

SKRIPSI

**PENERAPAN DATA MINING PADA MASALAH
PENGENALAN POINT OF INTEREST**



Gian Martin Dwibudi

NPM: 6181801015

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

UNDERGRADUATE THESIS

**APPLICATION OF DATA MINING ON THE PROBLEM OF
RECOGNIZING POINT OF INTEREST**



Gian Martin Dwibudi

NPM: 6181801015

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

LEMBAR PENGESAHAN

PENERAPAN DATA MINING PADA MASALAH PENGENALAN POINT OF INTEREST

Gian Martin Dwibudi

NPM: 6181801015

Bandung, 16 Januari 2023

Menyetujui,

Pembimbing

Digitally signed by
Kristopher David Harjono

Kristopher David Harjono, M.T.

Ketua Tim Penguji

Digitally signed by Natalia

Natalia, M.Si.

Anggota Tim Penguji

Digitally signed by Luciana
Abednego

Luciana Abednego, 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:

PENERAPAN DATA MINING PADA MASALAH PENGENALAN POINT OF INTEREST

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 16 Januari 2023



F5598AJX598203566

Gian Martin Dwibudi

NPM: 6181801015

ABSTRAK

Point of Interest (POI) merupakan sebuah lokasi geografis yang memiliki fungsi tertentu dan dikenali orang. POI dapat berupa tempat seperti toko, *mall*, restoran, dan tempat-tempat lainnya. Beberapa POI memiliki bagian penting yang unik terhadap POI tersebut. Bagian penting tersebut biasanya berupa logo atau bagian lainnya yang mudah terlihat dan berbeda dari bagian-bagian POI lain. Bagian penting dari POI tersebut dapat digunakan untuk mengenali sebuah POI dengan mudah. Seseorang dapat mengenali sebuah POI hanya dengan melihat bagian tersebut saja.

Komputer dapat mengenali POI dari sebuah gambar dengan menggunakan fitur lokal dalam gambar. Salah satu teknik yang dapat digunakan untuk mengenali gambar adalah (*Object Instance Recognition*) OIR. Teknik OIR mengenali gambar dengan memasang fitur lokal dari gambar tersebut dan gambar-gambar di *dataset*. Sebuah gambar dapat mengandung sangat banyak fitur lokal, sehingga proses pemasangan dapat memakan waktu yang sangat lama. Proses dapat dipercepat dengan menggunakan hanya fitur lokal dari bagian POI yang penting saja.

Penelitian ini bertujuan untuk menyediakan cara menyaring fitur lokal dan mengambil yang hanya berasal dari bagian penting POI saja. Penyaringan fitur lokal berdasarkan pada asumsi bahwa bagian penting dari POI adalah bagian yang sifatnya konsisten dan unik terhadap POI tersebut. Bagian penting tersebut bersifat konsisten dan unik maka akan menghasilkan fitur lokal yang sifatnya konsisten dan unik juga. Fitur lokal yang konsisten dan unik dapat ditentukan dengan menggunakan metode *clustering*.

Metode *clustering* digunakan untuk membagi fitur-fitur lokal menjadi kelompok-kelompok. Kelompok-kelompok tersebut akan digunakan untuk menentukan tingkat kekonsistenan dan keunikan fitur-fitur lokal anggotanya. Tingkat kekonsistenan dan tingkat keunikan tersebut lalu digunakan untuk menyaring fitur lokal yang digunakan dalam OIR.

Penyaringan fitur lokal berdasarkan nilai konsistensi dan nilai keunikan pada penelitian ini dapat mengurangi waktu yang diperlukan untuk melakukan OIR. Pada *dataset* yang digunakan di penelitian waktu proses OIR dapat berkurang hingga 73.97% dengan akurasi yang hanya menurun sebesar 10%.

Kata-kata kunci: *Point of interest*, fitur lokal, OIR, *clustering*, konsistensi, keunikan

ABSTRACT

Point of Interest (POI) is a geographical location with certain usage or function and mostly recognized by people. A POI can be a shop, mall, restaurant, or any other kind of places. Some POI have an important part in its shape that is unique to that POI. That important part may be a logo or another part that stands out and can be easily distinguishable from the parts of other POI. The important part of a POI can be used to identify a POI, one can easily identify a POI just by looking at that important part.

