

REPRESENTASI PENGETAHUAN DENGAN GRAF SPASIAL TEMPORAL PADA PEMAHAMAN SCENE

RINGKASAN DISERTASI

Marzuki

NIM: 33214302

(Program Studi Doktor Teknik Elektro dan Informatika)



Institut Teknologi Bandung
Juli 2021

REPRESENTASI PENGETAHUAN DENGAN GRAF SPASIAL TEMPORAL PADA PEMAHAMAN SCENE

Disertasi ini dipertahankan pada Sidang Terbuka Sekolah
Pascasarjana sebagai salah satu syarat untuk memperoleh gelar
Doktor Institut Teknologi Bandung

Juli 2021

Marzuki

NIM: 33214302

(Program Studi Doktor Teknik Elektro dan Informatika)



Promotor : Prof. Dr. Carmadi Machbub
Ko-promotor : Dr. Techn. Ary Setijadi P, ST., MT
Ko-promotor : Dr. Rinaldi Munir

Institut Teknologi Bandung
Juli 2021

REPRESENTASI PENGETAHUAN DENGAN GRAF SPASIAL TEMPORAL PADA PEMAHAMAN SCENE

Marzuki
NIM: 33214302

1. Latar Belakang

Persepsi mesin adalah suatu istilah yang digunakan untuk mengidentifikasi kemampuan mesin menginterpretasi informasi yang dihasilkan oleh sensor dengan cara yang alami (seperti cara manusia). Persepsi mesin secara teori telah diusulkan oleh Lotfi A. Zadeh (Zadeh, 2002) yang mengatakan “From Computing with Numbers to Computing with Words - From Manipulation of Measurements to Manipulation of Perceptions”. Secara umum komputasi dalam persepsi mesin berfokus pada manipulasi angka dan simbol. Sebaliknya, komputasi dengan kata-kata (Computing with Words), adalah metodologi dimana objek komputasinya adalah kata-kata dan proposisi yang diambil dari linguistik.

Beberapa sistem otomatis seperti pada robot, kendaraan cerdas, sistem keamanan, surveillance dan lain-lain sangat bergantung kepada pemahaman terhadap lingkungannya. Agar dapat memetakan dirinya dan membuat keputusan yang rasional, sistem-sistem otomatis tersebut harus secara komprehensif memahami ruang lingkup keberadaannya pada lingkungannya tertentu. Keberhasilan pemahaman tersebut sangat bergantung kepada bagaimana sistem memahami tempat beraktivitas dan bernavigasi secara visual (scene) melalui mekanisme tertentu dan menautkan objek-objek di sekitarnya. Sementara pada tataran pengolahan citra, beberapa proses yang dibutuhkan dalam pemahaman terhadap scene antara lain: (a) segmentasi; (b) pengenalan objek yang berbeda; dan (c) pemahaman terhadap hubungan elemen-elemen yang ada dalam scene, sehingga banyak penelitian yang tersebar pada 3 (tiga) tingkatan pengolahan tersebut.

Peningkatan yang signifikan pada penelitian pemahaman scene terkini (state of the art) adalah melalui pengembangan rekognisi dan klasifikasi kategorisasi scene berbasis dataset (Patterson dkk., 2014; Xiao dkk., 2010, 2012).

Xiao dkk., (2010) mengembangkan basis data (dataset) pemahaman scene dengan tujuan menyediakan informasi tentang lingkungan, tempat dan benda-benda di dalamnya kepada para peneliti dalam bidang visi komputer, persepsi manusia, pengenalan, neurosains, pembelajaran mesin, penambangan data, grafika komputer dan robotika. Xiao menggunakan sekumpulan besar citra beranotasi mencakup berbagai macam scene. Mereka melakukan pengembangan menggunakan kosa kata dibawah WordNet, mengumpulkan citra yang termasuk dalam setiap kategori scene

menggunakan mesin pencari citra secara daring dengan melakukan kueri untuk setiap kategori scene, selanjutnya memberi anotasi objek dalam gambar secara manual.

Dataset SUN dikembangkan menggunakan struktur graf pohon (tree) untuk menggambarkan hierarki dari scene yang memiliki data 908 kategori scene, 131.072 citra, 249.522 objek yang sudah disegmentasi dan 3.819 kategori objek. Dataset ini selanjutnya disebut dataset SUN yang dapat diunduh melalui tautan <https://groups.csail.mit.edu/vision/SUN/hierarchy.html> dalam 3 (tiga) format data berbeda yaitu berupa sekumpulan citra, lembar kerja (worksheet) yang menggambarkan struktur graf dari setiap tingkatan hierarki dan data anotasi objek dalam format Extensible Markup Language (XML) yaitu bahasa markup yang diciptakan oleh konsorsium World Wide Web (W3C).

Pemahaman scene secara visual pada lingkungan yang kompleks memiliki permasalahan adanya kemungkinan benda-benda yang saling menghalangi satu sama lain, sehingga mesin mengalami kesulitan untuk melakukan penalaran dan klasifikasi. Model yang dikembangkan oleh (Choi dkk., 2013, 2015) merupakan perluasan state-of-the-art pada bidang pemahaman scene, dengan memasukkan unsur-unsur tambahan yang mencoba mengatasi permasalahan tersebut yaitu 3DGP (Three Dimensional Geometric Phrases) yang secara manual dibuat sebagai model relasi sekumpulan objek pada kategori scene dalam ruangan sehingga setiap scene yang berbeda memiliki model 3DGP yang berbeda yang menyebabkan penalaran hanya bisa dilakukan pada scene yang sudah memiliki model.

Perkembangan kategorisasi scene terkini selain Choi dkk., (2015) yang juga mengaitkan pengetahuan objek dan scene menggunakan pendekatan pembelajaran dalam (deep learning) yang melakukan induksi skala objek dan scene dalam arsitektur jaringan saraf konvolusi multi-skala (Herranz dkk., 2016). Menurut Herranz dkk., (2016), karena scene terdiri dari bagian objek, akurasi pengenalan scene membutuhkan pengetahuan tentang scene dan objek. Dalam makalah tersebut mereka membahas dua masalah utama yaitu: (1) bias kumpulan data yang diinduksi skala dalam arsitektur jaringan saraf konvolusi multi-skala, dan (2) bagaimana menggabungkan pengetahuan scene-centric dan object-centric secara efektif (yaitu Places dan ImageNet) di CNN. Premis utama dalam (Herranz dkk., 2016) adalah objek dalam scene memiliki rentang skala tertentu. Hasil eksperimen menunjukkan bahwa akurasi pengenalan sangat bergantung pada skala, dan kombinasi multi-skala yang sederhana namun dipilih dengan cermat dari ImageNet-CNN dan Places-CNN, dapat meningkatkan akurasi pengenalan scene pada dataset SUN397 hingga 66,26% dan bahkan 70,17% dengan arsitektur yang lebih dalam pada CNN. Pendekatan ini memiliki akurasi yang baik, namun masih membutuhkan peningkatan akurasi jika digunakan pada sistem nyata dan penentuan skala pada pendekatan ini harus diatur berdasarkan data latih yang digunakan.

Perkembangan lain yang sangat terkait dengan pemahaman scene adalah deteksi objek. Perkembangan pendeteksian objek dua dekade terakhir berdasarkan survey yang dilakukan (Zou dkk., 2019) dibagi pada dua era yaitu; era sebelum tahun 2014 dan setelah tahun 2014 yang disebut metode deteksi tradisional dan metode deteksi berbasis pembelajaran dalam.

L. Liu dkk., (2020) menyatakan teknik pembelajaran yang mendalam telah muncul sebagai kekuatan pada deteksi objek pada tahun 2012 yaitu Deep Convolutional Neural Networks (DCNN) oleh (Krizhevsky dkk., 2012) dengan metode yang mempelajari representasi fitur langsung dari data latih dan telah menghasilkan terobosan luar biasa di bidang deteksi objek. Deteksi objek generik yang ideal adalah memiliki akurasi tinggi dan efisiensi tinggi, yaitu secara akurat melokalisasi dan mengenali objek dalam citra dengan tahapan deteksi berjalan secara real time dan kebutuhan memori (penyimpanan) yang dapat diterima (L. Liu dkk., 2020).

Pada era tradisional seperti Scale-Invariant Features (Lowe, 1999, 2004), deteksi wajah (P. Viola dan Jones, 2001; Paul Viola dan Jones, 2004), Histogram of Oriented Gradients (HOG) yang dipublikasikan pada tahun 2005 oleh N. Dalal dan B.Triggs (Dalal dan Triggs, 2005), Deformable Part-based Model (DPM) yang merupakan pengembangan dari HOG. Pendekatan DPM ini dipublikasikan oleh R. Girshick (Felzenszwalb dkk., 2010). Sebagian besar algoritma deteksi objek masa awal ini dibangun berdasarkan fitur buatan, karena kurangnya representasi citra yang efektif pada waktu itu. Namun demikian deteksi wajah Viola Jones sudah mampu diaplikasikan pada waktu nyata.

