yang terdiri dari durasi dan posisi entitas dapat dikumpulkan menjadi meningkat pesat. Hal ini membuka peluang untuk melakukan berbagai analisis pada data lintasan. Salah satu analisis yang dapat dilakukan pada data lintasan adalah pengelompokan data lintasan. Pengelompokan data lintasan berguna untuk mengetahui kelompok-kelompok entitas yang bergerak bersama-sama.

Kelompok entitas adalah dua atau lebih individu yang bergerak dalam sebuah bidang gerak yang sama dan menunjukkan kemiripan karakteristik lintasan dalam konteks spasial. Dua buah lintasan yang berada dalam sebuah kelompok akan menunjukkan karakteristik yang serupa. Hal ini disebabkan dua buah individu berkelompok pada umumnya memiliki tujuan bergerak yang serupa sehingga akan melakukan gerakan yang serupa pula.

Pengelompokan entitas dapat dibagi menjadi 2 berdasarkan data lintasannya. Pertama, suatu entitas dapat menjadi anggota kelompok-kelompok yang berbeda dan dapat berubah identitas kelompoknya seiring berjalannya waktu. Penelitian ini akan berfokus kepada tipe yang kedua di mana sebuah entitas hanya dapat menjadi anggota dari satu kelompok yang sama selama durasi lintasannya. Lebih jauh lagi, penelitian ini hanya memperhatikan data lintasan dari data pergerakan pejalan kaki.

Pada penelitian ini, pengelompokan dilakukan dengan cara membandingkan ukuran karakteristik lintasan. Ukuran karakteristik lintasan dibagi menjadi 3 kategori yaitu karakteristik waktu, intrinsik, dan spasial. Masing-masing kategori digunakan sebagai parameter pembandingan lintasan untuk tahapan *clustering* spesifiknya. Masing-masing *clustering*nya adalah *Highly Connected Subgraph*, *Anisotropic DBSCAN*, dan *DBSCAN* secara berurutan. Proses *clustering* akan dilakukan secara berurutan dimana hasil *clustering* tahap sebelumnya akan di *cluster* ulang menggunakan karakteristik selanjutnya, membagi kelompok lintasan secara bertahap.

Selanjutnya, metodologi pengelompokan yang diusulkan akan diuji dengan berbagai data lintasan pejalan kaki nyata yang didapatkan dari penelitian sebelumnya. Hasil eksperimen menunjukkan bahwa secara keseluruhan perangkat lunak yang diimplementasikan, mampu mengelompokkan lintasan-lintasan kedalam kelompok yang bermakna dan relevan ketika dibandingkan dengan hasil identifikasi manusia pada kondisi dan tipe data yang ideal. Relevansi hasil *clustering* juga ditemukan pada skenario-skenario yang pada penelitian sebelumnya kurang berhasil dikelompokkan, sebagai contohnya, 2 buah lintasan dengan bentuk, kecepatan, dan posisi spasial yang sama dapat dibedakan berdasarkan arah gerak maupun posisi jendela waktunya.

Kata-kata kunci: lintasan, *clustering*, karakteristik temporal, kedekatan spasial, kemiripan lintasan, evaluasi *clustering*

ABSTRACT

Trajectory is a route created by a moving object/entity. The rapid development of position tracking technology has led to a significant increase in the quantity (number) and quality (accuracy) of path data, consisting of entity duration and position, that can be collected. This opens up opportunities to perform various analyses on path data. One of the analyses that can be done on path data is clustering. Clustering path data is useful for identifying groups of entities that move together.

Groups of moving entity can be described as two or more individual that move in the same spatial space and exhibit similarity for both its trajectory characteristic and spatial closeness between two trajectory left behind by the moving entity. This similarity between two entity can be attributed to the fact that both entity are most likely have the same goal of doing that movement to start with, therefore is more likely to do the same kind of movement to achieve their common goal.

There are two types of entity clustering based on path data. In the first type, an entity can be a member of different groups. This research focuses on the second type, where an entity can only be a member of the same group during its path duration. Furthermore, this research only considers path data from pedestrians.

In this study, clustering is performed by comparing characteristic measures possessed by trajectory. Characteristic measures of paths can be divided into three main categories: time characteristics, intrinsic characteristics, and spatial characteristics. Each category is used as a comparison parameter for clustering in its specific clustering stage. Each clustering is as follow, Highly Connected Sub-graph for time component, Anisotropic DBSCAN for intrinsic characteristic component, and DBSCAN for spatial component in order. The clustering process is performed sequentially, where the results of the previous stage clustering are re-clustered using the next characteristics, thus dividing the group further.