Computer can identify a POI from an image with the help of local features in the image. One of the technique to identify an image is Object Instance Recognition (OIR). OIR can identify an image by matching local features from that image and local features from images in the dataset. An image can contains tons of local features that may cause feature matching to take a long time. In such cases, the feature matching process can be sped up by filtering the local features and take only the important ones.

This research aims to provide a way to filter local features and take only the important local features from a POI. The filtering is based on the assumption that an important parts of a POI are the parts that are both consistent and unique to that POI. The important parts being consistent and unique, would also produce local features that are consistent and unique. The consistency and uniqueness of a local feature can be determined using clustering.

Clustering method can be used to group local features into several groups. The groups can then be used to determine the consistency value and uniqueness value of its member. The consistency value and the uniqueness value will then be used to filter local features in the OIR dataset.

The filtering of local features by the consistency value and the uniqueness value in this research can reduce OIR processing time. In the dataset used in this research OIR processing time goes down by up to 73.97% with only 10% lower accuracy.

Keywords: Point of interest, local feature, OIR, clustering, consistency, uniqueness

*Untuk semua yang telah membantu saya dalam bentuk apapun
selama pengerjaan skripsi ini*

KATA PENGANTAR

Puji syukur kepada Tuhan yang Maha Esa karena atas berkat dan rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Data Mining pada Masalah Pengenalan Point of Interest” ini. Penulis juga mengucapkan terima kasih kepada semua pihak yang telah membantu penulis selama proses pengerjaan skripsi ini. Terima kasih terutama diucapkan kepada:

- Bapak Kristopher David Harjono, M.T. sebagai pembimbing selama pengerjaan skripsi ini, yang telah mengajari penulis bukan hanya tentang hal-hal yang berhubungan dengan skripsi dan kuliah tetapi juga tentang hal-hal umum yang akan berguna untuk bekal masa depan.
- Teman-teman yang saya temui dan kenal selama masa-masa kuliah, diantaranya Michael, Yalvi, dan Mario.
- Teman-teman yang saya kenal dan menjadi lebih mengenal selama waktu saya magang sebagai admin LabKom, yaitu Vincent, Dimas, Lord Patrick, dan Nadia, serta Gas dan Ivan.

Bandung, Januari 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xxi
DAFTAR ISTILAH	xxv
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Batasan Masalah	4
1.5 Metodologi	4
1.6 Sistematika Pembahasan	5
2 LANDASAN TEORI	7
2.1 Object Instance Recognition	7
2.2 SIFT (Scale Invariant Feature Transform)	10
2.2.1 Pencarian Extrema	10
2.2.2 Penentuan Skala	13
2.2.3 Penentuan Orientasi	13
2.2.4 Pembuatan Deskriptor	15
2.2.5 SIFT di OpenCV	16
2.3 ORB (Oriented FAST and Rotated BRIEF)	16
2.3.1 Pencarian Keypoint	16
2.3.2 Penentuan Skala	17
2.3.3 Penentuan Orientasi	18
2.3.4 Pembuatan Descriptor	19
2.3.5 ORB di OpenCV	21
2.4 BSIS (Best Score Increasing Subsequence)	22
2.4.1 Pairing	22
2.4.2 Verification	23
2.4.3 Scoring	24
2.5 Kd-Tree	24
2.6 Locality-Sensitive Hashing	25
2.7 Agglomerative Clustering	27
2.8 Dataset Stanford Mobile Visual Search	28
3 ANALISIS & EKSPERIMEN	31
3.1 Analisis Pengenalan POI	31
3.1.1 Ide Dasar Analisis	31
3.1.2 Tahapan Analisis	31