Era setelah tahun 2014 tonggak sejarah metode pembelajaran dalam diawali DCNN dengan metode deteksi berbasis pembelajaran dalam yang menggunakan representasi fitur langsung dari data latih. (Krizhevsky dkk., 2012) melatih jaringan saraf konvolusi yang besar dan dalam untuk mengklasifikasikan 1,2 juta citra yang memiliki resolusi tinggi ke dalam 1000 kelas objek yang berbeda. Pada tahun 2014, R. Girshick dkk., (2014) memperkenalkan R-CNN (Region based Convolutional Neural Networks). R-CNN yang menerapkan mekanisme yang disebut pencarian selektif untuk mengekstrak sekitar 2000 daerah yang diinginkan (Regions Of interest). Regions Of interest berupa area persegi panjang yang mewakili batas kelas suatu objek pada citra menjadi masukan pada CNN untuk menghasilkan fitur keluaran. Fitur luaran CNN selanjutnya diklasifikasikan untuk menentukan kelas objek. R-CNN mencapai rata-rata presisi 53,7% pada PASCAL VOC 2010 (Girshick dkk., 2014).

Sejak Tahun 2014 itu deteksi objek mulai berkembang dengan kecepatan yang belum pernah terjadi sebelumnya. Pada tahun 2015 kecepatan deteksi meningkat signifikan dengan dipublikasikannya YOLO tahun 2016 (Redmon dan Farhadi, 2016) yang mengklaim memiliki kecepatan proses 155 fps yang diaplikasikan pada

Dataset VOC07, dan terus dikembangkan pada tahun-tahun berikutnya (Redmon dan Farhadi, 2017, 2018).

Penelitian yang dilakukan di dalam disertasi ini adalah (1) merepresentasikan dataset SUN yang dikembangkan oleh Xiao dkk., (2010) menjadi basis pengetahuan pemahaman scene yang dapat digunakan oleh mesin pada waktu nyata dengan model graf properti yaitu suatu model graf yang dapat merepresentasikan struktur graf berlabel (memiliki label) yang dinotasikan dengan $G=(V,E,P)$, dalam hal ini G adalah notasi graf yang memiliki simpul (V), sisi (E) dan properti (P). Dengan demikian, untuk setiap simpul dan tepi dapat memiliki informasi berupa properti yang terdiri dari nama properti (label) dan nilai properti. (2) membangun algoritma penalaran graf dengan pendekatan probabilistik untuk menentukan kategori scene berdasarkan observasi yang dilakukan oleh sistem (3) mengembangkan algoritma pemetaan dan penjejakan objek pada bidang spasial berdasarkan spesifikasi spasial (Qiu dkk., 2003) dan algoritma Bochsinski dkk., (2017).

Dengan mempertimbangkan kecepatan proses deteksi, penelitian ini menggunakan pendeteksi objek YOLOv3 (Redmon dan Farhadi, 2018) sebagai komponen observasi dan sistem basis data graf sehingga memungkinkan diimplementasikan pada sistem waktu nyata seperti robotik, surveilans dan lain-lain.

2. Tujuan dan Sasaran Penelitian

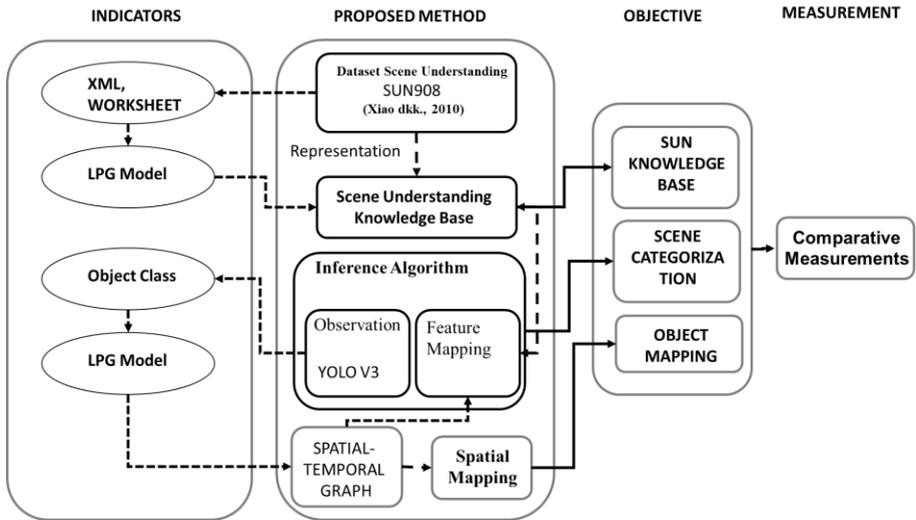
Penelitian ini bertujuan untuk menjawab beberapa persoalan sebagai berikut:

1. Merepresentasikan dataset SUN dengan Labelled Property Graphs (LPG) menjadi basis pengetahuan bagi pemahaman mesin agar dapat diaplikasikan dan digunakan oleh mesin sebagai pengetahuan pada waktu nyata dengan mengembangkan algoritma scene understanding knowledge base generator.
2. Mengembangkan model penalaran dalam pemahaman mesin dengan penalaran graf berbasis probabilitas untuk menentukan kategori scene dan mengusulkan algoritma kategorisasi scene.
3. Memperluas pemahaman scene pada citra dinamis dengan pemetaan dan penjejakan objek berdasarkan spesifikasi arah dan topologi yang diusulkan oleh Qiu dkk., (2003) dan penjejakan objek IOU yang dikembangkan oleh Bochsinski dkk., (2017).

3. Metode Penelitian

Tahapan penelitian yang dilakukan pada disertasi ini secara garis besar dibagi menjadi tiga tahapan utama yaitu menentukan indikator penentu permasalahan yang muncul pada pemahaman scene, memetakan dan mengusulkan pendekatan yang dimungkinkan dapat menyelesaikan permasalahan yang ada, dan mengukur serta mengevaluasi pendekatan yang diusulkan.

Tahapan penelitian ini selanjutnya divisualisasikan melalui kerangka berfikir pada Gambar 1 yang dapat dijelaskan sebagai berikut:



Gambar 1. Kerangka berfikir pada metode penelitian

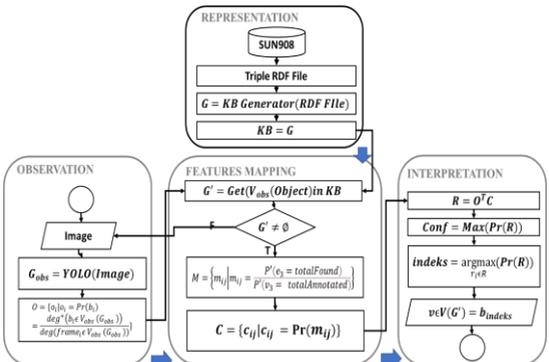
Indikator adalah parameter-parameter yang menyebabkan permasalahan pada penelitian dan parameter yang dapat digunakan sebagai solusi penyelesaian permasalahan. Dalam penelitian ini yang menyebabkan dataset SUN tidak dapat digunakan pada waktu nyata oleh mesin karena lambat saat penambangan data melalui proses traversal pada simpul-simpul data disebabkan format data yang dipublikasikan pada dataset SUN menggunakan format data triple RDF sedangkan format data untuk pengujian (diunduh) memiliki format data XML (eXtensible Markup Language) dan lembar kerja (worksheet) untuk merepresentasikan relasi pada kategori scene. Parameter yang diusulkan untuk permasalahan tersebut adalah model data LPG.

Parameter lain yang menyebabkan menurunnya akurasi algoritma kategorisasi scene dan penjejakan objek pada aplikasi nyata, karena dunia nyata memiliki kerumitan dan ritme perubahan yang tinggi (dinamis) yang disebabkan karena sebaran objek (object class) pada kategori scene tertentu dapat berubah berdasarkan waktu dan tempat yang berbeda, sedangkan algoritma yang berkembang saat ini sebagian mengabaikan informasi dinamis pada sebaran objek yang terdeteksi. Parameter yang diusulkan untuk permasalahan tersebut adalah model data LPG agar dapat menyimpan informasi-informasi spasial dan temporal.

Metode yang diusulkan adalah pendekatan yang dikembangkan dalam penelitian ini untuk mengatasi permasalahan penelitian. Dalam penelitian ini diusulkan tiga pendekatan yaitu merepresentasikan dataset SUN menggunakan model data LPG sebagai basis pengetahuan pada mesin yang dapat digunakan pada pemahaman scene pada waktu nyata, mengembangkan algoritma penalaran (inferensi) pada mesin berbasis penalaran graf pengetahuan dan memperluas pemahaman scene pada pemetaan, dan penjejakan objek pada bidang spasial.