Furthermore, the clustering technique is tested with various pedestrian path data. The experimental results show that the implemented software is capable of clustering paths into meaningful and relevant groups compared to human identification. The relevance of clustering results is also found in scenarios that were less successful in previous research.

Keywords: trajectory, clustering, temporal characteristic, spatial proximity, trajectory similarity, clustering evaluation

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan yang Maha Esa, yang karena rahmat dan karunianya memungkinkan penulis untuk menyelesaikan skripsi ini hingga tuntas. Penulisan skripsi dengan judul "Clustering Data Lintasan Berdasarkan Karakteristiknya" ini dilakukan sebagai bentuk pemenuhan kewajiban dan syarat tempuh gelar sarjana program studi teknik informatika di Universitas Katolik Parahyangan.

Penulis menyadari bahwa proses pengerjaan dan penulisan skripsi ini tidak dapat dilakukan tanpa bantuan dan kontribusi berbagai pihak. Penulis ingin mengungkapkan ucapan terima kasih yang sebesar-besarnya kepada setiap individu yang telah mendukung dan memberikan kontribusi baik secara moral maupun material selama proses pengerjaan skripsi ini, terlebih untuk individu-individu berikut :

1. Kedua orang tua yang telah memberikan peluang bagi penulis untuk menempuh studi serta memberikan dukungan selama proses pengerjaan skripsi ini.
2. Bapak Lionov selaku dosen pembimbing yang telah dengan sabar dan tekun membimbing penulis melalui proses penelitian dan penulisan skripsi meski banyak kekurangan dari penulis.
3. Ibu Dr. Veronica Sri Moertini dan Bapak Keenan Adiwijaya Leman, sebagai dosen penguji skripsi yang telah memberikan kritik serta saran konstruktif untuk perbaikan skripsi ini.
4. Bapak Pascal Alfadian Nugroho serta rekan-rekan di DNArtworks tempat penulis bekerja, yang memberikan ruang dan waktu untuk penulis menyelesaikan skripsi disamping pekerjaannya.
5. Teman-teman seperjuangan penulis di jurusan teknik informatika yang selalu ada saat penulis membutuhkan bantuan, saran, atau sekedar sebagai teman bicara kala mengerjakan skripsi, yaitu Reynaldi Lukas, Nicholas Aldo, Premananda Setyo, Vincent Kurniawan, Dimas Kurniawan, Stanislaus Dendrio, Kevin Sunjaya, serta masih banyak lagi yang penulis tidak bisa sebutkan satu persatu.
6. Teman-teman bermain penulis di grup Power Supply yang selalu menyediakan tempat bagi penulis untuk melepas penat dan membawa suasana menyenangkan ditengah-tengah tuntutan skripsi, serta selalu memberikan saran serta motivasi bagi penulis untuk menyelesaikan skripsi ini. Diantaranya adalah Kevin, Kelvin, William, Richard, Felix, Alfaza, Henrico, dan masih banyak lagi.
7. Para debater PEDS yang senantiasa menyemangati penulis selama proses pengerjaan skripsi serta memberikan ruang bagi penulis untuk fokus menyelesaikan skripsi.

Penulis sadar sepenuhnya bahwa karya skripsi ini masih jauh dari kesempurnaan serta tidak luput dari kesalahan. Maka dari itu penulis berharap dapat diberikan masukan dan kritis konstruktif yang dapat digunakan baik oleh penulis, pembaca, serta pihak lain yang hendak menggunakan atau mengembangkan penelitian pada skripsi ini kedepannya.

Bandung, Juli 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan	5
1.4 Batasan Masalah	5
1.5 Metodologi	5
1.6 Sistematika Pembahasan	6
2 LANDASAN TEORI	7
2.1 Pergerakan	7
2.1.1 Objek Bergerak	7
2.1.2 Lintasan	8
2.2 Homografi	12
2.3 <i>Collective Movement Pattern</i>	12
2.4 Analisis lintasan	16
2.4.1 <i>Clustering</i> Lintasan	17
2.5 <i>Graph-based Clustering</i> dengan algoritma HCS	18
2.6 <i>Density-Based Clustering</i> Dengan ADCN	19
2.7 Metode Perhitungan Jarak Antar Lintasan	20
2.7.1 Fréchet Distance	21
2.7.2 Proksemik antar lintasan	23
2.8 <i>Evaluasi Performa Secara Kualitatif</i>	24
2.8.1 <i>F1 score</i>	24
2.8.2 <i>Rand Index Score</i>	25
3 ANALISIS	27
3.1 Analisis Permasalahan penelitian sebelumnya	27
3.1.1 Perbedaan Arah	28
3.1.2 Data dengan jalur utama	28
3.1.3 shape difference	29
3.2 Usulan Penggunaan Parameter Tambahan	29
3.3 Metodologi Penyelesaian	30
3.3.1 Temporal Clustering	30
3.3.2 Attribute Similarity Clustering	31
3.3.3 Spatial proximity Clustering	31
3.4 Algoritma Pengelompokan Data Pergerakan	31
3.5 Algoritma Anisotropic DBSCAN	32