3.1.3	Implementasi	33
3.1.4	Hasil Analisis	33
3.2	Analisis Fitur Lokal dari Logo dan Bagian yang Konsisten	34
3.3	Dataset yang Digunakan	36
3.3.1	Dataset SMVS	36
3.3.2	Dataset GSV	37
3.4	Analisis Sifat Konsisten pada Fitur Lokal dari Gambar POI	40
3.4.1	Ide Dasar dan Tahapan Analisis	40
3.4.2	Implementasi	42
3.4.3	Hasil	42
3.5	Analisis Sifat Unik pada Fitur Lokal dari Gambar POI	44
3.5.1	Ide Dasar dan Tahapan Analisis	44
3.5.2	Implementasi	45
3.5.3	Hasil	46
3.6	Analisis Tingkat Keunikan dan Kekonsistenan Fitur Lokal pada Gambar POI	48
3.6.1	Ide Dasar Analisis	48
3.6.2	Tahapan Analisis	49
3.6.3	Implementasi	51
3.6.4	Hasil Analisis	52
3.6.5	Visualisasi Nilai Konsistensi dan Keunikan Fitur Lokal	53
3.7	Analisis Penentuan Nilai Threshold untuk Konsistensi dan Keunikan	56
3.7.1	Ide Dasar dan Tahapan Analisis	59
3.7.2	Metode Scoring	59
3.7.3	Hasil Analisis	61
3.8	Analisis Penggunaan Nilai Konsistensi dan Nilai Keunikan untuk OIR dengan BSIS	63
3.8.1	Data Train dan Test	64
3.8.2	Metode BSIS	65
3.8.3	Tahapan Analisis	66
3.8.4	Tahapan Implementasi	67
3.8.5	Hasil Analisis	69
3.9	Analisis Metode Ekstraksi Fitur Lokal ORB	74
3.9.1	Hasil Pengujian	74
4	PERANCANGAN	77
4.1	Rancangan Alur Program	77
4.2	Rancangan Implementasi Metode Clustering Pembuatan Model	77
4.2.1	Rancangan Struktur Folder	78
4.2.2	Rancangan Proses Clustering	78
4.3	Rancangan Kelas ClusterModel	81
4.4	Rancangan Kelas Util	82
4.5	Rancangan Kelas-kelas Implementasi BSIS	83
4.6	Rancangan Perangkat Lunak BSIS	86
5	IMPLEMENTASI & PENGUJIAN	87
5.1	Implementasi Perangkat Lunak BSIS	87
5.2	Pengujian Metode Clustering untuk Identifikasi dengan BSIS	90
5.2.1	Ide Analisis	90
5.2.2	Penentuan Threshold dan Hasil Analisis Metode SIFT	91
5.2.3	Penentuan Threshold dan Hasil Analisis Metode ORB	95
5.2.4	Analisis Hasil Pengujian	100
6	KESIMPULAN & SARAN	103