Pengukuran adalah tahapan untuk mengukur akurasi pendekatan yang diusulkan dengan membandingkan pendekatan-pendekatan yang dikembangkan terkini (state of the art) sebagai dasar simpulan pada disertasi ini melalui beberapa eksperimen:

1. Mengidentifikasi tingkat akurasi pendekatan yang diusulkan dengan mengukur tingkat akurasi penalaran mesin menggunakan representasi pengetahuan dengan LPG berdasarkan label kategori citra masukan dari basis data SUN yang diunduh dari tautan <https://groups.csail.mit.edu/vision/SUN/hierarchy.html>,
2. Mengukur dan mengidentifikasi tingkat akurasi pendekatan yang diusulkan dengan membandingkan akurasi antara algoritma yang dikembangkan oleh Choi dkk.,(2015) dan algoritma yang dikembangkan oleh Herranz dkk.,(2016) dengan mesin yang menggunakan basis pengetahuan dan algoritma kategorisasi scene yang diusulkan pada penelitian ini. Citra yang digunakan adalah seluruh pada citra CVPR13, SUN397 dan ImageNet.
3. Mengidentifikasi algoritma pemetaan objek pada bidang spasial pada citra statis dan citra dinamis (video) berdasarkan transformasi laman web pada spesifikasi arah yang diusulkan oleh Mei Kang Qiu, (2003) dan algoritma penjejakan yang diusulkan oleh Bochinski dkk, (2017).



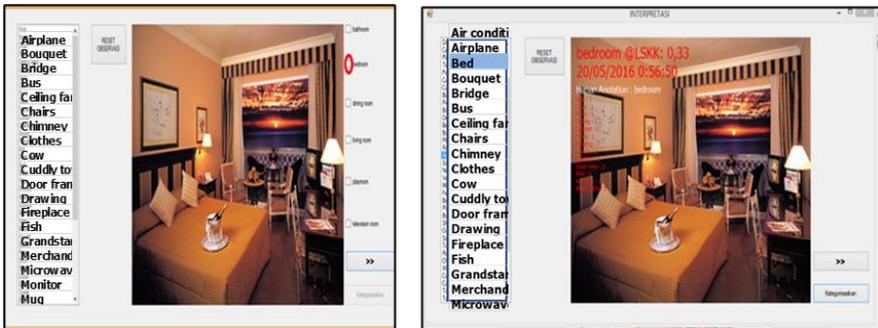
Gambar 2. Diagram alir metode penalaran

4. Hasil dan Pembahasan

4.1 Representasi Pengetahuan dan Anotasi Objek Manual

Representasi pengetahuan yang dimaksud dalam disertasi ini adalah menyusun ulang taksonomi SUN (Xiao dkk., 2010) ke dalam model LPG dengan menghubungkan objek-objek yang tersebar pada setiap kategori scene. Dengan mendefinisikan dataset SUN sebagai $G=(V,E,P)$, yaitu sebuah graf yang memiliki properti (P) maka dataset ini selanjutnya menjadi sebuah graf properti yang tersusun sedemikian rupa yang merepresentasikan hierarki dataset SUN (tree).

Pada studi kasus penalaran, eksperimen yang dilakukan pada penelitian ini untuk mengidentifikasi tingkat akurasi sistem yang menggunakan model yang dikembangkan dengan menggunakan citra uji dari SUN397, selanjutnya mengembangkan sebuah aplikasi sederhana dengan graphical user interface seperti pada Gambar 2 untuk mengimplementasikan model yang dikembangkan yang melibatkan 10 observer sebagai annotator dengan menggunakan 1.045 citra.



Gambar 3. Antarmuka aplikasi a). Antarmuka *annotator*; b). Luaran sistem berupa citra yang memiliki kategori *bedroom*

Untuk menguji representasi pengetahuan dan kategorisasi scene menggunakan penalaran graf properti maka pengujian menggunakan data citra SUN397 dan kategori scene yang digunakan adalah bathroom, bedroom, dining room, living room, playroom dan television room dengan jumlah citra pada masing-masing kategori seperti tertera pada Tabel 1.

Tabel 1 Jumlah citra pada setiap kategori scene

No	Kategori scene	Jumlah Citra
1	bathroom	280
2	bedroom	286
3	dining room	286
4	living room	154
5	playroom	37
6	television room	2

Eksperimen yang dilakukan pada penelitian tahap ini bertujuan untuk mengidentifikasi tingkat akurasi kategorisasi scene yang menggunakan basis pengetahuan berdasarkan informasi dari SUN dengan masukan label objek yang ditemukan/terdeteksi pada gambar dan dilakukan secara manual oleh observer. Jumlah citra yang diuji dari dataset SUN397 pada masing-masing observer dideskripsikan pada Tabel 2.

Tabel 2 Distribusi citra pada setiap observer

Annotator	bath room	bed room	dining room	living room	play room	television room	Jumlah citra
1	26	26	26	13	7	2	100
2	26	26	26	26	7		111
3	26	26	26	26			104
4	26	26	26	26			104
5	26	26	26	13	10		101
6	26	26	26	26			104
7	36	26	26	13	6		107
8	39	26	26	11	7		109
9	26	39	39				104
10	23	39	39				101

Dari jumlah seluruh gambar yang diobservasi, hanya 1.012 gambar yang teranotasi dari 1.045 citra sehingga ada 33 citra yang hilang atau tidak teranotasi dengan hasil akhir seperti pada Tabel 3.

Tabel 3 Jumlah gambar teranotasi

TOTAL IMAGES DISTRIBUTION											
bathroom		bedroom		dining room		living room		playroom		television room	
Anotator	Sistem	Anotator	Sistem	Anotator	Sistem	Anotator	Sistem	Anotator	Sistem	Anotator	Sistem
24	25	25	21	25	43	14	6	7	3	5	2
28	28	23	23	27	35	23	14	6	5	5	7
26	26	22	16	25	28	22	18	2	2	3	8
24	26	22	26	26	28	24	20	0	0	4	0
26	25	25	26	26	31	12	6	10	9	1	3
26	26	32	33	28	38	11	3	1	0	2	0
36	35	30	26	33	38	1	1	0	0	0	0
30	34	26	15	23	26	12	9	7	5	2	11
26	22	39	35	35	43	0	0	0	0	0	0
20	23	38	31	37	43	4	2	0	0	1	1

Pada pengenalan pola dan temu kembali informasi, precision dan recall adalah dua perhitungan yang banyak digunakan untuk mengukur kinerja dari sistem. Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem, sedangkan recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Ekspirimen yang dilakukan pada penelitian ini menggunakan penalaran graf pada basis pengetahuan untuk menentukan kategori scene berdasarkan anotasi kelas objek secara manual oleh observer. Performansi sistem diperoleh berdasarkan pengukuran akurasi yang didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

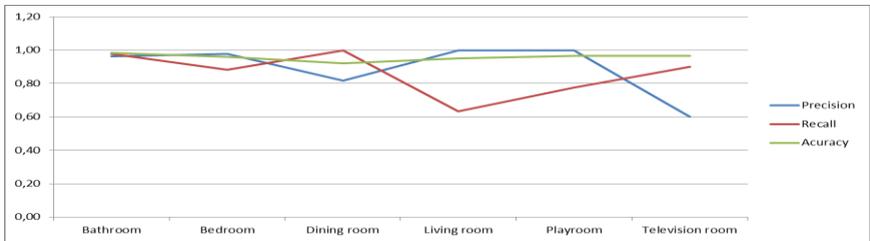


Gambar 3. Luaran sistem berupa citra yang sudah teranotasi berdasarkan basis pengetahuan dan penentuan kategori *scene*

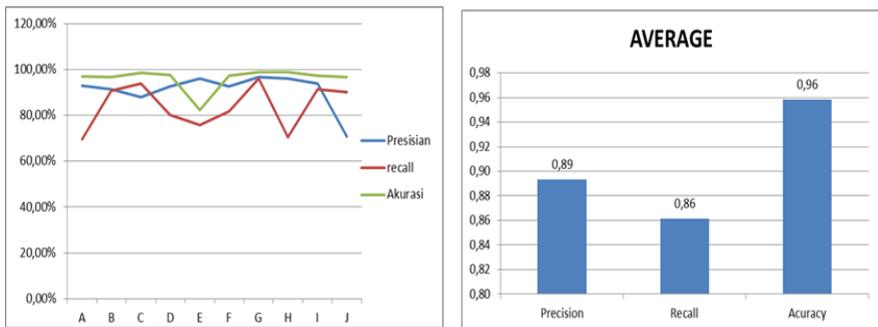
Model penalaran yang dikembangkan dalam penelitian ini menggunakan teori peluang yaitu suatu nilai yang digunakan untuk mengukur tingkat terjadinya suatu kejadian yang acak dan sering disebut dengan peluang atau kemungkinan atau probabilitas. Berdasarkan luaran sistem, selanjutnya data diolah dan dikelompokkan kembali berdasarkan kategori scene pada masing-masing observer dan dihitung presisi, recall dan akurasi seperti tertera pada Tabel 3 dan Gambar 4.