4	PERANCANGAN	35
4.1	Fungsionalitas Perangkat Lunak	35
4.2	Spesifikasi Input dan Output	37
4.2.1	bentuk data lintasan	37
4.2.2	bentuk data parameter kemiripan lintasan	38
4.2.3	bentuk data identifikasi manusia (<i>Ground-Truth</i>)	38
4.2.4	bentuk keluaran hasil <i>clustering</i>	38
4.3	Modul Masukan dan Penerjemah	38
4.4	Modul Pembangunan Karakteristik	39
4.5	Modul <i>Clustering</i> Temporal	40
4.6	Modul <i>Clustering</i> Kemiripan Atribut Intrinsik Lintasan	40
4.7	Modul <i>Clustering</i> Spasial Lintasan	41
4.8	Modul Visualisasi video	41
4.9	Modul Evaluasi	41
5	IMPLEMENTASI	43
5.1	Lingkungan Implementasi	43
5.2	Modul Masukan dan Penerjemah	43
5.3	Modul Pembangunan Karakteristik	45
5.4	Modul <i>Clustering</i> Temporal Lintasan	47
5.5	Modul <i>Clustering</i> Kemiripan Atribut Lintasan	48
5.6	Modul <i>Clustering</i> Spasial Lintasan	48
5.7	Modul Visualisasi	48
5.8	Modul Evaluasi	50
6	EKSPERIMEN	51
6.1	Data Pergerakan	51
6.1.1	ETH Walking Pedestrian	51
6.1.2	ETH hotel	52
6.1.3	Galleria Vittorio Emanuele II (GVEII)	52
6.2	Evaluasi Kuantitatif	53
6.3	Evaluasi Kualitatif	54
6.4	Hasil Eksperimen	54
7	KESIMPULAN	75
7.1	Pencapaian Tujuan	75
7.2	Jawaban Atas Masalah Penelitian	76
7.3	Saran Penelitian Lanjutan	77
	DAFTAR REFERENSI	79
	A KODE PROGRAM	81
	B CONTOH DATA LINTASAN	91

DAFTAR GAMBAR

1.1	Contoh pergerakan manusia	1
1.2	Contoh lintasan dalam bidang 2 dimensi	2
1.3	Contoh lintasan yang dapat kita temui di dunia nyata	3
1.4	potongan video data lintasan pejalan kaki yang akan digunakan sebagai data eksperimen	4
2.1	Lintasan dengan interval panjang (kiri) dan lintasan dengan interval singkat (kanan)	11
2.2	visualisasi cara kerja <i>Graph-based Clustering</i>	18
2.3	visualisasi <i>graph</i> persentase tumpang tindih jendela waktu lintasan dengan nilai batas 80%	19
2.4	<i>Density-Based Clustering</i>	19
2.5	perbedaan DBSCAN dan ADCN	20
2.6	2 buah lintasan dengan jumlah titik yang berbeda	21
2.7	visualisasi Fréchet	22
3.1	2 buah lintasan yang berdekatan namun memiliki arah yang berlawanan	28
3.2	data dengan fitur jalur utama	29
3.3	2 lintasan yang berdekatan dengan bentuk yang berbeda	29
4.1	Diagram Konteks	35
4.2	Diagram alir data tingkat pertama	37
4.3	Diagram flowchart modul masukan dan penerjemah	40
4.4	Diagram data flow untuk modul evaluasi	42
6.1	salah satu frame dari dataset ETH	52
6.2	salah satu frame dari dataset ETH-hotel	52
6.3	salah satu frame dari dataset GVEII	53
6.4	salah satu frame dari dataset GVEII	54
6.5	salah satu frame dari hasil visualisasi dataset eth	66
6.6	rute utama dan hasil <i>clustering</i> jarak	66
6.7	contoh lintasan yang berdekatan namun berlawanan arah	67
6.8	Nilai batas perbedaan kecepatan rata-rata yang terlalu ketat menyebabkan sebuah <i>cluster</i> terpecah menjadi 2	67
6.9	Informasi mengenai data lintasan	68
6.10	contoh 2 <i>cluster</i> yang menempati rute identik berhasil dipisahkan oleh perangkat lunak	69
6.11	contoh lintasan yang berdekatan namun berlawanan arah	69
6.12	entitas 59 dengan bentuk lintasan yang serupa dan berada dalam proksemik interaksi sosial entitas 56 dan 67 dapat dibedakan berdasarkan kecepatan rata-rata objek bergerak	70
6.13	Parameter perbedaan arah yang terlalu ketat menyebabkan beberapa rombongan tidak teridentifikasi oleh perangkat lunak	70
6.14	Informasi mengenai data lintasan	71