6.1 Kesimpulan	103
6.2 Saran	104
DAFTAR REFERENSI	105
A KODE PROGRAM	107
B HASIL EKSPERIMEN	133

DAFTAR GAMBAR

1.1	Salah satu contoh POI.	1
1.2	Contoh POI yang memiliki logo unik. POI seperti ini dapat dikenali dengan melakukan identifikasi pada logo tersebut.	2
1.3	Empat gambar di atas menunjukkan permasalahan yang dihadapi pada penelitian ini. Gambar (a) dan (b) merupakan gambar dari dua POI yang berbeda tetapi memiliki logo dengan bagian sudut yang mirip, sudut yang mirip tersebut kemungkinan akan menghasilkan fitur lokal yang mirip juga. Sedangkan pada gambar (c) dan (d) merupakan dua gambar dari POI yang sama tetapi banyak objek latar yang berbeda sehingga akan memunculkan fitur lokal yang tidak konsisten.	3
2.1	Contoh variasi pada gambar <i>cover</i> buku yang dapat menyebabkan masalah pada OIR	8
2.2	Ilustrasi pasangan fitur lokal yang konsisten dan tidak konsisten.	9
2.3	Kurva Gaussian dan bentuk representasi <i>matrix</i> -nya.	11
2.4	Kurva dan <i>matrix</i> Gaussian pada nilai σ yang berbeda.	11
2.5	Efek nilai σ pada hasil gambar konvolusi.	12
2.6	Operasi DoG pada gambar	12
2.7	Penggunaan DoG pada SIFT	12
2.8	Oktaf pada proses konvolusi SIFT	13
2.9	Ilustrasi pembobotan pada <i>Gaussian Weighting</i> . Titik tengah merupakan <i>keypoint</i> yang diperiksa sedangkan setiap kotak merupakan <i>pixel-pixel</i> di sekitar <i>keypoint</i> . Tanda panah pada tiap kotak menunjukkan <i>magnitude</i> dan orientasi <i>pixel</i> tersebut, panjang panah merupakan nilai <i>magnitude</i> dan arahnya merupakan orientasi	14
2.10	Histogram untuk menentukan orientasi dari <i>keypoint</i> . Bin dengan nilai tertinggi (tanda panah biru) akan digunakan sebagai orientasi dari <i>keypoint</i> . Untuk bin lain yang jumlahnya berada dalam rentang 80% dari bin tertinggi (tanda panah hijau) digunakan untuk membuat <i>keypoint baru</i>	15
2.11	Ilustrasi tahapan pembuatan <i>descriptor</i> pada SIFT.	15
2.12	Ilustrasi penentuan <i>keypoint</i> pada ORB. Setiap kotak menunjukkan sebuah <i>pixel</i> pada gambar dan angka di dalamnya merupakan nilai intensitasnya. <i>Pixel</i> di tengah gambar (<i>pixel</i> bernilai 4) merupakan <i>pixel</i> yang akan diperiksa apakah merupakan <i>keypoint</i> . <i>Pixel-pixel</i> yang ditandai dengan kotak putih merupakan 16 <i>pixel</i> yang digunakan untuk memeriksa apakah <i>pixel</i> di tengah merupakan <i>keypoint</i>	17
2.13	<i>Image Pyramid</i> pada ORB	18
2.14	Ilustrasi penentuan orientasi pada ORB.	19
2.15	Ilustrasi penghitungan <i>Integral Image</i> . <i>Matrix I</i> merupakan <i>matrix</i> awal dan <i>Matrix I_Σ</i> merupakan <i>Integral Image</i> dari <i>I</i>	20
2.16	Penghitungan nilai total sebuah daerah dengan menggunakan <i>Integral Image</i>	21
2.17	Ilustrasi penghitungan jarak Hamming antara dua vektor.	21
2.18	Tahapan verifikasi pada BSIS.	24
2.19	Contoh pohon struktur Kd-Tree. Angka di sebelah kanan menunjukkan elemen vektor yang diperiksa pada level tersebut.	25
2.20	Ilustrasi proses <i>banding</i> dan penggunaan <i>bucket</i> sebagai indeks dalam LSH.	26