Tabel 3 Tingkat presisi, *recall* dan akurasi sistem

OBSERVER	Bathroom			Bedroom			Dining room			Living room			Playroom			Television room		
	Precision (%)	Recall (%)	Acuracy (%)	Precision (%)	Recall (%)	Acuracy (%)	Precision (%)	Recall (%)	Acuracy (%)	Precision (%)	Recall (%)	Acuracy (%)	Precision (%)	Recall (%)	Acuracy (%)	Precision (%)	Recall (%)	Acuracy (%)
A	96	100	99	100	84	96	58.14	100	82	100	42.86	92	100	42.86	96	100	40	97
B	100	100	100	100	100	100	77.14	100	92.86	100	60.87	91.96	100	83.33	99.11	71.43	100	98.21
C	100	100	100	100	81.82	96	89.29	100	97	100	81.82	96	100	100	100	37.5	100	95
D	92.31	100	98	84.62	100	96	92.86	100	98	100	83.33	96	0	0	0	0	0	0
E	100	96.15	99	96.15	100	99	83.87	100	83.87	100	50	94	100	90	90	33.33	100	98
F	100	100	100	96.97	100	99	73.68	100	90	100	27.27	92	0	0	0	0	0	0
G	100	97.22	99	100	86.67	96	86.84	100	95	100	100	100	0	0	0	0	0	0
H	88.24	100	96	100	57.69	89	88.46	100	97	100	75	97	100	71.43	98	18.18	100	91
I	100	84.62	96	100	89.74	96	81.4	100	92	0	0	0	0	0	0	0	0	0
J	86.96	100	97	100	81.58	93	86.05	100	94	100	50	98	100	0	0	100	100	100
Rata-rata	96.4	97.8	98.4	97.8	88.2	96.0	81.8	100.0	92.2	100.0	63.5	95.2	100.0	64.6	80.5	60.1	90.0	96.5



a



Gambar 4. a) presisi, recall dan akurasi masing-masing observer dan kategori b) rata-rata tingkat presisian, recall dan rata-rata akurasi sistem

a

4.2 Penalaran Kategori Pemandangan Berdasarkan Observasi Mesin

4.2.1 Studi Kasus pada Dataset ImageNet dan SUN2012

Data eksperimen yang digunakan pada tahap ini menggunakan data masukan berupa citra yang diunduh dari ImageNet dan SUN397. ImageNet adalah basis data gambar yang disusun menurut hierarki yang tertera pada WordNet dan setiap node pada hierarki digambarkan oleh ratusan atau ribuan gambar dari setiap label citra. Namun, ImageNet tidak memiliki hak cipta dari citra. ImageNet hanya menyediakan thumbnail dan tautan citra dengan cara yang mirip dengan apa yang dilakukan oleh mesin pencari. Menurut (Deng dkk., 2009), hal ini adalah upaya penelitian berkelanjutan untuk memberikan kemudahan para peneliti dan akademisi yang melakukan penelitian pada visi komputer dan pengolahan citra yang membutuhkan data skala besar.

Deng dkk., (2009) mengilustrasikan kegunaan ImageNet melalui tiga aplikasi sederhana dalam pengenalan objek, klasifikasi gambar, dan pengelompokan objek otomatis, sehingga berdasarkan skala, akurasi, keragaman, dan struktur hierarki dari ImageNet serta tautan citra yang dapat digunakan bagi para peneliti di komunitas

visi komputer dan kegunaan lain. Saat makalah mereka dipublikasikan tersedia 12 *sub tree* dengan 5.247 *synsets* yaitu antar muka yang terkait penalaran NLP pada WordNet dan dengan total 3,2 juta citra.

Pada studi kasus ini tautan ImageNet yang digunakan adalah data gambar yang ada pada www.flickr.com menggunakan pada 4 (empat) kategori *scene* yaitu *bath room*, *dining room*, *living room* dan *bed room*. Sebagai pembandingan, selain gambar yang diunduh dari ImageNet, eksperimen ini juga menggunakan citra dari basis data SUN, yaitu sebuah proyek pengembangan database gambar yang menyediakan data untuk para peneliti dalam bidang visi komputer, persepsi manusia, kognisi dan *neuroscians*, pembelajaran mesin dan penambangan data, komputer grafik dan robotika, dengan koleksi gambar beranotasi yang mencakup berbagai macam gambar *scene* lingkungan, tempat dan benda-benda di dalamnya.



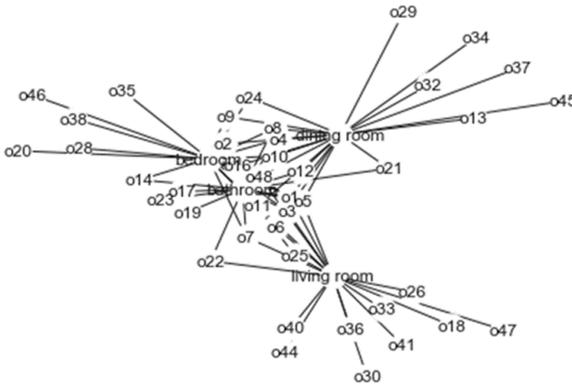
Gambar 5 Contoh data uji dari ImageNet



Gambar 6. Sampel dataset SUN

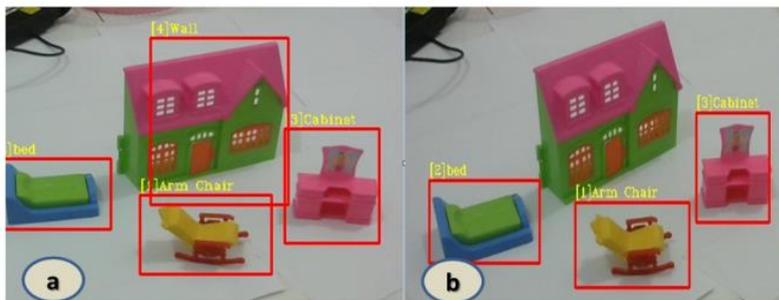
SUN database mengembangkan hierarki data dengan menghitung semua entri yang berhubungan dengan nama *scene*, tempat, dan lingkungan. (Xiao dkk., 2012) mengumpulkan gambar milik masing-masing kategori *scene* menggunakan mesin pencari gambar secara daring dan memberikan anotasi objek pada citra *scene* secara manual. Data yang digunakan dari SUN juga dikelompokkan dalam empat kategori gambar *bath room*, *dining room*, *living room* dan *bed room* yang diunduh dari tautan <https://groups.csail.mit.edu/vision/SUN/>.

Kategori *scene* yang digunakan pada pengujian basis pengetahuan yang dikembangkan dalam eksperimen penelitian tahap ini hanya pada empat yaitu *bath room*, *bed room*, *dining room*, dan *living room*, maka basis pengetahuan menjadi $G' = (V', E', P')$ sebagai sub graf dari G yang dinotasikan $G' \subseteq G, \text{if } V' \subseteq V, E' \subseteq E$. Selanjutnya interpretasi label *scene* dengan mencari nilai probabilitas tertinggi dari kemungkinan G_{obs} adalah suatu kategori *scene* tertentu pada G'



Gambar 7. Graf G' merupakan graf pengetahuan untuk pengujian

Pada eksperimen tahap ini, dilakukan pengujian dan perbandingan sistem pendeteksi objek. Pengujian awal yang dilakukan pada tahap ini adalah menggunakan data latih 9 (sembilan) jenis data latih (*object*) yang diambil dari berbagai sudut pandang menggunakan kamera HD Webcam C270 dengan resolusi 1.280 x 960 setara dengan 1.2 Mega Pixel (Gambar IV.6). Penamaan (*labelling*) objek yang yang digunakan pada data latih disesuaikan dengan kamus data objek pada dataset SUN yang dibatasi hanya pada kategori hierarki *Home or Hotel*. Pengenalan objek yang digunakan pada tahap awal ini adalah ekstraksi fitur menggunakan HOG dan SVM sebagai metode klasifikasi.



Gambar 8. Ketahanan rekognisi a) Observasi seluruh obyek dengan sudut pandang awal; b) Menurunnya ketahanan rekognisi dengan bergesernya kamera ± 150

Pengujian selanjutnya menggunakan YOLO yaitu suatu model pengenalan objek yang dikembangkan oleh (Redmon dkk., 2015)



Gambar 9. Pendeteksian objek secara *real time* menggunakan YOLO

Dengan menggeser arah kamera $\pm 25^{\circ}$ hasil luaran sistem masih stabil seperti yang diilustrasikan pada Gambar 9, maka untuk menunjang eksperimen ini pendekatan YOLO digunakan sebagai sistem rekognisi pada observasi. Selanjutnya data luaran YOLO didefinisikan menggunakan 6 parameter yang dideskripsikan pada Tabel 4.

Tabel 4. Parameter luaran YOLO

No	Parameter	Keterangan
1	Confidence	nilai keyakinan objek terdeteksi
2	Width	lebar area pembatas objek terdeteksi
3	Height	tinggi area pembatas objek terdeteksi
4	Type	kelas objek terdeteksi
5	X	kordinat x axis kiri atas area
6	Y	koordinat y axis kiri atas area

Dengan demikian maka luaran observasi selanjutnya pada tahap pemrosesan awal (*preprocessing*) dan didefinisikan sebagai himpunan pasangan berurutan tidak kosong M dengan himpunan N yang didefinisikan sebagai graf $G_{obs} =$

$(V_{obs}, E_{obs}, P_{obs})$ Maka, graf G_{obs} adalah representasi himpunan M dengan himpunan N yang dinyatakan sebagai himpunan pasangan berurutan (x,y) sebagai relasi pada G_{obs} dengan $x \in M$ dan $y \in N$.