6.15	contoh lintasan yang berdekatan namun berlawanan arah	72
6.16	entitas 99 dan 100 yang sempat terpecah dan kembali bersatu tidak dipecah oleh perangkat lunak	72
6.17	contoh skenario banyak entitas yang berdekatan secara spasial dibedakan berdasarkan perbedaan arah geraknya	73
6.18	contoh-contoh kasus dimana <i>cluster</i> besar terbentuk untuk entitas yang berada dalam rute yang sama namun lokasinya berjauhan	74

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pergerakan sebuah entitas atau benda adalah sebuah fenomena yang telah diamati dan didokumentasikan sepanjang sejarah manusia. Sehari-hari kita menemui objek yang bergerak, mulai dari kendaraan yang melaju, orang yang berjalan, migrasi binatang di alam liar, bilah kipas yang berputar dan masih banyak lagi. Namun hanya baru-baru ini, kita akhirnya dapat mengumpulkan dan mengekstraksi data dengan skala yang besar dan tingkat akurasi yang tinggi dari objek bergerak contohnya adalah data pergerakan pejalan kaki. Peningkatan kuantitas dan kualitas data yang dapat diekstrak dari sebuah objek bergerak secara mayoritas disebabkan oleh meledaknya perkembangan dan kebutuhan teknologi terutama dalam sektor *quality of life* dan keamanan yang menggunakan Global Positioning System dan juga kamera CCTV.



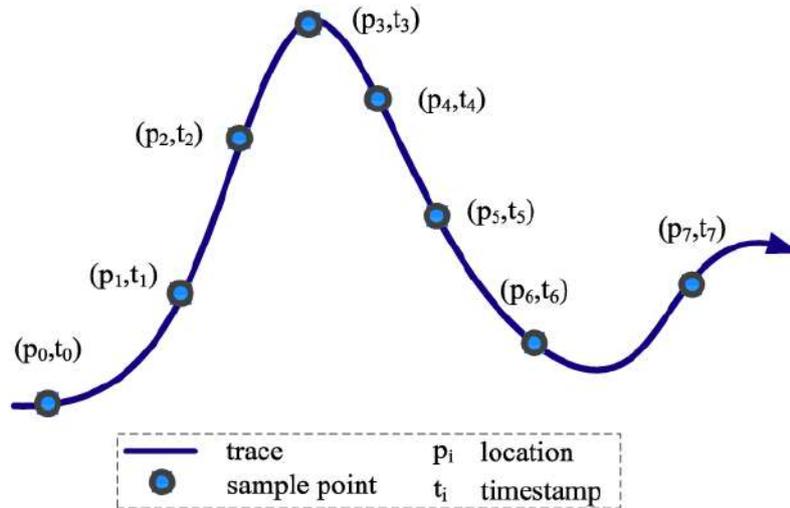
Gambar 1.1: Contoh pergerakan manusia¹

Data pergerakan, seperti data gerak pejalan kaki diklasifikasikan sebagai data deret waktu yang mana mengandung posisi objek (ruang) yang diamati dan juga *timestamp* dari setiap data lokasi tersebut (temporal). Data pergerakan biasanya didapatkan dari ekstraksi kumpulan poin data dari rekaman satu atau lebih entitas menggunakan teknologi tertentu. Data lokasi yang merepresentasikan posisi objek biasanya direkam dalam bentuk koordinat baik itu 2 dimensi maupun 3 dimensi, menyesuaikan dengan teknologi yang digunakan dan kebutuhan dari pengambilan data tersebut. Selain data koordinat, data pergerakan juga memiliki atribut waktu yang menandakan urutan atau marka waktu saat data posisi dihasilkan dan direkam.