3.1	Tahapan analisis pengenalan POI.	32
3.2	Dua gambar yang digunakan untuk melakukan analisis pengenalan POI dengan logo. Gambar yang di sebelah kiri adalah Gambar <i>Q</i> dan yang di sebelah kanan adalah Gambar <i>T</i>	32
3.3	Pasangan <i>keypoint</i> dari Gambar <i>Q</i> dan Gambar <i>T</i> yang kuat.	34
3.4	Beberapa sudut pengambilan dan waktu pengambilan berbeda dari suatu POI yang sama.	35
3.5	POI yang memiliki logo dengan sudut yang mirip.	36
3.6	Beberapa contoh gambar yang digunakan pada analisis untuk melihat sifat konsisten pada fitur lokal.	41
3.7	Histogram sebaran jumlah gambar unik tiap <i>cluster</i>	43
3.8	Contoh <i>keypoint</i> yang konsisten menurut analisis ini.	44
3.9	Beberapa contoh gambar yang digunakan pada analisis untuk melihat sifat unik pada gambar POI.	45
3.10	Histogram sebaran jumlah kelas gambar yang unik tiap <i>cluster</i>	47
3.11	Contoh <i>keypoint</i> yang unik menurut analisis ini.	48
3.12	<i>Flowchart</i> tahapan analisis <i>clustering</i> untuk mencari fitur lokal yang konsisten dan unik.	49
3.13	Contoh beberapa gambar yang digunakan pada analisis ini.	50
3.14	Histogram sebaran nilai keunikan.	53
3.15	Histogram sebaran nilai keunikan.	53
3.16	Beberapa contoh gambar POI beserta <i>keypoint</i> yang fitur lokalnya sudah diberi warna sesuai dengan nilai konsistensi dan keunikannya.	55
3.17	Contoh objek yang sifatnya konsisten dan unik dalam POI.	57
3.18	Contoh objek yang sifatnya konsisten dan unik dalam POI.	58
3.19	Contoh tampilan aplikasi Labelimg.	60
3.20	Skor ketepatan untuk tiap <i>threshold</i> pada nilai konsistensi.	62
3.21	Skor ketepatan untuk tiap <i>threshold</i> pada nilai keunikan.	63
3.22	Contoh gambar pada <i>dataset book_covers</i> . Keempat gambar tersebut berada dalam satu kelas yang sama.	64
3.23	Contoh gambar referensi dari <i>dataset book_covers</i>	65
3.24	Contoh gambar referensi dari <i>dataset book_covers</i> yang telah ditransformasi.	65
3.25	Modifikasi pada metode BSIS yang dilakukan pada analisis ini.	66
3.26	Tahapan yang dilakukan dalam analisis untuk menguji penggunaan <i>clustering</i> pada BSIS.	66
3.27	Sebaran nilai konsistensi (<i>consistency</i>) untuk Book Covers 400.	69
3.28	Sebaran nilai keunikan (<i>uniqueness</i>) untuk Book Covers 400.	69
3.29	Perbandingan sebaran waktu Book Covers 400.	70
3.30	Sebaran nilai keunikan (<i>uniqueness</i>) untuk Book Covers 600.	71
3.31	Sebaran nilai konsistensi (<i>consistency</i>) untuk Book Covers 600.	72
3.32	Perbandingan sebaran waktu Book Covers 600.	72
4.1	Rancangan alur program pada penelitian ini.	77
4.2	Rancangan struktur <i>folder</i> untuk pembuatan model.	78
4.3	Tahapan proses <i>clustering</i> hingga didapat nilai keunikan dan konsistensi.	79
4.4	Diagram kelas <code>ClusterModel</code>	82
4.5	Diagram kelas <code>Util</code>	83
4.6	Diagram <i>Package</i> <code>bsis</code>	84
4.7	Rancangan tampilan perangkat lunak untuk BSIS.	86
5.1	Tampilan awal aplikasi perangkat lunak BSIS.	87
5.2	Tampilan aplikasi perangkat lunak BSIS setelah memilih gambar <i>input</i>	88

5.3	Tampilan aplikasi perangkat lunak BSIS setelah didapat hasil identifikasi gambar <i>input</i> .	89
5.4	Histogram sebaran nilai keunikan pada <i>dataset</i> GSV 400.	91
5.5	Histogram sebaran nilai keunikan <i>dataset</i> GSV 400.	91
5.6	Sebaran waktu pengujian metode SIFT <i>dataset</i> GSV 400.	92
5.7	Histogram sebaran nilai keunikan pada <i>dataset</i> GSV 600.	93
5.8	Histogram sebaran nilai keunikan <i>dataset</i> GSV 600.	94
5.9	Sebaran waktu pengujian metode SIFT <i>dataset</i> GSV 600.	95
5.10	Histogram sebaran nilai keunikan pada <i>dataset</i> GSV 400.	96
5.11	Histogram sebaran nilai keunikan <i>dataset</i> GSV 400.	96
5.12	Sebaran waktu pengujian metode ORB <i>dataset</i> GSV 400.	97
5.13	Histogram sebaran nilai keunikan pada <i>dataset</i> GSV 600.	98
5.14	Histogram sebaran nilai keunikan <i>dataset</i> GSV 600.	99
5.15	Sebaran waktu pengujian metode ORB <i>dataset</i> GSV 600.	100