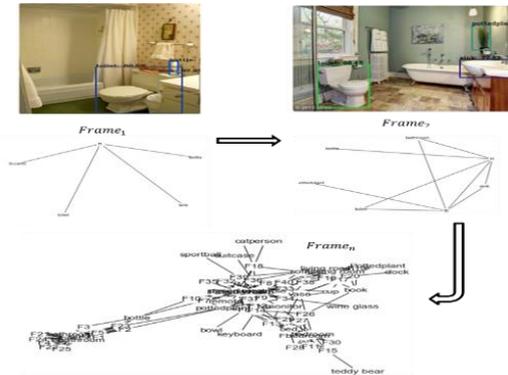
G_{obs} adalah graf properti, yaitu sebuah graf yang mengidentifikasi benda-benda yang berhasil dideteksi oleh sistem dengan pola baru menggunakan skema pada Gambar 10.

Gambar 10. Ilustrasi graf G_{obs} yang merupakan graf pada kejadian tertentu $frame_t$



Gambar 11. *Sportball* dan *remote* adalah objek yg tidak terdapat pada graf

Pada studi kasus ini selain mengukur tingkat akurasi sistem dengan luaran rata-rata akurasi juga dihasilkan pola graf yang berbeda dengan pola graf observasi citra tunggal karena $F = \{frame_1, frame_2, \dots, frame_t\}$ sehingga memungkinkan pendekatan ini digunakan pada citra dinamis seperti video atau kamera seperti ilustrasi Gambar 12.



Gambar 12. Graf luaran dengan iterasi 1 ke n pada proses interpretasi kategori

Pola graf yang didapat menggambarkan beberapa hal, antara lain: (1) objek yang sama pada suatu *scene* seperti kursi, meja, orang dan lain-lain dapat ditemukan pada *frame* yang berbeda dan setiap *frame* dapat memiliki objek yang berbeda, (2) *frame* pada waktu tertentu terelasi langsung terhadap *frame* berikutnya (citra dinamis) dan

(3) satu atau lebih *frame* dapat terelasi kepada satu label *scene*, akan tetapi satu *frame* hanya dapat terelasi pada satu label *scene*.

Tabel 6. Deskripsi graf pada *frame* menunjukkan bahwa objek tertentu dapat ditemukan pada *frame* yang berbeda dan ditemukan pada *scene* yang berbeda

No	Object	Frame	Scene
1	bed	6, 11, 12, 13, 14, 15, 26, 27, 28, 29, 30	dining room, bedroom, dining room
2	book	16, 19, 20, 26	living room, bedroom
3	bottle	1, 3, 5, 7, 10	Bathroom, dining room
4	bowl	6	dining room
5	cat	18	dining room
6	chair	6, 7, 8, 9, 10, 12, 14, 16, 18, 27, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40	dining room, living room, bedroom
7	clock	16	living room
8	cup	17, 33, 34	living room, dining room
9	diningtable	6, 7, 8, 9, 12, 14, 16, 31, 32, 34, 39, 40	dining room, living room
10	keyboard	14	dining room
11	person	18	dining room
12	pottedplant	2, 6, 7, 9, 16, 19, 20, 23, 26, 29, 31, 34, 36, 37, 40	Bathroom, dining room, living room, bedroom, dining room
13	remote	13, 39	bedroom, dining room
14	sink	1, 2, 3, 4, 5, 21, 22, 23, 24	bathroom
15	sofa	16, 17, 18, 19, 20, 36, 38, 39, 40	living room, dining room
16	sportball	39	dining room
17	suitcase	36	dining room
18	teddy bear	15	bedroom
19	toilet	1, 2, 3, 4, 5, 21, 22, 23, 24, 25	bathroom
20	tvmonitor	6, 13, 14, 38	dining room, bedroom, living room
21	vase	8, 9, 17, 20, 26, 29, 33, 36, 37	dining room, living room, bedroom
22	wine glass	34	dining room

nilai *Conf* dihasilkan oleh modul interpretasi yang merupakan nilai tingkat kepercayaan terbesar didapatkan melalui formulasi $Conf = \max(Pr(R))$ dikelompokkan berdasarkan *R* selanjutnya $index = \operatorname{argmax}(Pr(R))$ mengacu pada $r_i \in R$ indek kategori *scene*. Tahapan untuk menentukan kategori *scene* yang memiliki *Conf*, adalah dengan pemetaan label pada *B* yang memiliki indeks yang sama dengan nilai *Conf* seperti tertera pada Tabel 7.

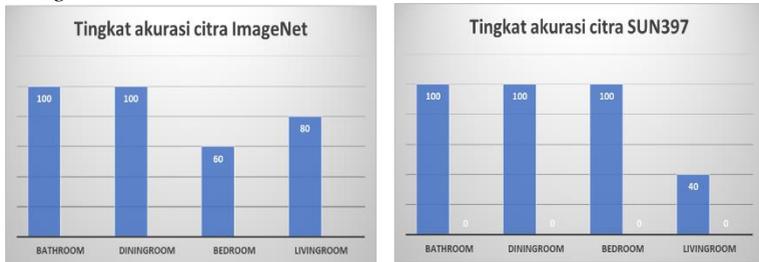
Tabel 7. Luaran sistem untuk masing-masing dataset

R						SUN2012					
1	31%	0%	0%	0%	bathroom	21	42%	0%	0%	0%	bathroom
2	19%	5%	3%	13%	bathroom	22	28%	0%	0%	0%	bathroom
3	37%	0%	0%	0%	bathroom	23	14%	4%	2%	10%	bathroom
4	19%	0%	0%	0%	bathroom	24	14%	0%	0%	0%	bathroom
5	37%	0%	0%	0%	bathroom	25	37%	0%	0%	0%	bathroom
6	0%	29%	33%	20%	dining room	26	0%	26%	6%	26%	bedroom
7	7%	13%	46%	32%	dining room	27	0%	75%	33%	14%	bedroom
8	0%	16%	72%	39%	dining room	28	0%	100%	0%	0%	bedroom
9	0%	17%	68%	40%	dining room	29	0%	38%	7%	25%	bedroom
10	18%	12%	50%	21%	dining room	30	0%	100%	0%	0%	bedroom
11	0%	100%	0%	0%	bedroom	31	0%	20%	81%	41%	dining room
12	0%	37%	55%	29%	dining room	32	0%	18%	80%	40%	dining room
13	0%	33%	0%	0%	bedroom	33	0%	18%	77%	36%	dining room
14	0%	25%	37%	20%	dining room	34	0%	15%	55%	32%	dining room
15	0%	50%	0%	0%	bedroom	35	0%	24%	100%	42%	dining room
16	0%	7%	15%	33%	living room	36	0%	14%	53%	37%	dining room
17	0%	0%	7%	31%	living room	37	0%	16%	55%	40%	dining room
18	0%	10%	43%	26%	dining room	38	0%	8%	33%	41%	living room
19	0%	5%	3%	36%	living room	39	0%	13%	58%	33%	dining room
20	0%	4%	4%	30%	living room	40	0%	13%	40%	42%	living room

Grafik akurasi pada Gambar 13 menunjukkan *scene* dengan label *Bedroom* dan *Living room* memiliki akurasi yang berbeda. Hal ini disebabkan:

1. Dataset COCO mengidentifikasi *table* sebagai *dining table* sementara pada data latih hanya mengenal *table*

2. Pada data latih *living room*, memiliki *chair* yang berbeda dengan *chair* pada *dining room*, yaitu *armchair* sehingga *living room* menjadi ambigu dengan *dining room*.

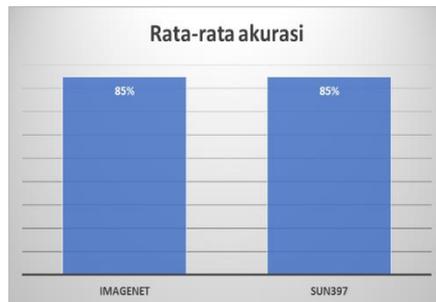


Gambar 13. Tingkat akurasi masing kategori pemandangan dengan sumbu x mendeskripsikan kategori *scene* dan sumbu y mendeskripsikan tingkat

Hasil eksperimen yang menggunakan data uji dari ImageNet, tingkat akurasi dari kategori *bathroom* dan *diningroom* memiliki akurasi 100%, sedangkan *bedroom* dan *livingroom* masing-masing 60% dan 80%. Tingkat akurasi dataset SUN397 (SUN2012) pada kategori *bathroom*, *diningroom* dan *bedroom* adalah 100% akan tetapi pada *livingroom* tingkat akurasinya hanya 40% (Gambar 13).

Penurunan akurasi pada dataset SUN397 disebabkan karena data latih yang digunakan pada pendeteksian objek adalah VOC pascal yang tidak membedakan *armchair* dengan *chair* sedangkan pada data uji yang digunakan *armchair* yang dikenali sebagai *chair*.

Berdasarkan luaran sistem tersebut di atas, selanjutnya dihitung rata-rata akurasi dari masing-masing dataset dan divisualisasikan pada Gambar 14.