Beriringan dengan meningkatnya kuantitas dan kualitas data pergerakan, bidang analisis data

¹Alexandra Tran, A Large Group of People Crossing the Street, 2022, <https://unsplash.com/photos/M95jQ5Nm5VI>

²Figure 1. [A trajectory is generated by sampling from a continuous trace]. Adapted from "A Survey on Trajectory Data Mining: Techniques and Applications" by Feng, Z., & Zhu, Y., 2016, IEEE Access, 4, p. 2056-2067.



Gambar 1.2: Contoh lintasan dalam bidang 2 dimensi²

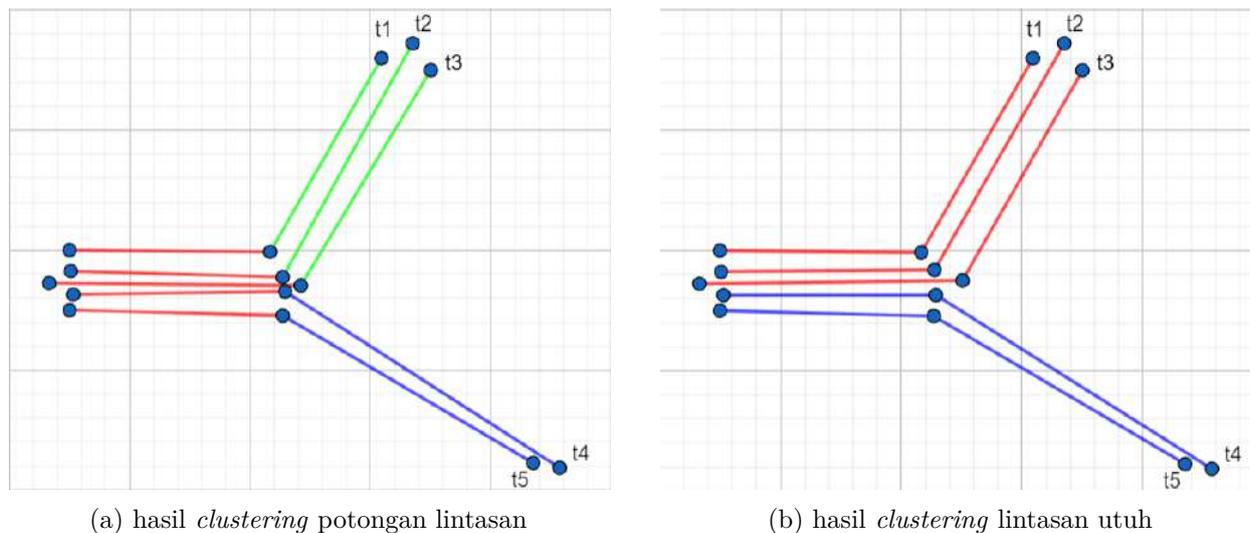
pergerakan juga mengalami kemajuan baik dalam hal metode maupun hasil analisis. Menggunakan data yang dihasilkan dari pengamatan dan rekaman menggunakan teknologi seperti : *Closed-Circuit Television (CCTV)*, *Global Positioning System (GPS)*, maupun *Radio Frequency Identification (FRID)*, Ada banyak penelitian mengenai analisa data pergerakan di berbagai bidang ilmu seperti *urban-planning*, *biodiversity research*, *weather prediction*, *crisis management*, dan masih banyak lagi. beberapa contoh konkret penggunaan data pergerakan dari penginderaan CCTV maupun GPS diantaranya :

1. Davide Antonio Maria COTA menggunakan data dari CCTV dan *deep learning* untuk mengenali apakah individu dalam video menggunakan masker dan menaati protokol kesehatan secara masal.[1]
2. Robin Steenweg et al. menggunakan penginderaan jarak jauh dan kamera untuk memantau keanekaragaman hayati di skala global.[2]
3. Studi mengenai penggunaan data CCTV sebagai basis informasi manajemen area urban juga dilakukan oleh Jung Hoon Han et al.[3]

ketersediaan data pergerakan berarti lebih banyak hal yang bisa dilakukan dan didapatkan dengan melakukan analisis data lintasan. Salah satu area yang banyak mendapat perhatian saat ini adalah pengelompokan atau *clustering* lintasan pejalan kaki [4] [5] [6]. Perhatian ini muncul akibat adanya kebutuhan manusia untuk mampu mengenali kelompok individu yang bergerak di sebuah area tertentu. Pengenalan kelompok dapat membantu bidang keamanan untuk mengenali dan menandai individu yang bergerak sendiri ataupun berkelompok sesuai dengan kebutuhan. salah satu contoh aplikasi yang mungkin berguna adalah pengenalan individu yang bergerak sendiri dan menunjukkan pola tertentu yang disinyalir memiliki tujuan yang tidak disegani di area sensitif seperti bandara, rumah sakit, dsb.