DAFTAR ISTILAH

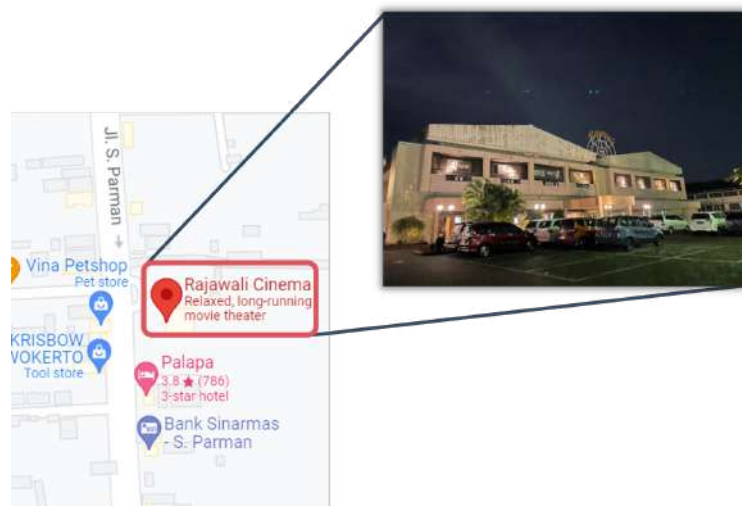
- Clustering** Teknik dalam penambangan data yang membagi data menjadi kelompok-kelompok sesuai dengan kemiripan fitur-fiturnya.
- Keypoint** Sebuah *pixel* yang merupakan titik penting dalam gambar. *Keypoint* biasanya berupa *pixel* yang merupakan titik perpotongan garis, atau *pixel* yang berada di daerah dengan tingkat kontras tinggi.
- Locality-Sensitive Hashing (LSH)** Teknik yang dapat digunakan untuk mencari pasangan terdekat dari vektor biner, dengan memanfaatkan sifat metode *hashing*.
- Object Instance Recognition (OIR)** Teknik pengenalan objek spesifik. OIR mengenali sebuah objek spesifik, bukan hanya kelas atau kategori dari objek tersebut.
- Point of Interest (POI)** Lokasi geografis yang memiliki fungsi atau kegunaan tertentu dan dikenali oleh orang secara umum.
- Best Score Increasing Subsequence (BSIS)** Teknik implementasi OIR. BSIS memodifikasi tahap *pairing* dan *verification* dari OIR
- Fitur Lokal** Fitur dalam gambar yang mendeskripsikan sebuah daerah tertentu dalam gambar. Fitur lokal dapat diperoleh dengan mencari *keypoint* dan melihat karakteristik daerah di sekitar *pixel* yang menjadi *keypoint* tersebut.
- Kd-Tree** Struktur data yang dapat digunakan untuk mencari pasangan vektor termirip dengan waktu proses yang relatif lebih cepat. Kd-Tree mencari pasangan terdekat dengan membandingkan masing-masing elemen pada posisi yang sama.
- Oriented FAST and Rotated BRIEF (ORB)** Metode ekstraksi fitur lokal yang membuat fitur dengan membandingkan nilai-nilai *pixel* di sekitar *keypoint*.
- Scale Invariant Feature Transform (SIFT)** Metode ekstraksi fitur lokal yang membuat fitur dengan menghitung nilai orientasi *pixel* di sekitar *keypoint*.
- Vektor Descriptor** Vektor berisi angka-angka yang mendeskripsikan sebuah fitur lokal. Vektor ini diperoleh dari daerah di sekitar *keypoint* dengan cara yang berbeda-beda untuk setiap metode.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Sebuah lokasi atau titik geografis yang memiliki fungsi atau manfaat tertentu bagi orang secara umum biasa disebut sebagai *Point of interest* (POI). POI dapat berupa tempat seperti restoran, bioskop, toko, rumah sakit, dan tempat-tempat lainnya. Tempat-tempat tersebut memiliki kegunaan dan dikenali oleh orang-orang secara umum. Salah satu contoh POI dapat dilihat pada Gambar 1.1



Gambar 1.1: Salah satu contoh POI.

POI-POI tertentu dapat dikenali dari logo tempat tersebut atau objek-objek lainnya yang terlihat secara langsung. Seperti ditunjukkan pada Gambar 1.2, kedua POI tersebut memiliki logo unik yang terlihat dengan jelas. Pada skripsi ini akan dibuat sebuah sistem untuk mengidentifikasi POI dari masukan yang berisi gambar POI tersebut. Proses identifikasi POI dilakukan dengan mendeteksi logo khusus atau bagian lain dari POI yang sifatnya unik.