Gambar 14. Tingkat rata-rata akurasi luaran dengan sumbu x adalah akurasi dari masing-masing dataset yang digunakan dan sumbu y mendeskripsikan tingkat akurasi

4.2.2 Penalaran Kategori Pematangan pada Dataset CVPR13

Pada studi kasus ini dengan memberikan masukan kepada komponen observasi berupa nama file data citra yang akan dianalisa untuk deteksi objek yang mengidentifikasi objek yang ditemukan, nama citra yang diolah, serta waktu pengolahan.



Gambar 15. Citra uji CVPR13

Citra yang dianalisis adalah citra yang diunduh dari CVPR13, yaitu data yang digunakan oleh banyak peneliti pada bidang visi komputer, data ini yang digunakan oleh (Choi dkk., 2013) dengan jumlah total 1.045 citra.

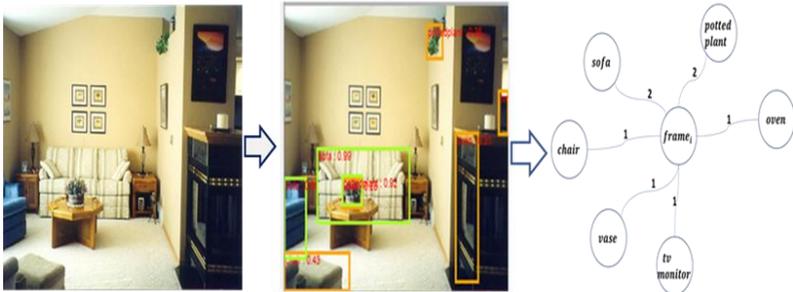
Tabel 8. Jumlah citra setiap Kategori

Kategori Scene	Jumlah Citra
bedroom	365
diningroom	365
livingrooom	365

Seperti yang telah dipaparkan sebelumnya, interpretasi adalah bagian dari proses yang dilakukan terhadap data sensor untuk mendapatkan pemahaman tertentu (persepsi). Pada penelitian ini interpretasi yang dilakukan berdasarkan pendekatan probabilitas bersyarat, yaitu suatu kejadian dapat ditentukan nilai peluangnya terjadi setelah sebuah kejadian lain diketahui. Untuk mengetahui kejadian pada waktu tertentu penelitian ini memberikan sebuah proses awal yang disebut observasi.

Observasi yang dimaksud pada penelitian ini adalah proses pendeteksian objek pada citra digital menggunakan algoritma deteksi objek YOLO, dengan hipotesa awal adalah sekumpulan objek yang terdeteksi pada proses observasi dapat digunakan untuk menentukan kategori suatu *scene* tertentu.

Pada tahap awal observasi sistem akan menerima masukan data citra yang akan di observasi, selanjutnya dilakukan deteksi dengan algoritma pendeteksi objek. Luaran algoritma ini adalah nama objek yang terdeteksi, koordinat kiri atas kotak pembatas, lebar kotak pembatas dan panjang kotak pembatas serta nilai probabilitas yang mewakili tingkat kepercayaan sistem (*confidence*) selanjutnya luaran direpresentasikan sebagai pengetahuan observasi dengan G_{obs} .

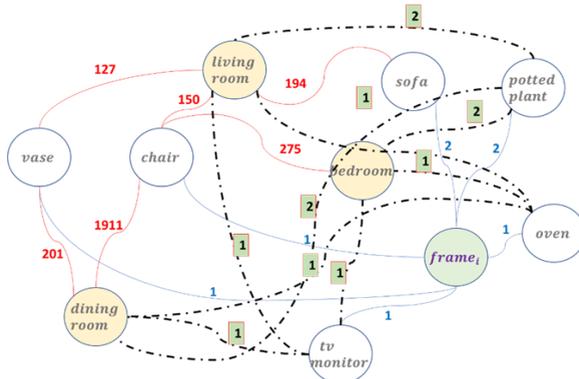


Gambar 16. Citra masukan (kiri) dan citra luaran yang telah diobservasi menggunakan pendeteksi objek (kanan)

Luaran observasi selanjutnya direpresentasikan menjadi model struktur data graf yang merepresentasikan keterkaitan objek: $B = P_{obs}(Object = name)$ yang terdeteksi dengan citra yang dilakukan observasi graf G_{obs} . Selanjutnya G_{obs} akan diekstrak ke struktur matriks $O = \{o_i | o_i = Pr(b_i) = \frac{deg^+(b_i \in V_{obs}(G_{obs}))}{deg^+(frame_i \in V_{obs}(G_{obs}))}\}$. Penentuan fitur selanjutnya adalah proses untuk mendapatkan pengetahuan yang mengandung kaidah peluang bersyarat berdasarkan fakta yang pernah ada graf pengetahuan (G) dengan menggunakan *query* Q berupa graf (V_{obs}, E_{obs}, f) di mana v_{obs} dan e_{obs} adalah kumpulan simpul dan tepi berurutan dari graf observasi (G_{obs}) untuk mendapatkan data sesuai label objek yang ada pada observasi (*chair, sofa, vase, sofa, pottedplant, oven, tvmonitor*) untuk objek yang tidak ditemukan pada G , maka untuk setiap G' akan menggunakan data pada G_{obs} , sehingga pengetahuan baru (G') yang terbentuk dari observasi pada Tabel 9 direpresentasikan dalam struktur data graf seperti pada Gambar IV.16.

Tabel 9. Pengetahuan baru berdasarkan observasi

Object/Scene category	bedroom	dining room	living room
Chair	275	1911	150
Vase	0	201	127
Sofa	0	0	194
Pottedplant	2	2	2
oven	1	1	1
tvmonitor	1	1	1



Gambar 17. Representasi pemutakhiran pengetahuan pada graf

Tahapan berikutnya adalah mencari rerata hitung (*mean*), yaitu untuk mencari rerata hitung dari jumlah (*n*) objek yang teranotasi berbanding jumlah citra (*N*) yang teranotasi dengan μ .

Tabel 10. Rerata hitung pengetahuan

Kategori Scene/ Obyek	Bedroom	Dining room	Living room
Chair	0.263663	5.082447	0.837989
Vase	0	0.534574	0.709497
Sofa	0	0	1.083799
Pottedplant	0.001918	0.005319	0.011173
Oven	0.000959	0.00266	0.005587
Tvmonitor	0.000959	0.00266	0.005587

Lalu dihitung berdasarkan Tabel 3 sehingga fitur dari pengetahuan diinterpretasikan sebagai *C*:

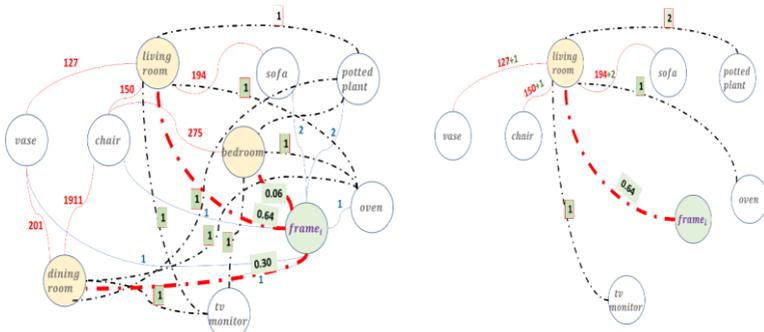
$$C = \begin{bmatrix} 0.04 & 0.82 & 0.14 \\ 0 & 0.43 & 0.57 \\ 0 & 0 & 1 \\ 0.10 & 0.29 & 0.61 \\ 0.10 & 0.29 & 0.61 \\ 0.10 & 0.29 & 0.61 \end{bmatrix}$$

Berikutnya adalah tahap interpretasi untuk menemukan nilai $R = O^T C$, perhitungan ini adalah bentuk lain dari mencari nilai total peluang dalam peristiwa jamak. Intinya dalam suatu peristiwa beberapa peristiwa ditemukan terjadi secara bersamaan. Dalam penelitian ini, pengamatan dalam *frame* dapat ditemukan di lebih dari satu objek dan memiliki probabilitas di lebih dari satu kategori *scene*.

$$\mathbf{R} = \begin{bmatrix} 0.125 \\ 0.125 \\ 0.25 \\ 0.25 \\ 0.125 \\ 0.125 \end{bmatrix}^T \begin{bmatrix} 0.04 & 0.82 & 0.14 \\ 0 & 0.43 & 0.57 \\ 0 & 0 & 1 \\ 0.10 & 0.29 & 0.61 \\ 0.10 & 0.29 & 0.61 \\ 0.10 & 0.29 & 0.61 \end{bmatrix} \\
 = [0.06 \quad 0.30 \quad 0.64]$$

Dan menentukan nilai probabilitas terbesar dari R dengan $conf = \max Pr(R) = [0.06, 0.30, 0.64] = 0.64$.