Clustering lintasan objek bergerak adalah sebuah metode analisis lintasan yang bertujuan mengelompokkan 2 atau lebih lintasan objek bergerak berdasarkan kemiripan lokasi spasial, bentuk, durasi, atau deskriptor lainnya. Secara umum, ada 2 tipe *clustering* lintasan, *clustering* potongan lintasan dan *clustering* lintasan penuh, yang mana pembedanya adalah apakah *clustering* dilakukan berdasarkan nilai dan atribut sebuah lintasan utuh atau yang sudah di segmentasi sebelumnya. *clustering* potongan lintasan sejauh ini merupakan metode yang paling sering digunakan dikarenakan kompleksitas waktu dan ruang yang lebih rendah dan menghasilkan *cluster* yang lebih mudah di jelaskan. Sayangnya segmentasi lintasan yang perlu dilakukan sebelum proses *clustering* menyebabkan hilangnya beberapa informasi kunci milik lintasan penuh. Selain hilangnya informasi, *clustering* potongan lintasan juga menimbulkan ketimpangan definisi kelompok, dimana dalam sebuah area pengamatan yang terbatas, individu yang bergerak dapat memiliki lebih dari 1 designasi kelompok

di waktu yang bersamaan. Maka dari itu, dalam penelitian ini akan digunakan metode *clustering* lintasan utuh dengan mengorbankan kompleksitas ruang dan waktu.



Gambar 1.3: Contoh lintasan yang dapat kita temui di dunia nyata

Melihat penjelasan diatas, kita dapat melihat bahwa dalam analisis lintasan objek bergerak, banyak metode dan algoritma yang dapat di implementasikan untuk menghasilkan pengenalan kelompok individu. Dalam penelitian sebelumnya seperti pada [7] [8], Jarak antar lintasan adalah dasar *clustering* yang paling umum digunakan untuk menentukan apakah 2 lintasan berada dalam kelompok yang sama. Jarak antar lintasan merepresentasikan kedekatan sebuah lintasan dengan lintasan lainnya. Pendekatan ini dalam kebanyakan kasus dapat menghasilkan kelompok yang memadai, sayangnya ada limitasi untuk kasus-kasus tertentu seperti adanya jalur padat dalam data, dimana hampir seluruh lintasan melalui jalur tertentu karena adanya jalan atau rintangan. kasus lain adalah saat 2 buah lintasan yang berdekatan memiliki arah atau kecepatan yang berbeda. Maka dari itu, tujuan penelitian ini adalah untuk melakukan eksplorasi dan implementasi pembuktian, untuk mengetahui apakah ada parameter lain yang dapat digunakan beriringan dengan jarak antar lintasan sebagai pembanding kemiripan dan kedekatan lintasan, beserta dengan relasi yang dimiliki atribut-atribut tersebut dengan satu sama lain. Pada umumnya, banyak penelitian yang beranggapan bahwa jika ada dua lintasan yang berdekatan berarti kedua lintasan dapat diasumsikan serupa.

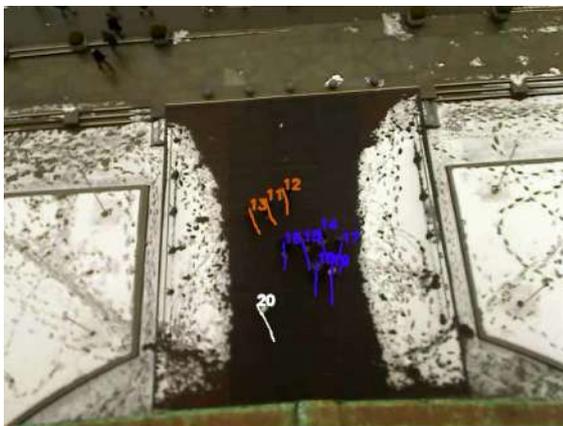
Dalam penelitian ini akan dilakukan *clustering* dengan memperhatikan baik kedekatan spasial antar lintasan, kemiripan aspek temporal lintasan, dan kemiripan antar atribut intrinsik lintasan. Kemiripan lintasan akan meliputi atribut-atribut intrinsik seperti arah gerak global objek bergerak, bentuk lintasan yang ditinggalkan objek bergerak, durasi keseluruhan lintasan, jarak tempuh objek, dan kecepatan objek bergerak dalam lintasan. Dimana arah gerak objek merupakan vektor gerak yang dilihat dari awal lintasan ke arah ujung akhir lintasan, bentuk lintasan nantinya akan dideskripsikan menjadi nilai-nilai seperti derajat penyimpangan, panjang lintasan, dan perubahan sudut gerak objek. Terakhir, untuk kecepatan objek akan digunakan nilai rata-rata kecepatan objek sepanjang lintasan dan total perubahan akselerasi objek bergerak.