Gambar 1.2: Contoh POI yang memiliki logo unik. POI seperti ini dapat dikenali dengan melakukan identifikasi pada logo tersebut.

Proses identifikasi POI akan dilakukan menggunakan teknik *Object Instance Recognition* (OIR). Teknik OIR sendiri merupakan teknik pengenalan objek secara spesifik. Sebuah algoritma OIR harus dapat mengatasi masalah seperti pencahayaan, sudut pengambilan, objek *background*. Objek harus tetap dapat diidentifikasi walaupun gambar memiliki gangguan-gangguan tersebut. OIR dapat dilakukan dengan memanfaatkan fitur lokal dalam gambar. Pada dasarnya OIR akan memasangkan fitur-fitur lokal dari gambar masukan dengan fitur-fitur lokal dari gambar-gambar *dataset*. Pada skripsi ini untuk mempercepat waktu proses pada tahap pencarian pasangan fitur lokal, pemasangan fitur lokal akan dilakukan dengan memanfaatkan *library* FLANN yang tersedia di OpenCV pada Python. FLANN sendiri merupakan *library* berguna untuk mencari perkiraan pasangan terdekat.

Fitur lokal merupakan fitur yang mendeskripsikan sebuah daerah penting pada gambar atau yang biasa disebut *keypoint*. Salah satu cara mendapatkan *Keypoint* adalah dengan mencari sudut-sudut atau perpotongan garis (*corner*) yang terdapat pada gambar. *Keypoint-keypoint* yang terdeteksi pada gambar akan memiliki sebuah vektor untuk mendeskripsikan daerah di sekitar *keypoint* tersebut yang disebut sebagai vektor *descriptor*. Proses pencarian fitur lokal pada penelitian ini akan dilakukan dengan menggunakan metode *Scale Invariant Feature Transform* (SIFT) dan *Oriented FAST and Rotated BRIEF* (ORB). Metode SIFT dan ORB akan menghasilkan vektor *descriptor* untuk tiap fitur lokal yang terdeteksi, vektor *descriptor* ini dapat digunakan untuk mengidentifikasi fitur lokal. Pemanfaatan metode SIFT dan ORB pada skripsi ini akan dilakukan dengan menggunakan implementasi OpenCV pada Python.

Salah satu masalah yang ada pada pengenalan POI adalah pada sebuah gambar POI tidak semua fitur lokal yang terdeteksi bersifat unik terhadap POI tersebut. Ada fitur lokal yang juga dimiliki oleh POI lain atau fitur lokal yang berasal dari objek di latar belakang yang sifatnya tidak konsisten. Masalah ini akan mempersulit pada proses OIR untuk mengidentifikasi POI yang tepat. Gambar 1.3 menunjukkan masalah-masalah ini, gambar 1.3a dan 1.3b menunjukkan fitur lokal yang mirip dari dua POI yang berbeda, sedangkan gambar 1.3c dan 1.3d menunjukkan objek-objek latar belakang yang tidak konsisten pada POI. Penelitian ini akan melakukan analisis untuk menemukan dan memisahkan fitur-fitur lokal tersebut agar tidak ikut diproses dalam pembuatan model POI.



(a)



(b)



(c)



(d)

Gambar 1.3: Empat gambar di atas menunjukkan permasalahan yang dihadapi pada penelitian ini. Gambar (a) dan (b) merupakan gambar dari dua POI yang berbeda tetapi memiliki logo dengan bagian sudut yang mirip, sudut yang mirip tersebut kemungkinan akan menghasilkan fitur lokal yang mirip juga. Sedangkan pada gambar (c) dan (d) merupakan dua gambar dari POI yang sama tetapi banyak objek latar yang berbeda sehingga akan memunculkan fitur lokal yang tidak konsisten.