Jika ditelusuri berdasarkan interpretasi tersebut dengan struktur graf, dapat dideskripsikan seperti pada Gambar 18.a bahwa semua *scene* pada pengetahuan memiliki nilai peluang berdasarkan graf observasi pada $frame_i$ dengan nilai probabilitas yang berbeda (garis putus-putus merah)



Gambar 18. Penalaran graf interpretasi pada $frame_i$

Pada Gambar 18 semua nilai probabilitas $Scene_r$ pada $frame_i$ adalah nilai yang ditempati dalam semua kategori *scene* yaitu *diningroom*, *bedroom*, dan *livingroom* (a), setelah nilai R terbesar diperoleh (b), sistem akan mendapatkan pemahaman bahwa $Pr(Scene_r | Object)$ berada pada *livingroom* dengan nilai kepercayaan = 0.64 berdasarkan struktur R dan nomor *index* dari *conf* maka label *scene* pada $frame_i$ adalah b_{index} berdasarkan nilai $con r_{index}$.

Tabel 10. Jumlah citra yang dikenali sistem

Label Citra CVR13	Luaran Sistem Penalaran			Total
	bedroom	diningroom	livingroom	
bedroom	316	18	31	365
diningroom	0	333	32	365
livingroom	1	23	291	315

Eksperimen pada penelitian ini diaplikasikan pada dataset CVPR13 yang digunakan oleh (Choi et al., 2015) yang memiliki 3 kategori *scene bedroom*, *diningroom* dan *livingroom*. Hasil eksperimen pada pengujian menggunakan dataset CVPR13 disajikan pada Tabel 10 menjelaskan bahwa;

- 1) Dari 365 citra *bedroom* yang uji, terdapat 31 citra yang terdeteksi sebagai *livingroom*, 18 citra sebagai *dining room*, dan 316 dikategorikan dengan benar (*bedroom*)
- 2) Pada kategori *diningroom* dengan jumlah citra 365, yang dikategorikan dengan benar oleh sistem terdapat 333 citra sedangkan 32 citra dikategorikan sebagai *livingroom* dan
- 3) Citra yang berlabel kategori *livingroom* berjumlah 315 citra, terdapat 1 citra sebagai *bedroom*, 23 sebagai *diningroom* dan 291 dikategorikan dengan benar yaitu *livingroom*.

4.3 Perluasan Metode Kategorisasi pada Pemahaman Scene

4.3.1 Pemetaan Objek pada Bidang Spasial

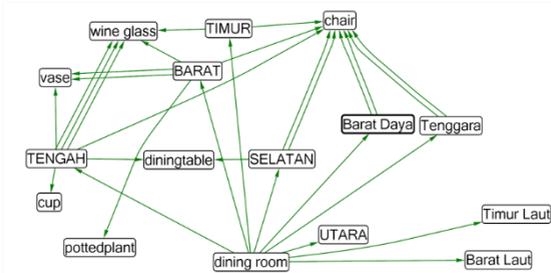
Penelitian ini selain mengenali tipe *scene* seperti ruang tamu, kamar tidur, ruang makan dan lain-lain, juga meningkatkan pengenalan *scene* ke tingkat pemahaman, yaitu sebuah teknik mengorganisasi kemampuan mesin untuk melakukan persepsi (identifikasi dan interpretasi) melalui data masukan sensor visual yaitu citra. Dengan persepsi mesin maka diharapkan mesin mampu menjelaskan melalui perspektif manusia mengapa mesin mengambil keputusan berdasarkan karakteristik yang diberikan kepada mesin.

Karakteristik yang diberikan pada pemahaman yang dimaksud dalam penelitian ini selain interpretasi probabilistik (sekumpulan benda memiliki nilai peluang untuk dapat menentukan kategori *scene*) melalui graf observasi dan graf pengetahuan juga memberikan karakteristik spasial yaitu karakteristik situs.

Situs (*Site*) adalah karakteristik yang menandai letak suatu objek terhadap objek lain di sekitarnya. Secara spesifik karakteristik yang diterapkan ini terkait posisi arah benda pada *scene* dan topologi benda yang satu dengan benda yang lain.

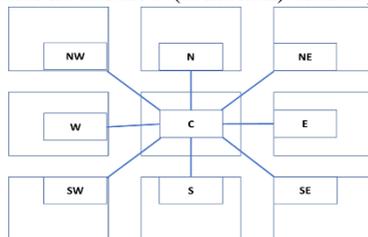
Contoh 4 : Sebagai ilustrasi pada Gambar 19 sistem diberi masukan sebuah citra statis, melalui algoritma yang dikembangkan dikenali citra tersebut memiliki kategori *scene Dining Room*. Jika dipetakan berdasarkan spesifikasi arah, maka ditemukan *dining tabel* berada di tengah citra, dan dikelilingi *chair* yang masing-masing dari arah barat daya 2, timur 1, barat 1, tenggara 2. Selanjutnya di area tengah terdeteksi *dining table*, 3 *wine glass* 1 *vase* dan satu *cup*.

Notasi spasial adalah spesifikasi yang menjadikan perbedaan antara *Graph Grammar Spasial* (GGS) dengan graf lain, secara umum spesifikasi notasi pada GGS oleh Mei Kang QIU (Qiu dkk., 2003) dibagi menjadi 3 (tiga spesifikasi) yaitu spesifikasi arah, topologi dan penjarangan. Pada eksperimen ini digunakan 2 (dua) spesifikasi yakni arah dan topologi;



Gambar 19. Relasi kategori pemandangan dengan tataletak objek pada bidang spasial.

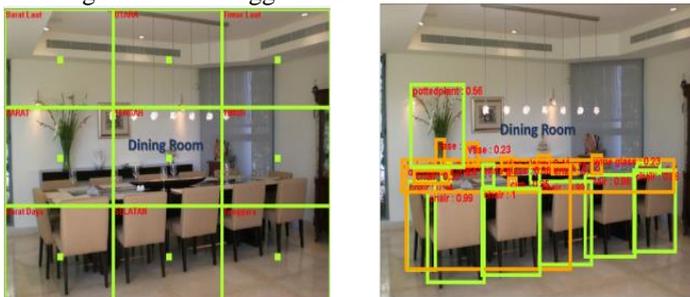
Spesifikasi arah adalah spesifikasi penempatan objek pada bidang spasial dengan membagi bidang spasial ke dalam 9 (sembilan) arah seperti pada Gambar 20.



Gambar 20. Arah bidang spasial

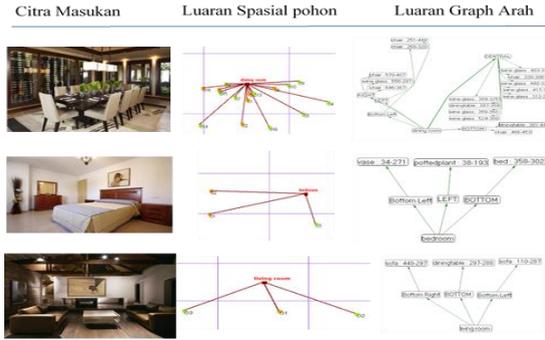
Pengelompokan objek menjadi graf berdasarkan tata letak pada arah bidang spasial dapat mewakili pemahaman *scene* (misalnya kursi berada pada sebelah kiri citra *scene*, meja berada ditengah citra ruangan, dll).

Pengelompokan sebaran objek pada pemahaman *scene* dilakukan membagi bidang spasial menjadi 9 (sembilan) blok arah dan mentransformasi tata letak objek pada bidang spasial menjadi graf spasial. Selanjutnya untuk setiap objek yang terdeteksi dipetakan ke arah berdasarkan pengukuran titik pusat objek dengan titik pusat arah menggunakan algoritma ketetanggaan terdekat.



Gambar 21. Pemetaan area arah pada citra kategori *scene*

Simpul daun pada graf yang terbentuk adalah label kelas sebuah objek yang dikenali pada *scene*, maka nama daun adalah label yang unik untuk membedakan kelas objek. Sebagai contoh, pada ruang makan dikenali 2 buah kursi yang sama tetapi lokasi pada citra berbeda. Untuk membedakan kedua objek kursi tersebut, maka nama objek diikuti nilai titik pusat sebagai simpul daun.



Gambar 22. Luaran sistem berupa graf spasial

4.3.3 Pemetaan Objek pada Citra Dinamis

Eksperimen pemetaan dan penjejakan objek pada studi kasus pada penelitian ini berjumlah 140 frame dengan person sebagai objek yang dijejaki bergerak pada area yang didefinisikan (A) pada waktu t_0 sampai t_n . Dalam hal ini t_0 adalah frame ke 1 dan $t_n = \text{jumlah frame} - 1$.

Citra yang digunakan pada eksperimen ini adalah video yang memiliki kecepatan 30 frame perdetik dengan durasi video (t) selama 4 detik dan resolusi citra adalah 1920×1080 piksel. Dengan demikian untuk $C^i = X.Y$, maka $X=1920$ dan $Y=1080$ sehingga jumlah piksel pada setiap sumbu horizontal (l^x) dan sumbu vertikal (l^y) adalah 640 piksel dan 340 piksel.