Pendekatan *clustering* data pergerakan atau lintasan dengan parameter yang meliputi kedekatan spasial, kemiripan aspek temporal dan kemiripan atribut intrinsik lintasan secara bertingkat sudah secara parsial pernah dilakukan dalam penelitian sebelumnya seperti pada [9] dengan masing-masing perbedaannya. Maka dari itu, penelitian ini tidak dapat dikatakan sepenuhnya baru, melainkan percobaan penggabungan banyak pendekatan yang diimplementasikan pada penelitian sebelumnya, menjadi sebuah alur proses baru dengan harapan dapat memberikan kontribusi, dorongan dan kemajuan dalam pengertian dan penelitian *clustering* data pergerakan.

Pada penelitian ini akan dilakukan implementasi perangkat lunak yang memisahkan kedekatan

dan kemiripan lintasan dan melakukan *clustering* menggunakan lebih dari 1 pendekatan (waktu, intrinsik dan spasial). Dalam penelitian ini, akan dibuat sebuah alur metodologi baru yang menggunakan parameter pergerakan kolektif seperti jendela waktu lintasan, arah gerak objek, kecepatan objek, yang harapannya akan memperluas definisi sebelumnya dari pengelompokan lintasan. Selanjutnya menggunakan rancangan metodologi yang baru, akan diimplementasikan algoritma-algoritma yang mampu mendeteksi kelompok yang terdapat dalam sebuah himpunan data lintasan.

Implementasi algoritma dan alur metodologi baru akan melalui proses eksperimentasi dan ujicoba menggunakan data lintasan pejalan kaki nyata yang diambil dari penelitian sebelumnya. Akan terdapat 3 buah data lintasan pejalan kaki nyata yang digunakan: data pejalan kaki ETH[10], data pejalan kaki ETH-hotel dan data pejalan kaki *Galleria Vittorio Emanuele II* (GVEII). Masing-masing data lintasan pejalan kaki memiliki karakteristik lintasan yang berbeda dan akan digunakan untuk menentukan kualitas dan relevansi algoritma dan metodologi yang baru diimplementasikan dalam menangani *clustering* terhadap data lintasan nyata. Kualitas dan relevansi hasil *clustering* oleh perangkat lunak akan ditentukan menggunakan evaluasi kuantitatif yaitu nilai *F1 score* dan *Rand Index* maupun kualitatif oleh manusia.



(a) Data ETH



(b) Data ETH-hotel



(c) data GVEII

Gambar 1.4: potongan video data lintasan pejalan kaki yang akan digunakan sebagai data eksperimen

Algoritma ini akan diimplementasikan dalam bentuk sebuah perangkat lunak yang menggunakan bahasa pemrograman python 3 dengan tampilan antar muka berupa *command line interface*. Perangkat lunak akan menerima masukan berupa himpunan data lintasan dalam bentuk sebuah

berkas termasuk anotasi dan video asli data lintasan, anotasi kelompok hasil identifikasi manusia, dan parameter batasan nilai atribut pembandingan lintasan. Perangkat lunak akan menghasilkan 2 buah keluaran yaitu nilai kualitas / relevansi hasil *clustering* data pergerakan dan berkas berupa video hasil clustering.

1.2 Rumusan Masalah

- Seperti apakah sifat dan karakteristik lintasan yang dapat digunakan untuk pendekatan pengelompokan lintasan?
- Bagaimana langkah-langkah pengelompokan lintasan objek bergerak berdasarkan kemiripan aspek lintasan?
- Bagaimana cara menilai akurasi dan kualitas algoritma pengelompokan lintasan ?

1.3 Tujuan

- Melakukan eksplorasi sifat dan karakteristik lintasan serta pendekatan pengelompokan lintasan.
- Membuat dan mengimplementasikan algoritma yang dapat mengelompokkan lintasan benda bergerak yang menggunakan beberapa aspek lintasan sekaligus.
- Mengimplementasikan algoritma evaluasi kuantitatif untuk menilai kualitas perangkat lunak yang dibuat.