Fitur-fitur lokal yang tidak unik dan tidak konsisten tersebut akan dipisahkan dengan menggunakan teknik *clustering*. Teknik *clustering* merupakan teknik pemrosesan data yang akan mengelompokkan data-data dengan sifat yang mirip ke dalam satu kelompok. Teknik *clustering* pada penelitian ini akan dilakukan menggunakan metode *Agglomerative*. Penelitian ini mengasumsikan fitur lokal yang merepresentasikan suatu POI adalah fitur lokal yang muncul secara konsisten di gambar POI tersebut dan relatif unik terhadap POI tersebut.

1.2 Rumusan Masalah

Skripsi ini memiliki rumusan masalah sebagai berikut:

1. Bagaimana membuat model pengenalan POI berdasarkan fitur lokalnya menggunakan teknik *data mining*?
2. Bagaimana mengidentifikasi POI dalam sebuah gambar berisi POI dengan memanfaatkan model pengenalan POI yang telah dibuat?

1.3 Tujuan

Skripsi ini memiliki tujuan sebagai berikut:

1. Membuat perangkat lunak yang akan menghasilkan model pengenalan POI berdasarkan dari *dataset* yang diberikan
2. Membuat perangkat lunak yang dapat melakukan identifikasi POI dari sebuah gambar POI dengan menggunakan model yang dihasilkan.

1.4 Batasan Masalah

Berikut batasan-batasan masalah dari skripsi ini:

- Pengambilan fitur lokal untuk analisis dilakukan menggunakan metode SIFT dan ORB dengan implementasi OpenCV pada Python.
- Pencarian pasangan fitur lokal dilakukan dengan menggunakan implementasi FLANN yang tersedia dari *library* OpenCV pada Python.
- Proses pelabelan daerah penting dalam POI dilakukan dengan menggunakan aplikasi Labelimg.

1.5 Metodologi

Skripsi ini akan memiliki metodologi sebagai berikut:

1. Melakukan studi literatur tentang metode OIR, teknik ekstraksi fitur lokal SIFT dan ORB, serta teknik-teknik *data mining* yang digunakan pada skripsi ini. Studi literatur dilakukan dengan mencari dan membaca *paper* atau buku yang berkaitan dengan topik tersebut.
2. Mengumpulkan *dataset* gambar POI yang diperlukan untuk penelitian dan pembuatan model identifikasi.
3. Melakukan analisis pada latar belakang masalah pengenalan POI, dengan melihat sifat-sifat fitur lokal pada gambar POI.
4. Menyusun rancangan perangkat lunak.
5. Melakukan implementasi perangkat lunak.
6. Menguji kinerja perangkat lunak.
7. Menulis buku skripsi.

1.6 Sistematika Pembahasan

Sistematika pembahasan yang digunakan pada penelitian ini adalah sebagai berikut:

1. Bab 1 Pendahuluan
Bab ini berisi tentang hal-hal yang menggambarkan skripsi ini secara garis besar. Hal yang dibahas merupakan latar belakang masalah, rumusan masalah, tujuan penelitian, batasan masalah, dan metodologi penelitian.
2. Bab 2 Landasan Teori
Bab ini berisi tentang dasar-dasar teori dari teknik atau metode yang digunakan dalam skripsi ini, yaitu OIR, metode SIFT, metode ORB, *Best Score Increasing Subsequence* (BSIS), Kd-Tree, teknik *clustering Agglomerative*, serta metode SIFT dan ORB di *library* OpenCV.
3. Bab 3 Analisis
Bab ini berisi analisis pada masalah yang dibahas pada skripsi beserta solusi yang digunakan untuk menyelesaikan masalah tersebut.
4. Bab 4 Perancangan
Bab ini berisi tentang perancangan baik dari metode *clustering* pada pemilihan fitur dan perancangan metode identifikasi POI dengan OIR. Bab juga akan berisi rancangan struktur *file* dan *folder* pada hasil akhir perangkat lunak.
5. Bab 5 Implementasi dan Pengujian
Bab ini berisi implementasi perangkat lunak pada metode pemilihan fitur dan identifikasi POI serta pengujian terhadap kinerja kedua metode tersebut.
6. Bab 6 Kesimpulan dan Saran
Bab ini berisi kesimpulan yang didapatkan dari hasil analisis serta keseluruhan implementasi dan pengujian yang dilakukan pada penelitian ini.