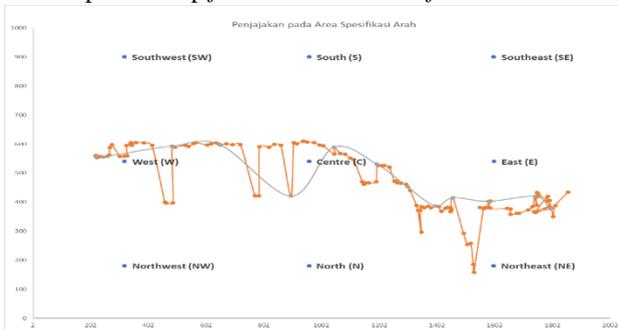
Berdasarkan koordinat setiap posisi arah pada bidang spasial pada Tabel IV.14, maka area target yang dijejaki (B^t) berdasarkan spesifikasi topologi pada himpunan arah (A) dipetakan dengan $A: \{ (B^t, A^w), (B^t, A^{nw}), (B^t, A^n), (B^t, A^{ne}), (B^t, A^e), (B^t, A^{se}), (B^t, A^s), (B^t, A^{sw}), (B^t, A^c) \}$.

Pemetaan objek yang dijejaki pada spesifikasi arah dilakukan dengan mencari titik pusat arah: $A_c^i = (A_{x_A}^i, A_{y_A}^i)$ dan dengan mencari titik pusat objek yang dijejaki (B_c^t) dengan:

$$B_c^t = (B_{x_B}^t, B_{y_B}^t) : x_B = \frac{x_1^t + x_2^t}{2}, y_B = \frac{y_1^t + y_2^t}{2}$$

Koordinat B_c^t selanjutnya dicari jarak kedekatan dengan koordinat A_c^i berdasarkan hitungan matriks *Euclidean* $D = \{d_i | d_i = D(B_c^t, A_c^i) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$.

Analisis pemetaan objek yang dijejaki pada spesifikasi arah berdasarkan spesifikasi topologi dilakukan pada setiap *frame* dari total 140 *frame*.

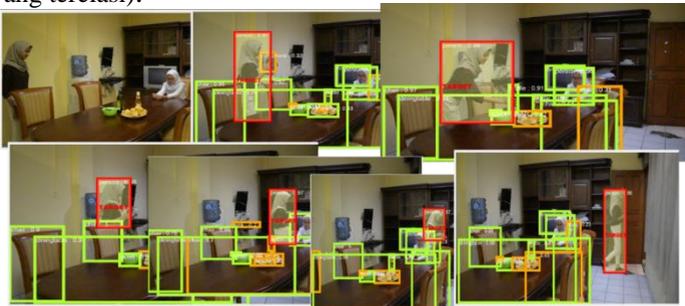


Gambar 23. Lokalisasi objek pada citra dinamis dengan penjejakan pada spesifikasi arah berdasarkan spesifikasi topologi

Contoh 2 : Graf observasi digunakan untuk menjejaki objek tunggal yang terkait dengan multi objek. Hal ini bertujuan untuk memperluas pemahaman mesin terhadap suatu *scene* dengan menganalisa kejadian pada suatu *scene* misalkan “orang A yang mengambil benda B”.

Jika diperhatikan kalimat “orang A yang mengambil Benda B” terdapat dua objek yang mungkin dijejaki untuk mendapatkan makna berdasarkan relasi graf, sehingga target yang dijejaki pun harus dinamis, yaitu mesin memiliki kemampuan menjejaki benda B atau orang A secara bergantian atau bersamaan dan menemukan makna kejadian pada waktu T .

Perbedaan antara eksperimen pada penjejakan pada video dengan objek tunggal dan eksperimen ini adalah target yang dijejaki bersifat statis (orang) dan dinamis (orang dan benda yang terelasi).



Gambar 24. Penjejakan objek orang dari arah barat (W) pada *dinning room* mengambil objek botol di meja makan

5. Kesimpulan atau Kontribusi Ilmiah

Terdapat beberapa kontribusi ilmiah yang dihasilkan dari disertasi ini antara lain:

1. Dataset SUN yang berisi koleksi lengkap citra beranotasi yang mencakup berbagai macam kategori scene dan benda-benda di dalamnya dapat direpresentasikan dengan model LPG sebagai basis pengetahuan untuk penalaran kategorisasi scene pada dataset lain yaitu ImageNet dan CVPR13.
2. Algoritma penalaran graf yang dikembangkan berdasarkan graf pengetahuan pada proses observasi dapat meningkatkan akurasi penalaran kategorisasi scene pada dataset CVPR13 dan dataset SUN397 secara berturut-turut sebesar 5,6% dan 14,83% dari (Choi dkk., 2015) dan (Herranz dkk., 2016).
3. Pemahaman scene dapat diperluas dengan pemetaan berdasarkan spesifikasi arah dan topologi dari Qiu dkk., (2003) dan penjejak objek IOU dari Bochinski dkk., (2017) pada bidang spasial dengan graf spasial-temporal menggunakan citra dinamis.

6. Tindak Lanjut

Penelitian ini tidak berhenti sampai disini. kelanjutan dari penelitian ini dapat dilanjutkan dengan :

1. Penalaran graf pada metode yang dikembangkan dalam penelitian ini menggunakan data eksperimen *scene* dalam ruangan (*indoor*) dengan menghitung peluang sekumpulan objek terelasi pada kategori pemandangan dalam ruangan. Pemahaman scene di luar ruangan (*outdoor*) dapat ditindaklanjuti dengan metode yang diusulkan berdasarkan data eksperimen di luar ruangan dengan menghitung peluang sekumpulan objek terelasi pada kategori pemandangan diluar ruangan.
2. Proses observasi yang dilakukan berdasarkan metode yang diusulkan memungkinkan pengetahuan baru muncul saat observasi yang belum ada pada dataset SUN, dalam penelitian ini disebut pengetahuan observasi. Pada pengembangan lebih lanjut, pengetahuan observasi yang belum ada pada basis pengetahuan sebelumnya dapat ditindaklanjuti dengan menambahkan pengetahuan observasi pada basis pengetahuan sehingga dimungkinkan basis pengetahuan dengan model LPG yang dikembangkan dapat ditambah dengan pengetahuan observasi.

Riwayat Hidup

Identitas Diri

Nama : Marzuki
NIM : 33214302
Email : marzuki@ubl.ac.id
Tempat Lahir : Lampung
Tanggal Lahir : 15 Juni 1973
Alamat : Mangkubumi Residence Blok E2 No.6 Gunung Agung
Langkapura Bandar Lampung

Riwayat Pendidikan

S1 : Ilmu Komputer, STIMK Tunas Bangsa
S2 : Ilmu Komputer, STMIK Eresha
S3 : Teknik Elektro, Institut Teknologi Bandung

Riwayat Pekerjaan

1. Dosen, ABA-AMIK Dian Cipta Cendikia, Bandar Lampung, Indonesia
2. Pengajar, Informatics, Lampung Indonesia
3. Ketua, Reserach of Information and Technology – UBL, Bandar Lampung, Indonesia
4. Ketua Program Studi, Informatika, FIK-UBL, Bandar Lampung, Indonesia
5. Peneliti, Pusat Penelitian Teknologi dan Informasi – ITB, Bandung, Indonesia

Daftar Publikasi

- Marzuki, A. Sukoco, M. S. Febrianti.(2015): Visual-based Machine Understanding Framework For Decision Making On Social Robot, 4th International Conference on Interactive Digital Media (ICIDM).
- Marzuki, Hidayat E. M. I, Munir. R, Prihatmanto, A. S, Machbub, C (2016): Scenes Categorization based on Appears Object Probability. *6th International Conference on System Engineering and Technology (ICSET)*.
- Marzuki, Munir. R, Prihatmanto, A. S, Machbub, C.(2021) Labeled Property Graphs Model in Machine Understandings for Scene Images, International Journal on Electrical Engineering and Informatics ITB.

Ucapan Terima Kasih

Penulis sangat berterima kasih pada Prof. Dr. Carmadi Machbub sebagai ketua Tim Pembimbing, atas segala saran, bimbingan dan nasehatnya selama penelitian berlangsung dan selama penulisan disertasi ini.

Penulis juga berterima kasih atas saran, kritik dan nasihat dari anggota Tim Pembimbing Dr.techn. Ary Setijadi Prihatmanto dan Dr. Rinaldi Munir.

Penulis juga berterima kasih atas saran, kritik dan nasihat dari Agus Pratondo, PHD dari Telkom University, Prof. Dr. Ir. Tati Latifah Erawati Rajab, Ir. Kridanto Surendro, M.Sc., Ph.D, Egi Muhammad Idris Hidayat, S.T, M.Sc., Ph.D., beserta seluruh dosen serta pegawai Laboratorium Sistem Kendali dan Komputer dan Pusat Penelitian Teknologi Informasi dan Komunikasi, Institut Teknologi Bandung pada khususnya, dan seluruh dosen Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung pada umumnya.

Terimakasih juga disampaikan kepada institusi tempat penulis beraktifitas, yaitu Universitas Bandar Lampung, atas kesempatan yang diberikan kepada penulis untuk menjalani studi S3 dan Kementerian Riset Teknologi dan Pendidikan Tinggi yang telah memberi bantuan pembiayaan pendidikan Doktorat selama tahun 2014-2017, sehingga penelitian Disertasi ini dapat terlaksana dengan baik.

Terima kasih disampaikan kepada keluarga dan rekan sejawat yang selalu mendukung dan menyemangati Penulis selama pendidikan program doktor ini..