1.4 Batasan Masalah

Penelitian ini dibuat berdasarkan batasan-batasan sebagai berikut:

- Data pergerakan yang akan digunakan dibatasi pergerakan pejalan kaki yang didapatkan oleh kamera CCTV atau sejenisnya
- Data yang digunakan berupa data beranotasi dari video yang telah diproses oleh pihak lain
- Anotasi dalam data yang digunakan akan diasumsikan benar dan tidak akan dilakukan perbaikan pada kesalahan yang ditemukan jika ada.
- Pengelompokan lintasan didasarkan pada data lintasan tercatat yang digunakan dan tidak mengkonsiderasikan faktor lain seperti interaksi individu dalam video, kemiripan lintasan setelah atau sebelum pencatatan data, dsb.

1.5 Metodologi

Berikut adalah metodologi yang digunakan dalam penelitian ini:

1. Melakukan studi pustaka mengenai data pergerakan dan karakteristik lintasan dari entitas yang bergerak,
2. Melakukan studi pustaka mengenai Definisi kelompok objek bergerak,
3. Melakukan studi pustaka mengenai perhitungan jarak dan algoritma perhitungan jarak,
4. Melakukan studi pustaka mengenai Analisis lintasan khususnya *clustering* lintasan,
5. Membuat algoritma yang mampu melakukan *clustering* data lintasan dengan parameter selain jarak
6. Mengimplementasikan algoritma *clustering* data lintasan menjadi perangkat lunak
7. Melakukan eksperimen dan menganalisa algoritma yang diimplementasikan menggunakan berbagai parameter dan algoritma dengan data pejalan kaki,
8. melakukan evaluasi dan memvisualisasikan kinerja algoritma baru.

1.6 Sistematika Pembahasan

Penelitian ini akan dibahas dalam tujuh bab yang masing-masing akan berisi :

1. Bab 1 Pendahuluan

Bab 1 akan berisi latar belakang penelitian mengenai pergerakan, pemanfaatan analisis lintasan, penelitian sebelumnya, dan masalah-masalah yang terdapat dalam analisis data pergerakan. Bab ini juga berisikan rumusan masalah, tujuan penulisan penelitian, batasan masalah, metologi penelitian, dan sistematika penulisan penelitian ini.

2. Bab 2 Dasar Teori

Bab 2 akan berisi dasar teori mengenai pergerakan, lintasan dan karakteristiknya, homografi, pergerakan kolektif, analisis lintasan, *clustering* lintasan, algoritma perhitungan jarak, serta evaluasi hasil *clustering*.

3. Bab 3 Analisis

Bab 3 akan berisi pembahasan mengenai permasalahan penggunaan parameter jarak sebagai satu-satunya pembeda lintasan pada penelitian sebelumnya. Bab ini juga akan mengusulkan penggunaan parameter lain yang dapat menjadi alternatif dasar pengelompokan lintasan. Selain itu bab ini juga akan membahas algoritma dan metodologi yang akan digunakan demi menyelesaikan masalah yang dikemukakan.

4. Bab 4 Perancangan

Bab 4 akan berisi pembahasan mengenai rancangan perangkat lunak *clustering* data pergerakan yang akan dibuat. Bab ini juga akan satu persatu membahas mengenai modul perangkat lunak yang akdibuat serta spesifikasi masukan dan keluaran dari perangkat lunak.

5. Bab 5 Implementasi

Bab 5 akan berisi pembahasan mengenai detail implementasi hasil rancangan perangkat lunak pada bab sebelumnya. Dalam bab ini akan dibahas mengenai paradigma dan pustaka yang digunakan untuk mengimplementasikan perangkat lunak *clustering* data pergerakan.

6. Bab 6 Eksperimen dan Hasil

Bab 6 akan berisi pembahasan mengenai tahap eksperimentasi dan uji coba fungsional perangkat lunak yang sudah diimplementasikan pada bab sebelumnya. Dalam bab ini akan dijelaskan bagaimana proses yang akan dilakukan dalam mengevaluasi hasil *clustering* dari perangkat lunak.

7. Bab 7 Kesimpulan dan Saran

Bab 7 akan berisi kesimpulan dari proses penelitian yang sudah dilakukan. Dalam bab ini, terdapat pembahasan mengenai pencapaian tujuan penelitian yang telah dinyatakan pada Bab 1 serta jawaban mengenai rumusan masalah penelitian. Bab ini juga akan membahas mengenai saran untuk penelitian lanjutan yang relevan dengan topik yang berada dalam bidang yang sama.